

New approaches to study excited states in density functional theory: general discussion

Jan Gerit Brandenburg,  Kieron Burke, Emmanuel Fromager, Matteo Gatti, Sara Giarrusso, Nikitas I. Gidopoulos, Paola Gori-Giorgi,  Duncan Gowland, Trygve Helgaker, Matthew J. P. Hodgson,  Lionel Lacombe, Gianluca Levi,  Pierre-François Loos, Neepa T. Maitra, Eduardo Maurina Morais,  Nisha Mehta, Filippo Monti, Manasi R. Mulay, Katarzyna Pernal,  Lucia Reining,  Pina Romaniello, Matthew R. Ryder,  Andreas Savin,  Dumitru Sirbu, Andrew M. Teale,  Alex J. W. Thom, Donald G. Truhlar,  Jack Wetherell and Weitao Yang

DOI: 10.1039/D0FD90026E

Kieron Burke opened discussion of the paper by Neepa Maitra: My question is, could one give a more quantitative answer to the suggestion: if orbital occupations stay close to ground state, does the adiabatic approximation work?

We have found it very helpful in our ground-state work (with Eunji Sim¹) to think very hard about how you decide if something is working. In that case, (almost) all efforts are focussed on the energy to use as a measure. We also focus exclusively on the density, not the potential, because it is so difficult to know if interesting features in the potential are relevant. In the end, TDDFT predicts the density.

1 M.-C. Kim, E. Sim and K. Burke, *Phys. Rev. Lett.*, 2013, **111**, 073003.

If you take the simplest case, such as just applying a uniform time-dependent E field with one frequency in time, and look at just $d(t)$, the time-dependent dipole moment produced, could you define some quantitative measure of how well TDDFT (or any other approximate scheme) is working? And if you could, maybe you could quantify your statement for at least a subset of cases (which should include of course linear response).

Lionel Lacombe replied: I agree that the density should be the main focus. I just wanted to add that actually the idea of “not working” we used in the paper could be quantified more directly, for example using the integral of the absolute of the difference between the exact and adiabatic time-dependent densities. I think the raw dipole is, at the same time, too “forgiving” for the details of the

density and too sensitive to the small amount of electronic density far from the nuclei.

Neepta Maitra answered: These are challenging questions but here are some things we can say. First, just to clarify, it is not how close the orbital occupation numbers are compared to the ground-state ones, but rather how close the KS occupation numbers are to the interacting occupation numbers. We can pick any initial KS state, so in a situation that the interacting system starts far from the ground state, we can pick an initial KS state that also starts far from its ground state; then our argument is that if we pick one where the initial KS occupation numbers are close to the initial interacting occupation numbers, then the accuracy of the adiabatic approximation then depends on how much in time the initial interacting occupation numbers evolve. The point I want to stress here is that this can explain why in some cases the adiabatic approximation gives accurate dynamics even when the interacting system is far from the ground state since when it stays close to a ground state then it is not unexpected that the adiabatic approximation works OK. Regarding what to measure, this is a very good point especially since we sometimes see some rapid and sharp structures in the potential that have relatively little effect on the density, and we completely agree that the measure should be based on the density. So, what we used to say whether adiabatic TDDFT was working well or not, was a locally-averaged density. If we just took the dipole moment $d(t)$ we found that this often gave a too positive picture, enhancing regions far from the center, and the underlying density was not necessarily close to the exact. To avoid reading too much into fluctuations in the pointwise density, we instead chose a “locally-averaged density” as a measure. But indeed we did not make a quantitative measure of this, although it could be possible. For linear response, I think our analysis is too coarse: our analysis is based on the density rather than perturbations of the density around a dominant ground state as in linear response, and I think our analysis may not be “fine” enough to make a statement about these perturbations. Our focus was really on the non-perturbative regime, way beyond the ground state, since this is where one naively expects adiabatic approximation to completely fail, yet sometimes it does not.

Sara Giarrusso enquired: At the end of section 4 of the paper (DOI: 10.1039/d0fd00049c), you mention that scaling relations could be helpful in simplifying or approximating your expression for the XC time-dependent potential that makes use of the time-dependent adiabatic connection formalism (eqn (15)). Do you have already some ideas or hints on how to use them or is it premature?

I ask because I have tried for quite a while to use scaling relations on expressions that are somehow similar to those in your paper in the sense that we also have a sort of what you call the “Berry-like term” but in our case it contains the derivative with respect to λ of the conditional amplitude (see ref. 1). In fact, at the local level, the term containing this derivative does not disappear even in the ground state (although, as is known, it does disappear when multiplied by the density and integrated over space, due to the Hellmann–Feynman theorem) and, in my experience, using scaling of the coordinates typically complicates the expressions, but maybe I have been looking at it in the wrong way.

- 1 S. Giarrusso and P. Gori-Giorgi, Exchange–correlation energy densities and response potentials: Connection between two definitions and analytical model for the strong-coupling limit of a stretched bond, *J. Phys. Chem. A*, 2020, **124**, 2473–2482.

Neepa Maitra answered: This is extremely interesting – I was not previously aware of your paper in *J. Phys. Chem. A*, but now I see it contains some beautiful analysis of ground-state potentials and their decompositions. It is very interesting you also find these similar terms with the lambda-derivative and that you found that scaling relations for the conditional amplitude did not help so far in your case. Indeed we were planning to look into the time-dependent scaling relations¹ to see if they could shed some insight or simplify something but we have not yet had a serious study of this. Ultimately approximations will be needed to get something useful, so one question is whether useful approximations will be suggested by the scaling procedure.

- 1 P. Hessler, J. Park and K. Burke, *Phys. Rev. Lett.*, 1999, **82**, 378.

Sara Giarrusso said: The “force-balance equation” (eqn (1)) is the cornerstone of your first line of investigation. This equation provides an exact expression for the exchange–correlation (XC) time-dependent potential in terms of interacting and non-interacting wave function quantities although somewhat complicated by the fact that the relation involves the divergence operator applied to the product of the density times the gradient of the potential. Such equation holds also in the Ground State theory.

However, for Ground State problems we are able to use a much simpler expression being quite close in spirit – namely, still providing an exact expression for the XC potential in terms of interacting and non-interacting wave function quantities – and having the advantage of equating directly the XC potential to something, compared to the force-balance equation as expressed in eqn (1). This simpler expression stems from a famous treatment based on the conditional amplitude (see ref. 1).

My (naïve) question is: would it not be possible to use a similar decomposition adopting the conditional amplitude in the time-dependent framework?

- 1 M. A. Buijse, E. J. Baerends and J. G. Snijders, Analysis of correlation in terms of exact local potentials: Applications to two-electron systems, *Phys. Rev. A*, 1989, **40**, 4190.

Neepa Maitra answered: This is an intriguing and thought-provoking suggestion. I really like the work that Baerends and co-workers have done in using their decomposition to analyze the ground-state Kohn–Sham (KS) potential. One thing that makes the time-dependent case a bit more complicated is the phase. In the ground-state case, the marginal is just the square-root of the density, while in the time-dependent case, the marginal has a phase, and this phase affects the potentials that appear in the effective equation for its motion. There is in general both a vector and a scalar potential. Then the question is whether this phase is the same for both the interacting and KS systems. It seems it would have to be if we would want to do the same trick as Baerends did in extracting the xc potential. If we require it to be the same, then there are two possibilities. One is that the KS current is equal to the true current, then the vector potentials in the effective equation coming from the interacting TDSE and the effective equation coming

from the KS TDSE would have to be the same, and subtraction can then yield a v_{xc} . But this then points to an inconsistency since in general for the TDKS current to equal the true current, an xc vector potential is needed, *i.e.* we would have to work within TDCDFT not TDDFT, and the KS equation from which we obtained the equation for the marginal would need a vector potential. If we stick with TDDFT and the possibility that the KS current is not equal to the true current, while the two marginals still have the same phase, then the two effective vector potentials are not the same and it would be unclear how then to extract v_{xc} . Finally, just to note that although our expression does not get the v_{xc} directly but rather gives $\text{div}(n \text{ gradient } v_{xc})$, the form of the equation is a Sturm–Liouville one for which I believe in the 3D case something like a generalized Poisson solver would work. (In 1D of course there is no problem). So we don't view that aspect of the expression as being a problem in itself.

Emmanuel Fromager commented: I am intrigued by the third term on the right-hand side of eqn (14). In the equation, an exact adiabatic connection formula is written for the time-dependent xc potential. The third term reminds me of the contributions that appear in Vignale's variational principle for TD-DFT (and that were not present in the original Runge–Gross variational principle). Those come from the fact that the wave function can vary at the initial and final times, even though the densities are fixed. If I remember right, these contributions fix the so-called “causality paradox”. I was wondering where this term comes from, in your case. Is there a connection with Vignale's variational principle?

Neepa Maitra replied: Indeed, Vignale's expression is also an exact expression for the xc potential, but there is a fundamental difference in that he is considering the action as a functional of the density, where the boundary-corrected stationary point (so to speak) gives the potential. (The boundary term does resolve the causality paradox, as you mentioned). The action that we use to derive our exact expression is simpler in that it is simply the Lagrangian action evaluated already at the solution point, *i.e.* it is zero. In other words, always we consider inputting into it, the solution of the TDSE. By doing this over the coupling constant integration, we are able to eliminate the troublesome kinetic term but at the expense of introducing other complicated terms, such as the one you mention. This term arises because when we take the lambda-derivative in the coupling-constant integral there are three terms, and simplifying the one involving the lambda-derivative of the ket by an integration by parts then introduces these boundary terms. The term has a derivative with respect to the coupling constant parameter projected onto the wavefunction at the time-boundaries, while the similar term in Vignale's expression that you noted involves a density-derivative at the time-boundaries. Your question raises a very interesting related question – whether doing a coupling-constant integral over Vignale's action gives a useful expression for the xc potential, also again transforming the difficult kinetic term into a coupling-constant integral. Lionel Lacombe had in fact looked into this, finding a different expression that is even more complicated, and it not clear yet whether it will yield something useful from the point of view of developing approximations.

Kieron Burke opened discussion of the paper by Pierre-François Loos: In the paper, you put in parameters α and β for H_2 . Are those general parameters, to be applied in any situation? Or are they system dependent? How about just looking at the initial slope, thereby avoiding the need to flatten the curve?

Pierre-François Loos responded: As shown in Fig. 3 of the paper (DOI: 10.1039/d0fd00059k), the coefficients α , β and γ associated with the weight-dependent exchange functional are system dependent as well as basis set and geometry dependent. They can even vary quite significantly even in the same system (H_2 at equilibrium and stretched H_2). However, these parameters are really trivial to determine (*via* a simple least square fit of the ensemble energy obtained at various weight values).

One can indeed only look at the slope around $w = 0$ if one is interested to perform ensemble calculation in the zero weight limit. This is actually what we do now as this is the sole information that one really needs to compute the infamous ensemble derivative term if the density is evaluated at $w = 0$.

Nikitas Gidopoulos asked: The non-linearity is due to the ghost interaction which is present already in the Hartree energy. Then, I think a natural way to correct the nonlinearity problem, would be to define the Hartree energy as a weighted sum of the Hartree energies of the constituent members of the ensemble, and also do the same with the exchange energy. In that case of course the ensemble HXC energy expression would become an implicit functional of the ensemble density and one would have to solve an OEP equation. Would this be a problem?

Pierre-François Loos answered: Yes, this is exactly true. We have eschewed this strategy due to the known problems associated with OEP equations but similar ideas are pursued by Gould and co-workers.^{1–3}

1 R. Garrick, A. Natan, T. Gould and L. Kronik, *Phys. Rev. X*, 2020, **10**, 021040.

2 T. Gould, *ChemRxiv*, DOI: 10.26434/chemrxiv.12382595.v1.

3 T. Gould, G. Stefanucci and S. Pittalis, *ChemRxiv*, DOI: 10.26434/chemrxiv.12456146.v3.

Lucia Reining remarked: Do you think it is easier to get excitation energies from the weight dependence in ensemble DFT than getting them from linear response, and if so, why? Is it because we ask for less information, *i.e.*, only for energies and not for intensities? Or is it maybe easier to implement exact constraints?

Pierre-François Loos answered: GOK-DFT is, in principle, less computationally demanding than linear response TD-DFT as one does not need to construct the TD-DFT Hamiltonian which spans the space of single excitations. The philosophy behind GOK-DFT is very similar to multiconfigurational wave function methods: one does include only the states that one is interested in.

Therefore, excitation energies are easier to extract from GOK-DFT. However, intensities are notoriously difficult to extract from such calculations and work in that direction is ongoing.

Lucia Reining questioned: The ensemble approach gives you also double excitations. Once you have them, do you think you could then conclude something about the frequency-dependent exchange–correlation kernel?

Pierre-François Loos replied: Quite possibly, one might be able to construct an effective Hamiltonian based on the individual states extracted from the ensemble calculation. In such a way, one would indeed be able to get insights about the frequency-dependence of the exchange–correlation kernel. We are planning to investigate such types of ideas in the future in the continuation of our recent work on dynamical kernels.¹

1 J. Authier and P.-F. Loos, *J. Chem. Phys.*, 2020, **153**(18), 184105.

Lucia Reining asked: Wouldn't you need also intensities to reconstruct the exchange–correlation kernel from the results of the ensemble approach?

Pierre-François Loos answered: One might be able to extract such intensities from the very same effective Hamiltonian. This has to be confirmed though.

Katarzyna Pernal commented: Is there an argument (theoretical or numerical) that a derivative of the exchange–correlation ensemble functional with respect to ensemble weight plays a more significant role in the description of double excitations (excitations which are missed in adiabatic approximation to TD-DFT) than for the single ones?

Pierre-François Loos responded: This is unclear and, I believe, very system specific. Our recent work on one-dimensional systems¹ has shown that the contribution of the correlation part of the ensemble derivative is quite significant for both single and double excitations (see Fig. 8 in ref. 1). It also varies significantly depending on the correlation regime (weak or strong).

1 P.-F. Loos and E. Fromager, *J. Chem. Phys.*, 2020, **152**, 214101.

Donald G. Truhlar communicated: Your paper points out that double excitations are completely absent from linear-response time-dependent Kohn–Sham density functional theory spectra with the adiabatic approximation for the exchange–correlation functional (for brevity, such calculations will simply be called TDDFT), and this is correct from a formal point of view. In practice though, TDDFT sometimes gives reasonable results for excitations that are classified by conventional means as double excitations. How can this be? We gave two possible answers.¹ First, some transitions conventionally labeled as double excitations are really cases where the ground-state has high multi-reference character, and the description as single or double excitations is oversimplified. A second factor is that a single-excitation method like TDDFT can yield an electron density similar to that obtained with explicit inclusion of doubly excited states. This provides a general caution against interpreting the non-interacting-electron reference wave function of Kohn–Sham DFT as if it were the wave function of the real system. Let me make an analogy. We are used to the fact that spin-polarized Kohn–Sham DFT can give a reasonable

potential curve for highly stretched singlet ground-state H_2 even though the Slater determinant wave function is a qualitatively incorrect wave function (because it predicts complete spin polarization whereas the true singlet has no spin polarization). In such a case we are used to accepting that we get a reasonable density and a reasonable energy even though the “structure of the wave function” is qualitatively incorrect. Similarly, perhaps we should accept that we can calculate reasonable densities and excitation energies for some excited states for which the “structure of the wave function” is qualitatively incorrect.

1 Y. Shu and D. G. Truhlar, Doubly Excited Character or Static Correlation of the Reference State in the Controversial 2^1A_g State of *trans*-Butadiene?, *J. Am. Chem. Soc.*, 2017, **139**, 13770–13778, DOI: 10.1021/jacs.7b06283.

Pierre-François Loos communicated in reply: Yes, this is an interesting comment. Such cases appear in butadiene and benzene where Gill and co-workers have clearly shown that, by taking into account orbital relaxation in the so-called doubly-excited state, one could end up re-classifying this excitation as a “single”.^{1,2} How to classify unambiguously the nature of a transition is, I think, an open question. In our recent work on double excitations (see ref. 3), we clearly observe different classes of double excitations.

The performance of wave function methods is clearly very dependent on the nature of the doubly excited states. For “fake” doubles (like in butadiene or benzene) or excitations with a large amount of single excitations, the error of methods like CC3 can be small. However, when one deals with genuine double excitations (where $\%T1 \sim 0$), errors can go up to 1 eV or more. So, there's definitely not one kind of double excitation.

1 G. M. J. Barca, A. T. B. Gilbert and P. M. W. Gill, *J. Chem. Theory Comput.*, 2018, **14**, 1501.

2 G. M. J. Barca, A. T. B. Gilbert and P. M. W. Gill, *J. Chem. Theory Comput.*, 2018, **14**, 9.

3 P.-F. Loos, M. Boggio-Pasqua, A. Scemama, M. Caffarel and D. Jacquemin, *J. Chem. Theory Comput.*, 2019, **15**, 1939.

Weitao Yang remarked: What you presented is an interesting approach to excitations. The treatment for ground state and excited state isn't on same footing as with TDDFT. Does this method suffer from the same problem that it'll miss the conical intersection between ground state and first excited state? (Or can it describe the topology of conical correctly?)

Pierre-François Loos answered: I would rather say that the treatment of ground and excited states is much more democratic within GOK-DFT than it is within TDDFT, especially in the case of an equi-ensemble where all the states have the same weight within the ensemble (similarly to a state-averaged CASSCF calculation). Of course, one can make one of the states predominant by increasing its weight. At zero-weight, the calculation indeed favours the lowest energy state. The faithful description of conical intersections in GOK-DFT will depend on the states present in the ensemble. If one includes the states relevant with the description of the conical intersection, we believe that the topology will be correct. After all, a GOK-DFT calculation is similar to a multiconfigurational calculation. Therefore, although it must be confirmed, we are very optimistic.

Andreas Savin opened discussion of the paper by Matteo Gatti: I would just like to check if I understood you correctly. If somebody would give you the exact Kohn–Sham potential for a given system, it would not be sufficient to generate your model potential.

Matteo Gatti replied: Yes, we argue that when we are interested in the trace of the frequency-dependent spectral function (as measured for example by angle-integrated photoemission) we would preferably use a more general auxiliary system with a frequency-dependent potential (see also ref. 1), since even the exact Kohn–Sham energies would not give the exact spectral function.

1 M. Vanzini, L. Reining and Matteo Gatti, 2017, arXiv:1708.02450 [cond-mat.mtrl-sci], <https://arxiv.org/abs/1708.02450>.

Andreas Savin asked: How about getting symmetry-forbidden transitions?

Matteo Gatti responded: An effective potential/kernel that is designed to give only a spectrum would be blind to symmetry-forbidden excitations as they do not contribute to the spectrum itself.

This observation is very general as it holds for any auxiliary system that targets only a specific quantity and is able to give only a limited amount of information. It is for the very same reason that it can be made easier to solve than the many-body problem (which would deliver any observable). In the framework of linear-response TDDFT a similar question has been discussed in ref. 1.

1 A. Heßelmann and A. Görling, *Phys. Rev. Lett.*, 2009, **102**, 233003.

Manasi Mulay enquired: Have you looked at metal oxide semiconductor systems? Any insights on the effect of bonding properties in metal oxide semiconductors on the frequency dependence of the scalar photoemission potential and effective absorption kernel?

Matteo Gatti answered: We have not applied the approach to metal oxides so far. We have focused on the feasibility of the computational scheme, so we have only considered prototypical sp metals/semiconductors. The approach is not mature yet for application to complex materials.

Manasi Mulay asked: Are you going to consider designing more complex systems such as semiconductor heterostructures? Any insights on the effect of heterostructure interface properties on the effective potential or kernel based design? What about ternary metal oxide (AB_2O_4) spinel structures such as zinc ferrite (ZnFe_2O_4) as an example? Do you anticipate any limitations of this method for these systems?

Matteo Gatti replied: We believe that when the approach is mature enough, it will be possible to use it for complex materials (surely with limitations due to unavoidable approximations). However, for the moment the approach has not reached that stage. In this project, our focus is presently on fundamental questions and methodological developments, rather than the simulation of materials properties.

Kieron Burke asked: Do you plan to apply this to finite systems, such as molecules? And if so, will you use the same models/connectors, or different ones?

Matteo Gatti answered: The definition of the auxiliary system is related to the observable of interest, so it can be the same for a finite or extended system. However, in the spirit of the connector theory approach (see ref. 1 and Lucia Reining's contribution, DOI: 10.1039/d0fd00068j), approximations based on importing results from model systems naturally lead to different choices for finite and extended systems.

In the present contribution (DOI: 10.1039/d0fd00067a), we have focused on extended systems. For the frequency-dependent potential for photoemission we have used the homogeneous electron gas (HEG) as the model and have imported tabulated results from the HEG to the materials of interest through a simple connector, the mean density of the material (see also ref. 2 – particularly, see footnote 55).

In finite systems such as molecules or clusters, instead, we would prefer a finite model (for example, a jellium sphere) rather than the HEG. Within the connector theory approach, we are indeed currently exploring models based on inhomogeneous and/or finite systems.

1 M. Vanzini, A. Aouina, M. Panholzer, M. Gatti and L. Reining, 2019, arXiv:1903.07930 [cond-mat.other], <https://arxiv.org/abs/1903.07930>.

2 M. Panholzer, M. Gatti and L. Reining, *Phys. Rev. Lett.*, 2018, **120**, 166402.

Lucia Reining opened a general discussion about double excitations: We can write linear response in a basis of transitions. If we diagonalize the resulting Hamiltonian, and if this Hamiltonian is static, then we can get only so many solutions as we have transitions in the basis. If the Hamiltonian is energy-dependent, instead, then we can get more solutions. The problem is, which are now single excitations and which are double excitations? I think this is not so straightforward to define, because you have to project on some basis, and this basis is a choice. According to the basis, we may arrive to a different conclusion. What is well defined, instead, is the question of the number of solutions: it is safe to say that only if the Hamiltonian is energy-dependent, we can have additional solutions.

Weitao Yang added a comment to the discussion on double excitations: If you stay within TDDFT, I agree that the kernel needs to be frequency dependent to describe double excitations. However, there is a different time dependent linear response theory, which is computationally more efficient or similar to TDDFT: the particle–particle random phase approximation (pp-RPA) for excitations, doing calculation from the ground state of $(N - 2)$ electrons. In pp-RPA, double excitation is a piece of cake, easy to get. Conical intersections as well. It can also start from $(N + 2)$ electron systems (the hole–hole random phase approximation). See ref. 1.

1 Y. Yang, H. van Aggelen and W. Yang, Double, Rydberg and charge transfer excitations from pairing matrix fluctuation and particle–particle random phase approximation, *J. Chem. Phys.*, 2013, **139**, 224105.

Neepta Maitra enquired: The pp-RPA is indeed elegant and practical too, and able to capture many states of double-excitation character, but not those which involve an excitation out of the HOMO–1, if I understand correctly. Can you confirm?

Weitao Yang replied: Indeed, particle–particle Random Phase Approximation (pp-RPA) is not perfect. While it solves many challenging problems associated with TDDFT, it is also limited with the orbitals involved. You have two orbitals you can pick. Two holes or two particles. If the excitation comes from HOMO–1 dominantly, the pp-RPA would not work well with the ground state of $(N - 2)$ electron system. But it still works well if the HOMO–1 contribution is less than 50%. This was documented in ref. 1. TDDFT is limited for different scenarios.

1 Y. Yang, D. Peng, J. Lu and W. Yang, Excitation energies from particle–particle random phase approximation: Davidson algorithm and benchmark studies, *J. Chem. Phys.*, 2014, **141**, 124104.

Kieron Burke added: The definition of the level of excitation is perfectly well-defined in KS-DFT and unambiguous, assuming states don't cross with the adiabatic connection parameter. It appears in my 2004 paper with Fan Zhang,¹ but is a simple consequence of Goerling–Levy perturbation theory.

Given some excitation of the true system, one simply follows that state's adiabatic connection path to $\lambda = 0$, where it becomes (usually) a single Slater determinant with some number of excitations from the Kohn–Sham ground-state. It can also be a mixture of more than one, in which case its character is mixed. But it is perfectly well defined.

In standard quantum chemistry and many-body theory, likely something similar can be done, but usually Hartree–Fock is used as the reference. This can mean the classification of a state can be quantitatively quite different in the two cases, but in each case it is well defined.

1 F. Zhang and K. Burke, *Phys. Rev. A*, 2004, **69**, 052510.

Pierre-François Loos responded: I fully agree with this comment. This is actually the method we use to characterize single and double excitations when we employ dynamical (*i.e.* frequency-dependent) kernels within linear response theory.¹ In this article, we indeed use a HF or GW reference but in that case one has to be extra careful (I think) as the density (or the Green's function) is not fixed along the adiabatic path.

1 J. Authier and P.-F. Loos, *J. Chem. Phys.*, 2020, **153**(18), 184105.

Nisha Mehta addressed Neepta Maitra: I am just curious how this approach performs compared to double hybrids?

Neepta Maitra answered: This is a good question. We have not tried using double hybrids for real-time propagation and I am not aware of any calculations by others who have. The mixing in of something like MP2 for correlation does represent a functional that would be nonlocal in space and time, through the orbital-dependence, but whether this nonlocality is the type that is important in

systems far from the ground state is not clear to me. I suspect probably not. One should also bear in mind that if we are sticking to pure KS DFT, the potential would need to be obtained *via* a TDOEP scheme, which gets complicated quite quickly.

Andrew M. Teale addressed Pierre-François Loos: Quite a long time ago I did some work with Prof. Helgaker and Dr Alex Borgoo,¹ where we used inversion techniques to look at weight dependence of the functional in the GOK formalism. We found this was quite difficult, and the outcome depended heavily on the basis set. We could get results similar to earlier ones shown by Prof. Burke but only for very large basis sets. For the practical approximations you present here do the results show any stronger sensitivity to the basis set size than typical density functional approximations?

1 A. Borgoo, A. M. Teale and T. Helgaker, *AIP Conf. Proc.*, 2015, **1702**, 090049.

Pierre-François Loos replied: The paper you refer to deals with the inversion of FCI quantities which is obviously more basis set dependent than density-based quantities. In our work, we have not observed strong basis set dependence (as it should within DFT). For example, excitations obtained with the aug-cc-pVTZ basis set were already extremely close to the complete basis set limit. In some cases, we even constructed the weight-dependent functional with a smaller basis set (aug-cc-pVTZ) and compute the excitation energies in a much larger basis (d-aug-cc-pVQZ). In short, basis set incompleteness error is not an issue within GOK-DFT.

Duncan Gowland communicated a question about the paper by Neepa Maitra: The authors present a “hybrid” approach where they show that mixing an adiabatic Hartree–Fock term dampens and stabilises dynamics of the density and potential (eqn (7), DOI: 10.1039/d0fd00049c). Firstly, are there other options for the adiabatic kernel used in a hybrid? Secondly, whilst the presented hybrid is limited in performance, can the authors comment on if there is an “exact” hybrid that exists? In contrast to this, can you comment on finding a “practical” hybrid potential which is stabilised long enough to capture some non-adiabatic dynamics of a real molecular system?

Neepa Maitra communicated in reply: Very interesting questions, to which I don't have complete answers, but here is what we can say. First, regarding other than Hartree–Fock for mixing in the hybrid, yes, for example one could try to use one of the density-matrix functionals from RDMFT, *e.g.* Mueller functional, BBC functionals, *etc.* Whether, once combined with the KS second-order DM, these would lead to a more stable propagation in the density-matrix coupled scheme, is a good question, and I am not sure of the answer. It seems that the instability occurs when a gentle minimum in the density (which need not be very near zero) becomes sharper and narrower in the approximate propagation (but not in the exact) with a sharp feature in the approximate potential that self-amplifies. It would be interesting to see whether a different hybrid would ameliorate this effect, I am not sure ahead of time. With your second question, the density-matrix coupled scheme does become exact if we had the exact ρ_2 as a functional of $\rho_{1,s}$ and ρ_1 : this is exact, but unknown. For a “practical” stable method, we'd love to find a paradigmatic system/model that gives a good approximation for the 1RDM

of the interacting system in terms of that of the KS system. This would enable us to bypass solving the equation for the 1RDM totally – we would simply use this directly in the TDKS equation.

Lionel Lacombe communicated in reply: Just to add to the first response from Neepa Maitra on the second part of the question: the ρ_1 is only used for the kinetic part of the potential, and this part only relies on the derivatives close to its diagonal. An “exact” functional does not need to provide a full exact ρ_1 and therefore any ρ_2 used for propagation that would give the exact derivatives on the diagonal could be considered as an exact functional.

Neepa Maitra communicated a question about the paper by Pierre-François Loos: The formulation presented includes the weight-dependence for the double excitation but if the weight-dependence is included for both the double and the single-excitation, then when a physical state has a very mixed character (like 50 : 50 single : double), do you expect the results to change significantly?

Pierre-François Loos communicated in reply: This is a tricky question. The excitation energies within GOK-DFT are supposed to be independent of the weights associated with the weights of the single and double excitations. The only quantity to model properly is the derivative of the xc ensemble functional. In our recent work on 1D systems,¹ we clearly see that it is harder to model properly single and double excitations when they strongly mix (like, for instance, in the strongly correlated regime at low density).

1 P.-F. Loos and E. Fromager, *J. Chem. Phys.*, 2020, **152**, 214101.

Weitao Yang communicated a question about the paper by Neepa Maitra: What you found and stated on the conditions for the potential in TDDFT is very interesting and useful. Are they consistent with being a functional derivative of an action functional (in the proper definition of action of Vignale, or other)? Is being a functional derivative important in your view? If not, how to think or define the potential and can the potential be unique? I also have the same set of questions for the kernel in TDDFT in its relation to the action functional (as the second-order derivative). Could you please also comment on the kernel?

Neepa Maitra communicated in reply: This is a great question. The expressions we use for the exact xc potential are indeed not derived from an action functional, although presumably there is a causality-corrected action of Vignale or an action defined along the Keldysh-contour (van Leeuwen) that underlies it, since it is the exact xc potential. Approximations that we build from the expression are not guaranteed to have this property. I am not sure if it is important but in fact I don't know, and it is an interesting question. The potential is certainly unique. It is defined from equating the second-time-derivative of the density of the true system with that of the KS system – this gives an expression that relates the potentials driving the two systems which has a unique solution for a given boundary condition, so we don't actually need an action functional to get a unique expression.

What we have found is certainly important are satisfaction of exact conditions like the zero force theorem, as there are direct physical consequences of violations

of that. As for the kernel, in principle it is a second functional derivative of the appropriate actions mentioned above, but again I am not sure if our approximations have that property, nor whether it is important. (Recently, Lionel Lacombe has derived a kernel for one of our approximations built from the exact expression, and is beginning to analyze what it gets.)

Matthew R. Ryder communicated a question about the paper by Matteo Gatti: Can you extend your approach to vibrational spectroscopy, such as infrared, Raman, and inelastic neutron scattering?

Matteo Gatti communicated in reply: In principle, the approach – being general – could be extended to vibrational spectroscopies. However, we have not investigated this possibility so far.

Donald G. Truhlar opened discussion of the paper by Gianluca Levi: The Perdew–Levy paper¹ shows that saddle points of $E(n)$ correspond to stationary densities. Your paper assumes, conversely, that all excited states are saddle points. Is it possible that some excited states are associated with local minima?

1 J. P. Perdew and M. Levy, Extrema of the density functional for the energy: Excited states from the ground-state theory, *Phys. Rev. B*, 1985, **31**, 6264–6272.

Gianluca Levi responded: As far as I know, it has not been proved yet that every excited state corresponds to a saddle point (and not a local minimum) of the KS functional. On the other hand, we note that in KS DFT, as in Hartree–Fock theory, the electronic Hessian in the canonical basis is diagonally dominant, with the dominant term of the diagonal elements consisting of the difference between the energy of a virtual orbital a and an occupied orbital i times the occupation number of i : $f_i(\varepsilon_a - \varepsilon_i)$. For each of the excited states included in our work, at the corresponding stationary point on the KS energy surface there is at least one occupied orbital higher in energy than a virtual orbital; therefore, the electronic Hessian is expected to have at least one negative eigenvalue for these cases. We have confirmed this by computing the Hessian numerically at the excited-state stationary points.

Pierre-François Loos asked: Usually single (singlet) excitations obtained with such approaches (*i.e.* using a single determinant) are very spin contaminated by the triplet state. Do you actually correct the energies somehow?

Gianluca Levi replied: It is correct that all (open-shell) singlet excited states are not pure-spin states in the calculations that we presented in the article, which use a single determinant. We do not apply any spin-purification correction of the energy of these states. This is because the intent of the calculations is to test the convergence properties of the direct optimization approach compared to SCF, and not to assess the accuracy of DFT with semi-local KS functionals for excited states of molecules.

Pierre-François Loos questioned: Do you see differences in convergence for single and double excitations? Hait and Head Gordon have studied single and double excitations with their technique which consists of optimizing the squared

gradient. In particular, they restricted their analysis to “single reference” double excitations and avoided on purpose “multireference” double excitations (like in the carbon dimer for instance). Is your method able to access those tricky doubly excited states?

Gianluca Levi answered: We have not assessed the convergence performance of the direct optimization (DO) MOM approach for double excitations yet. The tests on singly excited states that we have performed so far (see also ref. 1) show that DO-MOM is able to converge multireference cases, such as the $\Delta(\pi \rightarrow \pi^*)$ states of carbon monoxide and dinitrogen, whereas standard approaches based on SCF-MOM usually fail and require tuning modifications, such as electronic smearing.¹ Based on this, we can hypothesize that DO-MOM would perform better than SCF-MOM for multireference doubly excited states.

1 G. Levi, A. V. Ivanov and H. Jónsson, *J. Chem. Theory Comput.*, 2020, **16**, 6968.

Alex J. W. Thom commented: I have spent a while trying to find these excited states in HF (ref. 1 and 2) and DFT (ref. 3 and 4) recently. One problem is that states sometimes don't exist: at some geometries they are saddle points and some they don't exist. Do you see these sorts of problems where the states you are looking for just don't converge? In this paper (DOI: 10.1039/d0fd00064g), you seem to have convergence failures, but in the recently presented data you don't seem to have these failures.

1 A. J. W. Thom and M. Head-Gordon, *Phys. Rev. Lett.*, 2008, **101**, 193001.

2 H. G. A. Burton and A. J. W. Thom, *J. Chem. Theory Comput.*, 2016, **12**, 167.

3 K. T. Jensen, R. L. Benson, S. Cardamone and A. J. W. Thom, *J. Chem. Theory Comput.*, 2018, **14**, 4629.

4 R. A. Zarotiadis, H. G. A. Burton and A. J. W. Thom, *J. Chem. Theory Comput.*, 2020, DOI: 10.1021/acs.jctc.0c00822.

Gianluca Levi replied: Indeed, in our latest investigation¹ we could converge all the excited states included in the chosen benchmark set when employing a new limited-memory SR1 (L-SR1) formulation of the quasi-Newton update. These calculations are performed at a single geometry for each molecule and use a basis of linear combination of atomic orbitals. We cannot exclude that some of the states do not exist at different geometries or when the orbitals are represented using a different basis.

1 G. Levi, A. V. Ivanov and H. Jónsson, *J. Chem. Theory Comput.*, 2020, **16**, 6968.

Alex J. W. Thom asked: I'm interested in results for the multireference cases; do these have excited states that are harder to find?

Gianluca Levi responded: Within the KS approach, electronic states are approximated with a single Slater determinant, and states with multireference character are in general difficult to obtain. The benchmark set that was used for the convergence tests include some multireference cases, such as the $\Delta(\pi \rightarrow \pi^*)$ states of carbon monoxide, dinitrogen and acetylene, where both the donor and the acceptor orbitals belong to degenerate pairs. We have found that the common SCF-MOM approach usually fails to converge in these cases unless tuning

modifications, such as electronic smearing, are used.¹ The direct optimization (DO) MOM approach is, instead, more robust and can converge these states also when employing integer occupation numbers, but requires on average more iterations than the cases with single-reference character (see Fig. 3 in the *Faraday Discussions* article (DOI: 10.1039/d0fd00064g) and Fig. 2 in ref. 1 and related discussion in the text). The problem with multireference states is that there are multiple closely spaced critical points and the optimization can oscillate between them. Gradient-based unconstrained optimization methods seem to be less affected by this issue than SCF methods based on diagonalization of the Hamiltonian matrix.

1 G. Levi, A. V. Ivanov and H. Jónsson, *J. Chem. Theory Comput.*, 2020, **16**, 6968.

Filippo Monti remarked: As an experimental spectroscopist, I am very interested in your paper because I normally use commercially available TD-DFT methods to investigate the photophysics of the molecules I am studying (typically metal complexes or systems for charge separation). In this context, I am aware of the fact that TD-DFT can lead to underestimation of charge-transfer state energies, if no long-range corrected functionals are used. Is your proposed method intrinsically safe from this point of view or is the reliability of the results still very functional dependent? Indeed, in the last slide of your presentation you showed that you applied a self-interaction correction to the functional.

Gianluca Levi responded: There is evidence that charge-transfer (CT) excitation energies obtained from variationally optimized excited states in DFT are less affected by systematic errors and are less functional dependent than those obtained from calculations based on practical implementations of TDDFT. For example, some studies show that the variational methods give satisfactory results for short-range CT excitations already at the level of local KS functionals (see for example ref. 1 and 2). One reason for this is that these methods include orbital relaxation, which is important for such excitations but is missing in practical TDDFT calculations. Also, the $1/R$ dependence of the excitation energy on the separation between donor and acceptor in long-range CT systems, which in TDDFT can only be recovered using long-range corrected functionals, is, instead, reproduced by variational methods with local or ("global") hybrid functionals.^{3–5}

1 H. R. Zhekova, M. Seth and T. Ziegler, *Int. J. Quantum Chem.*, 2014, **114**, 1019–1029.

2 G. Levi, E. Biasin, A. O. Dohn and H. Jónsson, *Phys. Chem. Chem. Phys.*, 2020, **22**, 748–757.

3 G. M. J. Barca, A. T. B. Gilbert and P. M. W. Gill, *J. Chem. Theory Comput.*, 2018, **14**, 1501–1509.

4 L. Zhao and E. Neuscamman, *J. Chem. Theory Comput.*, 2019, **16**, 164–178.

5 J. Liu, Y. Zhang, P. Bao and Y. Yi, *J. Chem. Theory Comput.*, 2017, **13**, 843–851.

Filippo Monti added: From an experimental point of view, TD-DFT calculations are often used not only to get excitation energies but also oscillator strengths. Could the excited-state DFT method you described also give the transition dipole moment associated to a specific excitation?

Gianluca Levi answered: Transition dipole moments (TDMs) can be accessed from variationally optimized excited states in DFT. However, in the case the optimized excited state of interest is not orthogonal to the ground state, as can

happen when the two states have the same symmetry, the non-zero overlap between the ground and excited states could cause a spurious enhancement of the TDM.¹ In that case, it is recommended to benchmark the quality of the TDMs with respect to methods where the states are orthogonal, as done for example in ref. 1.

1 A. T. B. Gilbert, N. A. Besley and P. M. W. Gill, *J. Phys. Chem. A*, 2008, **112**, 13164–13171.

Filippo Monti said: As far as I could understand from the paper, you carried out frozen-geometry calculations (*i.e.*, single-point). In the perspective of performing excited-state geometry optimisations, do you think that analytical gradients could be easily implemented for your DO-MOM method?

Gianluca Levi replied: Analytical atomic forces are readily available for variationally optimized excited states in DFT using the routines for calculations of ground-state forces. This is because states corresponding to stationary points of the energy functional obey the Hellmann–Feynman theorem, which is used to obtain analytical forces. Therefore, optimization of geometries and molecular dynamics calculations can easily be performed with the excited-state DFT method that we present in the article. Several previous studies have used excited states obtained from variational DFT methods to perform excited-state geometry optimization and Born–Oppenheimer molecular dynamics in vacuum and in solution (see for example ref. 1–4). More recent studies extend the excited-state molecular dynamics simulations to the nonadiabatic regime.⁵

1 M. W. D. Hanson-Heine, M. W. George and N. A. Besley, *J. Chem. Phys.*, 2013, **138**, 064101.

2 G. Levi, M. Papai, N. E. Henriksen, A. O. Dohn and K. B. Møller, *J. Phys. Chem. C*, 2018, **122**, 7100–7119.

3 G. Levi, E. Biasin, A. O. Dohn and H. Jónsson, *Phys. Chem. Chem. Phys.*, 2020, **22**, 748–757.

4 E. A. Briggs, N. A. Besley and D. Robinson, *J. Phys. Chem. A*, 2013, **117**, 2644–2650.

5 M. Mališ and S. Lubert, *J. Chem. Theory Comput.*, 2020, **16**, 4071–4086.

Trygve Helgaker remarked: I am in fact surprised that you can converge these difficult saddle point optimizations using a first-order method. Have you considered implementing a second-order method? Modifications of the trust-region methods with level shift for step control (length and direction) may be very useful, in the same manner as for MCSCF energy minimizations. This may be valuable in difficult cases, depending less critically on the starting guess providing a faster local convergence. Second-order methods may perhaps also help you characterize the solutions.

Gianluca Levi answered: Indeed, a second-order optimization method would be more robust and would allow to characterize the obtained stationary solution. We will consider the implementation of a second-order method, which could be used as a backup when dealing with particularly difficult cases. On the other hand, it would still be desirable to use a first-order method on a regular basis because second-order methods are more computationally expensive. Regarding the cost of the direct optimization method presented in the article, the current (first-order) quasi-Newton implementation is competitive with SCF. This is achieved by employing limited-memory inverse Hessian updates, which scale linearly, and scaling and squaring algorithms for evaluating the matrix exponential, which have a cost that does not exceed matrix diagonalization.

Dumitru Sirbu opened discussion of the paper by Pina Romaniello: In your paper, are the 2D monolayers of h-BN and MoS₂ considered to be in a vacuum?

Pina Romaniello answered: Yes, in the sense that to describe 2D monolayers we use a 3D supercell with in principle an infinitely large separation in the direction perpendicular to the monolayer plane to avoid interlayer interaction. Therefore the resulting macroscopic dielectric function of the supercell contains the combined dielectric response of the 2D material and the vacuum. We extract the dielectric function of the 2D material by invoking the principle of equivalent capacitance as suggested by Laturia *et al.*¹ This leads to eqn (10) and (11) in our paper (DOI: 10.1039/d0fd00073f).

1 A. Laturia, M. L. Van de Put and W. G. Vandenberghe, *npj 2D Mater. Appl.*, 2018, 2, 6.

Dumitru Sirbu asked: I am wondering if this removes the effect of dielectric confinement. Working in vacuum will give you a huge dielectric contrast at the interface. Normally the materials, no matter how clean, are covered in a thin layer of dirt of organic nature that will have a dielectric constant larger than vacuum. So can your model be adapted to a semiconductor layer sandwiched between two insulator layers? You seem to be repeatably getting a blueshift as compared to the experimental data. This might give results closer to the experimental values and will also be of interest for quantum well structures.

Pina Romaniello replied: This is a good point. To model the two insulator layers we should introduce them in the supercell, and then employ the same procedure we use now. It would be interesting to check if this will improve the agreement with experiment. For example, using an analytical model dielectric function for 2D semiconductors, Trolle *et al.*¹ predicted substrate screening to strongly influence the exciton binding energy of supported 2D semiconductors.

1 M. L. Trolle, T. G. Pedersen and V. Vénier, *Sci. Rep.*, 2017, 7, 39844.

Jan Gerit Brandenburg enquired: In your work, there is quite some difference between the two materials h-BN and MoS₂. Can you anticipate if there are problems for treating materials like graphene, which might have non-trivial response due to its *K*-point conductance.

Pina Romaniello responded: In principle there is no problem, because our approach can deal also with metals and semimetals, due to a proper treatment of the intraband response. The question is if the Polarization functional will produce an exciton or not. This depends on the dielectric constant of the correspondent 3D bulk system, which is graphite. If the dielectric constant is large, we will not see any exciton, as in MoS₂.

Paola Gori-Giorgi asked: In eqn (5) of your paper (the GLLB approximation), you have the constant K_{xc} , which is derived from the uniform electron gas (exchange-only actually). Have you tried to see how your results change if you change the numerical value of this constant? Are they very sensitive to it? And is this value the one that gives the most sensible results or not?

Pina Romaniello answered: No, we have not tried to change this constant. We took its value from the uniform electron gas in which the GLLB exchange response potential becomes exact. Théo Jaffrelo Inizan and Evert Jan Baerends studied the influence of this parameter on the calculated band gaps: an enhanced K seems to improve the band gaps, at least for the systems they studied.¹ In our case, the GLLB functional tends to overestimate the band gaps with respect to experiment, except for monolayer MoS₂. A better description of the band gap would red-shift the theoretical spectrum towards the experimental one. Also the binding energies would improve, but the improvements are not significant from what we could observe in h-BN by using the GW band gap.

1 T. J. Inizan, Master thesis, Vrije Universiteit Amsterdam, 2018, unpublished.

Matthew Hodgson asked: My question regards the nice link you have found between the exchange–correlation (xc) derivative discontinuity of density functional theory (DFT), *i.e.* the difference between the fundamental gap and the Kohn–Sham gap, and the xc kernel of time-dependent DFT: is there some aspect of the kernel which introduces the electron–hole interaction which corresponds to the difference between the fundamental gap and the optical gap?

Pina Romaniello replied: In fact we separate the xc kernel into two parts, $f_{xc}^{(1)}$ and $f_{xc}^{(2)}$ as introduced by Sottile *et al.*;¹ only $f_{xc}^{(1)}$ is related to the difference between the fundamental gap and the Kohn–Sham gap, whereas $f_{xc}^{(2)}$ takes into account excitonic effects. The two parts together allow for the correct description of the optical onset. This separation is arbitrary, and one could also think of other possible separations which could facilitate the approximation of the full f_{xc} .

1 F. Sottile, V. Olevano and L. Reining, *Phys. Rev. Lett.*, 2003, **91**, 056402.

Matthew Hodgson said: The exchange–correlation (xc) derivative discontinuity (DD) corresponds to a charged excitation, whereas optical absorption corresponds to uncharged excitations. In this sense the link between the xc kernel and the DD is a little surprising. There is a uniform shift in the exact xc potential which is analogous to the DD that occurs for uncharged excitations – it is the difference between the optical gap and the Kohn–Sham gap.¹ Could there be a link between this shift and the xc kernel?

1 M. Levy, *Phys. Rev. A*, 1995, **52**, R4313.

Pina Romaniello answered: Yes, it might seem surprising the link between the derivative discontinuity (DD), which is related to charged excitations, and the xc kernel, which is related to neutral excitations, but, again, this link is between the DD and only a part of f_{xc} , and depends on the kind of separation we have chosen for the f_{xc} kernel. There is surely a link between the uniform shift in the exact xc potential in the case of neutral excitations and the xc kernel. Following a derivation similar to the one we proposed for the link between the DD and $f_{xc}^{(1)}$ one can find a link between the optical DD and the full f_{xc} .

Lucia Reining commented: Just to add to the explanation: in principle we can write the exchange–correlation kernel as a sum of two pieces, where we choose the

separation arbitrarily. Traditionally, however, we come from the Bethe–Salpeter equation for linear response, where we first calculate a product of two one-body Green's functions (an electron and a hole, in our approximations). The one-body Green's functions contain the information about electron addition and removal. Then we correct for the fact that we do not add or remove electrons, but we excite them, so this is expressed as an electron–hole interaction. The separation that is used for the exchange–correlation kernel follows this logic, so the first part simulates going from Kohn–Sham to electron addition and removal energies, and the second part simulates the electron–hole interaction.

Weitao Yang added: This is a very nice interpretation for the Bethe–Salpeter equation for linear response. I would state that the picture remains true for density functional theory. In recently developed density functional approximations with minimal delocalization error (ref. 1), the G_0 directly from the generalized Kohn–Sham calculations already has shown accuracy similar or better than the GW Green's function approaches, based on extensive molecular tests, for quasiparticle energies, or electron removal and addition energies, beyond just the HOMO and LUMO (ref. 2). The particle–hole interaction is what the TDDFT kernel is for.

- 1 C. Li, X. Zheng, N. Q. Su and W. Yang, Localized orbital scaling correction for systematic elimination of delocalization error in density functional approximations, *Natl. Sci. Rev.*, 2018, **5**, 203–215.
- 2 Y. Mei, C. Li, N. Q. Su and W. Yang, Approximating Quasiparticle and Excitation Energies from Ground State Generalized Kohn–Sham Calculations, *J. Phys. Chem. A*, 2019, **123**, 666–673.

Neepa Maitra asked: The epsilon in your functional uses the static RPA, but do you have some idea what you might get if you instead replaced it with the frequency-dependent RPA dielectric function?

Pina Romaniello responded: Inserting eqn (7) of the paper (DOI: 10.1039/d0fd00073f) with a frequency-dependent RPA dielectric function into eqn (6), one can verify that one obtains the RPA dielectric function plus 1, hence the imaginary part would be identical to that of the RPA. Therefore there are no excitons. If one would like to have a frequency-dependent kernel one can use the original Polarization functional,¹ which together with the static part reported in eqn (7) has also an imaginary part coming from the Vignale–Kohn functional. This imaginary part produces the Drude tail in the absorption of metals and gives a broadening of the excitonic peak in gapped materials. In this work we have used only the static part of the Polarisation functional, hence the broadening of the exciton is introduced by hand.

- 1 J. A. Berger, *Phys. Rev. Lett.*, 2015, **115**, 137402.

Alex J. W. Thom continued the discussion of the paper by Gianluca Levi: With regard to orthogonality, in small systems with symmetry, excited states are often of different symmetries to the ground state so must be orthogonal. In larger systems without symmetry, excited states are normally not orthogonal, and one has to diagonalise a Hamiltonian matrix in the basis of these states to get meaningful orthogonal states. In small systems, therefore, orthogonality will not be a problem, but in larger ones it will be more so.

Pierre-François Loos asked: You seem to have designed a robust algorithm which is able to find saddle points in multi-dimensional surfaces. Have you thought about applying this kind of scheme to coupled cluster theory in the future? One can use the same kind of algorithm by replacing the coefficients by the coupled cluster amplitudes. In such a way, one can potentially access excited states at the coupled cluster level.

Gianluca Levi replied: While the direct optimization (DO) MOM algorithm has been presented in the context of KS DFT, the method is of more general applicability: it can be used to find optimized non-Aufbau determinants of different energy functionals given the appropriate expression for the gradient (for KS functionals the gradient expressions are eqn (15)–(17) in the article (DOI: 10.1039/d0fd00064g)). DO-MOM could be used to converge HF non-Aufbau reference determinants for excited-state coupled-cluster calculations. As recently shown by Lee *et al.*,¹ excited-state coupled-cluster solutions can be targeted by employing a non-Aufbau determinant optimized at Hartree–Fock (HF) level as a reference state. In this way one avoids the use of linear response or equation-of-motion formalisms to access excited-state properties at coupled-cluster level.

1 J. Lee, D. W. Small and M. Head-Gordon, *J. Chem. Phys.*, 2019, **151**, 214103.

Donald G. Truhlar communicated: The article by Gianluca Levi (DOI: 10.1039/D0FD00064G) says that due to the lack of double excitations, the description of conical intersections between ground and excited states given by TDDFT is qualitatively incorrect – with a citation of the paper by Levine *et al.*,¹ which points out that TDDFT gives an incorrect topology at conical intersections between the ground state and the first excited state due to lack of double excitations from the ground state. We proposed a dual-level procedure that gives the correct topology at conical intersections without explicit consideration of double excited states; in this procedure, one calculates the orbitals with one functional and the response with another.² I have been asked how one can justify this formally. We proposed the dual-functional procedure as a practical procedure, not as a formal one. However, a path to a formal justification can proceed as follows. The problem is that the ground and first excited states (call them S_0 and S_1) should interact, causing the intersection to avoid in $F - 2$ dimensions, where F is the number of internal coordinates ($3N - 6$). However, linear-response TDDFT does not have any coupling between S_0 and S_1 if S_1 is a single excitation; one could consider this to be a generalization of Brillouin's theorem to TDDFT. The reason that adding double excitations gives the correct topology at conical intersections is that they couple to both S_0 and S_1 . If one uses functional $F1$ to obtain the ground-state orbitals, but $F2$ to calculate the response, one does get a coupling between S_0 and S_1 because the singly excited states of functional $F1$ look like a mixture of singly excited and double excited states from the point of view of functional $F2$ (and *vice versa*). Therefore, this is a way to incorporate the topological effect of double excitations without treating them explicitly. Furthermore, we found that vertical excitation energies of seven organic molecules were equally as accurate as with single-functional calculations, and in the region of a conical intersection in NH_3 , we found that not only did we obtain the correct topology but also we obtained potential energy contours in good agreement with MS-CASPT2.³

- 1 B. G. Levine, C. Ko, J. Quenneville and T. J. Martínez, *Mol. Phys.*, 2006, **104**, 1039–1051.
- 2 Y. Shu, K. A. Parker and D. G. Truhlar, Dual-Functional Tamm–Dancoff Approximation: A Convenient Density Functional Method that Correctly Describes S_1/S_0 Conical Intersections, *J. Phys. Chem. Lett.*, 2017, **8**, 2107–2112, DOI: 10.1021/acs.jpclett.7b00594.
- 3 Y. Shu, K. A. Parker and D. G. Truhlar, Dual-Functional Tamm–Dancoff Approximation with Self-Interaction-Free Orbitals: Vertical Excitation Energies and Potential Energy Surfaces Near an Intersection Seam, *J. Phys. Chem. A*, 2017, **121**, 9728–9736, DOI: 10.1021/acs.jpca.7b11400.

Weitao Yang opened discussion of the concluding remarks lecture by Andreas Savin: I would like to ask you on your thoughts for the future of DFT? Can one meet the demand on accuracy and computational cost, given challenges from theories like tensor networks and other accurate wavefunction approaches that can calculate things accurately on one hand, and machine learning with great efficiency on the other hand? DFT is in the middle. You want a theory that can adapt itself when it is applied in different domains – a “universal” theory, scale-wise.

Andreas Savin responded: My feeling is that for the large systems that are of much interest now, a brute force solution of the Schrödinger equation is too slow. Usually, we assume that experience and intuition tells us how to separate, see, *e.g.*, “divide-and-conquer”.¹ Next, one can use a multi-scale approach, having different methods for different regions, see, *e.g.*, the “domain separation” of ref. 2.

As systems get larger, I am afraid that intuition will not be able to lead us to find the best way to decide about the division of space, and the method to use for it. For example, W. Kohn speaks about “nearsightedness of electronic matter”,³ but lists cases when it does not apply. I believe that we need methods that define the domains, and are chosen for them such that they are as fast and as accurate as needed. For example, in the “inter-exchange theory” of H. Nakatsuji and H. Nakashima,⁴ electrons belonging to distant atoms in a molecule, only partial antisymmetrization or even no antisymmetrization is effectuated.

I think that this can be generalized. Not our intuition, but there should be some automatic/parametric recognition of what method can be applied. When and how do we switch from one level to another? When is a region apt to be described by machine learning? When is the semi-classical regime valid? Classical DFT is very fast. When are we allowed to use it without being penalized?

The automatic, physical selection will also guide us in understanding the results. Producing such an approach will take a lot of time, and a lot of manpower, but I think it is worth pursuing it.

- 1 W. Yang and T.-S. Lee, *J. Chem. Phys.*, 1995, **103**, 5674.
- 2 M. A. Mosquera, L. O. Jones, C. H. Borca, M. A. Ratner and G. C. Schatz, *J. Phys. Chem. A*, 2020, **124**, 5954.
- 3 E. Prodan and W. Kohn, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 11635.
- 4 H. Nakatsuji and H. Nakashima, *J. Chem. Phys.*, 2015, **142**, 194101.

Kieron Burke remarked: Related to the previous comments, I have been using machine learning (ML) for a while now and I am preparing a lecture course about it to go online. I think it does represent a paradigm shift and it's based on the acceleration in the amount of computational power available and extreme collection of data, mostly over the past 10 years. It will change things, producing a new way of doing calculations in the same way computational physics and chemistry came online 30–40 years ago. It is a way of finding patterns. It came up

in Jack Wetherell's talk (DOI: 10.1039/d0fd00061b). Neural nets can be extremely wasteful, lots of parameters, huge calculation, don't care about overfitting, but allows you with deep neural networks to find underlying patterns. The Schrödinger equation is a difficult differential equation to solve, with many different scales depending on the situation. What ML can do is set it up to find density matrices very fast, once trained, for lots of situations. Much faster than you can do an accurate solution directly. Once done, can use it if needed for a realistic calculation, *e.g.* interatomic potentials, allowing you to get to places faster/quicker than before. Now ML is working to get simpler patterns back out (autoencoders). Then you can work on your intuition again, to find models that capture what this pattern recognition has done automatically for you.

Andreas Savin replied: ML can be very useful. I think that what we need are techniques where we recognize that simple, fast methods can be used safely. We would need adjustable parameters that tell us when we are in this regime. For example, in your approach, to know when we are in a semi-classical regime. Or, can we switch between quantum and classical density functional theory? I mean something in the spirit of ref. 1, relax the antisymmetry rule.

1 H. Nakatsuji and H. Nakashima, *J. Chem. Phys.*, 2015, **142**, 194101.

Jack Wetherell added: I agree there should be a focus to use ML as a tool towards human learning, rather than to be seen as a replacement. As you suggest this transfer of knowledge from human to machine can be done in many ways. Human → machine in the form of integrating with something like the connector approach, what functional is the best to learn, how should we present the data to be learned *etc.* And from the machine → human by providing many more examples to study than was possible before, and by using deep learning tools like autoencoders to “unpack” the black box.

Kieron Burke said: I think there's still lots of room for human theory and understanding within DFT. Maybe more than ever before. There's a huge variety of situations we're applying it to, and we find out how useful it is in so many cases. So many of those applications are in different fields, each with different ways of thinking. It can be quite difficult to cross some of these disciplinary boundaries. It can also be hard to develop theory that will lead to practical applications. We have so many numerical successes so far. I think we should spend a little more time on fundamental theory, especially cross-disciplinary areas. All molecules, liquids, and solids, satisfy the same Schrödinger equation and fundamental theorems of DFT!

Andreas Savin added: By machine learning a large mass of results will be produced easily. What is the danger of producing wrong results, and that these are accepted just by the weight of its mass? For example, there was a time when Density Functional Theory was generally considered “bad”, while a decade later I heard people say: “Since I started using Density Functional Theory my life has changed. I can do so many applications using it.”

Kieron Burke replied: Indeed, a large number of papers are being produced. There are multiple dangers:

(a) It can be very hard to check a calculation because methods are new and not standardized.

(b) Even if a specific calculation is reproducible, the tricky part is always the generalizability of the method. Many proofs-of-principle run into difficulties when applied to broader, more useful problems. With machine learning, it is very difficult to know if a specific method will work for a given class of problems in advance.

(c) Even when things work and are useful, it can be very difficult to understand why.

And, yes, this list applies exactly to DFT about 20 years ago (and somewhat, today). But even in all this noise, there is good stuff being developed, and it tends to win out in the long term. I cite your own pioneering work on range separation as a perfect example.¹ I expect 10 years from now, many folks will be using machine learning, but many at our level of “seniority” will not enjoy it.

1 J. Toulouse, F. Colonna and A. Savin, *Phys. Rev. A*, 2004, **70**, 062505.

Kieron Burke continued the general discussion: If you analyse it carefully, the only reason we know our modern density functionals work as well as they do and fail where they do is because of traditional quantum chemistry. The G2 dataset already existed, already the comparison had been made between theory and experiment (full *ab initio*) that was the crucial point that allowed DFT to take off by showing it got reasonable results with that data set. This can be extrapolated all the way to materials science, where highly accurate experimental numbers for thermochemistry are very difficult to generate. Really accurate quantum chemical calculations tell us whether our functionals are any good, and also tell us how bad they are when they fail.

Donald G. Truhlar answered: The G2 training set consisted entirely of experimental data, not results from traditional quantum chemistry. The G2 method does not correspond to full configuration interaction; it is a composite method based on MP4 and QCISD(T) with basis set extrapolation and a two-parameter empirical high-level correction. Accurate high-level calculations, when possible, are very useful, but, although the situation is improving recently, when one considers the kind of complex systems of most interest for practical work, so far there is more testing of DFT against experiment than against high-level calculations. It is often hard to estimate the accuracy of even high-level quantum mechanical calculations for complex systems, and it is sometimes hard to estimate the reliability of experimental data. Furthermore, it often requires additional analysis and approximations to extract the most useful data from experiments; for example, molecular spectra often consist of broad peaks and extracting vertical or adiabatic excitation energies involves further assumptions. The selection of data to make useful training sets or testing sets for DFT for either molecules or materials is often very difficult because the reliability of the available data is often uncertain, but I agree that it is very important.

Andreas Savin commented: In the time when Density Functional Theory was generally considered “bad”, I was told: “Density Functional people publish results only when they are good”. So I went around suggesting people to do systematic tests. Dennis Salahub who by that time had published already many interesting

applications answered me: "I know that Density Functional Theory is good. I don't need to convince other people."

Eduardo Maurina Morais continued discussion of the concluding remarks lecture by Andreas Savin: General comment from an experimentalist who recently started working with DFT. I noticed that there is a huge gap between what is discussed in these kinds of meetings and the things that in practice matter for experimentalists. How do you perceive this gap? For instance, the number of people that have been using DFT and similar techniques is increasing, and at some point in the future, it might become akin to NMR in the sense that a lot of people use NMR, but not many understand how it fundamentally works. I always do my best to understand the limitations of my knowledge and my DFT results, but a lot of people don't, so they end up over-interpreting their results. What do you think about this? Will DFT one day be used in a similar way to NMR?

Andreas Savin replied: True, when Density Functional Theory started to become popular, terrible results were published: people were putting bars in the tables when they did not know how to converge their calculations.

Mel Levy told me a story from these times. He met a colleague who told him proudly that he was doing BLYP calculations. Mel said: "So, you are now doing DFT." The colleague answered: "No! I am using BLYP, not DFT."

However, this improved much with time. I have seen experimentalists knowing very well what they were doing. Those who want learn to know what they are doing, like you, they can do it. In this respect, I have an optimistic attitude. I know someone who said during his postdoc: "I use the computers like others use an apparatus." He got so much interested in understanding what he did, that he later got a chair in theoretical chemistry, and is now developing methods.

The question is what do we want to know, to understand, and how deep do we have to get into knowing what the tools we use really are. I use a computer every day, but I never open it to see what is inside it. I use a fork and if I use it properly, it's fine. Not if I scratch my face with it. I think that experimentalists are now using DFT well. They don't use the fork for eating soup.

Eduardo Maurina Morais commented: Regarding what experimentalists want. In my case, at this moment DFT is mostly a tool. So if at some point I could use machine learning in the same way that I use DFT, as long as it gives me good predictions, that's all I want as an experimentalist, *e.g.* spectrum of compounds of interest or other properties. I think that experimentalists don't necessarily care about the method, they just care about the results.

Eduardo Maurina Morais added: The problem is that there is a long way between learning about the existence of the fork and knowing how to properly use it. I think that computational chemistry is a bit more complicated than a fork!

Andrew M. Teale said: A remark on the idea that density-functional methods are becoming as widely used by non-experts as tools like NMR in chemistry. All theories should aspire to this status. They should be as accurate and as simple to use as possible. By definition, of course, their practical realisations are

approximate and so will necessarily fail sometimes and can be applied to situations beyond their capabilities. It is, however, a remarkable testament to density-functional theory that simple practical methods based upon it are used so often with success and are so widespread. Of course, to address the areas/systems that are still out of reach requires further development of new practical approaches, ideas, and technical expertise of the type displayed by speakers at this discussion.

Nikitas Gidopoulos remarked: To put my brief comment in context, Mr Eduardo Maurina Morais (experimentalist) made a comment that there is a gap between what was discussed in the meeting (progress in theoretical and methodological development) and what interests people who are not experts but still use DFT without understanding well the results. He said that he was interested in DFT as a tool mainly but still wanted to develop an understanding of the results of a calculation and its limitations and asked how he could do this. My comment is that it would be helpful for everybody in the scientific community if DFT users who are not experts, like Eduardo, developed physical intuition and understanding in their own field, which would then allow them to assess critically the results of DFT calculations independently, from their own point of view.

Andreas Savin added: What I think is important for the interaction with experimentalists is not to compute the numbers they measure, but help them to answer the questions they asked themselves, motivating the measurement.

Nikitas Gidopoulos communicated: If I may, I would like to ask the following question to all participants. On the issue of fundamental theory that Prof. Burke raised in his general question, I get the feeling from reading the literature in recent years that standard KS theory with a local multiplicative potential may be running out of steam and that generalised KS theory is the future, because the effective potential in GKS theory has greater flexibility. In short the view might be described as “what a local potential can do a nonlocal (or frequency dependent) potential can do better”. I have not made up my mind on this issue yet and I admit I like the minimalism and elegance of a local multiplicative potential in standard KS theory. What is the view of the participants?

Weitao Yang communicated in reply: This is a very interesting question. First, in any DFT calculations, the quality of the prediction, in comparison with experiments, is the direct reflection of the density functional approximation (DFA). Second, with a chosen DFA, the methods of minimization of the functionals, whether using KS(OEP) with a local potential or GKS with a nonlocal potential, play a minor role in calculating the total energies. However, one needs to be careful with other properties, such as the orbital energies. For functionals explicitly dependent on the non-interacting density matrix, such hybrid functionals, the GKS HOMO/LUMO energy has been shown to be equal to the chemical potential of electron removal/addition, and hence the DFA prediction for $-I/-A$ (ref. 1). In the introductory lecture (DOI: 10.1039/D0FD00102C), we have a section addressing this (section 2.4 “The comparison of GKS with OEP for density matrix functionals”). No such connection can be derived with the HOMO/LUMO orbital energies in the OEP calculations of the same DFA, although the HOMO energy in OEP calculations can be related to $-I$, if the correct asymptotic behavior of the

potential is imposed. One can still use OEP information to calculate the chemical potentials associated with the DFA – they are just not simply the eigenvalues of OEP (see ref. 1). And the total energy is invariant with respect to a constant shift of the OEP potential and hence eigenvalues.

1 A. J. Cohen, P. Mori-Sánchez and W. Yang, Fractional charge perspective on the band gap in density-functional theory, *Phys. Rev. B*, 2008, 77, 115123.

Lucia Reining communicated in reply: I think we have to distinguish two questions, and although we all know this, I would like to insist because in discussions sometimes things are mixed up.

First, suppose we want to calculate a given observable directly from an auxiliary system with an effective potential. According to the observable, this potential must then be more or less flexible in principle. For example, to get the density it is enough in principle to have a local and static potential (the Kohn–Sham potential). But if you want the interacting density matrix, at zero temperature even a static non-local potential is in general not enough. If you want the HOMO and the LUMO energies, you can imagine having a local potential that is different for the two states, for example, and so on.

If we want a spectral function, in general we move to a non-local and frequency-dependent self-energy. Matteo Gatti's paper (DOI: 10.1039/d0fd00067a) works all this out. So, if we insist on calculating observables other than the density and related quantities directly as the solution of an auxiliary system, the answer is dictated by the nature of the observable.

Second, even if you can get your observable in principle from a static local potential, it may be difficult to find good approximations in practice. So for example, we may find it difficult to get a good description of the density of quite localized electrons from a static and local Kohn–Sham potential, whereas the result is improved using non-local hybrid potentials. We could maybe also use local OEP potentials, but they turn out to be complicated: a local potential does not necessarily have to be simple, so its “minimalism and elegance” may hide some really horrible things inside. In this case we can ask the question: how much “outside” complexity do we have to admit in order to get rid of these “inside” complications? For example, the Fock exchange is non-local but looks much simpler than the EXX OEP local potential. Another example, we have approximations to the non-local and frequency-dependent self-energy, for example from many-body perturbation theory, that are relatively simple and can also be used to calculate charge densities. So in this second case the question is: for a static and local potential, meant to reproduce the density, do we still have hope to approximate the “inside” in an increasingly precise, yet simple, way? My personal answer is yes, with the hope that something like our connector approach, which takes pre-calculated pieces from elsewhere, will work right as the LDA hides much complexity in the pre-calculated HEG potential. My bet is that if we cannot make such a trick work, we will simply slowly move towards self-energies, by adding non-locality, or frequency dependence, and eventually both.

Conflicts of interest

There are no conflicts to declare.