Challenges for large scale simulation: general discussion

Jan Gerit Brandenburg, [ID] Kieron Burke, Bartolomeo Civalleri, [ID]
Daniel J. Cole, [ID] Gábor Csányi, Grégoire David, Nikitas I. Gidopoulos,
Duncan Gowland, Trygve Helgaker, Michael F. Herbst, [ID]
Ben Hourahine, Tom J. P. Irons, [ID] Christoph R. Jacob, [ID]
Pierre-François Loos, Nisha Mehta, Manasi R. Mulay,
Johannes Neugebauer, Katarzyna Pernal, [ID] Aurora Pribram-Jones,
Pina Romaniello, Matthew R. Ryder, [ID] Andreas Savin, [ID] Dumitru Sirbu,
Chris-Kriton Skylaris, Donald G. Truhlar, [ID] Jack Wetherell
and Weitao Yang

Grégoire David opened discussion of the paper by Johannes Neugebauer: In your paper and the presentation, you say the Broken-Symmetry determinant is a superposition of the singlet and the triplet with equal contributions, and we have an expected $\langle S^2 \rangle = 1$.

In my opinion, it is not true since there is no reason to expect such an exact equal mixing of both spin states. The $\langle S^2 \rangle$ of the BS determinant is expected to be close to 1 in the context the Heisenberg–Dirac–van Vleck Hamiltonian since this model Hamiltonian stands for well-localised magnetic centres and a value close to 1 should represent this situation. However, nothing indicates that $\langle S^2 \rangle$ must be equal to 1.

Johannes Neugebauer replied: We agree. Our statement in the form mentioned in the paper only refers to certain idealized cases, but does not hold in general. It does hold if we assume that the orbitals in singlet and triplet states are the same, and that all but the two singly occupied orbitals are doubly occupied and not spin-polarized. We will include a "Note added in proof" to point the readers' attention to this fact and its clarification in this discussion.

If we consider, *e.g.*, the verdazyl radical dimers studied in our paper, we may approach such a near-ideal situation with an S^2 expectation value close to 1 for large distances, where the orbitals of the subsystems are essentially unaffected by the presence of the respective other subsystem. That means that triplet and BS-state spin densities are essentially obtained from the same set of (spatial) orbitals. If, in addition, spin polarization in the other orbitals is negligible, we would come close to this idealized value.

Grégoire David asked: Your use of the Yamaguchi formula with the density-based evaluation of the expectation of S^2 is based on the fact that the over-estimation in both the triplet and the BS determinant is of the same magnitude compared to the orbital-based evaluation. Do you know or did you check if it is the case for other combinations of high-spin state/broken-symmetry determinant? For example systems with more unpaired electrons like the chromium cluster presented in your paper?

Johannes Neugebauer responded: We have also tested the difference in the S^2 expectation value between the orbital and density based evaluation for the chromium cluster. There we found differences of 0.29 to 0.33 between the two formalisms for all investigated spin states. Therefore, the magnitude of the overestimation of the density based formalism is very similar (difference: 0.04) for the investigated spin states.

Additionally, we would like to point out that, in view of the large discrepancies observed for the energy differences, the overestimation of the S^2 expectation values will most probably not be the dominant error source in sDFT-type calculations as presented in our paper.

Christoph Jacob enquired: You are using the Wang–Becke–Smith (WBS) XLDA functional[1] for calculating the expectation value of S^2 from the spin density. Such an S^2-functional would also be very interesting in the context of our paper (DOI: 10.1039/D0FD00060D), as it would allow to enforce a specific value of $\langle S^2 \rangle$, which would be one way of setting up a spin-state specific functional. As you discuss in your paper, with the WBS-XLDA functional negative regions of the spin density lead to deviations from the ideal $\langle S^2 \rangle$ value. However, it is known that for pure S^2 eigenstates the spin density of the true, interacting system will have negative regions.[2,3] Therefore, I think that the WBS-XLDA functional does not approximate $\langle S^2 \rangle$ of the interacting system but $\langle S^2 \rangle$ of the Kohn–Sham determinant. I think this is related to the neglect of the cumulant term in your eqn (17). Do you have any thoughts on the physical meaning of $\langle S^2 \rangle$ provided by the WBS-XLDA functional?

1 J. Wang, A. D. Becke and V. H. Smith Jr., *J. Chem. Phys.*, 1995, **102**, 3477–3480.
2 J. A. Pople, P. M. W. Gill and N. C. Handy, *Int. J. Quantum Chem.*, 1995, **56**, 303–305.
3 C. R. Jacob and M. Reiher, *Int. J. Quantum Chem.*, 2012, **112**, 3661–3684.

Johannes Neugebauer replied: We agree that for pure S^2 eigenstates the spin density of the true, interacting system can have negative regions and that the evaluation of S^2 expectation values with the WBS-XLDA functional would lead to deviations from the ideal values. However, here we use the WBS-XLDA approach only as a diagnostic tool and are merely interested in the differences of S^2 expectation values between the high-spin and broken-symmetry state. We obtained differences for S^2 expectation values which are of the same magnitude as those obtained with the (standard) orbital-based approach. Therefore, we think that the use of the WBS-XLDA approach as a diagnostic tool is justified for our purpose.

Andreas Savin remarked: I find the idea to deal with subsystems good. However, with broken symmetry we have a problem. Let me give two examples.

(1) The O_2 molecule is in a triplet state. If you dissociate it, each atom is in a triplet state, say $M_S = 1$. If you use broken symmetry, each atom has $M_S = 1$, or $M_S = -1$. The total M_S is either 2, or 0, or -2, never 1. (2) The CO_2 molecule is in a singlet state. Each of the atoms is in a triplet state. Three atoms with $M_S = 1$, or $M_S = -1$ never give $M_S = 0$.

You either get the right spin-density of the molecule and a too high energy, or the right energy and wrong spin-density. You can get out of this dilemma, *e.g.* by giving up having orbitals being eigenfunctions of S_z, as K. Yamaguchi *et al.*[1] do, or using on-top pair densities.

1 K. Yamaguchi and T. Fueno, *Chem. Phys.*, 1977, **19**, 35–42.

Johannes Neugebauer answered: Our starting point is the broken-symmetry approach in conjunction with the Heisenberg–Dirac–van Vleck (HDVV) Hamiltonian. This effective Hamiltonian is based on the idea of local spin sites or magnetic subsystems. This concept is not without problems (consider, *e.g.*, the formation of covalent bonds between two radical centers). Nevertheless, we are mainly interested in the magnetic interaction of organic radical pairs, where each monomer can be regarded as one well-defined spin site, even though the spin density can be delocalized over the entire molecule. The main point we wanted to make is that if you accept this formal framework, sDFT offers an easy way to access the broken-symmetry state.

Tom J. P. Irons commented: You described the calculation of magnetic exchange coupling constants using the full non-interacting kinetic energy of the super-system, obtained by inversion of the Kohn–Sham equations to construct the potential and Kohn–Sham orbitals pertaining to the sum of subsystem densities. In calculating these exchange coupling constants, did you evaluate $\langle S^2 \rangle$ from these super-system orbitals using the standard UHF expression? Whilst this is not an exact expression for $\langle S^2 \rangle$ in Kohn–Sham theory, for the non-interacting system with the same density as the interacting system and constructed with, for example, the Wu–Yang method it should be exact.

Johannes Neugebauer replied: We have tested this for one case, and obtained a difference of S^2 expectation values between the triplet and BS state of 1.018 with the density based formalism and 1.069 with the orbital based formalism using the reconstructed orbitals. The difference between these two approaches is thus minor (0.05) in this first test case.

Donald Truhlar said: This comment concerns the discussion of $\langle S^2 \rangle$ and its use in the Yamaguchi formula. Consider the case where a state may either be an open-shell singlet (*i.e.*, a diradical) or a triplet and where there are two singly occupied orbitals with all the rest of the orbitals being non-spin-polarized and doubly occupied. A single Slater determinant with this occupancy and $M_S = 0$ may give an expectation value of 1.0 (halfway between the singlet value of 0 and the triplet value of 2). In such a case, with currently available functionals, it might be reasonable to assume that the energy calculated for this determinant is the average of the singlet and triplet energies if the orbitals are the same in the singlet and triplet. The assumption that the orbitals are the same is the key assumption

behind the formula, whereas some of the discussion here centered instead on whether the two singly occupied electrons are located on symmetrically related centers, which is a separate issue. A complication comes in, however, if we consider the unknown exact functional. The unknown exact functional should give the ground-state energy, not an energy part way between the singlet and triplet. The accuracy of the Yamaguchi formula has been tested, and it does not uniformly improve the results compared to assuming that one has calculated the ground-state energy, although for many functionals there is a small improvement in the average absolute error of the spin splitting.[1]

1 S. Luo, Density Functional Theory of Open-Shell Systems. The 3d-Series Transition Metal Atoms and Their Cations, *J. Chem. Theory Comput.*, 2014, **10**, 102–121.

Johannes Neugebauer responded: We agree. Please also see our answer to the previous question concerning the S^2 expectation value of the BS state.

Nikitas Gidopoulos opened discussion of the paper by Michael Herbst: Could you please expand a little on how error estimation can help determine whether we have reached convergence or self-consistency in a calculation?

Michael Herbst replied: Thanks. This is an interesting question. In order to answer it, let me first approach the subject from a slightly different angle. The aim of our efforts is not to change the way self-consistent field (SCF) convergence is determined, much rather we want to provide a mathematically rigorous answer to the question: "What should I do to get closer to the exact Kohn–Sham result?" Let me elaborate on this:

If you are thinking about a standard SCF scheme, it is typically composed of two coupled iterations, one for iterating the density and one inner iteration for computing the Kohn–Sham orbitals for each density. It is common practice to not use the same convergence threshold for the inner iteration, but to tighten it more and more as the SCF converges. So in the beginning, when the density is still far off from self-consistency, then we might only converge to a tolerance of 10^{-2}, say. In such a case employing double precision in the eigensolver is most likely an overkill and one might as well go with single precision as the cheaper option. This, however, may change at the next SCF iteration step when the SCF might be better converged, such that a respectively lower eigensolver tolerance is required. Thus one ought to better select double precision. But of course it could also still be OK to not increase the eigensolver tolerance or to not switch to higher precision and thus exploit the faster eigensolver/single-precision arithmetic for a few more steps.

So the questions are: How do I choose the tolerance of the inner eigensolver in line with the error of the self-consistency? How do I choose the floating point type in line with the eigensolver tolerance? Another question one might ask equally well is: When do I have to increase the basis size? Because, clearly, the error in the density (*versus* the exact Kohn–Sham density at infinite basis size) will at some point during the SCF no longer be dominated by the lack of self-consistency in the density, but much rather by the finite basis size I have chosen for the calculation.

What *a posteriori* error analysis now does is that it gives us mathematically rigorous upper bounds for the error caused from each of the knobs I mentioned

(convergence tolerance, basis size, floating-point precision). It therefore allows to determine at each step of the SCF which error term makes up the largest contribution. Thus it further allows to determine – rigorously and not heuristically – what knob one has to turn and by how much one has to turn it to get closer to the exact Kohn–Sham answer in the next SCF step.

Now I return to your original question. If suitable *a posteriori* error estimates for all sources of error in Kohn–Sham DFT were known, the recipe to reach a targeted accuracy in the self-consistent density would become as follows: Keep iterating the SCF and at each step follow the recommendation of the error estimates until all sources of error are below my desired accuracy. Then stop and call the calculation converged. This criterion includes the "traditional" notion of self-consistency of the density, but it actually goes beyond it, because it also ensures, for example, that the error due to the basis is smaller than the self-consistency error.

Chris-Kriton Skylaris remarked: Whether you are doing traditional SCF diagonalisation, or using some kind of energy minimisation, say conjugate gradient, you will see different iterations, some kind of convergence, you will see energy converging to whatever number of decimals, what I'm not sure is what we take as the converged result to a certain precision, and if it really is converged to the precision we observe, or whether this precision is an artefact of, say, numerical errors that accumulate. For example, if you run in parallel, you will see some noise when you change the number of cores, parallelisation strategy, even though this might be much less than the chemical part of the calculation, it's still there. So even if you forget about convergence in the Brillouin zone, and use only the gamma point, what do you take as converged precision *vs.* artefact?

Michael Herbst responded: Yes I can only agree. That is indeed a good question. In practice of course you can do some things to check your result, *e.g.* evaluate a residual. But when you are doing that, you are again making at least an arithmetic error. Certainly one can imagine cases where the numerical error is larger than the residual value you get. For example if your numerical error is 10^{-3} and the residual norm is 10^{-14}, than of course your calculation may not be converged to 1×10^{-14}. In such cases in fact you can't really say anything about the accuracy of your result beyond the numerical error of 1×10^{-3}. In fact this is also one of the motivations for our work, namely to be able to really pinpoint error contributions and provide a guaranteed error, *i.e.* an error to which one can guarantee the calculation has converged to.

Chris-Kriton Skylaris asked: Is it possible in your Julia code to raise the precision arbitrarily high?

Michael Herbst answered: Yes, since any floating point type (including arbitrary precision types) can be used in the code.

Andreas Savin commented: I have a question related to nomenclature. You use the term *discretisation error* for the error produced by using a basis set. But there are two sources of errors in it. One is the range, *e.g.*, the cutoff in a plane wave basis. And the other is what I would call *discretisation*; for plane waves, it is the

choice of k-points. Don't you think it would be useful to distinguish between the two?

Michael Herbst answered: I completely agree with you that these are indeed orthogonal sources of error, which is the fact why we have denoted them as two steps in our list of errors on the first page of our manuscript (DOI: 10.1039/d0fd00048e).

Both steps combined, however, are needed to bring the infinite-dimensional continuous Kohn–Sham problem to the finite-dimensional discrete problem, which is eventually solved numerically. If one just employs a finite number of k-points and keeps the infinite plane-wave cutoff, then the part of the Kohn–Sham operator corresponding to a particular k-point is still infinite-dimensional, thus not discrete. So in agreement with the term *discretisation* referring to the full process continuous → discrete, we have termed the combined error of both steps the *discretisation error*.

Pierre-François Loos remarked: I don't know if you know it but there is a beautiful paper by Peter Gill and collaborators from 2010 named "Density functional triple jumping",[1] where they found there are 3 main "directions" in which one can improve a DFT calculation: basis set, quadrature grid, and functional.

For the basis set, one can perform a dual basis-type calculation in order to estimate the basis set incompleteness error, and one can perform a dual grid calculation in order to estimate the numerical integration error. Finally, one can play the same game with the functionals by performing for example a LDA calculation and then a (cheap) calculation with a better functional in order to estimate the error associated with the functional.

For each direction, this scheme gives you a rough idea of the error, and by combining them one gets a pretty good idea of how good the calculation is. Is your protocol/code able to do this? Maybe in fewer directions?

1 J. Deng, A. T. B. Gilbert and P. M. W. Gill, *Phys. Chem. Chem. Phys.*, 2010, **12**, 10759–10765.

Michael Herbst replied: The paper you mention deals with molecular DFT calculations based on Gaussian basis sets. In contrast to Gaussian-based DFT calculations, the grid one typically uses for the calculation of the XC term is directly related to the plane-wave grid (*i.e.* the basis) in the case of plane-wave DFT. In this sense the two directions due to the quadrature grid and the basis set fall on top of each other in our case. Regarding the functional direction of Gill *et al.* you mention, this is not yet considered by our work as we have so far excluded the treatment of the XC term. But with respect to the other errors we have worked on, namely the discretisation error, the algorithm error and the arithmetic error, they are similarly obtained as three independent directions. And since we have three independent terms for all these errors, we can also point in the direction of which knob to turn to make the final result more accurate.

Christoph Jacob commented: This is very nice work, I think looking at uncertainties and putting bounds on the errors of calculations is a very important topic. In your paper, you are looking at the absolute errors for the individual

eigenvalues. In many cases it is more important to consider the error for energy differences (*e.g.* band gaps), and I would expect that the individual errors are correlated such that the error is actually much smaller for the energy differences. In many cases quantum chemistry relies on such error cancellations. Is it possible to consider this in your framework?

Michael Herbst answered: You're right that of course looking at individual eigenvalues has not much use in chemistry and it is more the differences of energies that matter. Within our current framework capturing such error cancellation effects is not considered explicitly so far.

For estimating the gap errors there are two approaches which come to my mind based on our current work. The simplest idea is to get a naive upper bound in the gap error by adding the errors in the individual eigenvalues. Alternatively one could use the lower bounds we obtain for the eigenvalues not only to obtain a lower bound on the gap (as we do now) but also an upper bound for it and in this way estimate the error in the gap. Since this latter approach involves some operator norms of the potential, which are currently estimated rather roughly, I do not think this will presently give an improved result over the naive bound in many cases.

Christoph Jacob added: In the context of error estimates for energy differences I was also thinking of the BEEF functionals of Nørskov, Jacobsen and co-workers,[1,2] which allow for estimating the error of the approximate exchange–correlation functional. This can be used to estimate errors in relative energies (*e.g.* reaction energies), which are much smaller than the errors of the total energies of the reactants. However, in contrast to your paper in which you derive provable error bounds they only obtain statistical error estimates.

1 J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Nørskov, J. P. Sethna and K. W. Jacobsen, *Phys. Rev. Lett.*, 2005, **95**, 216401.
2 J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard and K. W. Jacobsen, *Phys. Rev. B*, 2012, **85**, 235149.

Michael Herbst responded: Indeed it is a very interesting idea to take some of the error into the functional. I am not yet completely familiar with their work, but it seems their efforts could be interesting to estimate the model error term we have so far neglected.

Christoph Jacob asked: Is it possible to extend your ideas for obtaining bounds for the basis set error from periodic calculations using plane wave basis sets to molecular calculations using Gaussian basis sets? I guess this will not be that easy, as the basis set is not determined by a single parameter. Do you have any ideas in this direction?

Michael Herbst answered: Indeed Gaussians are a bit more tricky. The main reason is that our error estimates rely on the properties of the potential outside of the chosen basis set X. Unlike for plane waves, where this amounts to ana-lysing the decay of the Fourier components of the potential for wave vectors larger than the cutoff, the structure of the complement space $X_{\perp}$ is more involved for Gaussians. In particular obtaining bounds on operator norms such

as $||P_{X\perp} H P_{X\perp}||_{op}$, *i.e.* the terms which couple the unknown part of the wave-function outside the basis *via* the Hamiltonian again to the contributions not covered by the basis, is an open research problem. This implies that obtaining lower bounds for the gap is currently out of reach, which in turn implies that the superior Kato–Temple bounds for the error in the eigenvalues cannot be employed for Gaussian basis sets.

It still turns out to be feasible to estimate the residual norm $||H\varphi - \varepsilon\varphi||$ of an eigenpair (ε, φ), however, because one can take a different route: if V is a sum of Gaussians (like in the pseudopotential case we consider) and φ is similarly a linear combination of Gaussians, so is the residual $H\varphi - \varepsilon\varphi$. As a result the residual norm can be obtained as a sum of analytic integrals. At least the Bauer–Fike-type bounds are therefore computable for Gaussian basis sets as well.

Ben Hourahine said: For systems with non-integer filling, such as a Fermi function at finite temperatures, there can be numerical noise in the resulting occupation numbers (another source of your type 5 error). This becomes more apparent when finding the Fermi level for a specific number of electrons in a small HOMO/LUMO gap or metallic systems (naive root finding at double precision evaluation often gives only single precision in the resulting total electron number, and fixing this with scaling or occupation number changes then deviates from the target distribution).

This can cause problems in the self-consistent solution of the problem and (weakly?) breaks convexity. Other than working to higher precision, there does not seem to be an obvious way to control this. Your examples are integer occupation cases, but have you considered the effect of this type of error?

Michael Herbst answered: We have not considered this type of error so far, because for our non-self-consistent model (with no density dependence in the Hamiltonian) the band energies as our quantity of interest do not depend on the Fermi level or the occupation numbers. Therefore our results are in fact even valid in metallic cases with non-integer occupation.

This changes if one considers, for example, the total energy in our models, which requires a Brillouin zone (BZ) integration for its computation. As a result additional sources of error need to be considered (BZ integration mesh, BZ quadrature, error in the Fermi level, smearing *etc.*). Keeping aside the discretisation error contributions we have classified as 1. in our manuscript, this indeed leaves an additional contribution to the arithmetic error, which could be determined using interval arithmetic and brought forward to the arithmetic error of the total energy.

In cases as you mention, where the occupation numbers/Fermi level is only numerically determined to single precision, this would thus reflect in a large arithmetic error in the total energy, telling you that the apparent issues with convexity are numerical noise.

Nisha Mehta asked: Can these ideas be extended to other basis sets?

Michael Herbst replied: While the discussion we presented in the paper (DOI: 10.1039/d0fd00048e) in particular is specific to plane-wave basis sets, the general idea of deriving *a posteriori* errors can be extended to other basis sets as well.

However, to obtain the Kato–Temple-type and Bauer–Fike-type bounds we presented the properties of both basis and potential were crucial to get our results. As I discussed in the context of Gaussian basis sets (see the answer to Christoph's question) for other basis types it may therefore be challenging to obtain bounds of a similar structure or one may need to follow a different route to derive them. In other words one really needs to look at the details of each type of basis anew to construct appropriate error bounds.

Kieron Burke opened discussion of the paper by Daniel Cole: Noticed that you had to use fairly stripped down basis MP2 for the QM parts because of system size. Have you considered using some of Grimme's methods that are designed to be as efficient as possible with simplest cost-effective calculations, especially if you want this accuracy of <1 kcal mol^{-1} in free energy? (Guess much of that is coming from the molecular mechanics.)

Daniel Cole answered: Thank you for the comments. We chose the QM method as a reasonable balance of accuracy and computational expense for this study, as we were more interested in the construction of the Gaussian Approximation Potential (GAP) itself. However, we certainly envisage using a spectrum of underlying QM methods in future, depending on the system size and accuracy requirements. Grimme's xTB method,[1] for example, would be a good addition at the lower cost end of this spectrum. It's worth pointing out that our short term goal now is to reduce the number of QM training configurations needed, so actually we could go to higher accuracy methods if required.

1 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.

Kieron Burke added: My question is whether one can trust the reference electronic structure calculation for geometries that are stretched by atoms moving 0.5 Å from equilibrium? And did these cause trouble for your electronic structure calculation?

Daniel Cole replied: Note that when we generate new configurations of the molecules using these 0.5 Å moves, this is only to generate representative configurations as basis functions to enter into eqn (1) of our paper (DOI: 10.1039/d0fd00028k). We do not compute QM energies or forces for these configurations, so there is no problem with converging the electronic structure calculations. These calculations are only run on physically reasonable structures extracted from molecular dynamics trajectories.

Manasi Mulay asked: The paper illustrates that the model based on the kernel regression machine learning technique is transferable to condensed phase and is applicable in drug delivery. For the drug delivery applications the diffusion properties of organic molecules are also of prime interest. Do you envisage the transferability of the technique used to build this model to estimating the diffusion properties of the long organic molecules?

Daniel Cole answered: Thank you for your comments. Yes, one could in theory run all sorts of simulations of this molecule in complex environments, as long as

the rest of the system is treated using MM. Since the calculations do take longer than standard MM, one should ensure that the gains in accuracy expected from improving the intramolecular energetics are worth the extra expense.

Manasi Mulay added: Can the described method of machine learning in this paper possibly be transferred to porous materials such as metal–organic frameworks (MOFs) or 2D materials such as molybdenum disulfide or graphene? What could be the potential limitations?

Gábor Csányi replied: Thank you for your question. Actually, these applications are closer to the materials-based systems that GAP was originally conceived for, and in some respects these applications are more straightforward as they do not have the flexibility seen in organic chemistry/biology. It is has been shown that GAP is very powerful for a number of systems like these, see *e.g.* ref. 1. There are a number of competing machine learning based frameworks for building force fields, and many of them perform similarly in expert hands.

1 A. P. Bartok, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csanyi and M. Ceriotti, *Sci. Adv.*, 2017, **3**, e1701816.

Dumitru Sirbu enquired: The machine learning seems to open many new possibilities to compute ever larger and more complicated systems. So is it possible to work with excited states, similarly to the results presented for a flexible molecule in a large environment? Or a flexible organic molecule at a complex lipid–water interface, like the cellular membrane?

Daniel Cole responded: Thank you for your comments and questions. We have not done it yet, but yes as long as one can extract QM energies and forces for the electronic excited state there is no reason why we cannot train a GAP on these. Actually, this would be a good example of going beyond what is possible using traditional molecular mechanics force fields. And yes, one could in theory run all sorts of simulations of this molecule in complex environments, as long as the rest of the system is treated using MM. Since the calculations do take longer than standard MM, one should ensure that the gains in accuracy expected from improving the intramolecular energetics are worth the extra expense.

Katarzyna Pernal opened discussion of the paper by Jack Wetherell: A given *v*-representable electron density determines a one-electron reduced density matrix (1RDM), which corresponds to the same external potential as the density. As we know, there may be more than one 1RDM, which yield the same *v*-representable electron density, if they correspond to different external potentials. How will your ML algorithm determine the "best" 1RDM for a given *v*-representable density?

Jack Wetherell replied: For a given Hamiltonian there is a one-to-one mapping from the charge density to density matrix, and this is the Hamiltonian (local external potential and non-local interaction) that we use to build the data-set and train. So this is the density matrix the ML model would predict. If we of course trained on the Kohn–Sham density matrix for example, it would predict that from

the density. It is the density matrices in the training data-set that define which density matrix we choose based on the Hamiltonian.

Katarzyna Pernal added: One of the properties of the $\gamma(r, r')$ function which is supposed to be an N-representable 1RDM is that in the spectral representation the diagonal elements – natural occupation numbers – are non-negative and not greater than two. This condition is not imposed in your algorithm. Does it mean that output 1RDMs, returned by your ML algorithm, will violate the condition?

Jack Wetherell responded: Yes, in this case we focussed on minimising the loss based on the absolute value of the density matrix. As this loss goes to zero it implies that the natural orbitals will approach the correct values due to the linearity of the eigen equation. Of course if the loss is sufficiently far from zero it is possible the natural orbitals break some fundamental conditions. If this was a consistent problem you could add the natural occupation numbers to your loss function, and guide the ML training to obey those conditions.

Kieron Burke asked: Related to that question, when you do your search you only look at density matrices that have come from an external potential. Is that correct? If so, the ones you find (assuming no interpolation error) have no representability issues, since they are actually ground-state 1-RDMs of some Hamiltonian.

Jack Wetherell answered: That is correct. Our training set is constructed by choosing a smooth, random (Fourier based) potential V, and from this computing the exact 1-RDM and density. The model is trained to produce the 1-RDM from the density. This ensures that the 1-RDM and density correspond to the ground state of a Hamiltonian containing V, and so contains no V-representability issues. Then, as you say, if the model fails to interpolate correctly it is not guaranteed that the predicted 1-RDM is representable.

Kieron Burke added: This is really important. A huge problem in density matrix functional theory is the infinite number of reduced density matrices that don't correspond to any Hamiltonian with a multiplicative/local potential. So in any search over those, you have questions of representability. If I'm understanding right, your scheme bypasses that because you're always interpolating within the manifold of matrices that are those of a Hamiltonian, and those are guaranteed implicity to satisfy all representability constraints. When you interpolate, the approximate interpolator could violate the constraints, but the interpolation is between density matrices that must satisfy them.

Jack Wetherell responded: Exactly yes. The dataset is designed to be a set of density matrices that have been constructed from the solution to systems defined by a local external potential. Thus guaranteeing the training points we are interpolating between are representable.

Kieron Burke asked: Define your measure of a "good" density matrix. Quantitatively.

Jack Wetherell replied: There are three distinct levels we must consider when assessing the accuracy of the machine-learned density matrix.

(1) The value itself. This is defined as the percentage integrated difference between the model and exact density matrix. Also the same difference between the first and second derivatives for example.

(2) Observables. This is defined by the percentage error between the model and exact observables computed *via* the functional of the density matrix. Such observables would be the kinetic energy, Hartree and exchange energies and potentials, and charge density.

(3) Properties. This is defined as the property calculated using these observables to an overall system property, such as the bond length of a covalent bond, its bond energy, ionisation energy, the Fisher point *etc*. These involve the computation of many observables for different sets of system parameters. It is important we compute accuracy of the machine learned density matrix as a functional of the density across all of these levels. In particular we must see how errors propagate up the levels, as it is paramount we understand how the errors behave, for example: do they exploit error cancellation, or is one constituent error dominating *etc*.

A description of the errors for our case is discussed in the article (DOI: 10.1039/d0fd00061b).

Aurora Pribram-Jones remarked: Related to Burke and Pernal's comments earlier: we've seen elsewhere that learned functionals can often have issues with their derivatives. Do you see any issues here with first and/or second derivatives? Or do your loss functions, constraints, and allowable density matrices smooth those things out for you?

Jack Wetherell answered: This would depend on how the model maker decided to train the model, and what was the most important quantity. If you were interested most in the kinetic energy for example, you would train the network to minimise the loss in the second derivative of the density matrix. If you were more interested in the exchange energy it would be the value itself. The loss function can also be a linear combination of these with the weights indicating the importance.

Aurora Pribram-Jones asked: Does de-noising help keep your targeted learning general while smoothing things out, sort of an alternative way of getting some smoothing without overly constraining things or making the procedure too specific to a specific use?

Jack Wetherell responded: In a sense, it is a way of designing a functional where the missing information is of a higher frequency (or more intricate detail) that the information well described. For example, the neglect non-linear terms may be treated as noise, or instead all correlation effect. This ensures that the information is presented to the machine in such a way that we can benefit most from the previous human learning, and leave as little to machine learn as possible.

Pina Romaniello said: I'd like to go back to the issue of the N-representability of the one-body density matrix. For the diagonal part of the density matrix (which is the density) you use (at least some of) the density N-representability conditions. I was wondering if the other constraints you use to compress the information of the density matrix into the density only are related to the N-representability conditions of a pure-state density matrix (*e.g.* by imposing a certain type of external potential).

Jack Wetherell replied: It is correct that by assembling our data-set from the exact solution of a randomised set of external potentials we guarantee that the densities and density matrices in question are V and $N = 2$ representable objects, and so any constraints uncovered will be applicable to objects of that type. How it would apply as a form of extrapolation outside this domain is something to be investigated.

Dumitru Sirbu opened discussion of the paper by Bartolomeo Civalleri: I am interested in the wider applicability of the CRYSTAL17 code, specifically for the study of nonlinear optical properties. Can it be used to accurately predict the second and third order susceptibilities $\chi^{(2)}$ and $\chi^{(3)}$ of materials? As the electron delocalisation becomes ever more important for second and third order hyper-polarizabilities is it able to correctly describe the system?

Bartolomeo Civalleri responded: The CRYSTAL code indeed allows users to compute electric susceptibilities up to $\chi^{(3)}$ at the static limit and up to $\chi^{(2)}$ for the frequency-dependent ones. See the CRYSTAL User's Manual for further details.[1] Accuracy depends on the choice of the exchange–correlation functional and basis set. Large basis sets of triple-zeta quality are usually needed.

1 R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D'Arco, M. Llunell, M. Causà, Y. Noel, L. Maschio, A. Erba, M. Rérat and S. Casassa, *CRYSTAL17*, Università di Torino, Torino, 2017.

Trygve Helgaker asked: Your composite scheme uses a double-zeta basis set and you note that, with the help of semi-classical corrections, you achieve triple-zeta quality. Are you able also to use triple-zeta basis sets in your calculations, without running into problems with linear dependencies? In what way do you modify your basis sets to remove the dependencies?

Bartolomeo Civalleri replied: Molecular basis sets are often not directly suit-able for solid state calculations. In fact, even the double-zeta basis sets employed in our composite methods were revised for solids.[1] This is even more delicate for triple-zeta (TZ) basis sets that contain low-exponent Gaussian functions that can lead to problems with linear dependencies. The first option is to remove functions based on the overlap matrix computed for the specific system under investigation, which is indeed possible in an automated way in CRYSTAL as well as most molecular quantum chemistry codes.

The second option is to redesign basis sets for solid state calculations as in many cases low exponent functions are not needed. Therefore a very crude approach to avoid linear dependency is to remove some of the most diffuse

functions from the basis set. Alternatively, for the examined system, one can tighten a bit the low-exponent functions by simply upscaling them or better by reoptimizing the corresponding exponents. We recently proposed a scheme to reoptimize the outermost exponents of valence shells by a constrained minimization of the energy of the system that includes an additional term that depends on the condition number of the overlap matrix.[2] This allowed us to use extended basis sets of TZ and also QZ quality for simple prototypical solids such as diamond, graphene sodium chloride, and LiH. Also, in the last years, a family of TZ quality basis sets has been devised by Bredow and co-workers[3–5] for calculations on solids through a systematic optimization of the outermost exponents averaged over a large set of solids.

Finally, one can tune computational parameters like integral thresholds and, eventually, activate an option to check the values of the eigenvalues of the overlap matrix and remove the eigenvectors whose eigenvalues are below a given threshold, as mentioned above, but the latter should be used very carefully. The use of a screened Coulomb range-separated hybrid functional as for HSEsol-3c can also partly reduce the problem of linear dependencies as those are most pronounced when long-range exchange contributions in small gap systems are present.

1 L. Doná, J. G. Brandenburg and B. Civalleri, *J. Chem. Phys.*, 2019, **151**, 121101.
2 L. E. Daga, B. Civalleri and L. Maschio, *J. Chem. Theory Comput.*, 2020, **16**, 2192–2201.
3 M. F. Peintinger, D. Vilela Oliveira and T. Bredow, *J. Comput. Chem.*, 2013, **34**, 451–459.
4 J. Laun, D. Vilela Oliveira and T. Bredow, *J. Comput. Chem.*, 2018, **39**, 1285–1290.
5 D. Vilela Oliveira, J. Laun, M. F. Peintinger and T. Bredow, *J. Comput. Chem.*, 2019, **40**, 2364–2376.

Manasi Mulay enquired: The paper includes examples of adsorption of several molecules on silica MCM-41 such as ibuprofen. The question is whether the cost of computation greatly varies when the adsorption of molecules is dissociative or molecular ?

Bartolomeo Civalleri responded: In the paper we did not report any geometry optimization, rather we analyzed the cost of the calculation of the energy through the SCF process and, then, of the nuclear and cell (for periodic systems) gradients (*i.e.* SCF+gradients). Basically, this is the cost of a single geometry optimization step. Regarding your question, I would expect that the cost would be similar in the two cases. The point is that the shape of the potential energy surface (PES) would be different thus leading to a different overall number of optimization steps and, of course, of the wall-clock time of the two calculations (*i.e.* dissociative or molecular)

Manasi Mulay commented: If the dissociative adsorption of molecules is considered, ring formation takes place to form, for example, a bidentate structure. It might require more steps for geometry optimisation than the molecular adsorption. Have you considered this possibility? Could the sol-3c hybrid methods still remain cost-effective?

Bartolomeo Civalleri replied: The cost effectiveness of the sol-3c DFT methods depends on the cost of the SCF+gradients calculation. The efficiency of the

geometry optimization depends on many different factors. The initial geometry is indeed one of them, but ultimately the efficiency of the geometry optimization depends on the shape of the potential energy surface. In our work, we analyzed the cost of the calculation in terms of the SCF+gradients wall-clock times. Therefore, the sol-3c hybrid methods still remain cost-effective. To have a rough estimate of the global cost of a geometry optimization one needs to know approximately the total number of steps needed to reach the minimum. In our experience the geometry optimizer implemented in CRYSTAL is quite efficient.[1]

1 R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, Y. Noël, M. Causà, M. Rérat and B. Kirtman, *Int. J. Quantum Chem.*, 2014, **114**, 1287–1317.

Manasi Mulay remarked: When there are various stable adsorption configurations possible, however with large energy differences, how valid would it be to claim a hybrid method (say the sol-3c methods in this paper) as being cost-effective for the wide range of family of structures of a same material system? Can we consider a possibility where in fact a method is cost-effective for some configurations of a material system and expensive for other configurations of the same material system?

Bartolomeo Civalleri responded: As answered before, the cost of a geometry optimization step depends on the cost of SCF+gradients. For a family of structures (*e.g.* different drugs adsorbed on the silica surface model), they are still cost-effective on average. Of course, the cost for different conformations of a given system (*i.e.* same number of atoms/basis functions) would not be exactly the same, but it should be rather similar.

Donald Truhlar commented: We obtained satisfactory results with CRYSTAL for bulk calculations by deleting the more diffuse Gaussian basis functions to avoid linear dependence. These basis functions are not needed in the bulk because functions on neighboring centers carry the weight. But we ran into large BSSE for calculating adsorption on surfaces where the deleted basis functions are not made up by neighboring atoms because the solid is replaced on one side by a vacuum. The paper uses a geometrical counterpoise correction to correct for BSSE with double zeta basis sets, and this is a very encouraging development. To get accurate adsorption energies, we usually use triple zeta basis sets. Are the geometrical counterpoise corrections available for triple zeta basis sets?

Jan Gerit Brandenburg answered: We indeed have a range of basis sets parametrized, which has been extended from their original publication;[1,2] *i.e.* its use with various DZ and TZ basis sets is possible. Your example of surface adsorption using extended basis sets for the molecule and compact ones for the bulk is more tricky. Mixing of basis sets is currently not possible with gCP and would require a reparametrization. However, in these situations our HSEsol-3c method as presented in this *Faraday Discussions* contribution (DOI: 10.1039/d0fd00066c) is a viable and presumably useful option.

1 H. Kruse and S. Grimme, *J. Chem. Phys.*, 2012, **136**, 154101.

2 J. G. Brandenburg, M. Alessio, B. Civalleri, M. F. Peintinger, T. Bredow and S. Grimme, *J. Phys. Chem. A*, 2013, **117**, 9282.

Bartolomeo Civalleri added: I agree with J. Gerit Brandenburg's comment. Removing basis functions from the parametrized gCP basis sets makes the parametrization invalid, *i.e.* a better approach might be to start out with a more compact basis set like sol-def2-mSVP (ref. 1) or def2-mTZVP (ref. 2) or pob-TZVP (ref. 3) and remove most of the inter- and intramolecular BSSE with the gCP correction.

1 L. Doná, J. G. Brandenburg and B. Civalleri, *J. Chem. Phys.*, 2019, **151**, 121101.
2 J. G. Brandenburg, C. Bannwarth, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2018, **148**, 64104.
3 M. F. Peintinger, D. Vilela Oliveira and T. Bredow, *J. Comput. Chem.*, 2013, **34**, 451–459.

Nisha Mehta asked: Did you consider using different variants of GCP corrections; *i.e.* DAMP-GCP *vs.* GCP?

Bartolomeo Civalleri answered: At the moment, the revised sol-3c composite methods are based on D3 and gCP semi-classical corrections, but they could be easily modified to include different variants for both the inclusion of dispersion energy (*e.g.* D4) and the removal of the basis set superposition error (*e.g.* DAMP-GCP). Of course, parameters that enter in the two corrections have to be refitted.

Nisha Mehta addressed Bartolomeo Civalleri and Jan Gerit Brandenburg: (1) Is this functional equally good for clusters? (2) Do you have plans to implement that in other QC codes, like ORCA or Turbomole?

Bartolomeo Civalleri responded: (1) The sol-3c hybrid composite methods have also been tested and validated on molecular systems, so we expect that they can be used to study clusters as well. For instance, in the present *Faraday Discussions* contribution (DOI: 10.1039/d0fd00066c) we reported benchmarking calculations on nanoparticles, in particular for the inorganic compound forsterite. Just, some attention must be paid when applying the methods to very low band gap systems and metallic clusters.

(2) In principle, our composite methods, even if targeted to solid state calculations, can be also implemented in molecular codes that allow the running of calculations with global and range-separated hybrid functionals. Of course, it would be interesting to have them available in other codes.

Jan Gerit Brandenburg added: (2) As we are using the screened exchange functionals (HSEsol) as the basis, it would be some implementation effort to do this as most of the molecular quantum chemistry codes (Orca, Turbomole, Psi4) to the best of my knowledge don't yet have an HSE implementation (including nuclear gradients). Once this is available, extension to those codes is trivial and we published all needed parameters and specifications in our original publication.[1]

The same holds for HSE-3c,[2] which can however be replaced in molecular calculations with the global hybrid variant PBEh-3c.[3]

1 L. Doná, J. G. Brandenburg and B. Civalleri, *J. Chem. Phys.*, 2019, **151**, 121101.
2 J. G. Brandenburg, E. Caldeweyher and S. Grimme, *Phys. Chem. Chem. Phys.*, 2016, **18**, 15519.
3 S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *J. Chem. Phys.*, 2015, **143**, 54107.

Jan Gerit Brandenburg continued the discussion of the paper by Daniel Cole: I have four questions and two comments.

Q1 Delta learning. Would it make sense to not learn the reference energy directly, but rather the difference with respect to the FF?

Q2 Inter *vs.* intra balance. FEP with FFs works to a large extent by error compensation of all energy terms. Can you infer if improving the intra terms alone yields more accurate binding energies?

Q3 Learning curves. For ML models it is common to look at learning curves (RMS *vs.* #training). Apart from the two GAP variants, have you studied the learning curves more systematically? From my experience you should be able to learn this potential with less than 3k training points.

Q4 Reference energies. MP2 is problematic for intramolecular dispersion.[1] Have you considered using DHDFs (*e.g.* DSD-BLYP-D4)?

C1: Competitor methods. A comparison to molecule specific FF would give a more balanced viewpoint. GFN-FF values[2] are shown in Table 1 and Fig. 1 here, analogous to Table 1 and Fig. 4 of the *Faraday Discussions* paper presented at this meeting (DOI: 10.1039/d0fd00028k).

C2: Dimension reduction. To get an understanding of how the SOAP descriptor distinguishes the sampled configurations, two-dimensional PCA maps using SOAP and a linear kernel as implemented in the ASAP tool[3] are shown here in Fig. 2.

1 S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105.
2 S. Spicher and S. Grimme, *Angew. Chem., Int. Ed.*, 2020, **59**, 15665.
3 B. Cheng, R.-R. Griffiths, S. Wengert, C. Kunkel, T. Stenczel, B. Zhu, V. L. Deringer, N. Bernstein, J. T. Margraf, K. Reuter and G. Csanyi, *Acc. Chem. Res.*, 2020, **53**, 1981, DOI: 10.1021/acs.accounts.0c00403.

Gábor Csányi responded: Thank you very much for your comments and questions.

Q1. We have not tried it for this system, but when we were training the model we did find that molecular dynamics using traditional force fields did not sample the correct configurational space, so it might be that force field energetics are too

Table 1 RMS errors (kcal mol^{-1}) in the total energies of configurations taken from MC simulations in three different environments relative to QM

RMS (kcal mol^{-1})	Water	Kinase	Hydrolase
GAP-v1	0.83	0.60	19.04
GAP-v2	1.42	0.95	1.13
OPLS^a	11.87	3.36	3.45
GFN-FF^a	**2.24**	**2.24**	**2.43**

^{a} Configurations were sampled from the GAP-v2 trajectories, and the MM and QM energies were shifted to align the mean energies.
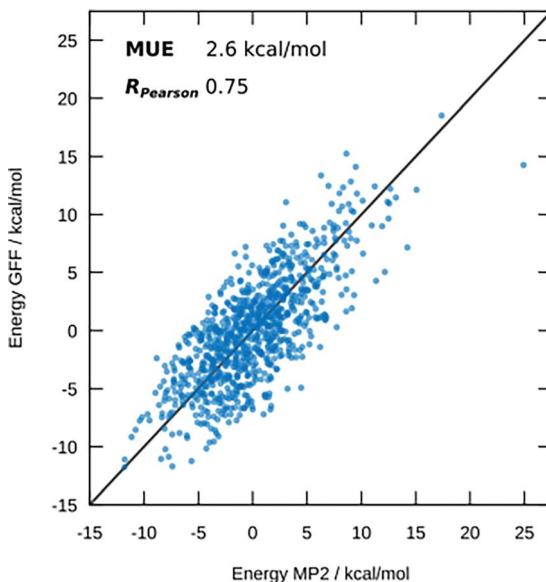
Faraday Discussions

Fig. 1 Correlation between GFN-FF and QM energies of 3BPA sampled from MC simulations. The mean energy of each distribution has been shifted to zero.
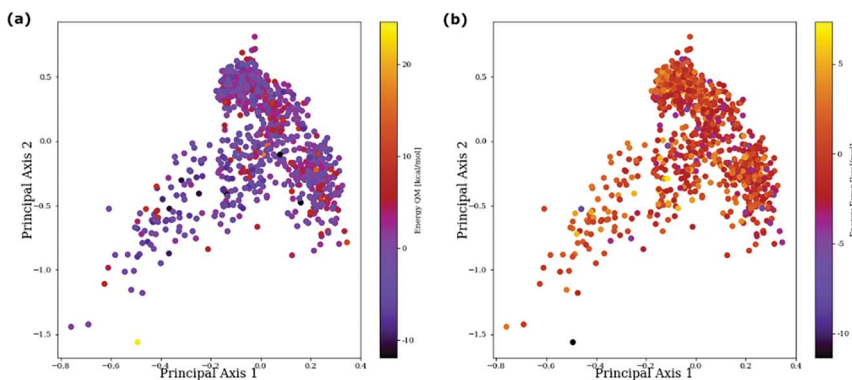

Fig. 2 Principal component analysis based on SOAP descriptor and linear kernal of structures sampled by MC simulations with GAP-v2. Color scale shows the (a) QM energy (kcal mol^{-1}) and (b) energy error (kcal mol^{-1}) of GFN-FF.

far from QM to serve as a useful reference. But it is certainly a good idea worth trying in all situations in which a good enough model exists that can be used as a base line. Using so-called "delta learning" is well established in the community, and in fact much predates the recent interest in data driven force fields. Pioneering work by Bowman and Braams[3] made use of electrostatic baselines for water on top of which data driven corrections were added for dimers and trimers to arrive at a very high quality water potential.

Q2. Since we submitted this paper, Rufa *et al.*[1] have indeed shown that errors in protein–ligand binding free energies are reduced from around 1 kcal mol^{-1},

using a standard MM force field, to around 0.5 kcal mol^{-1}, using a ML/MM approach where the intramolecular energetics are described by the ANI2x model. So yes, while there is a lot of work to do, we do believe that improving the intramolecular terms on their own can provide large gains in accuracy. This does not of course mean that intermolecular terms should not be improved, and there are some ML efforts in this direction too.

Q3. Yes we agree that 3k training points are not really needed for a system of this size and complexity. Our experience with using different molecular descriptors since we submitted this paper has shown that we can comfortably reduce the training set to around 500 configurations and still have high accuracy. These tests will be the subject of an upcoming benchmark publication, which will include full training curves and comparisons to other ML based force fields.

Q4. We chose the QM method as a reasonable balance of accuracy and computational expense for this study, as we were more interested in the construction of the GAP itself. However, we certainly envisage using a spectrum of underlying QM methods in future, depending on the system size and accuracy requirements. And indeed optimising this balance will be key for use of these methods in a high throughput manner.

C1. These are very interesting data, and it makes sense to us that a molecule-specific force field should be more accurate than the transferable one (OPLS), but less accurate than the GAP, which is not restricted by the choice of functional form. Indeed, we are also interested in developing molecule-specific force fields within the standard functional form.[2] If this were shown in a wider variety of cases, we would have access to a hierarchy of fast intramolecular potentials, which would be very beneficial to the molecular modelling field.

C2. Interesting plot! To our eyes, it shows that the GFN-FF is not particularly biased in this projection, its error is roughly homogeneously in PCA space.

1 D. A. Rufa, H. E. Bruce Macdonald, J. Fass, M. Wieder, P. B. Grinaway, A. E. Roitberg, O. Isayev and J. D. Chodera, Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning / molecular mechanics potentials, bioRxiv 2020.07.29.227959, DOI: 10.1101/2020.07.29.227959.
2 J. T. Horton, A. E. A. Allen, L. S. Dodda and D. J. Cole, *J. Chem. Inf. Model.*, 2019, **59**, 1366–1381, DOI: 10.1021/acs.jcim.8b00767.
3 B. J. Braams and J. M. Bowman, *Int. Rev. Phys. Chem.*, 2009, **28**(4), 577–606.

Andreas Savin asked: How does one deal with the following two problems?

(1) The input to a ML model is some quantity. However, it may not properly characterize the system. There may be several systems that are described by a similar input, but should produce a quite different output.

(2) The training set can be inhomogeneous, with some parts better described than others. Although one can identify these "outliers", and force them into the model, see, *e.g.*, ref. 1, it may be useful to identify some feature common to the "outliers", for example for generating data for new systems to be included into the training of the ML model.

1 P. Pernot, B. Huang and A. Savin, *Mach. Learn.: Sci. Technol.*, 2020, **1**, 035011.

Daniel Cole answered: Thank you very much for your comments.

(1) Properly choosing the molecular descriptors for these molecules to optimise accuracy and reduce training data is very much an area of active research.

We've chosen one representation here, based on vectors of interatomic distances and the squared exponential kernel, as a proof of principle, but are also actively investigating others (*e.g.* SOAP (ref. 1)) as well as comparing with others' work (*e.g.* ANI (ref. 2) and sGDML (ref. 3)).

(2) As you have seen in the paper, we do indeed find some configurations to be outliers that are not well modelled by GAP-v1. At the moment, we identify these "by eye", but one could certainly imagine automating this process such that new training data is generated on-the-fly and built back into the model.

1 A. P. Bartok, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csanyi and M. Ceriotti, *Sci. Adv.*, 2017, **3**, e1701816.
2 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203.
3 S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, *Nat. Commun.*, 2018, **9**, 3887.

Andreas Savin continued the discussion of the paper by Bartolomeo Civalleri: You have shown us that we can do calculations for 3000 atoms on relatively small computers accessible to most labs. If you use the computing capacities that exist today, you certainly go much further. But is there not the risk that the essential information gets lost when systems get bigger and bigger? Would it not be necessary to develop tools that extract the interesting information for large systems? Or maybe tell you what part of the system is interesting, so that you can do more accurate calculations for this part?

Bartolomeo Civalleri responded: Andreas Savin's question reminds me of the J. L. Borges "*Del rigor en la ciencia*" short-story on a map that becomes so exact to be of the same size and complexity of the territory itself, thus making the map useless. Therefore, I agree that the larger the system the more difficult it is to get insights in terms of simple models. In particular:

(1) Extracting information → Yes, this indeed gets more challenging the larger the system. For example, while for small molecules one can check the individual molecular orbital occupation numbers, this is not instructive for 3000-atom systems. Here, even the calculation of properties as the electron charge density or the electrostatic potential becomes a big issue. For instance, in CRYSTAL we have a parallel module to compute such properties for big systems.[1] Along with that, the development of analyzing and visualization tools is also of crucial importance (see for instance ref. 2).

(2) Model building → This is an important first step prior to any *in-silico* modelling. However, for certain systems, a reduced model cannot describe the required chemistry and physics; *e.g.* an active pharmaceutical ingredient adsorbed in a metal–organic framework needs a fully periodic description of the whole system. In particular, when lateral interactions of the molecule with the framework take place. In that case, the use of simplified cluster models would be complicated and even more difficult for multisite interactions or higher loadings.

(3) Multilevel modelling → Depending on the property to estimate, different computational levels might be used. For instance the introduced composite DFT methods tested in our contribution (DOI: 10.1039/d0fd00066c) are ideal for the fast generation of reliable geometries, more elaborate methods might be used on those geometries for higher quality calculations. Also, a multiscale approach would be useful as in a QM/MM scheme.

1 S. Casassa, A. Erba, J. Baima and R. Orlando, *J. Comput. Chem.*, 2015, **36**, 1940–1946.
2 G. Beata, G. Perego and B. Civalleri, *J. Comput. Chem.*, 2019, **40**, 2329–2338.

Andreas Savin continued the discussion of the paper by Jack Wetherell: We don't need to compute properties that are already known from experiment, except for checking or training methods. An apparatus can do it. We can also use a reliable method in a predictive way. But, as Prof. E. R. Davidson put it at a dinner table: "The chemist wants the computer in his head". I think that it is important for the "chemist's" creativity. Can we extract from calculations on big systems, *e.g.*, ML models information that the "chemist" can put into their head, and use without running a program on the computer?

Jack Wetherell replied: Yes I agree. Regarding a use of ML to build intuition *via* a large range of solvable cases: far more possible than without the powerful ML.

Andreas Savin added: As I understand ML can give fast and accurate answers to specific questions. What happens if my question is slightly modified, *e.g.*, if the output is slightly different from what the ML model has been trained for? Can they provide something to help humans think? Are there some "meta-ML" models?

Jack Wetherell responded: This is an interesting point that is popular in the ML community as "transfer learning". The meta-ML you describe is described in detail in the nice paper: "Learning What and Where to Transfer" (2019, Jang, Lee, Hwang, Shin).[1] This should be a priority when considering the application of ML models to cases not directly within the training space.

1 Y. Jang, H. Lee, S. J. Hwang and J. Shin, Learning What and Where to Transfer, 2019, arXiv:1905.05901.

Jack Wetherell remarked: A comment on the general aims we should have when applying machine learning approaches to the many-body problem. We should be careful to not rely on brute-force machine learning as a method to replace human knowledge if it simply serves as a black box from observable to value. The analogy would be if we had ML methods at the time of the Young's Double Slit experiment. If we would have trained a brute-force ML model to take the set-up as an input and prejudice the diffraction pattern as an output it would "solve" the problem of prediction, but without offering any insight into the underlying quantum nature of matter. ML could be instead thought to be a tool to gain human insight be being able to perform simulations faster by efficiently recycling knowledge from data-sets into models.

Andreas Savin added: Let me give an example of what I mean. From the experience with molecular orbitals Roald Hoffmann uses rules to explain chemistry without making calculations. I assisted once at a talk given by an experimentalist. Roald Hoffmann asked him about the evidence for the orientation of a molecule on a surface. He explained, based on his knowledge of molecular orbitals, that he expected a different orientation for the molecule. The

experimentalist answered that he just made a drawing, he did not know how the molecule was oriented.

Are there approaches that would allow to extract from machine learning simple humanly usable rules for humans?

Jack Wetherell responded: I think this question has two possible approaches. The first is that the machine learning can help us solve for cases that were otherwise intractable, thus adding to our intuition by collecting more known examples. In your case of molecular orbitals, this would only be possible (I imagine) due to the availability of the shape of molecular orbitals made possible by solving the single-particle Schrödinger equation. This set of known cases (the single-particle orbitals) allowed for this reasoning. Now we can use ML to extend to solving more complex problems, thus adding to this library of available known concepts. The second is that we can use ML techniques to in a way "open the black box" of neural networks by looking at the convolution kernels *etc*. This is far more open ended and speculative, as this is itself a new field within fundamental ML research.

Weitao Yang asked: Regarding your machine learning approach for the density matrix from density as input, do you plan to use it in connector theory? Density is normalised to the number of electrons. How do you deal with the dependence on the number of electrons? Ignore it, look locally? Or do you have to know, by integrating, to get N?

Jack Wetherell answered: Indeed we intend to integrate the connector and machine learning approaches to develop approximations to $\rho[n]$. In particular this work serves as a benchmark to machine learning the entire matrix itself. In the future the connector approach will contain some parameters that have to be approximated. The approach will be chosen such that these parameters are favourable to machine learning. This will allow us to inject as much human learning as possible, and only machine learn the smallest part. In terms of the dependence of the number of electrons N there are many options to consider. It is worth noting that in this work we dealt entirely with a fixed electron number $N = 2$. The connector approach itself more naturally addresses this, as it imports from model systems whose density and density matrix span the full range of values (similar to the LDA from the HEG). As discussed before the parameter to machine learn would depend on the density everywhere (for example) and so the model would have access to N through the integral.

Weitao Yang added: In your approach, could it be a challenge if you're doing this for a local theory such as using near-sightedness in the connector approach? In a local theory, global information is not available or used. Continuous interpolation between N may be an issue, N is discrete, and there are various discontinuities at the integers. Interpolation will not be easy?

Jack Wetherell replied: The functional presented can in principle capture any functional behaviour due to the convolutional layers in the de-noising autoencoder connecting every element of the density matrix to the whole density in an arbitrarily non-linear way. It is important that we make a sensible choice of how to

present the data to the machine while training. We would not want to include N as a free parameter due to the pathological dependence that you describe. There are a few possible options: the brute force approach of training a model for each and every N (not efficient). Or including N as a categorical variable in the ML model and restricting it to integer values, as the model is shown the whole density and so can compute the integral to obtain N. Referring to my previous answer, integrating ML with the connector approach is also a viable way to approach the N dependency problem.

Duncan Gowland communicated a question about the paper by Daniel Cole: In the manuscript in section 2.1 just above eqn (2), there is a comment regarding the example molecule (3BPA) having only eight permutation symmetry elements when presenting the Gaussian kernel. For this low symmetry molecule, what are the sources of these symmetries and how do they affect learning the intra-molecular potential? Similarly, are there implications for the potential effective-ness as well as the computational expense for training kernels for very low symmetry systems – such as large molecules with multiple chiral centres?

Gábor Csányi communicated in reply: The typical symmetries that we mentioned are the rotation of a CH_3 group or the reflection of an NH_2 group. The ML force fields work perfectly well without explicitly baking in these symmetries, but they do work better if we do. In the way we have done in this paper, adding more symmetry increases the cost, because the entire molecular representation is symmetrized. In the future, we will move to molecular descriptors that are separable, and so the cost will grow linearly with molecule size irrespective of the number of these small symmetric side groups.

Matthew R. Ryder communicated a question about the paper by Bartolomeo Civalleri: Have you tested the reliability of PBEsol0-3c and HSEsol-3c for vibra-tional mode calculations? We've recently found that HSE-3c and PBE-3C in CRYSTAL17 give quite good results for MOFs and molecular crystals (drug molecules), but often with an offset compared to experiment and typical DFT approaches.

Bartolomeo Civalleri communicated in reply: Usually, the offset of the peaks is due to the fact that vibrational frequencies are calculated in the harmonic approximation. In addition, the use of small-to-medium basis sets and a high amount of Hartree–Fock exchange have been shown to lead to too overestimated frequencies for the original composite methods.[1–4] Authors then recommended the use of the following scaling factors: HF-3c 0.86; HSE-3c/PBEh-3c 0.95, and B97-3c 1.0. They also tested them on molecular heat capacities and also checked some phonon modes of molecular crystals (squaric acid), where this scaling seemed to be rather consistent.

We tested the sol-3c methods on α-quartz, but frequencies are below 1200 cm^{-1}, so the agreement is not bad in that case.[5]

In a recent work with hybrid DFT sol-3c methods we simply scaled the vibra-tional frequencies to a reference peak from measurements to have a proper comparison with the experimental spectrum.[6] Incidentally, the scaling factor (*i.e.* 0.95) is consistent with the original methods.

1 R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672–1685.

2 S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *J. Chem. Phys.*, 2015, **143**, 054107.

3 J. G. Brandenburg, E. Caldeweyher and S. Grimme, *Phys. Chem. Chem. Phys.*, 2016, **18**, 15519–15523.

4 J. G. Brandenburg, C. Bannwarth, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2018, **148**, 064104.

5 L. Doná, J. G. Brandenburg and B. Civalleri, *J. Chem. Phys.*, 2019, **151**, 121101.

6 B. E. Souza, L. Donà, K. Titov, P. Bruzzese, Z. Zeng, Y. Zhang, A. S. Babal, A. F. Möslein, M. D. Frogley, M. Wolna, G. Cinque, B. Civalleri and J.-C. Tan, *ACS Appl. Mater. Interfaces*, 2020, **12**, 5147–5156.

Conflicts of interest

There are no conflicts to declare.