Are elementary crystals really sufficient test case for confidence in dft methods / codes? Many common DFT issues occur in for example highly correlated oxides, adsorption of CO on metal surfaces, etc, etc

⇨ The DFT issues you are referring to are probably physics issues (DFT not agreeing with experiment), not implementation issues. We use this test set in the first place to detect bugs in codes or the quality of pseudopotentials, without referring to any experiment.

Why are you comparing only pairwise ? Why wouldn't you compare over more codes and extract an average curve/value that is set as a "standard" and all the codes are compared to this standard ?

⇨ It could have been done in that way too, but then that average curve would get in the mind of people a status of being 'the true result', which it isn't. A new code that in the future would get the absolutely perfect numerical solution, would then have a nonzero delta with respect to that reference curve, which is not really intuitive.

Did any code got worse (in terms of precision) with time?

⇨ Not among the cases we could inspect. If that would happen now, the developers would probably examine the issue before releasing that code.

Some DFT codes have license conditions which make benchmarking legally problematic (a few years ago, notoriously one such code). What can be done? Should the scientific community simply boycott codes which prohibit / obstruct validation? Should peer-reviewers be skeptical of papers based on results of software that cannot be independently validated?

⇨ In my opinion, there is an easy solution to this. If every code that does agree being benchmarked effectively participates in a benchmark effort (as for the Delta-project), then then few codes that do not will make themselves spontaneously suspicious. In the end, they will be forced to participate as well. No boycott needed.

Where can I find the results of your comparison with the parameters you used?

⇨ The supplementary material of that Science paper (nearly 200 pages…) has every setting for every test. At molmod.ugent.be/deltatcodesdft, you can retrieve that information as well (pdf per test, at the very right of each line).

A really bad performance in a few systems may not show up in the overall error, but is important to flag. Will your system highlight such outliers as a warning to developers / users?

⇨ The Delta-values per element, prior to the averaging, are available. In the supplementary material of the paper in table form, as well as in periodic table form at molmod.ugent.be/deltacodesdft (click on the Delta-value for a specific test, and the period table pops up).

Are you getting good involvement from developers of codes other than those in Vienna? Obviously the greatest value is in having a wide diversity of softwares.

⇨ As you can see both in the paper on at molmod.ugent.be/deltacodesdft, nearly all mainstream DFT codes have participated – with enthusiasm. This was a community-wide effort, by far not restricted to Vienna.

Would you expect less packed structures to have a different Delta behavior between codes with different basis sets? E.g. larger deviations for some codes.

⇨ This will be examined in the fcc-hcp-diamon-sc elemental crystal set part of the new test set. I expect there will be no big surprises there, as these are 'regular' crystal structures for most elements. Some of the binary oxides have unusual interatomic distances and bond types, and therefore test pseudopotentials beyond the common environments.

Is comparing the ground state energy enough?

⇨ It's never enough. We started with the total energy, because that is a property that is for sure computed by every DFT code. Comparing DOS, band structure, densities,… would be a next step.

From what I understand from dft with the right exchange potential the ground state energy should be correct. People nowadays are using dft as a base for example transport equations. Are you planing to compare other features of the bandstructure between the codes.

⇨ It's at least on our wish list to do systematic comparisons of band structures, although there is no active work on this right now. At EPFL there is definitely work going on in this direction.