

Guido Falk von Rudorff

Exploration of Chemical Space Using Quantum Alchemy

Materials or compound design at heart is searching for the needle in the haystack and faces two major challenges: properly identifying what exactly qualifies as needle (property estimation) on the one hand and the sheer size of the haystack on the other hand. The latter is often called chemical (compound) space and is meant to encompass all (stable) systems. Common estimates for the size of subspaces thereof are 10^{60} molecules relevant for pharmaceutical applications [1] and 10^{29} ternary face-centered-cubic supercells with less than 30 atoms [2]. Even if a very well trained chemist could consider one million compounds a second and even if everybody on earth would be such a prolific domain expert, it would take 10^{26} times the age of the universe to go through this list of molecules. Essentially that illustrates that the speed at which one can perform property estimation does not matter: no matter how fast a high-throughput experiment can be, or how quickly a computational estimate can be produced, we either need systems of favourable properties to be very common or we require approaches that can consider parts of chemical space as a whole without enumerating systems one-by-one.

When considering such approaches, it is most useful to go back to the original problem: solving Schrödinger's equation. The only ingredients of the common Hamiltonian are the atomic positions and nuclear charges, the net charge, and the net spin. The net charge is often small to allow for stability at all, so there are only few (easily enumerable) independent cases for a given system. The same argument applies to the net spin. For the nuclear charges and coordinates however, they interact and span a combinatorially growing space of possible systems. For spatial coordinates, vibrations commonly explore a wide configurational space and functional groups or atoms can easily be replaced and still form stable molecules, which is commonly exploited by generating candidate molecules via graph enumeration [3–6], stochastic modifications of bond networks [7, 8] or generative neutral networks [9, 10]. Therefore, the key to cover a large volume of chemical space is by exploiting trends and correlations regarding nuclear positions and charges.

One family of such approaches is Quantum Alchemy. It is a theoretical framework which includes the derivatives of properties with respect to the change in nuclear charges of the atoms in-

volved. This is valid from a quantum chemistry standpoint as the nuclear charges only enter Schrödinger's equation by virtue of the external field, i.e. non-integer nuclear charges are conceptually allowed even though they cannot be realised in the lab. The first few orders of such derivatives then often allow to predict energies and electron densities of new systems with sufficient accuracy [11] using a Taylor or Padé approximant [12]. A Taylor series is a local approximation to a function as a polynomial the coefficients of which come from the derivatives of the approximated function while a Padé approximant uses the ratio of two polynomials to achieve the same and can be obtained from a Taylor series. Mathematically, this approach is similar to using the force acting on a molecular configuration and the Hessian matrix to estimate the minimum energy geometry, a well-established procedure for geometry optimizations. Since the individual derivative contributions to a Taylor polynomial are purely additive, the alchemical gradients can be calculated once for each atom and then combined to form millions of new systems [11].

Including mixed derivatives of geometry and alchemical charges allows to not only estimate the effect of substituting atoms with those of other elements ("vertical change") but also how the geometry of the system will change due to that substitution ("relaxed change") [13], useful for inverse design using such derivatives [14]. Calculations with arbitrary precision math have confirmed that the Taylor series from alchemical derivatives converges for surprisingly large changes in nuclear charges, e.g. for predicting the energy, orbital energies and density matrix elements of CO, BF, BeNe from derivatives evaluated at N only [12]. With the same method, the dimerization of a chain of Hydrogen atoms is predicted correctly from alchemical perturbations starting from a chain of Helium atoms where there is no dimerization. This exemplifies the fact that in the context of quantum alchemy, the perturbative effect of newly added sites can be estimated reliably, relevant e.g. for protonation energies [15, 16] where the common approach [17] of estimating protonation from the electrostatic potential is actually the zeroth order expansion in quantum alchemy. Quantum Alchemy has been developed from different perspectives starting with Wilson's four dimensional electron charge [18] which can be extended to estimate molecular energies [19, 20] via electrostatic potentials and perturbation expansion [21]. The virtual crystal approximation (VCA) which aims at modeling fractional occupancy of an atomic site by interpolating elements to produce a mean effect is closely related [22] but much less accurate as local structure is not resolved. In conceptual Density Functional Theory (DFT), it is natural to consider response functions to an external change, here of the external potential [23–28]. In both the Hartree-Fock and DFT worlds, coupled

Prof. Dr. Guido Falk von Rudorff
Universität Kassel
Computational Chemistry of Nanomaterials
Heinrich-Plett-Str. 40, 34132 Kassel
vonrudorff@uni-kassel.de
DOI-Nr.: 10.26125/jmf0-3f32

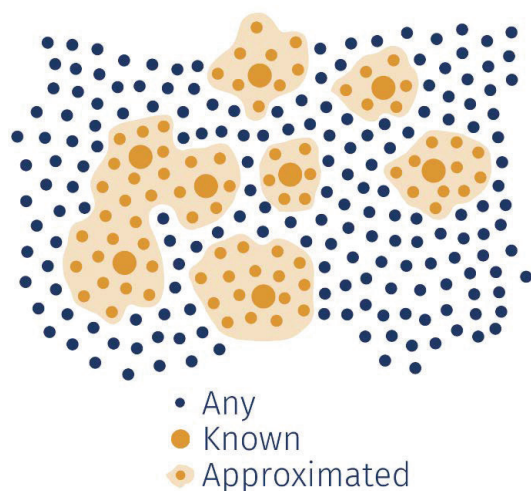


Fig. 1: Quantum alchemy allows to approximate large domains in chemical space from a single calculation. Blue dots denote any valid system, e.g. any molecule. Large orange dots denote selected systems for which an appropriate quantum chemistry method has been evaluated. Orange areas denote the regions which can be approximated from alchemical perturbations.

perturbed approaches have been used to develop closed form expressions for the derivatives [29] improving upon earlier combinations of finite differences and the Hellmann-Feynman theorem [30]. Most recently, automatic differentiation (AD) has been applied to quantum alchemy in a bid to simplify derivation of the expressions and the implementation thereof [31]. Since AD does not require to manually derive the derivatives in question, this approach is likely to be of significant transformative power in quantum chemistry. Thermodynamic cycles can help building error correction schemes [32] for some applications such as catalysis [33, 34].

Whenever a comparably rigid scaffold is involved, alchemical perturbations commonly require about three orders to predict a system doped with heteroatoms or replaced functional groups to sufficient accuracy, i.e. comparable to standard DFT calculations, which are the most commonly used calculations [35]. This is in part due to a large convergence radius of the perturbation series in those cases which is highly beneficial for exploring large chemical spaces: if the convergence radius is e.g. two nuclear charges, then any atom with nuclear charge can be replaced with any of the five elements from $Z-2$ to $Z+2$. If N such sites in a system exist, then the same few derivatives access 5^N . For a single coronene, this would equal about 10^{16} systems ignoring symmetry. Similar cases have been investigated for both covalent and non-covalent interactions e.g. for BN-doping fullerenes [26], finding extremal cases of non-covalent interactions of doped coronene [11], catalyst discovery [33, 34, 36], screening of ionic crystals [37] or to uncover design rules for doped polycyclic aromatic hydrocarbons [38].

Less rigorously, the perturbative approach in principle can be applied to design problems where the property in question depends smoothly on e.g. the atomic charges. For example, “vertical” changes in the periodic table (e.g. As-Sb) have been explored in the design of materials [39]. Even though the formal change of the nuclear charge is huge, the electronic structure between the elements is similar as they share the

number of valence electrons. Any case where such an argument can be made, a perturbative approach akin to Quantum Alchemy is likely to be efficient. The closed form expression for the property over the region of chemical space covered by the perturbation series is convenient: for many properties no such closed form exists and they have to be derived, implemented and calculated from quantum chemistry methods. At the same time, this closed form is highly interpretable as each term is a direct consequence of some change to the system. Interactions between different such changes can be investigated and analysed directly, which e.g. has been used to identify approximate symmetries in chemical space which can be used to reduce the search space [38] or to provide a conceptual structure to machine learning models [40].

Several challenges remain: currently, calculating derivatives w.r.t. nuclear charges requires non-standard approaches and is highly technical. One major obstacle is that many computational chemistry methods ignore the smoothness of properties between elements and rather parametrize each of them from scratch [41], although some procedures obtain variationally optimal parametrizations [42] which lowers the error. In practice, any general parametrization leads to inconsistencies and discontinuities which require careful interpolation. This is particularly relevant for periodic systems such as materials which typically rely on heavy parametrization to reduce computational cost. Furthermore, reliable error estimates are needed to automatically detect the trust radius, i.e. by how much nuclear charges and geometries can be changed without obtaining inaccurate predictions. With these topics addressed, the design problem of finding the right needle in the haystack of all possible systems can be substantially simplified in a physically rigorous manner.

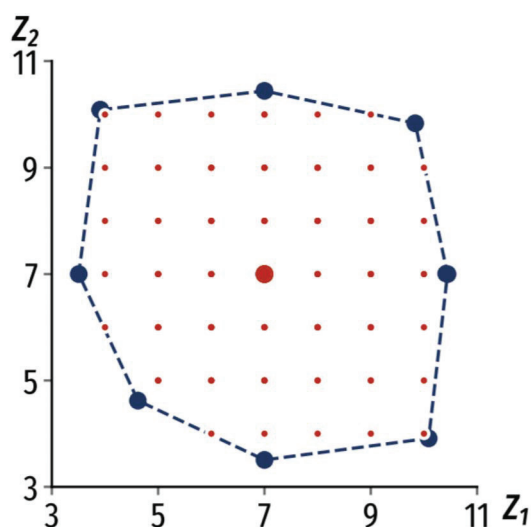


Fig. 2: Convergence radius of the alchemical perturbation series as Padé approximant around the Nitrogen dimer. Either nuclear charge Z can be changed by several integer values for this particular simulation. For two atoms, this yields a quadratic number of systems that can be assessed without extra cost (red dots). As discussed in the text, this effect is the stronger the more atoms are involved. In practice, plus and minus one nuclear charge per atom are feasible and sufficiently accurate with few leading orders.

References

- [1] R. S. Bohacek, C. McMartin, and W. C. Guida, "The art and practice of structure-based drug design: A molecular modeling perspective," *Med. Res. Rev.*, vol. **16**, no. 1, pp. 3–50, Jan. 1996, doi: 10.1002/(sici)1098-1128(199601)16:1<3::aid-med1>3.0.co;2-6.
- [2] K. Shinohara, A. Seko, T. Horiyama, M. Ishihata, J. Honda, and I. Tanaka, "Enumeration of nonequivalent substitutional structures using advanced data structure of binary decision diagram," *The Journal of Chemical Physics*, vol. **153**, no. 10, p. 104109, Sep. 2020, doi: 10.1063/5.0021663.
- [3] M. A. Yirik, M. Sorokina, and C. Steinbeck, "MAYGEN: An open-source chemical structure generator for constitutional isomers based on the orderly generation principle," *Journal of Cheminformatics*, vol. **13**, no. 1, Jul. 2021, doi: 10.1186/s13321-021-00529-9.
- [4] B. D. McKay, M. A. Yirik, and C. Steinbeck, "Surge: A fast open-source chemical graph generator," *Journal of Cheminformatics*, vol. **14**, no. 1, Apr. 2022, doi: 10.1186/s13321-022-00604-9.
- [5] J. E. Peironcely et al., "OMG: Open molecule generator," *Journal of Cheminformatics*, vol. **4**, no. 1, Sep. 2012, doi: 10.1186/1758-2946-4-21.
- [6] C. Benecke, T. Grüner, A. Kerber, R. Laue, and T. Wieland, "MOlecular structure GENeration with MOLGEN, new features and future developments," *Fresenius Journal of Analytical Chemistry*, vol. **359**, no. 1, pp. 23–32, Aug. 1997, doi: 10.1007/s002160050530.
- [7] M. Krenn, F. Häse, A. Nigam, P. Friederich, and A. Aspuru-Guzik, "Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation," *Machine Learning: Science and Technology*, vol. **1**, no. 4, p. 045024, Oct. 2020, doi: 10.1088/2632-2153/aba947.
- [8] K. Karandashev et al., "Evolutionary monte carlo of qm properties in chemical space: Electrolyte design." arXiv, 2023. doi: 10.48550/ARXIV.2307.15563.
- [9] Łukasz Maziarka, A. Pocha, J. Kaczmarczyk, K. Rataj, T. Danel, and M. Warchol, "Mol-CycleGAN: A generative model for molecular optimization," *Journal of Cheminformatics*, vol. **12**, no. 1, Jan. 2020, doi: 10.1186/s13321-019-0404-1.
- [10] R. Gómez-Bombarelli et al., "Automatic chemical design using a data-driven continuous representation of molecules," *ACS Central Science*, vol. **4**, no. 2, pp. 268–276, Jan. 2018, doi: 10.1021/acscentsci.7b00572.
- [11] G. F. von Rudorff and O. A. von Lilienfeld, "Alchemical perturbation density functional theory," *Physical Review Research*, vol. **2**, no. 2, p. 023220, May 2020, doi: 10.1103/physrevresearch.2.023220.
- [12] G. F. von Rudorff, "Arbitrarily accurate quantum alchemy," *The Journal of Chemical Physics*, Nov. 2021, doi: 10.1063/5.0073941.
- [13] S. Fias, S. Chang, and O. A. von Lilienfeld, "Alchemical normal modes unify chemical space," *J. Phys. Chem. Lett.*, Nov. 2018, doi: 10.1021/acs.jpclett.8b02805.
- [14] T. Shiraogawa and J.-y. Hasegawa, "Optimization of general molecular properties in the equilibrium geometry using quantum alchemy: An inverse molecular design approach," *The Journal of Physical Chemistry A*, vol. **127**, no. 19, pp. 4345–4353, May 2023, doi: 10.1021/acs.jpca.3c00205.
- [15] G. F. von Rudorff and O. A. von Lilienfeld, "Rapid and accurate molecular deprotonation energies from quantum alchemy," *Physical Chemistry Chemical Physics*, vol. **22**, no. 19, pp. 10519–10525, 2020, doi: 10.1039/c9cp06471k.
- [16] M. Muñoz, A. Robles-Navarro, P. Fuentealba, and C. Cárdenas, "Predicting deprotonation sites using alchemical derivatives," *The Journal of Physical Chemistry A*, vol. **124**, no. 19, pp. 3754–3760, Apr. 2020, doi: 10.1021/acs.jpca.9b09472.
- [17] S. A. Grimm and M. Reiher, "The electrostatic potential as a descriptor for the protonation propensity in automated exploration of reaction mechanisms," *Faraday Discussions*, vol. **220**, pp. 443–463, 2019, doi: 10.1039/c9fd00061e.
- [18] E. B. Wilson, "Four-dimensional electron density function," *J. Chem. Phys.*, vol. **36**, no. 8, pp. 2232–2233, Apr. 1962, doi: 10.1063/1.1732864.
- [19] P. Politzer and R. G. Parr, "Some new energy formulas for atoms and molecules," *J. Chem. Phys.*, vol. **61**, no. 10, pp. 4258–4262, Nov. 1974, doi: 10.1063/1.1681726.
- [20] P. Politzer and M. Levy, "Energy differences from electrostatic potentials at nuclei," *J. Chem. Phys.*, vol. **87**, no. 8, pp. 5044–5046, Oct. 1987, doi: 10.1063/1.452823.
- [21] M. Levy and Y. Tal, "Atomic binding energies from fundamental theorems involving the electron density, $\langle r^{-1} \rangle$, and the z^{-1} perturbation expansion," *J. Chem. Phys.*, vol. **72**, no. 5, pp. 3416–3417, Mar. 1980, doi: 10.1063/1.439527.
- [22] C. Eckhardt, K. Hummer, and G. Kresse, "Indirect-to-direct gap transition in strained and unstrained snGe alloys," *Physical Review B*, vol. **89**, no. 16, p. 165201, Apr. 2014, doi: 10.1103/physrevb.89.165201.
- [23] N. Sablon, F. D. Proft, P. W. Ayers, and P. Geerlings, "Computing second-order functional derivatives with respect to the external potential," *Journal of Chemical Theory and Computation*, vol. **6**, no. 12, pp. 3671–3680, Nov. 2010, doi: 10.1021/ct1004577.
- [24] P. Geerlings, S. Fias, Z. Boisdenghien, and F. D. Proft, "Conceptual DFT: Chemistry from the linear response function," *Chemical Society Reviews*, vol. **43**, no. 14, p. 4989, 2014, doi: 10.1039/c3cs60456j.
- [25] R. Balawender, M. Lesiuk, F. D. Proft, C. V. Alsenoy, and P. Geerlings, "Exploring chemical space with alchemical derivatives: Alchemical transformations of h through ar and their ions as a proof of concept," *Physical Chemistry Chemical Physics*, vol. **21**, no. 43, pp. 23865–23879, 2019, doi: 10.1039/c9cp03935j.
- [26] R. Balawender, M. Lesiuk, F. D. Proft, and P. Geerlings, "Exploring chemical space with alchemical derivatives: BN-simultaneous substitution patterns in c60 ," *J. Chem. Theory Comput.*, vol. **14**, no. 2, pp. 1154–1168, Jan. 2018, doi: 10.1021/acs.jctc.7b01114.
- [27] C. Cárdenas, W. Tiznado, P. W. Ayers, and P. Fuentealba, "The fukui potential and the capacity of charge and the global hardness of atoms," *The Journal of Physical Chemistry A*, vol. **115**, no. 11, pp. 2325–2331, Feb. 2011, doi: 10.1021/jp109955q.
- [28] T. Gómez, P. Fuentealba, A. Robles-Navarro, and C. Cárdenas, "Links among the fukui potential, the alchemical hardness and the local hardness of an atom in a molecule," *Journal of Computational Chemistry*, vol. **42**, no. 23, pp. 1681–1688, Jun. 2021, doi: 10.1002/jcc.26705.
- [29] M. Lesiuk, R. Balawender, and J. Zachara, "Higher order alchemical derivatives from coupled perturbed self-consistent field theory," *The Journal of Chemical Physics*, vol. **136**, no. 3, p. 034104, Jan. 2012, doi: 10.1063/1.3674163.
- [30] O. A. von Lilienfeld, "Accurate ab initio energy gradients in chemical compound space," *J. Chem. Phys.*, vol. **131**, no. 16, p. 164102, Oct. 2009, doi: 10.1063/1.3249969.

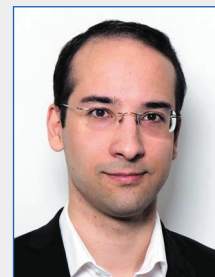
- [31] M. F. Kasim, S. Lehtola, and S. M. Vinko, "DQC: A python program package for differentiable quantum chemistry," *The Journal of Chemical Physics*, vol. **156**, no. 8, p. 084801, Feb. 2022, doi: 10.1063/5.0076202.
- [32] E. A. Eikev, A. M. Maldonado, C. D. Griego, G. F. von Rudorff, and J. A. Keith, "Quantum alchemy beyond singlets: Bonding in diatomic molecules with hydrogen," *The Journal of Chemical Physics*, vol. **156**, no. 20, May 2022, doi: 10.1063/5.0079487.
- [33] K. Saravanan, J. R. Kitchin, O. A. von Lilienfeld, and J. A. Keith, "Alchemical predictions for computational catalysis: Potential and limitations," *The Journal of Physical Chemistry Letters*, vol. **8**, no. 20, pp. 5002–5007, Sep. 2017, doi: 10.1021/acs.jpcclett.7b01974.
- [34] C. D. Griego, K. Saravanan, and J. A. Keith, "Benchmarking computational alchemy for carbide, nitride, and oxide catalysts," *Advanced Theory and Simulations*, vol. **2**, no. 4, p. 1800142, Nov. 2018, doi: 10.1002/adts.201800142.
- [35] B. Huang, G. F. von Rudorff, and O. A. von Lilienfeld, "The central role of density functional theory in the AI age," *Science*, vol. **381**, no. 6654, pp. 170–175, Jul. 2023, doi: 10.1126/science.abn3445.
- [36] C. D. Griego, J. R. Kitchin, and J. A. Keith, "Acceleration of catalyst discovery with easy, fast, and reproducible computational alchemy," *International Journal of Quantum Chemistry*, vol. **121**, no. 1, Jul. 2020, doi: 10.1002/qua.26380.
- [37] A. Solovyeva and O. A. von Lilienfeld, "Alchemical screening of ionic crystals," *Phys. Chem. Chem. Phys.*, vol. **18**, no. 45, pp. 31078–31091, 2016, doi: 10.1039/c6cp04258a.
- [38] G. F. von Rudorff and O. A. von Lilienfeld, "Simplifying inverse materials design problems for fixed lattices with alchemical chirality," *Science Advances*, vol. **7**, no. 21, p. eabf1173, May 2021, doi: 10.1126/sciadv.abf1173.
- [39] K. Y. S. Chang and O. A. von Lilienfeld, "AlxGa1-xAs crystals with direct 2 eV band gaps from computational alchemy," *Physical Review Materials*, vol. **2**, no. 7, Jul. 2018, doi: 10.1103/physrevmaterials.2.073802.
- [40] C. D. Griego, L. Zhao, K. Saravanan, and J. A. Keith, "Machine learning corrected alchemical perturbation density functional theory for catalysis applications," *AIChE Journal*, vol. **66**, no. 12, Oct. 2020, doi: 10.1002/aic.17041.
- [41] G. Domenichini, G. F. von Rudorff, and O. A. von Lilienfeld, "Effects of perturbation order and basis set on alchemical predictions," *The Journal of Chemical Physics*, vol. **153**, no. 14, p. 144118, Oct. 2020, doi: 10.1063/5.0023590.
- [42] F. Weigend, F. Furche, and R. Ahlrichs, "Gaussian basis sets of quadruple zeta valence quality for atoms hKr," *The Journal of Chemical Physics*, vol. **119**, no. 24, pp. 12753–12762, Dec. 2003, doi: 10.1063/1.1627293.

ZITATE

"Albert Einstein: So here we are, hmm? Lost in your quantum world of probabilities, and needing certainty." — Christopher Nolan, *Oppenheimer: The Official Screenplay*

"Prediction is difficult, especially for the future." — Niels Bohr

Prof. Dr. Guido Falk von Rudorff



Guido Falk von Rudorff is professor for Computational Chemistry of Nanomaterials at University of Kassel and an associate member of the CINSaT institute. His lab, nablachem.org, explores the calculation of property gradients and derivative information to build physics-based approximations and to identify fundamental laws of chemical space.

In 2014, he obtained a MSc Physics from Freie Universität Berlin, followed by a PhD from University College London in 2017 using classical alchemical potentials in conjunction with quantum chemistry methods on iron-oxide/water interfaces. He then worked as postdoctoral researcher at University of Basel and University of Vienna on quantum alchemy and machine learning models before joining University of Kassel in 2022.

ZITATE

„Man kann auch ganz burleske Fälle konstruieren. Eine Katze wird in eine Stahlkammer gesperrt, zusammen mit folgender Höllenmaschine (die man gegen den direkten Zugriff der Katze sichern muss): in einem Geigerschen Zählrohr befindet sich eine winzige Menge radioaktiver Substanz, so wenig, dass im Laufe einer Stunde vielleicht eines von den Atomen zerfällt, ebenso wahrscheinlich aber auch keines; geschieht es, so spricht das Zählrohr an und betätigt über ein Relais ein Hämmerchen, das ein Kölbchen mit Blausäure zertrümmert. Hat man dieses ganze System eine Stunde lang sich selbst überlassen, so wird man sich sagen, dass die Katze noch lebt, wenn inzwischen kein Atom zerfallen ist. Der erste Atomzerfall würde sie vergiften haben. Die Psi-Funktion des ganzen Systems würde das so zum Ausdruck bringen, dass in ihr die lebende und die tote Katze (s.v.v.) zu gleichen Teilen gemischt oder verschmiert sind. Das Typische an solchen Fällen ist, dass eine ursprünglich auf den Atombereich beschränkte Unbestimmtheit sich in grobsinnliche Unbestimmtheit umsetzt, die sich dann durch direkte Beobachtung entscheiden lässt. Das hindert uns, in so naiver Weise ein „verwaschenes Modell“ als Abbild der Wirklichkeit gelten zu lassen. An sich enthielte es nichts Unklares oder Widerspruchsvolles. Es ist ein Unterschied zwischen einer verwackelten oder unscharf eingestellten Photographie und einer Aufnahme von Wolken und Nebelschwaden.“ – Erwin Schrödinger (*Naturwissenschaften*. 48, 807; 49, 823)