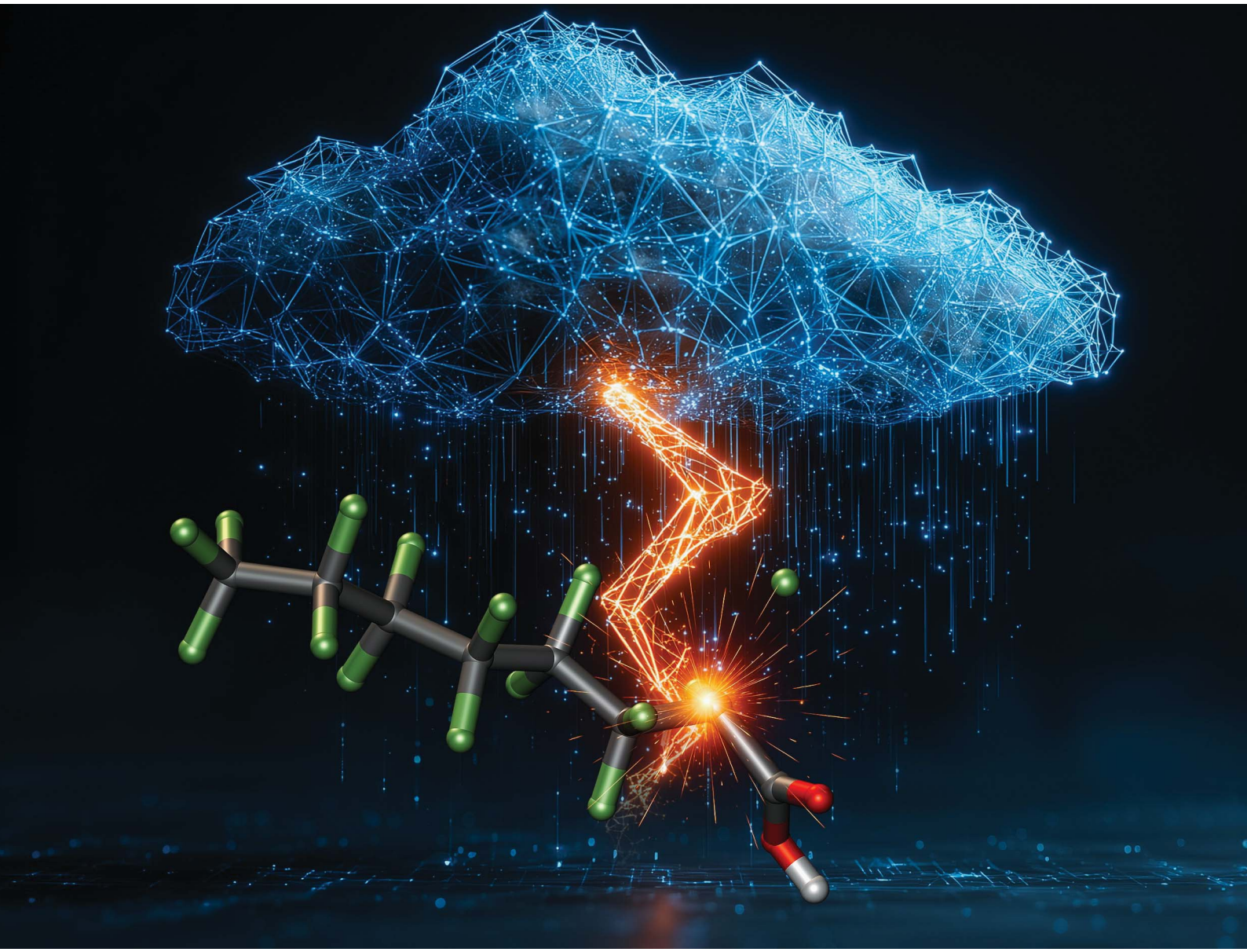
Volume 16
Number 41
7 November 2025
Pages 18961–19470

Chemical Science

rsc.li/chemical-science



ISSN 2041-6539

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

Rudi Plesch, Takeshi Yamazaki *et al.*

Breaking down per- and polyfluoroalkyl substances (PFAS): tackling multitudes of correlated electrons

15 YEARS ANNIVERSARY

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Sci.*, 2025, **16**, 19099

All publication charges for this article have been paid for by the Royal Society of Chemistry

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Breaking down per- and polyfluoroalkyl substances (PFAS): tackling multitudes of correlated electrons

Alan E. Rask,[a] Lee Huntington,[a] SungYeon Kim,[a] David Walker,[a] Andrew Wildman,[a] Rodrigo Wang,[a] Nicole Hazel,[a] Alan Judi,[b] James T. Pegg,[a] Punit K. Jha,[a] Zara Mayimfor,[b] Carl Dukatz,[c] Hassan Naseri,[c] Ilan Gleiser,[d] Maxime R. Hugues,[d] Paul M. Zimmerman,[ae] Arman Zaribafiyan,[a] Rudi Plesch *[a] and Takeshi Yamazaki *[a]

A tractable approach to solving the exact many-body electronic wavefunction has long remained an elusive goal in quantum chemistry. Accurate computation of the electronic structure and related properties of molecules would unlock a trove of precise information, but the exponential scaling of exact methods has impeded this possibility. In this work, we report that the combination of an incremental expansion of electronic correlation with effective utilization of modern cloud compute environments can overcome the long-standing barrier of intractability. As a demonstration of viability, we investigate the bond breaking of per- and polyfluoroalkyl substances (PFAS), a class of strongly correlated systems with pressing environmental relevance. Using the incremental Full Configuration Interaction (iFCI) method, we decompose the many-body wavefunction into independently computable units and distribute them across a record-breaking one million simultaneous cloud vCPUs. Calculations for perfluorooctanoic acid, the largest PFAS studied, involve correlating 150 electrons in 330 orbitals, corresponding to a wavefunction dimension of $\sim 10^{151}$ configurations and producing the most accurate correlation energies and electron densities to date. iFCI reveals an electron localization transition in the electronic state during bond dissociation, which standard quantum chemical methods (*e.g.*, DFT) fail to capture due to insufficient treatment of correlation. The union of a highly parallel cloud-enabled infrastructure with a polynomial-scaling method that treats static and dynamical correlation on equal footing lays the foundation for further development of novel protocols for the degradation of PFAS, alongside potential for characterization and screening of advanced molecules and materials in a wide range of chemistries.

Received 25th April 2025
Accepted 16th August 2025

DOI: 10.1039/d5sc03019f

rsc.li/chemical-science

Introduction

Quantum mechanics and quantum chemistry provide the fundamental rules for describing chemical processes. Our basic understanding of chemistry, material science, and life science rests on these foundations, and underpins our exploration of important physical phenomena relevant to the many facets of chemistry. In general, quantum chemical approaches simulate these phenomena by solving the many-body Schrödinger equation.[1–3] Unfortunately, computing the exact solution to this problem is only possible for the smallest molecular systems due to extremely high computational demand.[4] Contemporary chemical problems cannot be approached at the highest level of accuracy with modern electronic structure theories, which are either unscalable or too approximate for practical use.

Per- and polyfluoroalkyl substances (PFAS) are one such example of relevant chemical systems, persistent pollutants that have been dubbed "forever chemicals" because of their resilience and accumulation in humans and the environment.[5–7] One of the most common PFAS, perfluorooctanoic acid (PFOA), would require approximately 10^{151} electron configurations to capture the exact energy with full configuration interaction (FCI). This benchmark-level many-body wavefunction would correlate all 150 valence electrons in 330 orbitals (assuming a polarized double-zeta basis), while the largest fully variational CI to date has considered a maximum of $\sim 10^{15}$ determinants.[8,9]

Typical approximations with reduced computational costs lose the ability to reliably treat electron correlation, or introduce significant uncontrolled approximations (*i.e.*, non-black-box input).[10] For example, conventional single-reference methods, such as coupled-cluster theory,[11] are not expected to properly characterize bond breaking. Non-black-box, multireference

[a]*SandboxAQ, 780 High Street, Palo Alto, CA, 94301, USA. E-mail: rudi.plesch@sandboxaq.com; takeshi.yamazaki@sandboxaq.com*

[b]*Good Chemistry Company, 200-1285 West Pender Street, Vancouver, BC, V6E 4B1, Canada*

[c]*Accenture, 1 Grand Canal Quay, Grand Canal Dock, Dublin, D02 P820, Ireland*

[d]*Amazon Web Services, 410 Terry Ave N, Seattle, WA, 98109, USA*

[e]*Department of Chemistry, University of Michigan, 930 N. University Ave, Ann Arbor, 48109, USA*

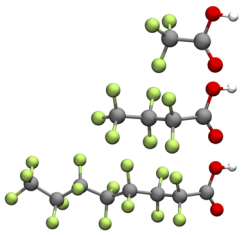
© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

(MR) approaches can correctly break bonds, but with quantitative results that depend on user-selected inputs. In the case of PFAS, simulating bond dissociation is further challenged by their complex electronic structure: non-dynamical correlation arises when a single-reference wavefunction fails during bond breaking, and dynamical correlation remains significant due to non-perturbative interactions among breaking bonds, lone pairs, and virtual orbitals. Accurately capturing these correlations requires simultaneous treatment of many electron configurations – involving 18 electrons per CF_2 unit – beyond what is amenable to CC or MR approaches. Furthermore, fluorine's electronegativity produces strong C–F bonds, setting PFAS apart from more-commonly studied systems, and necessitating a thorough investigation of the dissociation manifold. In recent years, progress has been made in inventing new chemistries to break down PFAS,[12–14] showing that they are not truly immortal chemicals, and motivating the development of computational tools that can guide such efforts. In order to address these problems, a method that is both scalable and highly accurate must be identified.

The incremental full configuration interaction (iFCI) method is a novel approach to tackle large-scale electronic structure problems with high accuracy.[15,16] For example, iFCI has previously been shown to generate smooth dissociation profiles for both single- and multi-bond breaking scenarios and can also be used to target excited states in transition metal complexes.[17–19] iFCI uses a many-body decomposition of the energy of a molecular system within a localized orbital basis to reproduce the FCI limit of energies and electron densities. This decomposition reduces the problem from exponential to polynomial scaling, in principle enabling simulation of much larger systems, and requires no specialized input such as the choice of an active space. Even with the polynomial cost of iFCI, however, a system such as PFOA would require order one million (10^6) many-body terms with an average of 10^8 dimensions per term. PFOA is therefore intractable to study with iFCI in its original implementation.[15–17] Here, we will show that a combination of highly parallel computation and correlation screening can dramatically reduce the costs and runtime of iFCI computations on larger systems. Within one day of walltime and distributed across over 1 million cloud vCPUs, the new iFCI algorithm approaches the FCI limit for PFOA (10^{151} dimensions).

In this work, we describe a massively parallel implementation of iFCI that can simulate the wavefunctions of molecules the size of PFAS within realistic time frames. Three representative PFAS – trifluoroacetic acid (TFA), perfluorobutanoic acid (PFBA), and perfluorooctanoic acid (PFOA) (Fig. 1) – were selected based on their environmental relevance and the difficulty of studying these systems with traditional electronic structure theory.[20] In particular, describing reactions involving the C–F bond is important for understanding natural and induced PFAS degradation pathways, as well as designing potential catalysts for PFAS remediation. This is the first quantum chemistry calculation to use the cloud to scale an accurate many-body wavefunction to systems of this size, opening up a wide scope of future possibilities for precise insight into strongly correlated molecules and materials.



Name	Active e⁻ / orbitals	FCI determinants
TFA: trifluoroacetic acid	42 / 96	10^{42}
PFBA: perfluorobutanoic acid	78 / 174	10^{78}
PFOA: perfluorooctanoic acid	150 / 330	10^{151}

Fig. 1 PFAS species under consideration in this work. Color scheme: green (fluorine), red (oxygen), gray (carbon), white (hydrogen).

Results and discussion

The formal iFCI approach used to describe PFAS is delineated in previous works[15–17] and in the Computational methods section. The core principle is to gather together small groups of electrons (for instance 2, 4, 6…electrons at a time), compute their exact correlation energies within each group, and sum up these standalone contributions to the total energy and electron density. Importantly, contributions from groups of 4 electrons are additive, and do not double-count contributions repeated from their 2-electron subsets, meaning that the method is formally size extensive. Higher-order terms involving many (>8) electrons are therefore vanishingly small, and can be neglected. Additionally, some lower-order terms may be systematically screened, focusing the iFCI computation on only those terms that significantly contribute to the total iFCI wavefunction. The Computational methods section describes this process in greater detail.

Despite the relative compactness of the iFCI wavefunction compared to the astronomically sized FCI wavefunction, an efficient parallel algorithm is required to scale iFCI up to compute the large PFAS molecules of this work. To enable this study, we developed a massively parallel implementation of iFCI within QEMIST Cloud (**https://www.sandboxaq.com/solutions/quantum-simulation**). QEMIST Cloud provides accurate methods to predict the electronic structure and properties of chemical systems *via* traditional *ab initio* quantum chemistry and machine learning methods, and now includes the iFCI method for analysis of strongly correlated systems. For this work, our software was implemented using Amazon Web Services (AWS), using Amazon Elastic Compute Cloud (Amazon EC2) Instances with Intel Xeon Scalable processors. Further details are available in the Computational methods section of this article. This new implementation allows iFCI to be performed on an almost arbitrary number of processors, up to and beyond the 1 million vCPUs employed in this work.

PFAS bond-breaking simulation

The challenging electronic characteristics of PFAS (see the Introduction) – including strong C–F bonds, dense networks of lone pairs, and delocalized electron density – require powerful electronic structure methods to accurately predict their chemical properties. Prior simulations have primarily utilized density

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

functional theory (DFT) to describe the bond breaking processes in PFAS, but the results vary widely across functionals.[21,22] High-accuracy electronic structure calculations have been limited to small molecules,[22] and the "gold standard", CCSD(T)[23] (coupled cluster with single and double excitations with perturbative triples), does not correctly treat stretched bonds or diradical systems.[10,11,24,25]

Multiconfigurational methods like MR-CI or CASSCF offer more flexibility for describing non-dynamical correlation, but they rely on predefined active spaces.[26,27] In systems like PFAS, where bond breaking induces collective correlation effects across many orbitals, any practical active space becomes either too small to be accurate or too large to be tractable. For example, a CASPT2(2,2) treatment might focus only on the breaking σ_{CF} bond, while neglecting the extended impact that the breaking bond may have throughout the molecule. Expanding the active space incrementally leads to an open-ended problem with no clear convergence, as new interactions continue to emerge with each addition.

In contrast, the iFCI approach includes all electrons and systematically incorporates correlation across the full virtual space, without requiring an arbitrary choice of the active space. This allows iFCI to capture the many-body, delocalized nature of PFAS bond dissociation with high fidelity. Enabled by a massively parallel implementation, iFCI offers a tractable yet fully correlated description of these chemically and computationally complex systems.

Bond dissociation energies and potential energy surfaces

The iFCI energies of the equilibrium and stretched geometries for three PFAS molecules were computed with iFCI up to the 4-body expansion ($n = 4$), which correlates all combinations of four occupied orbitals with the full set of virtual orbitals. The 4-body expansion of iFCI has previously been shown to closely approach the FCI limit, with the total $n = 4$ contribution less than 10 mHa for both test systems and transition metal complexes, well within chemical accuracy.[15,19] These results therefore provide an accurate bond dissociation energy for the rigid-body stretch of the bond between a fluorine and the α-carbon on the carboxylic acid group, among other chemical properties. The full potential energy surface along the bond dissociation coordinate of TFA and the equilibrium and dissociated geometries of PFBA and PFOA were computed and are summarized in Fig. 2 (see also SI, Table S2). The BDEs of each species show a similar trend with increasing n, and the relatively small change in BDE from $n = 3$ to 4 reflects good internal convergence of iFCI (see SI, Fig. S5).

Direct experimental measurements of BDEs for the specific PFAS studied here are not available in the literature, which limits direct comparison. However, data from related systems offer useful reference points. Typical C–F bonds fall in the range of 110 to 116 kcal mol^{-1}, with fluorocarbons reported to have a C–F bond strength of up to 127 kcal mol^{-1}.[28–31] These values are in good agreement with our iFCI results for all three PFAS and confirm that our results reflect expected bond strengths of comparable systems.

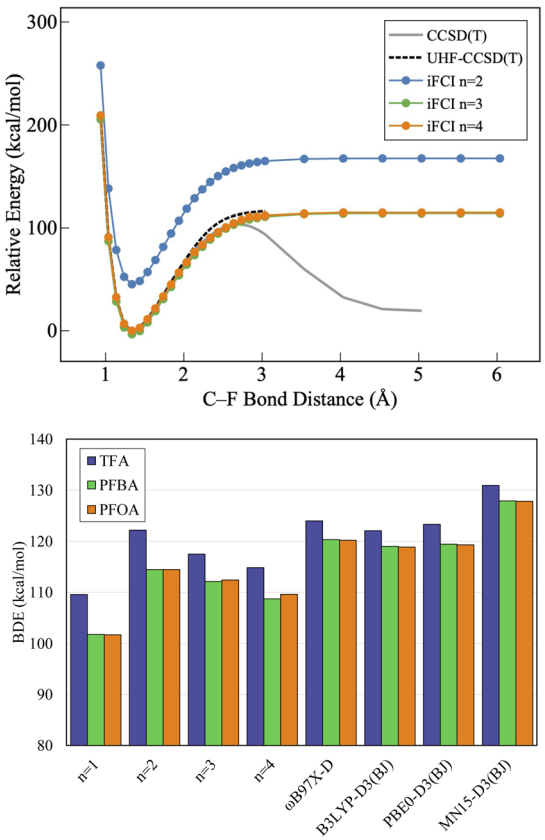


Fig. 2 Potential energy surface of TFA C–F dissociation (top) and bond dissociation energies for all three PFAS from $n = 1$–4 and various DFT functionals (bottom). All energies on the PES are displayed relative to the iFCI $n = 4$ equilibrium geometry energy. The PES for each DFT functional is shown in Fig. S2 of the SI.

Among the PFAS tested, TFA expresses the largest BDE, while PFBA and PFOA have similar BDEs at all n. The similarity of BDE between the longer-chain PFAS indicates that additional CF_2 units have relatively small effects on the dissociation of a distant C–F bond. This observation is reasonable given that the local geometry in these two molecules is similar for the chosen dissociated bond, and the electronic environment is therefore similar as well. However, a small effect of the chain length on the BDE arises at higher n, with PFOA's BDE at $n = 4$ being almost 1 kcal mol^{-1} above PFBA's. This may be evidence of a slight stabilization of the radical within longer-chain PFAS.

The BDE of various functionals using unrestricted DFT are compared to iFCI in Fig. 2. Overall, DFT agrees with the general trend of the longer-chain PFAS possessing lower BDEs, but all functionals consistently overestimate the BDE to varying degrees compared to iFCI. The lower BDEs of longer-chain PFAS is consistent with experimental studies, where longer-chain PFAS have been found to exhibit faster decomposition rates.[32–34] However, the difference between the BDE of short- *versus* long-chain PFAS estimated by DFT is much smaller than iFCI, probably due to the semi-local treatment of exchange-correlation in DFT, which does not treat long-range correlation effects. In a similar fashion, the correlation retrieved in

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

iFCI $n > 2$ results in a slight widening of the difference between PFBA and PFOA, while DFT resembles $n = 2$ with PFBA and PFOA having nearly the same BDE.

The overestimation of the BDE by DFT can be contextualized in terms of half-life using the Arrhenius equation,[35–37] under the simplifying assumption that the BDE and activation energy are equivalent. For example, the 10.6 kcal mol^{-1} difference between DFT (ωB97x-D) and iFCI ($n = 4$) for PFOA equates to a predicted half-life of bond breaking to be $10^{8.5}$ times longer for DFT compared to iFCI (at 273 K). The difference in relative strength of the C–F bond and overall overestimation of the BDE by DFT, along with the varying range of BDE across functionals, may impact the study and design of PFAS remediation strategies.

TFA's iFCI potential energy curve for C–F dissociation shows physically correct behavior for a stretched bond for all n (Fig. 2), a feature usually found only in multireference methods with careful selection of active space. The $n = 3$ and 4 curves agree well with CCSD(T) and UHF-CCSD(T) around the equilibrium region. Since the CCSD(T) curve becomes unphysical upon dissociation, a meaningful comparison of the BDE cannot be made. As reported in literature,[10,24] the UHF-CCSD(T) method provides marked improvement over RHF-CCSD(T). However, UHF-CCSD(T) still exhibits a large difference compared with iFCI (and CCSD(T)) in the intermediate region as the bond begins to dissociate. The accuracy of the UHF-CCSD(T) energy improves as the bond is stretched further toward dissociation, but it was not possible to converge the UHF-CCSD equations – and therefore also not possible to obtain UHF-CCSD(T) energies – for geometries beyond $R_{C-F} = 3.0$ Å. Additionally, we were not able to obtain UHF-CCSD(T) results at any geometry for PFOA due to the high computational demands of a system of this size. For PFBA and PFOA, iFCI calculations of the equilibrium and dissociated geometries were used to determine the BDE, though we did not compute the full potential energy profiles. These species are expected to express similar behavior in the dissociation region when compared to TFA (see Fig. 2). Comparing each method, iFCI $n = 4$, CCSD(T), and UHF-CCSD(T) exhibit similar energies at equilibrium (all within 0.2 kcal mol^{-1} of each other), but both CC methods deviate from iFCI by >1 kcal mol^{-1} at 1.7 Å and higher. Further comparison between iFCI, CCSD(T), and UHF-CCSD(T) can be found in the SI.

A metric to indicate the accuracy of the iFCI calculations is the non-parallelity error (NPE), which measures differences (maximum *vs.* minimum error) between potential energy curves. As shown in Table 1, the NPE systematically decreases with increasing order n of the n-body expansion, indicating good convergence of iFCI. The NPE between $n = 3$ and $n = 4$ is smaller than a few kcal mol^{-1}, suggesting the 4-body expansion approaches the FCI result. For TFA, the maximum and minimum errors were located near the equilibrium and dissociated geometries, and therefore the NPE for PFBA and PFOA may be estimated given the error at their respective equilibrium and dissociated geometries. The longer-chain PFAS also follow a similar trend to that of TFA with increasing n, indicating a comparable convergence of iFCI for the larger species.

The NPEs of other methods when compared to the FCI solution are typically on the order of tens of kcal mol^{-1}, with

Table 1 Non-parallelity error (NPE) analysis of PFAS potential energy curves (kcal mol^{-1}), each made with iFCI ($n = 4$) as a reference. iFCI exhibits good convergence with increasing n. The NPE of UHF-CCSD(T) is comparatively small, but does not converge at larger bond distances. Although the generally large value for all DFT results is as a result of being compared with $n = 4$, they vary greatly between both functionals and molecules

Method		TFA	PFBA	PFOA
iFCI	$n = 1$	11.7	6.9	7.9
	$n = 2$	9.1	5.7	4.8
	$n = 3$	3.6	3.4	2.8
CCSD(T)		95.1	71.4	64.2[a]
UHF-CCSD(T)		8.8	—	—
ωB97x-D		16.9	11.6	10.6
B3LYP-D3(BJ)		12.5	10.3	9.3
PBE0-D3(BJ)		16.8	10.7	9.7
MN15-D3(BJ)		19.4	19.2	18.2

[a] DF-CCSD(T).

methods using a restricted reference experiencing large errors at dissociation.[10,24,38] For example, the NPE of CCSD(T) is 95 kcal mol^{-1} compared to iFCI $n = 4$ (see Table 1) due to its failure to produce a qualitatively correct electronic state at dissociation. Restricted DFT expresses similar problems. Unrestricted methods handle dissociation more accurately – UHF-CCSD(T) and UHF-DFT demonstrate NPEs of <20 kcal mol^{-1} – but are nowhere near chemical accuracy. The maximum errors in UHF-CCSD(T) arise from the unphysical behavior in the intermediate region, and the maxima for UHF-DFT are in the dissociation region (see SI, Fig. S2). Only iFCI treats all regions on close to equal footing.

Analysis of orbital interactions

In addition to an accurate description of bond breaking, iFCI provides the means to assess the impact of individual molecular orbital interactions on the overall electron state. Although iFCI does not construct a single, global wavefunction in the traditional sense, it constructs a collection of converged many-body expansions of all physically relevant observables within a given basis set, including energies, density matrices, and other expectation values. The one- and two-particle reduced density matrices produced by iFCI are equivalent to those of full configuration interaction (FCI) at convergence, and thus share the same interpretability as the exact non-relativistic Born–Oppenheimer wavefunction in that basis. Here, we analyze correlation contributions in a localized molecular orbital basis to provide chemically intuitive interpretations of bonding changes during dissociation. As with all localization schemes, the choice of basis is somewhat arbitrary, but the resulting orbitals align with conventional bonding descriptions, such as σ_{CF} and π bonds.

Each n-body term is uniquely identified by a combination of occupied orbitals (X), and the correlation energy for each term (ϵ_X in eqn (1)) is a measure of the interaction of electrons within these orbitals. The localized orbital basis allows for a direct comparison between terms of the equilibrium and dissociated

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

geometries. For example, the correlation from a 1-body term defined by the C–F σ bond on the α-carbon can be represented as $\epsilon_{equi}(\sigma_{CF})$. Although dissociation of this bond changes its appearance, the orbital retains the same σ_{CF} character, and therefore the contribution to the BDE from this bond can be calculated through simple subtraction, $\epsilon_{diss}(\sigma_{CF}) - \epsilon_{equi}(\sigma_{CF})$.[18,19]

Orbital interaction analysis was performed by grouping the localized orbitals of PFOA into general characterizations, or classes, with chemically relevant features: σ_{CF}, σ_{CC}, and lp_F (orbitals present on the carboxyl group are excluded from this analysis for brevity, as they do not provide a substantial difference to the total BDE). Contributions to the bond dissociation energy of PFOA from these orbital classes are shown in Fig. 3, where each class (or combination of classes for $n > 1$) represents the summed contribution to the BDE for all terms that fit that description. For example, σ_{CF} represents the contribution from 1-body terms that correspond to all σ bonds between C and F. Likewise, lp_F/σ_{CF} represents the contribution from all 2-body terms that correspond to the interactions between one of the lone pairs of a fluorine atom and a C–F σ bond.

As expected, the σ_{CF} class shows the most significant contribution to the BDE by far (-197 kcal mol^{-1}), a natural result of the dissociating C–F bond. However, in n-body terms with combinations of σ_{CF} and other orbitals, the BDE is increased by some and decreased by others. For example, the sum of lp_F/σ_{CF} 2-body interactions widens the BDE by $\sim$9 kcal mol^{-1}, while the sum of $lp_F/\sigma_{CF}/\sigma_{CF}$ narrows the BDE by $\sim$2 kcal mol^{-1}. Overall, the BDE is determined by a subtle balance between orbital interactions, where the presence of notable contributions to the BDE (>1 kcal mol^{-1}) at $n > 2$ indicates that highly-excited configurations – which are not present in other common computational methods, such as CCSD(T) – significantly influence the electronic state. Further analysis of orbital-specific contributions from the many-body expansion may provide future insights in rational design of PFAS remediation processes, such as identifying relative strengths of bonds to calculate the energetics of PFAS degradation.[22,39,40]

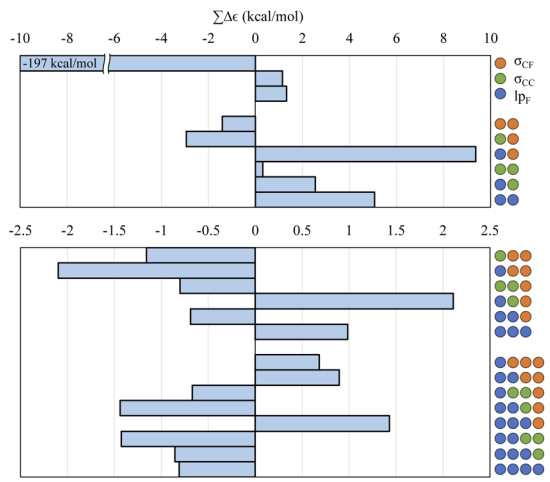


Fig. 3 Total contributions to the BDE from grouped combinations of orbitals for PFOA. Orbital classes are paired with colors to aid in readability: σ_{CF} is orange, σ_{CC} is green, lp_F is blue. Combinations of orbitals that result in $|\sum\Delta\epsilon| < 0.5$ kcal mol^{-1} have been removed for clarity. Positive sums result in a larger overall change to the BDE (*e.g.*, lp_F/σ_{CF} interactions cumulatively raise the BDE by 9.4 kcal mol^{-1}).

Electronic localization transition with C–F bond breaking

The central challenge of describing bonding changes in PFAS molecules is the simultaneous presence of dynamical and non-dynamical correlation. Both types of correlation result in qualitatively different structures of natural orbitals (NOs), which are the orbitals that diagonalize the 1-electron density matrix of a wavefunction. In this section, it will be shown that the NOs undergo a transition from delocalized to localized structures upon bond breaking, and that this transition is the natural outcome of the full treatment of correlation by iFCI.

According to valence bond theory, a single covalent bond comes as a pair of one bonding orbital with a corresponding anti-bonding orbital. The NOs of TFA and PFOA's equilibrium geometry fit this description for iFCI at $n = 1$ (Fig. 4), *i.e.*, as a σ/σ^* pair. At $n > 1$, correlation between this bond and all other bonds causes the bonding picture to become more delocalized, representing the close coupling between many pairs of the $n = 1$ localized orbitals. At least at the equilibrium geometry, individual and distinct σ_{CF} bonds are representative of simple theories (like the GVB-PP reference state) that are relatively uncorrelated. iFCI recovers the correct picture of orbitals with mixed, noninteger occupancy of delocalized orbitals across the molecule.

The electronic state for a molecule with a dissociated bond, *i.e.*, a diradical, has a combination of the delocalized, weakly

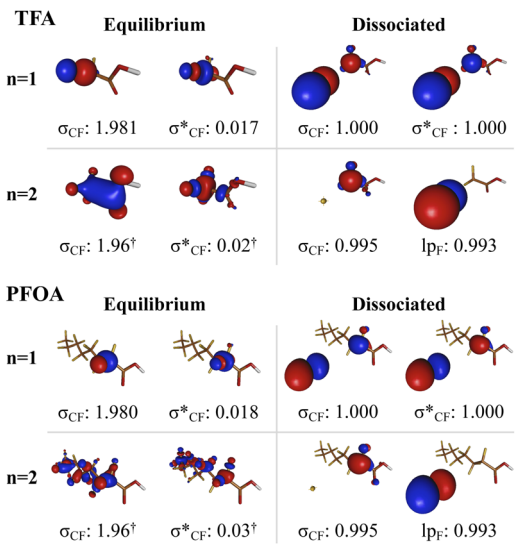


Fig. 4 Natural orbitals and occupation numbers of the dissociated σ_{CF} bond in TFA and PFOA at $n = 1$ and 2. Orbital spaces are mixed at $n = 2$ due to the correlation of more than one orbital at a time at $n > 1$, leading to a fully delocalized σ_{CF} space in the equilibrium geometry. For brevity, only one of these delocalized σ_{CF} NOs are shown and their occupation numbers are the average across the NOs exhibiting this bond (indicated by the † symbol). The difference in appearance of singly occupied NOs between $n = 1$ and 2 is a result of different $\pm$ combinations of electron densities, where the former appears as a σ/σ^* pair and the latter as a p/lp pair.

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

correlated states and localized, strongly correlated portion. Computational methods such as DFT have difficulty spanning the two regimes, as they prefer to have the same level of delocalization regardless of electronic state (see SI, Fig. S3). The iFCI $n = 1$ NOs also do not show the correct behavior (two localized $1e^-$ states) because they are not correlated with electrons outside of the σ/σ^* pair, and their portrayal is therefore a holdover from the reference orbitals. However, once correlation across all orbitals is introduced at $n > 1$, the NOs spanning the dissociated bond become localized with one unpaired electron on each fragment, which is the qualitatively correct description of the correlated state. (Their occupation numbers deviate slightly from 1 due to weak correlation with neighboring lone pairs and bonds.) As shown in Fig. 4, the covalently bonded orbitals of the dissociated PFAS fragment remain delocalized at $n = 2$, as expected. In all $n = 2$ NOs, iFCI provides the correct description of the two regimes – and together with the smooth potential energy profile of Fig. 2 – a smooth transition from one distribution of the electronic localization to the other. (Note, orbital spaces are further correlated at $n > 2$ but the degree of delocalization remains constant at higher n, and we therefore limit our discussion to just $n = 1$ and 2.)

Conclusions

This work demonstrates the viability of tackling the full spectrum of dynamical and non-dynamical correlation in realistic PFAS molecules using a massively parallel cloud-based implementation of the iFCI method, involving an orchestration of over 1 million simultaneous vCPUs. iFCI provides a physically correct depiction of all intermediate states involved in the reaction process, a feature lost when using single reference methods (such as DFT or CC theories). Careful screening of long-distance, many-body correlations to reduce the workload of iFCI is combined with scaling up *via* effective use of cloud resources to achieve these previously unobtainable results. The ability to bring together any number of processors (for instance the 1 million vCPUs of this work) makes the continued use of iFCI tenable.

These calculations constitute the first quantum chemical method to make effective use of the incredible scaling capabilities of the cloud, and are the most accurate electronic structure calculations on PFAS species to date. Calculations of this kind lay the foundation for further development of novel protocols for the degradation of PFAS, with the potential next step being the integration of this method with other pivotal methodologies in the study of PFAS bond-breaking chemistry (*e.g.*, reaction path prediction, machine learning based screening, surface reactions, solvation, *etc.*).

With the emerging availability of scalable iFCI simulations, it will be possible to develop highly accurate benchmark datasets that include weakly and strongly correlated molecular systems. These datasets will provide valuable training data for the development of universal machine-learning force fields[41–44] that could cover a wide range of molecular systems and chemical behaviors that conventional methods are not able to accurately describe. This difference is expected to be dramatic, as most ML force fields to date are trained on DFT data. In the long term, the availability of this high-quality training data will mean the fast and accurate force fields for strongly correlated systems could become commonplace, exponentially increasing the impact of the output from iFCI simulations.

Computational methods

Incremental full configuration interaction (iFCI)

iFCI calculations were performed with QEMIST Cloud using the cc-pVDZ basis set. The incremental full configuration interaction approach[15–19] makes use of an incremental or many-body expansion of the correlation energy in terms of occupied orbitals labeled by $i, j, k, l,\ldots$

$$E_c = \sum_i \epsilon_i + \sum_{i>j} \epsilon_{ij} + \sum_{i>j>k} \epsilon_{ijk} + \sum_{i>j>k>l} \epsilon_{ijkl} + \ldots, \quad (1)$$

in which ϵ_i, ϵ_{ij}, ϵ_{ijk}, ϵ_{ijkl} are respectively, one-, two-, three- and four-body incremental energies defined as

$$\epsilon_i = E_c(i) \quad (2)$$

$$\epsilon_{ij} = E_c(ij) - \epsilon_i - \epsilon_j \quad (3)$$

$$\epsilon_{ijk} = E_c(ijk) - \epsilon_{ij} - \epsilon_{ik} - \epsilon_{jk} - \epsilon_i - \epsilon_j - \epsilon_k \quad (4)$$

$$\epsilon_{ijkl} = E_c(ijkl) - \epsilon_{ijk} - \epsilon_{ijl} - \epsilon_{ikl} - \epsilon_{jkl} - \epsilon_{ij} - \epsilon_{ik} - \epsilon_{il} - \epsilon_{jk} - \epsilon_{jl} - \epsilon_{kl} - \epsilon_i - \epsilon_j - \epsilon_k - \epsilon_l. \quad (5)$$

Here, $E_c(X)$ are the correlation energies obtained from the space of X occupied orbitals correlated with the full set of virtual orbitals. For example, $E_c(ij)$ indicates solving CISDTQ in a space of i, j, and all virtual orbitals. By extension, four-body incremental energies provide correlations for up to four occupied orbitals (eight electrons) at a time. This notation is simplified throughout this work with n representing the maximum number of occupied orbitals that are correlated in a computation. For example, $n = 3$ refers to an energy that includes the summation of all combinations of three occupied orbitals that are correlated with the full set of virtual orbitals ($\sum_i \epsilon_i + \sum_{i>j} \epsilon_{ij} + \sum_{i>j>k} \epsilon_{ijk}$). We also discuss contributions from individual orbitals, combinations of orbitals, or classes of orbitals (*e.g.*, ϵ_{ijk}), and refer to them as n-body terms.

In order to make calculations tractable for larger systems, the incremental correlation energies $E_c(i)$, $E_c(ij)$, $E_c(ijk)$,… are computed using the Heat-Bath Configuration Interaction (HBCI) approach,[45] which is an efficient selected configuration interaction method that constitutes as an approximation to full configuration interaction. In this method, inspired by heat bath sampling, a number of the most important determinants are selected for a variational configuration interaction treatment. The determinant space is then expanded and second-order Epstein–Nesbet perturbation theory is performed to correct for the missing correlation effects. Two parameters are used to control the selection of determinants for the variational and perturbative calculations, ε_1 and ε_2 respectively. For the calculations reported herein, we used values of $\varepsilon_1 = 1$ mHa and $\varepsilon_2 = 1$ μHa.

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

To further increase the efficiency of calculations without sacrificing accuracy, we make use of Summation Natural Orbitals (SNO),[17] which facilitate a selection of the virtual orbitals for each n-body term and discard a number of them that have little effect on the correlation energy. Namely, for each 1-body term, the virtual space is spanned by natural orbitals of the exact CISD 1-body RDM (*i.e.*, CISD is exact for 2-electron systems such as a 1-body term) and the natural orbitals for which the occupation numbers fall below a threshold $10^{-\eta}$ are discarded. For higher n, a density is formed as the sum of the one-particle CISD density matrices for each occupied orbital included in the given n-body term. For example, a 3-body term would form a density as the sum of the 1-body densities with the same set of occupied orbitals in the form

$$\rho_{ijk} = \rho_i + \rho_j + \rho_k. \quad (6)$$

Then in analogous fashion to the 1-body case, the virtual space for the 3-body term is spanned by natural orbitals of this SNO density and the natural orbitals with occupation numbers falling below a threshold $10^{-\eta}$ are discarded. The final HBCI calculation is then performed in the truncated virtual space. Herein, we used $\eta = 8.5$ as a middle ground of accuracy and efficiency, as 9.5 did not show significantly different energies for TFA (*e.g.*, correlation retrieved at $n = 3$ differed by <3 mHa between $\eta = 8.5$ and 9.5).

An additional way to reduce the computational cost of calculations without having a significant effect on accuracy is to screen certain higher order terms entering the incremental expansion of eqn (1). As the size of the system increases, the number of terms entering eqn (1) at $n \leq 3$ starts to become large and many of the incremental energies have negligible contributions to the total correlation energy, especially as n increases. Hence, we can use the connectivity arguments discussed in detail in ref. 18 and 19 to discard the less important contributions to the incremental expansion for $n \geq 3$. We chose a value of $10^{-4.6}$ Ha for the screening parameter (*i.e.*, the connected n-body screening procedure is applied to $n = 3$ and 4 with a value of $\mathscr{C} = 10^{-4.6}$Ha) as a good middle-ground because correlation remains to be captured at looser cutoffs, while tighter cutoffs introduce significantly more terms with negligible change to the total energy. Further discussions on this connected n-body (CnB) truncation procedure are provided in the SI.

Our implementation of iFCI makes use of a zeroth order reference wavefunction from Generalized Valence Bond, Perfect Pairing (GVB-PP) theory.[46–48] This constitutes one of the simplest multi-configurational wavefunctions to include static (non-dynamical) correlation and has been shown to be a good reference wavefunction for iFCI calculations on strongly correlated systems[15] and transition metal complexes.[18] As mentioned in ref. 15, GVB-PP provides an excellent localized basis for the incremental expansion of eqn (1), and since the PP ansatz includes correlation beyond Hartree–Fock (*i.e.*, $E_{\text{GVB-PP}} < E_{\text{HF}}$), it also provides a better starting point for the incremental expansion in iFCI.

In our initial GVB-PP calculations on TFA, we found it difficult to converge to the same state for different geometries at larger C–F separation due to a discontinuity in the potential energy surface. This was mostly due to the presence of many degenerate electronic states in this region, evidenced by a change in orientation of the bonding and lone-pair orbitals on the dissociating fluorine as the bond is stretched. Therefore, we introduced a small perturbation in the form of a point charge (charge of −1 a.u.) along the dissociating C–F bond direction at a constant distance of 5 Å from the F atom, which successfully corrected the orientation of the lone-pair orbital, provided a smooth potential energy surface across the full range of geometries, and changed the GVB-PP energy by less than a few mHa. The perturbation was removed in the calculation of the reference energy for the iFCI expansion – the energy of a single determinant computed in the basis of the converged GVB-PP MOs – and the one-electron operator used in the calculation of the single determinantal energy was unperturbed, though the converged GVB-PP MOs used for this calculation were still perturbed to retain the proper orientation of the dissociating fluorine orbitals. Similarly, the perturbation was not included in the final iFCI calculation, even though the GVB-PP orbitals used in the calculation were perturbed to retain the proper orientation of orbitals of the dissociating fluorine atom. We confirmed that this procedure had negligible effect on the final iFCI ($n = 3$ and 4) energy for the equilibrium region of the potential energy surface. This is not surprising, as for a given starting set of localized MOs (*e.g.*, GVB, localized HF MOs, *etc.*), the iFCI expansion will converge to the HBCI result as n is increased. This procedure was utilized for all three molecules considered in this work.

QEMIST cloud architecture design

The massively parallel framework built for the iFCI method was implemented within QEMIST Cloud (**https://www.sandboxaq.com/solutions/quantum-simulation**), a computational chemistry software as a service (SaaS) that provides accurate methods to predict the electronic structure and properties of chemical systems with both traditional *ab initio* quantum chemistry and machine learning methods. For this work, our software was implemented using Amazon Web Services (AWS) as the platform of cloud-based computing resources, using Amazon Elastic Compute Cloud (Amazon EC2) Instances with Intel Xeon Scalable processors. An overview of the QEMIST Cloud architecture is shown in Fig. 5.

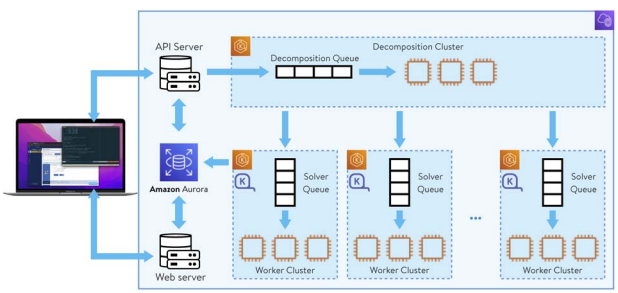


Fig. 5 QEMIST cloud architecture overview. The laptop represents a client, stacked white boxes represent queues, and orange CPU symbols represent individual workers. Each dashed box is a group of components that was deployed into a specific cluster.

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Before describing the technologies used to implement this software architecture at scale, we will first introduce some terminology. In modern cloud computing environments, physical hardware is transformed into "virtual machines", where a vCPU (virtual central processing unit) is a software-defined representation of a physical CPU core. An availability zone is a set of one or more data centers in a particular region. Nodes and instances refer to discrete allocations of hardware resources, and workers refer to processes that use some or all of these resources. A cluster is a collection of nodes, and a spot instance is some spare capacity on AWS that is available at up to a 90% discount compared to on-demand prices, but may be reclaimed by AWS before the calculation is finished (with a warning issued two minutes prior to reclamation).[49] The total cost of running iFCI depends on the number of n-body terms in the calculation and computational complexity of each term, as dictated by the number of virtual orbitals and the HBCI convergence threshold (more details are in the Methods section). As described in the last paragraph of this section, it is beneficial to optimize the usage between the on-demand and spot instances. Individual compute tasks were performed with standardized units of software, referred to as containers. Kubernetes performed distribution and execution of the compute tasks and containers.[50] Kubernetes was deployed using the AWS managed service, Amazon Elastic Kubernetes Service.

The architecture supporting the iFCI implementation is structured around the independence and natural parallelism of the subproblems generated by the method of increments.[17] In iFCI, the n-body correlation energy is composed of a summation of independent calculations, each correlating unique sets of n occupied orbitals. Due to the natural independence of each calculation and very low communication between calculations, traditional message passing interface parallelism is not required for these calculations. Instead, when our platform receives an iFCI calculation request, a dedicated decomposition worker is assigned and generates all required subproblems for the current n. The decomposition worker then submits subproblems to solver queues, where solver workers pick up and complete each subproblem, writing their results into a central database once finished. The number of solver workers scales dynamically so that larger calculations have more subproblems being solved simultaneously. Once all subproblems for a given n are completed, the decomposition worker sums the total n-body correlation energy and moves to the next subsequent n until the desired highest order of n is computed (usually not higher than $n = 4$). Each n is completed sequentially before moving on to $n + 1$ to reduce the computational cost *via* a screening procedure (see the Methods section).

To reach the scale required for the largest PFAS in this study (1 million simultaneous vCPUs, or 62 500 workers with 16 vCPUs each), this infrastructure required numerous enhancements. In general, these improvements fell into three categories: more efficient node recruitment, enhanced database connections, and widening of the network to accommodate a large number of nodes. A summary of the final architecture is given in Fig. 5.

In order to reach a given scale, enough nodes must be available. While cloud infrastructure providers generally have more than enough to reach the scale aimed for in this work, no one availability zone is guaranteed to have the required amount. To ensure adequate node availability, the solver queues and solver workers were moved from a single cluster into multiple clusters, which distributed the work of provisioning nodes and enabled node recruitment on a larger scale. All clusters were allowed to provision nodes from multiple availability zones on the infrastructure provider to ensure a deep pool of spot instances. Distributing across availability zones does not degrade the performance of the algorithm because there is low communication between nodes during the calculation. While this multi-cluster design allows for recruitment on a larger scale, it adds complexity to the queuing and monitoring systems, where each cluster's queue must be periodically rebalanced so that all clusters remain active.

While the scale of node recruitment was addressed by the previous modifications, the speed of recruitment also needed to be improved in order to reach large scale queuing in a short time. For this, the Karpenter open-source autoscaler (**https://karpenter.sh**) was used *in lieu* of the standard Kubernetes cluster autoscaler. Karpenter removes the need for an autoscaling group, allowing for node provisioning and removal to occur at a rapid rate. An additional benefit of changing to Karpenter was the ease of scheduling multiple workers on a variety of nodes, which improved spot instance availability across all clusters due to diversification of instance type and size.

In order to efficiently store the outputs from a massive number of worker nodes, Amazon Aurora was used in a "serverless" fashion with a highly scalable proxy in between (RDS proxy) that allows pooling and sharing of the database connections between nodes. This allowed us to handle large numbers of simultaneous connections (*e.g.*, when 62 500 or more workers are active). We optimized communication between nodes and the database by decreasing their size and frequency, preventing data transfer from becoming a hindrance to computation. Additionally, leveraging 1 million simultaneous vCPUs required a widened IP address range, achieved by deploying IPv6 networking for all components, which is novel for some of the components used.

All subproblems of order $n > 1$ were run on spot instances and were occasionally evicted prior to completing. We therefore implemented a mechanism to automatically re-queue interrupted subproblems, where each is maintained by a decomposition worker that reschedules them in the event that a compute node stops before completing.

PFAS simulations

Preparation of molecular geometries. The equilibrium geometries of TFA, PFBA, and PFOA were optimized using the DF-MP2 method in an aug-cc-pVTZ basis set. The other geometries used for computing the BDE and PES of each PFAS were generated by performing a linear rigid-body stretch of one of the C–F bonds on the α-carbon to the carboxylic acid group. We chose to follow the dissociation of the C–F bond with the

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

smallest dihedral angle with the C=O to maintain a consistent and comparative analysis between the different PFAS.

Coupled-cluster calculations. All coupled-cluster calculations reported herein were performed in the Psi4 package.[51] The RHF-CCSD(T) calculations on PFOA were performed using the density-fitting CCSD(T) code in Psi4, which is described in ref. 52. The RHF-CCSD(T) and UHF-CCSD(T) calculations for TFA and PFBA were performed without the use of density fitting. Stability analysis was employed to obtain the broken symmetry UHF reference wavefunction.

DFT calculations. All DFT calculations were performed with QChem,[53] where we used the ωB97x-D, B3LYP, PBE0 and MN15 exchange–correlation (XC) functionals in a spin-unrestricted formalism with the cc-pVDZ basis set. Grids were set to their default values. The D3(BJ) dispersion correction was applied to all XC functionals except for ωB97x-D.

Author contributions

CD, HN, AZ, and TY conceived the project. HN, IG AZ, RP, and TY coordinated the project. AER, LH, RW, JTP, PKJ, and TY performed quantum chemistry calculations. AER, LH, JTP, and TY analyzed the quantum chemistry calculation results. LH and JTP improved quantum chemistry aspect of QEMIST Cloud. SK, DW, NH, AJ, and ZM improved the infrastructure and networking of QEMIST Cloud. SK, DW, and ZM implemented logging and monitoring tools in QEMIST Cloud. SK, DW, AW, RW, NH, AJ, and RP optimized application code of QEMIST Cloud. DW and NH optimized data transfer and storage in QEMIST Cloud. SK, DW, NH, AJ, and RP operated the QEMIST Cloud by monitoring the progress of one million core run. MRH provided support and advice regarding the QEMIST Cloud architecture. PMZ provided support and advice regarding the quantum chemistry calculations. AER, LH, DW, AW, RW, PMZ, RP, and TY wrote the manuscript. All authors contributed to discussions and revision of the manuscript to its final version.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the SI.

The supplementary information contains details of the CnB truncation procedure, total number of n-body terms computed, tables of bond dissociation energies, analysis of DFT and CC results, convergence of iFCI, computational performance on one million vCPUs, total energies, and molecular geometries. See DOI: https://doi.org/10.1039/d5sc03019f.

Acknowledgements

We thank Barry Bolding at Amazon Web Service for his support for this work and Angela Wilson at Michigan State University for the valuable discussions and insight. We are grateful to Intel for their support. We thank Stephen Harper at Accenture for his technical support and Philip Ifrah at SandboxAQ for facilitating communication. We also thank Adam Lewis, Ang Xiao, Martin Ganahl, Atashi Basu, and Nadia Harhen at SandboxAQ for their revisions of the manuscript and overall support.

References

1 J. A. Pople, *Rev. Mod. Phys.*, 1999, **71**, 1267–1274.
2 R. A. Friesner, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6648–6653.
3 J. Řezáč and P. Hobza, *Chem. Rev.*, 2016, **116**, 5038–5071.
4 M. Head-Gordon and E. Artacho, *Phys. Today*, 2008, **61**, 58.
5 M. G. Evich, M. J. Davis, J. P. McCord, B. Acrey, J. A. Awkerman, D. R. Knappe, A. B. Lindstrom, T. F. Speth, C. Tebes-Stevens, M. J. Strynar, *et al.*, *Science*, 2022, **375**, eabg9065.
6 B. C. Crone, T. F. Speth, D. G. Wahman, S. J. Smith, G. Abulikemu, E. J. Kleiner and J. G. Pressman, *Crit. Rev. Environ. Sci. Technol.*, 2019, **49**, 2359–2396.
7 L. Ahrens, *J. Environ. Monit.*, 2011, **13**, 20–31.
8 K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi and W. A. De Jong, *J. Chem. Phys.*, 2017, **147**, 184111.
9 A. Shayit, C. Liao, S. Upadhyay, H. Hu, T. Zhang, E. D. III, C. Yang and X. Li, *Breaking the Quadrillion Determinant Barrier in Numerically Exact Configuration Interaction*, 2025, https://arxiv.org/abs/2505.20375.
10 M. L. Abrams and C. D. Sherrill, *J. Chem. Phys.*, 2004, **121**, 9211–9219.
11 D. I. Lyakh, M. Musiał, V. F. Lotrich and R. J. Bartlett, *Chem. Rev.*, 2012, **112**, 182–243.
12 S. Verma, T. Lee, E. Sahle-Demessie, M. Ateia and M. N. Nadagouda, *Chem. Eng. J. Adv.*, 2023, **13**, 100421.
13 R. Kumar, T. K. Dada, A. Whelan, P. Cannon, M. Sheehan, L. Reeves and E. Antunes, *J. Hazard. Mater.*, 2023, **452**, 131212.
14 J.-M. Arana Juve, B. Wang, M. S. Wong, M. Ateia and Z. Wei, *Curr. Opin. Chem. Eng.*, 2023, **41**, 100943.
15 P. M. Zimmerman, *J. Chem. Phys.*, 2017, **146**, 104102.
16 P. M. Zimmerman, *J. Phys. Chem. A*, 2017, **121**, 4712–4720.
17 P. M. Zimmerman, *J. Chem. Phys.*, 2017, **146**, 224104.
18 A. E. Rask and P. M. Zimmerman, *J. Phys. Chem. A*, 2021, **125**, 1598–1609.
19 A. E. Rask and P. M. Zimmerman, *J. Chem. Phys.*, 2022, **156**, 094110.
20 E. Dimitrov, G. Sanchez-Sanz, J. Nelson, L. O'Riordan, M. Doyle, S. Courtney, V. Kannan, H. Naseri, A. G. Garcia, J. Tricker, M. Faraggi, J. Goings and L. Zhao, *Pushing the Limits of Quantum Computing for Simulating PFAS Chemistry*, 2023, http://arxiv.org/abs/2311.01242.
21 M. J. Bentel, Y. Yu, L. Xu, Z. Li, B. M. Wong, Y. Men and J. Liu, *Environ. Sci. Technol.*, 2019, **53**, 3718–3728.
22 A. H. S. Filho and G. L. De Souza, *Phys. Chem. Chem. Phys.*, 2020, **22**, 17659–17667.

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

23 K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479–483.
24 A. Dutta and C. D. Sherrill, *J. Chem. Phys.*, 2003, **118**, 1610–1619.
25 Y. Ge, M. S. Gordon and P. Piecuch, *J. Chem. Phys.*, 2007, **127**, 174106.
26 M. W. Schmidt and M. S. Gordon, *Annu. Rev. Phys. Chem.*, 1998, **49**, 233–266.
27 P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka and R. Shepard, *Chem. Rev.*, 2012, **112**, 108–181.
28 C. D. Vecitis, H. Park, J. Cheng, B. T. Mader and M. R. Hoffmann, *Front. Environ. Sci. Eng. China*, 2009, **3**, 129–151.
29 Z. Du, S. Deng, Y. Bei, Q. Huang, B. Wang, J. Huang and G. Yu, *J. Hazard. Mater.*, 2014, **274**, 443–454.
30 H. Javed, C. Lyu, R. Sun, D. Zhang and P. J. Alvarez, *Chemosphere*, 2020, **247**, 125883.
31 D. Shetty, I. Jahović, T. Skorjanc, T. S. Erkal, L. Ali, J. Raya, Z. Asfari, M. A. Olson, S. Kirmizialtin, O. A. Yazaydin and A. Trabolsi, *ACS Appl. Mater. Interfaces*, 2020, **12**, 43160–43166.
32 S. C. Panchangam, A. Y. C. Lin, K. L. Shaik and C. F. Lin, *Chemosphere*, 2009, **77**, 242–248.
33 H. Tian and C. Gu, *Chemosphere*, 2018, **191**, 280–287.
34 E. Banayan Esfahani, F. Asadi Zeidabadi, S. Zhang and M. Mohseni, *Environ. Sci.: Water Res. Technol.*, 2022, **8**, 698–728.
35 J. C. Bevington, *Nature*, 1970, **227**, 419.
36 K. J. Laidler, *J. Chem. Educ.*, 1984, **61**, 494–498.
37 M. Altarawneh, M. H. Almatarneh and B. Z. Dlugogorski, *Chemosphere*, 2022, **286**, 131685.
38 W. Purwanto, W. A. Al-Saidi, H. Krakauer and S. Zhang, *J. Chem. Phys.*, 2008, **128**, 114309.
39 J. Liu, D. J. Van Hoomissen, T. Liu, A. Maizel, X. Huo, S. R. Fernández, C. Ren, X. Xiao, Y. Fang, C. E. Schaefer, C. P. Higgins, S. Vyas and T. J. Strathmann, *Environ. Sci. Technol. Lett.*, 2018, **5**, 289–294.
40 Y. Zhang, A. Moores, J. Liu and S. Ghoshal, *Environ. Sci. Technol.*, 2019, **53**, 8672–8681.
41 Y. Li, H. Li, F. C. I. Pickard, B. Narayanan, F. G. Sen, M. K. Y. Chan, S. K. R. S. Sankaranarayanan, B. R. Brooks and B. Roux, *J. Chem. Theory Comput.*, 2017, **13**, 4492–4503.
42 O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Chem. Rev.*, 2021, **121**, 10142–10186.
43 I. Poltavsky and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2021, **12**, 6551–6564.
44 H. Yu, M. Giantomassi, G. Materzanini, J. Wang and G.-M. Rignanese, *Mater. Genome Eng. Adv.*, 2024, **2**, e58.
45 A. A. Holmes, N. M. Tubman and C. J. Umrigar, *J. Chem. Theory Comput.*, 2016, **12**, 3674–3680.
46 W. A. I. Goddard and R. C. Ladner, *J. Am. Chem. Soc.*, 1971, **93**, 6750–6756.
47 W. A. I. Goddard, T. H. J. Dunning, W. J. Hunt and P. J. Hay, *Acc. Chem. Res.*, 1973, **6**, 368–376.
48 F. W. Bobrowicz and W. A. Goddard, in *The Self-Consistent Field Equations for Generalized Valence Bond and Open-Shell Hartree—Fock Wave Functions*, Springer US, Boston, MA, 1977, pp. 79–127.
49 Why Amazon EC2 Spot Instances?, **https://aws.amazon.com/ec2/spot/**, Accessed: July 15, 2025.
50 E. A. Brewer, *Proceedings of the sixth ACM symposium on cloud computing*, 2015, pp. 167.
51 D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Phys.*, 2020, **152**, 184108.
52 A. E. I. DePrince and C. D. Sherrill, *J. Chem. Theory Comput.*, 2013, **9**, 2687–2696.
53 E. Epifanovsky, A. T. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z. Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. D. Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P. T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. Hanson-Heine, P. H. Harbach, A. Hauser, M. F. Herbst, M. H. Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Á. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y. P. Li, J. Liang, M. Liebenthal, H. H. Lin, Y. S. Lin, F. Liu, K. Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S. P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. Menger, J. M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y. C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham,

© 2025 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 21 August 2025. Downloaded on 2/7/2026 12:32:16 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J. D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. Deprince, R. A. Distasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C. P. Hsu, T. C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. Thom, A. Tkatchenko, D. G. Truhlar, T. V. Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. Gill, M. Head-Gordon, J. M. Herbert and A. I. Krylov, *J. Chem. Phys.*, 2021, **155**, 084801.