

Automated Computational Thermochemistry for Butane Oxidation:

A Prelude to Predictive Automated Combustion Kinetics

Murat Keceli,¹ Sarah N. Elliott,² Yi-Pei Li,³ Matthew S. Johnson,³ Carlo Cavallotti,⁴ Yuri Georgievskii,¹
William H. Green,³ Matteo Pelucchi,⁴ Justin M. Wozniak,⁵ Ahren W. Jasper,¹ Stephen J. Klippenstein¹

¹ Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, IL, 60439, USA

² Center for Computational Quantum Chemistry, University of Georgia, Athens, GA, 30602, USA

³ Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

⁴ Department of Chemistry, Materials and Chemical Engineering “G. Natta”, Politecnico di Milano, Milan, Italy

⁵ Math and Computer Science Division, Argonne National Laboratory, Argonne, IL, 60439, USA

Corresponding Author:

Stephen J. Klippenstein

9700 S. Cass Ave.

Argonne National Laboratory

Argonne, IL, 60439

USA

e-mail: sjk@anl.gov

Colloquium: Kinetics

Paper Length: 6152 Words (Method 1)

Main Text (4668); References (507), Figures (977); Fig. 1 (379); Fig. 2 (412); Fig. 3 (186).

Abstract: Large scale implementation of high level computational theoretical chemical kinetics offers the prospect for dramatically improving the fidelity of combustion chemical modeling. As a first step toward this goal, we have developed a procedure for automatically generating the thermochemical data for combustion of an arbitrary fuel. The procedure begins by producing a list of combustion relevant species from a specification of the fuel and combustion conditions of interest. Then, for each element in the list of species, the procedure determines an internal coordinate z-matrix description of its structure, the optimal torsional configuration via Monte Carlo sampling, key rovibrational properties for that optimal geometry (including anharmonic corrections from torsional sampling and/or vibrational perturbation theory), and high level estimates of the electronic and zero-point energies via arbitrarily defined composite methods. This data is then converted first to partition functions, then to thermodynamic properties, and finally NASA polynomial representations of the data. The end product is an automatically generated database of electronic structure results and thermochemical data including representations in appropriate format for combustion simulations. The utility and functioning of this predictive automated computational thermochemistry (PACT) software package is illustrated through application to the automated generation of thermochemical data for the combustion of n-butane. Butane is chosen for this demonstration as it is of reasonably manageable size for debugging level computations, while still presenting most of the key challenges that need to be surmounted in the consideration of larger fuels. Furthermore, its low temperature chemistry is representative of that occurring with larger alkanes.

Keywords: Computational Thermochemistry, Butane Oxidation

1. Introduction

The ever-expanding computational capabilities accessible to combustion scientists yield exciting new opportunities for high fidelity first principles computational modeling. Within the US there is currently a drive towards exascale computing with major resources being invested in the Exascale Computing Project (ECP) [1]. One of the projects funded under the ECP, led by Jackie Chen of Sandia National Laboratories, has the aim of “Transforming Combustion Science and Technology with Exascale Simulations.” It is structured around providing a state-of-the-art exascale ready simulation capability for first-principles (DNS) and near-first principles (DNS/LES hybrids) simulations of turbulence-chemistry interactions for multiple-component fuels in conditions relevant to practical combustion devices, including turbulence, mixing, spray vaporization, low-temperature ignition, flame propagation, and soot/radiation.

Historically, such computational fluid dynamics CFD simulations have been the focus of most high performance computing applications of relevance to combustion [2,3]. However, high performance computing also offers exciting and as yet unrealized prospects for combustion chemistry. In particular, currently available computational resources and theoretical kinetics methodologies offer the possibility for accurately predicting more or less all of the parameters that underlie current chemical mechanisms for combustion. To do so requires the automation of a large number of tasks that are currently performed on a case-by-case basis, with the treatment of each species and each reaction effectively considered as a separate computational task. Furthermore, the actual prediction of the rate coefficients for even a single reaction (or thermochemical properties for a single species) currently involves the use of a variety of different codes, which span the gamut from high quality commercial packages, to in-house research level codes developed by investigators.

The utility of many CFD simulations relies on the accuracy of the underlying chemical mechanisms. For example, under the low temperature conditions of **of relevance to compression ignition engines including novel technologies such as RCCI and HCCI [4]**, combustion is exquisitely sensitive to the auto ignition process and consequently to the detailed mechanism of fuel chemistry. To this end, an Argonne based team (Klippenstein, Bair, Jasper, Wagner, and Wozniak (all at ANL), together with Cavallotti **(Politecnico di Milano)**, Green (MIT), Lu (UConn), and Zador (Sandia)) **is working** to develop codes for automatically generating high fidelity mechanisms through exascale level predictive automated combustion kinetics (PACK) calculations. This effort involves a combination of mechanism generation, theory based calculations of rate constants, thermochemistry, and transport properties, and mechanism reduction **algorithms**, all designed to be implemented automatically, and with uncertainty analyses driving the **specific** choice of simulations to be performed. The goal is to automatically produce chemical mechanisms for arbitrary fuels with a well-defined level of accuracy. Ultimately, this PACK effort will enable the use of high fidelity drop-in chemical models in the exascale simulations, which would greatly mitigate **the** risks due to the limited accuracy of current literature chemical mechanisms, particularly for applications outside their range of validation.

The evaluation of partition functions is a central feature of thermodynamics as well as of chemical rate theory. Indeed, thermodynamic properties are readily obtained from derivatives of the partition function, while, within conventional transition state theory, high-pressure limit rate constants can be obtained from ratios of partition functions for the transition state and the reactants. The evaluation of partition functions generally centers around the determination of the energies and rovibrational properties of stationary points on the potential energy surface. More advanced theoretical treatments require some consideration of the properties along the reaction path, but such determinations have many analogies with partition function determinations. The consideration of pressure dependent effects on the rate constants depends on microcanonical rather than canonical partition functions, but the primary input parameters for such evaluations are closely related.

Thus, as a first step in this process, we have automated the prediction of the thermochemical properties for arbitrary sets of combustion related species. Furthermore, by coupling to the RMG automated mechanism generation software [5] we derive a procedure for automatically predicting the thermochemical properties of relevance to an arbitrary fuel. In this manuscript, we illustrate the utility and functioning of this predictive automated combustion thermochemistry (PACT) code through sample calculations for the combustion of n-butane.

As demonstrated some time ago by Hughes et al. [6], uncertainties in the heat of formation are a major component of the uncertainties in combustion modeling. Furthermore, the general paucity of experimental thermochemical data at combustion relevant temperatures suggests that other properties like the heat capacities and entropies may be even more uncertain and have similarly important impacts on modeling predictions. This problem is especially severe for the properties of ephemeral species like QOOH radicals. A recent study of Bugler et al. [7] highlighted these problems for the oxidation of pentane, with a demonstration of order of magnitude changes in the ignition delay with updated thermochemistry.

Currently, many combustion mechanisms employ group additivity [8] to estimate the thermochemical properties for many species. High-level theoretical methods can provide more quantitative predictions than group theory [9], at least for smaller molecules. Such calculations can also be used to refine and improve the group values, as in the recent work of Burke et al. [10], which facilitates the continued extension of higher accuracy to more and more fuels. Exascale level computations will greatly expand the size of molecules for which high-level methods may be implemented. They will also permit the study of the large numbers of species present in current mechanisms.

Our development of the PACT software has used n-butane as a target for debugging and illustration purposes. This fuel is of interest, as it is the smallest alkane to provide significant low temperature oxidation chemistry that closely mimics that of larger alkanes [11]. Its small size allows for

the relatively facile implementation of high-level electronic structure methods and leads to reasonably manageable mechanism sizes. Nevertheless, its mechanism presents many of the complications present for larger fuels such as numerous torsional degrees of freedom, the presence of cis/trans and other stereochemical isomers, ring species, multiple unsaturated bonds, and a variety of heavily oxidized species of limited stability.

A simple survey of literature mechanisms (see, e.g., Refs. [12,13]) suggests that current thermochemical treatments for n-butane still contain significant uncertainties. A graphic illustration of these uncertainties can be obtained by substituting the thermochemistry employed in one mechanism into another mechanism. For example, the variation in the predicted speciation for butane oxidation at 1 atm in a jet stirred reactor is illustrated in Fig. 1 for a case where the thermochemistry from [12] is substituted into the mechanism from [13]. Clearly, the remaining uncertainties in the thermochemistry are an important component of modeling uncertainties.

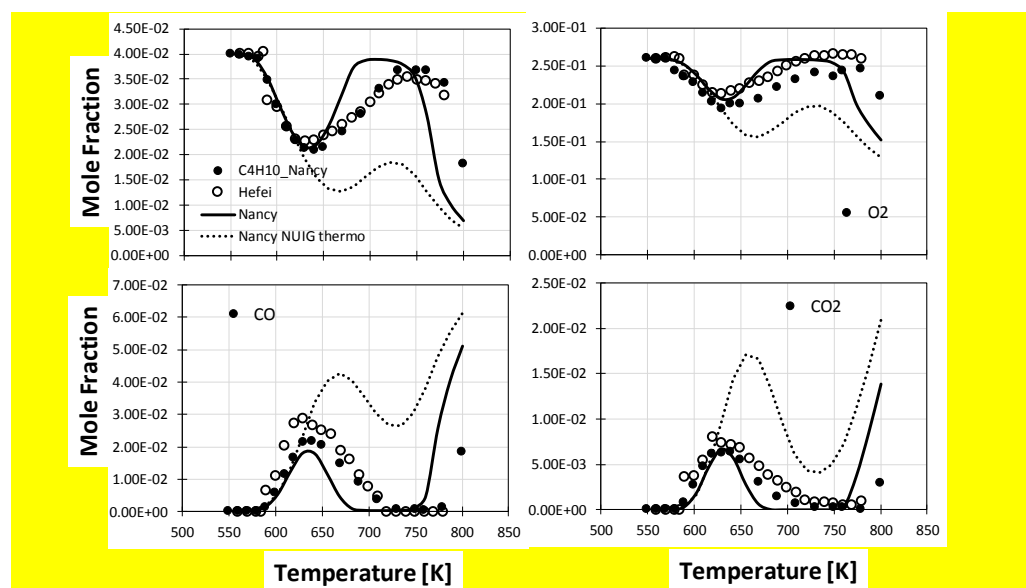


Figure 1: Plot of the mole fraction versus temperature from jet-stirred reactor experiments of Herbinet et al. [14]. The solid lines denote simulations using the original thermochemical parameters, while the dotted lines denote simulations employing the thermochemical data substituted from [12].

Recent automated thermochemistry studies from other groups provide useful indications of the scale of calculations that are now feasible. For example, Ramakrishnan et al. [15] predicted the thermodynamic properties of 134,000 molecules with up to 9 heavy atoms at the B3LYP/6-31G(2df,p) level. A smaller subset of 6,095 C₇H₁₀O₂ molecules was examined at the higher level G4(MP2) level. More recently, Ghahremanpour et al. [16] applied the CBS-QB3, G2, G3, and G4 methods to the study of 2000 molecules with up to 47 atoms, while a smaller set of 650 molecules with up to 16 atoms were studied with two variants of the W1 method. Within the combustion community, a very recent study of 920 species at the B-97 and B3LYP levels demonstrated the utility of error cancelling balanced reactions [17]. Meanwhile, a program (KiSTheIP) for predicting thermochemical properties and rate constants from quantum chemistry results was recently described by Canneaux et al. [18]. The present work differs from these prior studies with its emphasis on the connection to combustion chemistry, the implementation of arbitrary high accuracy composite methodologies, the explicit consideration of anharmonic effects, and the focus on **proceeding towards** exascale level computations.

2. Methodology

2.1. Overview

An overview of the flow of information and the codes we use to accomplish the tasks and their coupling is provided in Fig. 1. The process begins with a specification of the fuel and the simulation conditions of interest. Implementation of the RMG mechanism generator [5] produces a list of species **(as SMILES or InChi strings)** of relevance to the combustion of the fuel for a range of low-dimensional combustion simulations. These simulations also produce sensitivity and importance estimates for each of the species. **For each species in this list (or for those species with highest sensitivity, etc.) we generate thermochemical polynomials through a sequence of calls controlled by the overarching QTC Python script.**

Predictive Automated Combustion Thermochemistry

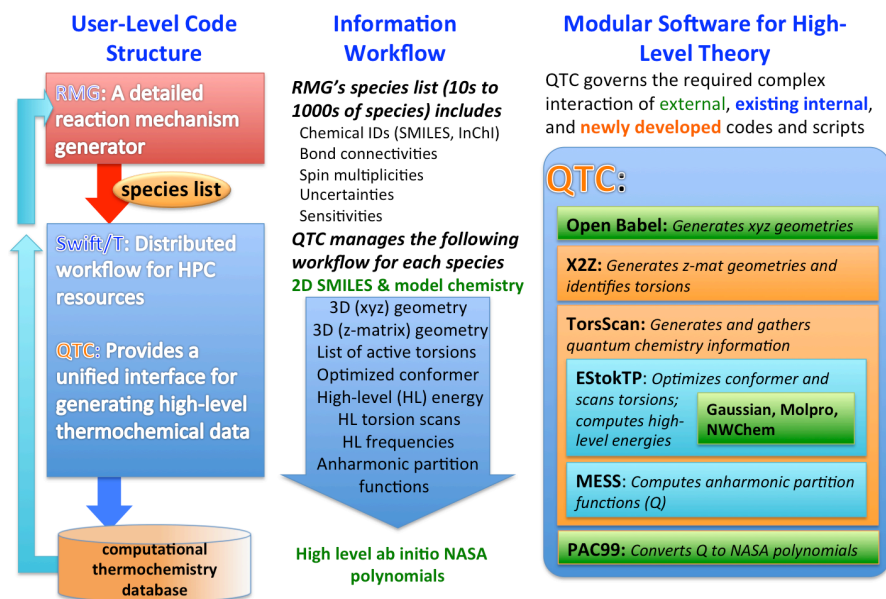


Figure 2: A schematic indication of the flow of information accomplished through the PACT implementation of automated thermochemistry.

The information flow proceeds from the SMILES name (passed to QTC by RMG) for a given species, to an xyz coordinate structure guess (via Open Babel [19]), to an internal coordinate z-matrix (via x2z). The next step involves the determination of an optimized torsional configuration and overall geometry, vibrational frequencies, and torsional potentials. This information is generally obtained through TorsScan driven calls to EStokTP, which in turns drives calls to electronic structure codes (currently Gaussian, Molpro, or NWChem). Alternatively, when no torsional optimization is required, QTC itself can direct the calls to the electronic structure codes codes. From these structures, high-level energy evaluations are obtained from QTC, TorsScan, or EStokTP driven calls to the electronic structure codes. Composite energy estimates, which are based on a combination of electronic structure methods, are obtained through multiple electronic structure calls, which generate a database of energies. The energies, torsional potentials, and vibrational frequencies are passed to the master equation code MESS, which includes a partition function evaluation module. The resulting partition function data is converted to NASA polynomial format via calls to first Therm and then PAC99. The PACT package is designed to

run effectively on high performance linux computing clusters and to be available as freeware. Further details of the code availability and citations are provided in the supplemental material.

2.2. Process Details

The process begins by using the Reaction Mechanism Generator (RMG) software [5] to generate a reference chemical mechanism from a specified initial set of reactants and reaction conditions, including temperature, pressure, and species compositions. RMG reacts the initial species in all possible ways based on known reaction family templates and simulates the time evolution of a batch reactor at the given reaction conditions. At each time step, RMG evaluates the flux at which each new chemical species is produced. If a flux attains a sufficiently high value, the new species and all the reactions producing that species are added to the mechanism and the simulation starts over. This process continues, and the mechanism grows incrementally with each iteration, until the user-specified termination time or reactant conversion are satisfied. The reaction rates and initial thermodynamic properties of the reacting species are taken from libraries or estimated using group additivity or rate rules [20].

As part of this effort, improvements to error management within RMG, particularly within the differential equation solver, were made to facilitate its operation without human assistance. Furthermore, revised selection criteria for the addition of new species and reactions were implemented to overcome inherent difficulties with picking up key, low flux, chain branching reactions. Previously, these reactions had to be added to models manually. These revisions improve the ability of RMG to develop comprehensive models for larger molecules, as needed for this project. Meanwhile, improvements to the RMG uncertainties module have improved the robustness of uncertainty calculations and make it possible to consistently estimate the uncertainties needed to analyze species importance within an arbitrary RMG model.

For each species selected from the RMG mechanism, theory based thermochemical parameters are obtained through the following sequence of steps. First, a reference starting geometry is generated

with Open Babel, which may then be improved with a low-level geometry optimization. The x2z code then transforms this Cartesian coordinate geometry to a set of internal coordinates; producing at the same time an indication of the relevant torsional degrees of freedom. The optimum torsional configuration is then obtained through geometrical optimization of a set of structures with randomly chosen torsional angles. From this optimal torsional configuration, one-dimensional torsional potentials are mapped for each of the torsional degrees of freedom. At each stage in the analysis, the methodology allows for arbitrarily specified electronic structure methods to be used. Furthermore, high-level energies at the final optimized geometry are obtained from arbitrary composite methods expressed as arithmetic combinations of arbitrary electronic structure methods.

Occasionally, the electronic structure calculations will fail for any one of a multitude of reasons. To facilitate the recovery from these errors the code automatically cycles through a system of preset electronic structure templates. For example, with the Gaussian calculations the templates are set up to do the calculations with or without the scf=qc option, with or without nosym and with internal or Cartesian coordinates. These templates allow for great flexibility in attempting to obtain convergence, and are readily modified by the user.

The partition function evaluations in this work are based on rigid rotor harmonic oscillator (RRHO) evaluations coupled with quantized one-dimensional hindered rotor torsional corrections. The reduced set of vibrational frequencies is obtained by projecting out the torsional modes [21]. Such RRHO + hindered rotor methodologies provide the basis for much of the thermochemical work in current literature. We have also incorporated two advanced treatments of anharmonicities: spectroscopic vibrational perturbation theory and multi-dimensional coupled torsional analyses. For simplicity, we restrict our attention here to the more standard one-dimensional torsional treatments.

In addition to generating the thermochemical data, the code also generates and organizes the calculated structures, energies, vibrational frequencies, and torsional parameters according to the specific set of electronic structure methods. Currently, this data takes the form of a structured system of

files or can be exported as a JSON file. These values can be mined to explore the efficacy of arbitrary combinations of electronic structure methods. For example, one might consider various methods for extrapolating to the CBS limit, for incorporating bond additivity corrections [22], etc. High accuracy values can be used to validate lower level approximations or to determine optimal group additivity expressions. The end result is a growing database representing the computational thermochemical genome for combustion.

The Swift/T code [23] facilitates parallelization of the overall process. Currently, this involves the trivial parallel running of QTC for each species, coupled with the threaded parallelization that is inherent to typical electronic structure codes. The QTC script also allows for further parallelization of the composite electronic structure methods through distribution of each of the component methods to a different node. At this level, the parallelization for a typical practical fuel, with $\sim 10^3$ chemically relevant species, would involve simultaneous utilization of $\sim 10^4$ processors. Notably, NWChem is specifically designed for massive cross node parallelization, which will facilitate high-level energy evaluations for larger molecules.

For some species there are multiple 3D structures with the same 2D structure, and with these structures differing in more than just the torsional configuration for a weakly hindered rotor. For example, there are a number of species with multiple enantiomers or with cis and trans isomers arising from geometrical differences across an unsaturated bond. Such isomers are effectively different species with their own thermochemical properties. For the case of cis/trans isomers about unsaturated bonds we expand the list of species in order to separately evaluate the thermochemical properties for each isomer and ultimately pass this information back to RMG. Otherwise, for now we simply note the particular enantiomer being considered, but do not expand out to separate mappings of all enantiomers.

3. Protocol for Butane Analysis

For the butane model, the RMG run used the **foundation fuel chemistry model** [24] and the rates reported by Burke et al. [25] as a starting submechanism and was specifically forced to include species relevant to key low temperature OH feedback loops. Other rate information was drawn from the work of Dooley et al. [26] or estimated using RMG's default rate families as the model was constructed. All the butane simulations started with the stoichiometric air-fuel ratio and the reactors used spanned a temperature range of 650-2000 K at a pressure of 1 bar. The termination criterion was 99% conversion of fuel. The exact RMG input file used is provided in the supplementary information.

The ω B97-XD/6-31G* method was used to determine the optimum torsional configuration through Monte Carlo sampling of the torsional angles. This Monte Carlo sampling employed the minimum of $(5+3^{n_{\text{rotor}}})$ and 100 configurations, where n_{rotor} is the number of non-methyl torsional modes for the given species. The M06-2X/cc-pVTZ method was then used to map out one-dimensional torsional potentials for each of the torsional degrees of freedom. Occasionally (i.e., for ~ 10 species) the 1-d torsional mapping discovered configurations that were lower in energy than the starting geometry. This shortcoming was generally related to a difference in the optimal configuration for the two electronic structure methods. Simply repeating the process with the M062X/cc-pVTZ method employed for both Monte Carlo sampling and the torsional mapping remedied such problem cases. The lack of any further discovery of lower energy configurations indicates that the Monte Carlo sampling procedure is generally sufficient. For reference purposes, the optimized geometries for each of the species are reported in the supplementary material.

This torsional analysis is followed by a B2PLYP-D3/cc-pVTZ geometry optimization and frequency analysis. **Very rarely this process yields imaginary frequencies (i.e., for three species), which is remedied here by repeating the whole the torsional optimization and mapping process at the B2PLYP-D3/cc-pVTZ level.** At these B2PLYP-D3 geometries, we have implemented a sequence of CCSD(T)-F12 and MP2-F12 calculations, incorporating the cc-pVDZ-F12 and cc-pVTZ-F12 basis sets for the CCSD(T) analysis and up to the cc-pVQZ-F12 basis set for the MP2 analysis. **The use of the cc-pVnZ-**

F12 series (n=D,T,Q) of basis sets allows us approach convergence with respect to basis set. In that regard, our recent detailed and wideranging computational thermochemical analysis for small combustion species [9] indicates that the complete basis set limit is approached to within about 3, 1, and 0.5 kcal/mol for n=D, T, and Q, respectively.

The primary purpose of this paper is to demonstrate the functioning of our approach. For this purpose, the particular choice of electronic structure methods is somewhat irrelevant. Nevertheless, we have chosen methods that should provide results that at least equivalent in accuracy to the methods commonly employed in current combustion thermochemical kinetics analyses. In particular, it is our experience that the B2PLYP-D3 method provides geometries and vibrational frequencies that are comparable in accuracy to CCSD(T) ones, which are generally taken to provide the gold standard. Meanwhile, the M06-2X method should provide a more accurate representation of the torsional potentials than the more commonly used, but now outdated, B3LYP method. Furthermore, it is our experience that the cc-pVTZ basis provides reasonably well-converged rovibrational properties for such density functional theory calculations.

4. Results and Discussion

The success/failure of the PACT approach depends on both its ability to automatically produce the requisite thermochemical information and its ability to produce such data at a useful level of accuracy. For butane, RMG indicates that 160 species are of significance to its combustion. The inclusion of cis/trans isomers expands this list to the 173 species provided in the Supplementary Material. The aforementioned PACT protocol has been applied to this full list of species. For illustrative purposes we report the results of calculations applying four separate composite electronic structure methods: (i) CCSD(T)-F12/cc-pVDZ-F12, (ii) CCSD(T)-F12/cc-pVDZ-F12 + MP2/cc-pVTZ-F12 – MP2/cc-pVDZ-F12, (iii) CCSD(T)-F12/cc-pVTZ-F12 and (iv) CCSD(T)-F12/cc-pVTZ-F12 + MP2/cc-pVQZ-F12 – MP2/cc-pVTZ-F12. In the following, for ease of notation, we label these methods as M1-

M4. Full sets of polynomials, enthalpies, entropies, heat capacities, geometries, and vibrational frequencies for these four methods are reported in the Supplementary Material.

The abovedescribed protocol within the PACT code successfully predicts the thermochemical properties for all but 10 species in the RMG generated n-butane list. Three of these species [HC(O)O, CH₃C(O)O, and OCHCHC(O)O] involve a radical bound to the C atom of a OCO group, for which the near degeneracies of two electronic states creates difficulties for the B2PLYP-D3 method. For the remaining 7 species [CH₂OOH, OCHOOCO, CH₃C(O)OOCH₂, CH₃CHOOC(O)CH₃, CH₃C(OOH)CH₂CH₂OOH, CH₃CH(OOH)CH₂CHOOH, and CH₃C(O)OOC(CH₃)CH₂CH₃] the protocol indicates that the molecules are not stable and instead directly dissociate to bimolecular species [CH₂O + OH, OCO + OCHO, CH₃C(O)O + CH₂O, CH₃C(O)O + CH₃CHO, CH₃C(O)CH₂CH₂OH + OH, CH₃CH(OOH)CH₂CHO + OH, CH₃C(O)O + CH₃CH₂C(O)CH₃]. These species involve radicals with OOC(O) and OOH groups on them and suggest the need for further refinement of RMGs rules for such species.

Although mostly automated, effective treatment of electronic degeneracies and symmetry factors required human intervention for some species. In particular, states with multiple degenerate electronic states (i.e., C(³P), O(³P), OH (²Π), and CH₃O) required a revision to the partition function input to incorporate such degeneracies. The proper inclusion of symmetry corrections is a complex topic. For 1-dimensional torsional treatments the proper symmetry number is effectively temperature dependent and difficult to determine a priori. However, in the high temperature limit, the correct symmetry factor is the same as that for a multi-dimensional torsional mapping. We have devised an automated procedure for determining that from the number of enantiomers indicated by OpenBabel, the internal rotor symmetries, and the external rotational symmetry. This automated procedure is ineffective for ring-based enantiomers, because OpenBabel does not indicate those. Thus, for about 10 species, we have incorporate revised symmetry factors by hand.

At low temperature, an approach that counts the number of degenerate minimum generated by the 1-dimensional mappings is often appropriate. Hthe torsional potential becomes irrelevant and t Our interest in combustion leads us to incorporate symmetry factors appropriate for this high temperature limit.

Our automated expansion of cis/trans isomers sometimes fails to capture the presence of such isomers for resonantly stabilized molecules with radical sites on an O atom. Thus, the trans isomers for OCHCHOH, CH₃CHC(O)CH₃ and OHCHC(O)OOH were added to the species list by hand. Finally, for the two largest species, CH₃CH₂CH(CH₃)OOC(O)CH₂ and CH₃CH₂CH(CH₃)OOC(O)CH₂OO certain computational limits precluded our implementation of methods M3 and M4.

A recent high level thermochemical analysis [9] for a set of core combustion species provides useful data for testing these PACT predictions. In particular, for the set of 45 overlapping species we find that the root-mean squared deviation (RMSD) between the M1 – M4 methods and the ATcT [27] results for the 0 K heat of formation are 1.10, 0.97, 0.61, and 0.61 kcal/mol, respectively. The 298 K RMSD are essentially identical. These results suggest that the cc-pVTZ-F12 results are reasonably well converged with respect to basis set. Furthermore, the MP2-F12 method is moderately effective in correcting for basis set limitations. Notably, for method 3 the 2 σ errors are about 1.2 kcal/mol.

The ANL0 computational method, which was seen to have 2 σ uncertainties of about 0.2-0.3 kcal/mol, provides an additional, somewhat larger, set of data for evaluating the accuracy of the current predictions. For this set of 63 overlapping species, the M1-M4 RMSD are 1.22, 1.07, 0.71, and 0.72 kcal/mol, respectively. Interestingly, the average errors for this set (-1.02, -0.92, -0.57, and -0.58 kcal/mol) are a significant fraction of the RMSD, which suggests that in future work it may be feasible to implement simple correction schemes as a means to reducing the errors. Notably, the RMSD values after shifting by the mean errors are only 0.65, 0.53, 0.39, and 0.39 kcal/mol. The plot in Fig. 2 illustrates the distribution of errors for each method.

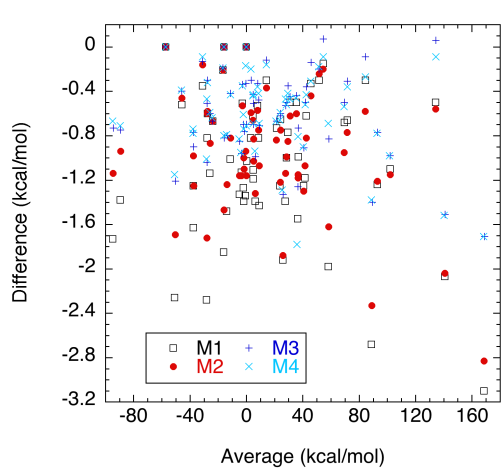


Figure 3: Difference graph for methods M1-M4 0 K heats of formation relative to ANL0 values. The average denotes the average of M_i and ANL0.

The recent review of Burke et al. [10] provides a useful set of data for testing the thermal aspects of the predictions. It includes molar entropies at 298 K and heat capacities at a variety of temperatures. The present M3 predictions for both the enthalpy and the entropy are in remarkably good agreement with the recommendations (cf. Table S1). For the enthalpy, the maximum discrepancy is 0.7 kcal/mol, which is well within the error bars of the predictions. For the entropy, the differences are similarly small (less than $1 \text{ cal K}^{-1} \text{ mol}^{-1}$) for all but $\text{CH}_3\text{CH}_2\text{CH}_2\text{OO}$, and $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2\text{OO}$. The larger deviations ($2\text{--}3 \text{ cal K}^{-1} \text{ mol}^{-1}$) for the two RO_2 species are likely indicative of uncertainties in the literature values. For the heat capacities, the deviations are also generally quite small (i.e., mostly $< 0.5 \text{ cal K}^{-1} \text{ mol}^{-1}$). However, for the RO_2 species and for butane, the deviations are much larger ($\sim 2 \text{ cal K}^{-1} \text{ mol}^{-1}$). These discrepancies, which deserve further detailed study, are likely indicative of the difficulties of treating multidimensional torsions and their coupling with umbrella modes. We do not believe it is an inherent failing of the PACT protocol.

It is perhaps worth noting that many of the recommendations from [10] are largely based on theoretical studies that employed methods similar to those used here. Thus, the agreement between the present predictions and the literature recommendations should not necessarily be taken to imply that the present predictions actually have this level of accuracy in an absolute sense. Rather, they provide strong confirmation that the overall PACT protocol is functioning correctly and at a level of accuracy that is at least comparable to current literature data.

A further comparison of the enthalpies, entropies and heat capacities with underlying RMG database values is provided in the Supplementary Materials. Notably, for about 10-20% of the cases, particularly the larger less well-studied species, there are some fairly dramatic differences in one or the other of the properties. These differences may be indicative of shortcomings in the RMG database.

5. Conclusion

The PACT approach developed here provides a seamless procedure for producing high-level thermochemical representations directly from the specification of the fuel and the reactor conditions of interest. The application of this approach to butane oxidation is seen to yield thermochemical parameters at an accuracy that is comparable to or better than existing literature data. It can be used to explore the thermochemical properties of arbitrary fuels at high levels of accuracy. Such a capability is particularly important for low temperature combustion where simulations show strong sensitivity to the thermochemistry. It can also be used to produce training sets for machine learning algorithms or to providing benchmark values for improving the accuracy of fitted lower level methods. Similarly, the high level data can be used to optimize bond additivity, group additivity, or other empirical corrections.

The ultimate goal of this effort is to derive high-accuracy thermochemical properties for a broad range of larger and larger species. The present analysis for butane used on the order of 10^5 CPU hours. With 10^7 CPU hours, which is readily available to us through our ECP project, it should be feasible to perform similar quality calculations for a complete heptane and/or isooctane mechanism. Meanwhile,

another order of magnitude CPU resources would facilitate similar quality calculations for dodecane, which is of particular interest to Chen's combustion simulation ECP team. Alternatively, with more modest resources, these codes can still be used to obtain high quality predictions through modest reductions in the methodologies and/or restricted application to a set of the most important species.

Acknowledgements

The authors gratefully acknowledge helpful and informative discussions with Branko Ruscic regarding the conversion of partition functions to thermochemical polynomials. This research was supported by the Exascale Computing Project (ECP), Project Number: 17-SC-20-SC, a collaborative effort of two DOE organizations, the Office of Science and the National Nuclear Security Administration, responsible for the planning and preparation of a capable exascale ecosystem including software, applications, hardware, advanced system engineering, and early test bed platforms to support the nation's exascale computing imperative. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This material is based on work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences at Argonne under Contract No. DE-AC02-06CH11357. CC acknowledges the financial support of the Chemical Sciences and Engineering Division of Argonne National Laboratories for his sabbatical. SNE gratefully acknowledges support from a DOE Computational Sciences Graduate Fellowship through grant number DE-FG02-97ER25308. MP acknowledges the support of Argonne National Laboratory's visiting graduate program.

Separate List of Figure Captions

Figure 1: Plot of the mole fraction versus temperature from jet-stirred reactor experiments of Herbinet et al. [14]. The solid lines denote simulations using the original thermochemical parameters, while the dotted lines denote simulations employing the thermochemical data substituted from [12].

Figure 2: A schematic indication of the flow of information accomplished through the PACT implementation of automated thermochemistry.

Figure 3: Difference graph for methods M1-M4 relative to ANL0 values. The average denotes the average of M_i and ANL0.

List of Supplemental Material

Supplementary Table S1 comparing thermochemistry properties predicted by methods M1-M4 with data from [10].

Code availability comments.

Input file for RMG.

RMG log file.

4 Excel files containing thermodynamic properties from methods M1-M4.

4 Text files containing ChemKin polynomials for methods M1-M4.

Cartesian coordinates and frequencies for B2PLYP-D3 calculations.

References

-
- [1] <https://www.exascaleproject.org/exascale-computing-project/>
- [2] J.H. Chen, Proc. Combust. Inst. 33 (2011) 99-123.
- [3] J.H. Chen, J.K. Chaves, K. Harms, P. Srivastava, S. Som, S. Quan, K. Richards, M. Garcia, J. Energy Res. Tech. Trans. ASME 138 (2016) 052203.
- [4] G.T. Kalghatgi, Proc. Combust. Inst. 35 (2015) 101-115.
- [5] C.W. Gao, J.W. Allen, W.H. Green, R.H. West, Comp. Phys. Comm. 203 (2016) 212-225.
- [6] K.J. Hughes, J.F. Griffiths, M. Fairweather, A.S. Tomlin, Phys. Chem. Chem. Phys. 8 (2006) 3197-3210.
- [7] J. Bugler, K.P. Somers, E.J. Silke, H.J. Curran, J. Phys. Chem. A 119 (2015) 7510-7527.
- [8] S. Benson, Thermochemical kinetics: Methods for the estimation of thermochemical data and rate parameters (Wiley-Interscience, New York, 1976).
- [9] S.J. Klippenstein, L.B. Harding, B. Ruscic, J. Phys. Chem. A 121 (2017) 6580-6602.
- [10] S.M. Burke, J.M. Simmie, H.J. Curran, J. Phys. Chem. Ref. Data, 44 (2015) 013101.
- [11] F. Battin-LeClerc, Prog. Energy Combust. Sci. 34 (2008) 440-498.
- [12] D. Healy, N.S. Donato, C.J. Aul, E.L. Petersen, C.M. Zinner, G. Bourque, H.J. Curran, Combust. Flame, 157 (2010) 1526-1539.
- [13] E. Hebrard, A. S. Tomlin, R. Bounaceur, F. Battin-Leclerc, Proc. Combust. Inst. 35 (2015) 607-616.
- [14] O. Herbinet, F. Battin-Leclerc, S. Bax, H. L. Gall, P.-A. Glaude, R. Fournet, Z. Zhou, L. Deng, H. Guo, M. Xie, F. Qi. Phys. Chem. Chem. Phys. 13 (2011) 296-308.
- [15] R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, Scientific Data 1 (2014) 140022.
- [16] M.M. Ghahremanpour, P.J. van Maaren, J.C. Ditz, R. Lindh, D. van der Spoel, J. Chem. Phys. 145 (2016) 114305.
- [17] P. Buerger, J. Akroyd, S. Mosbach, M. Kraft, Combust. Flame 187 (2018) 105-121.
- [18] S. Canneaux, F. Bohr, E. Henon, J. Comp. Chem. 35 (2014) 82-93.
- [19] N.M. O'Boyle, C. Morley, G. R. Hutchison, Chem. Central J. 2 (2008) 5.
- [20] S.W. Benson, J.H. Buss, J. Chem. Phys. 29 (1958) 546-572.
- [21] S. Sharma, S. Raman, W.H. Green, J. Phys. Chem. A 114 (2010) 5689-5701.
- [22] C.F. Goldsmith, G.B. Magoon, W.H. Green, J. Phys. Chem. A 116 (2012) 9033-9057.
- [23] J.M. Wozniak, T.G. Armstrong, M. Wilde, D.S. Katz, E. Lusk, I.T. Foster, "Swift/T: Scalable Data Flow Programming for Distributed-Memory Dataflow Processing," Proc. CCGrid (2013) 95-102.
- [24] G.P. Smith, Y. Tao, H. Wang, Foundational Fuel Chemistry Model Version 1.0 (FFCM-1) <http://nanoenergy.stanford.edu/ffcm1>.
- [25] M.P. Burke, M. Chaos, Y. Ju, F.L. Dryer, S.J. Klippenstein, Int. J. Chem. Kinet. 44 (2012) 444-474.
- [26] S. Dooley, M.P. Burke, M. Chaos, Y. Stein, F.L. Dryer, V.P. Zhukov, O. Finch, J.M. Simmie, H.J. Curran, Int. J. Chem. Kinet. 42 (2010) 527-549.
- [27] B. Ruscic, R.E. Pinzon, G. von Laszewski, D. Kodeboyina, A. Burcat, D. Leahy, D. Montoya, A.F. Wagner, J. Phys. Conf. Ser. 16 (2005) 561-570.