

Abstract

BENNETT, MICHAEL CHANDLER. Electronic Structure of Heavy Element Systems and Many-Body Constructions of High-Accuracy Effective Core Potentials. (Under the direction of Lubos Mitas).

The electronic structure of atomic, molecular, and solid state systems provides key insights into their rich variety of properties. Fundamental to understanding the electronic structure of these systems are the solutions to the time-independent Schrödinger equation under the Born-Oppenheimer approximation. However, due to the inter-particle interactions of the electrons, an exact polynomial scaling algorithm for solving the equation does not exist. As a result, a vigorous pursuit of simplifications and approximations, that keep the predictive power of the equation intact, has been underway since the birth of quantum theory in the 1920s. This pursuit, in tandem with exponentially increasing computational power, has led to the development of a number of high-accuracy methodologies. Of particular prominence, is the diffusion Monte Carlo method which benefits from low-order polynomial-scaling, applicability over a wide range of system sizes – up to hundreds of electrons, and a small number of fully specified errors.

We have applied diffusion Monte Carlo to various heavy transition metal systems. These systems are attractive from a theoretical point of view due to the presence of localized *d*-electrons which consequently results in significant amounts of correlation and therefore are challenging for electronic structure methods in general. We study trends in the errors of diffusion Monte Carlo under the fixed-node approximation in order to better understand the accuracy limits of the method for these systems in various environments. Additionally, we investigate a specific error that is present within diffusion Monte Carlo studies of heavy elements due to the unavoidable introduction of pseudopotentials. From coupled-cluster calculations, we find that the errors of tabulated pseudopotentials within molecular settings can be significant relative to all-electron results for high-accuracy many-body techniques especially when the molecules are out of equilibrium configurations. We develop a strategy to generate pseudopotentials that incorporates near-exact correlated many-body spectra and properties and as a result lead to high accuracy over a test bed of molecules – both within and out of equilibrium. We use this strategy to develop pseudopotentials for 1st row elements, the entire set of 2nd row elements, and the full 3*d* transition metal series.

© Copyright 2019 by Michael Chandler Bennett

All Rights Reserved

Electronic Structure of Heavy Element Systems and Many-Body Constructions of
High-Accuracy Effective Core Potentials

by
Michael Chandler Bennett

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Physics

Raleigh, North Carolina

2019

APPROVED BY:

Jerzy Bernholc

Alexander Kemper

Jerry Whitten

Lubos Mitas
Chair of Advisory Committee

Dedication

Dedicated to my parents who made it possible
And to Cassidy who drove me to see it through
And to Monroe who made it most worthwhile.

Biography

My interest for physics developed midway through my undergraduate studies. I was drawn specifically to physics after reading popular physics literature out of mild curiosity. These qualitative treatments of classical and modern theories proved to be unimaginably seductive. Instantly, I recognized the importance in humanity's pursuit of a deeper understanding of the universe. As a consequence I felt urged to play a part in this pursuit – I had to contribute. At the time I had nearly fulfilled all requirements for a bachelor's in computer science. Given that computation was, as it continues to be, an integral component of numerous research topics in physics, I was confident that continuing to refine my skills in programming while also inundating my academic schedule with courses in core areas of physics would provide me with a unique set of talents which I could then apply to the further discovery of new ideas in the field. After completing my undergraduate studies I was accepted into NCSU's Master of Computer Science program. During my second semester there I enrolled in the first part of a two-part sequence of upper-level undergraduate classical mechanics (at the level of Marion). Upon completion of this course, my interest in the subject was affirmed and I grew hungry for more. I continued to consume further topics in physics through coursework, while also continuing to complete the master's requirements, and I began to acquire more specific interests in the field. As a consequence of taking Dr. Lubos Mitas' course in computational physics I developed an interest in the study of nanoscience and materials physics. In this course, I recognized how beneficial the marriage of theory and machine can be – the ability to extract unexpected properties and new understanding of these systems from simulation were benefits I found exceptionally satisfying not to mention beautiful. After graduating with my master's, I was invited into NCSU's physics PhD program where I joined Professor Mitas' research group enabling me to take part in materials simulation research directly at the professional level which lead to the work presented in this thesis.

Acknowledgements

I am particularly grateful to my advisor, Professor Lubos Mitas. Throughout my PhD, he has been supportive and encouraging while also cultivating an atmosphere that has driven me to continually improve in all aspects of my work. His enthusiasm for science is contagious and I have been lucky to have him as a mentor these last few years.

I am also grateful to Luke Shulenburger and Thomas Mattsson, who provided me with the opportunity to take part in research within a laboratory setting during my PhD studies. This experience was vastly enriching and strengthened my abilities as a researcher. I greatly appreciate the chance to work along side them.

I would also like to thank the current and past members of the Mitas group. Special thanks to Adem Kulahlioglu and Kevin Rasch, each being particularly helpful mentors when I first entered the group. I am also grateful for the many invaluable discussions with Cody Melton and have enjoyed being colleagues during our time in the group.

Finally, I am tremendously thankful to all the love and support from my family and friends during my studies.

Table of Contents

List of Tables	ix
List of Figures	xviii
1 Introduction	1
1.1 Electronic Structure	2
1.2 Hartree-Fock	3
1.3 Configuration Interaction	6
1.4 Coupled Cluster	8
1.5 Density Functional Theory	8
2 Quantum Monte Carlo	12
2.1 Monte Carlo Integration	12
2.2 Metropolis Algorithm	14
2.3 Variational Monte Carlo	15
2.4 Diffusion Monte Carlo	17
3 Quantum Monte Carlo Study of mono(benzene)TM and bis(benzene)TM	24
3.1 Introduction	25
3.2 Computational Details	26
3.3 Results and Discussion	28
3.3.1 MoBz and MoBz ₂ systems	28
3.3.2 WBz and WBz ₂ systems	29
3.4 Conclusions	33
4 Quantum Monte Carlo with Variable Spins	35
4.1 Introduction	36
4.2 Fixed-Phase Diffusion Monte Carlo	37
4.2.1 Fixed-phase upper bound property	38
4.2.2 Fixed-phase as a special case of the fixed-node	38
4.2.3 Importance sampling	39
4.3 Spin Orbit Interactions	41
4.3.1 AREP and SO Operators	41
4.3.2 Variational property of the fixed-phase method for nonlocal, complex, Hermitian operators	45
4.4 Spin Representation and Sampling	47
4.4.1 Spin Representations	47
4.4.2 Trial Wave Functions	49
4.4.3 Evaluation of the Pseudopotential and Importance Sampling	52
4.4.4 Spin Sampling	55

4.5	Time-step Errors and Approximations	56
4.6	Applications	61
4.6.1	PbH	61
4.6.2	Sn and Sn ₂	62
4.6.3	Electron Affinities	64
4.7	Conclusions	66
5	A New Generation of Effective Core Potentials for Correlated Calculations	67
5.1	Introduction	68
5.2	Desired properties	71
5.3	Effective ECP Hamiltonian: Isospectrality on a subspace of valence states . .	72
5.3.1	ECP Form	73
5.4	Optimization methods and constructions	74
5.4.1	Objective function with atomic spectral discrepancies only	74
5.4.2	Objective function with spectral and spatial density matrix discrepancies	76
5.4.3	Optimization methods	77
5.4.4	Constructed, Combined and Iterated Schemes	78
5.5	Results	79
5.5.1	Boron	79
5.5.2	Carbon	82
5.5.3	Nitrogen	84
5.5.4	Oxygen	86
5.5.5	Sulfur	88
5.6	Transferability	92
5.7	Conclusions	94
6	A New Generation of Effective Core Potentials from Correlated Calculations: 2nd row elements	101
6.1	Introduction	102
6.2	ECP Atomic Correlation Energies	104
6.3	Construction	106
6.3.1	Many-body Energy Consistency	107
6.3.2	Single-body Norm Conservation	108
6.3.3	Weighted Combination and Core-Valence Partitioning	108
6.4	ECP Form	110
6.5	Valence basis sets	112
6.6	Results	113
6.6.1	Sodium	113
6.6.2	Magnesium	114
6.6.3	Aluminum	116
6.6.4	Silicon	117

6.6.5	Phosphorus	121
6.6.6	Sulfur	123
6.6.7	Chlorine	123
6.6.8	Argon	125
6.6.9	Molecular binding parameters, total energies and core radii	127
6.7	Conclusions	129
7	A New Generation of Effective Core Potentials from Correlated Calculations:	
	3d Transition Metal Series	133
7.1	Introduction	134
7.2	Methods	136
7.2.1	ECP Parametrization	136
7.2.2	Objective Function and Optimization Protocols	137
7.3	Results	142
7.3.1	Atomic Spectra	145
7.3.2	Scandium	148
7.3.3	Titanium	148
7.3.4	Vanadium	150
7.3.5	Chromium	152
7.3.6	Manganese	152
7.3.7	Iron	155
7.3.8	Cobalt	157
7.3.9	Nickel	157
7.3.10	Copper	160
7.3.11	Zinc	163
7.3.12	Average molecular discrepancies	163
7.4	Conclusions	167
8	Projector quantum Monte Carlo with averaged vs. explicit spin-orbit effects:	
	applications to tungsten molecular systems	170
8.1	Introduction	171
8.2	Fixed-Phase Diffusion Monte Carlo	172
8.3	Spin-Orbit Interactions and Dynamic Spins	174
8.4	Results	177
8.4.1	Tungsten Oxide	178
8.4.2	Tungsten Dimer	181
8.5	Conclusions	183
9	Conclusion	185
	Bibliography	187

Appendices	201
A Supplementary Material: A new generation of effective core potentials for correlated calculations	202
A.1 Basis Set Extrapolation	202
A.2 Correlation Consistent Basis Sets for the ECPs	203
B Supplementary Material: A new generation of effective core potentials from correlated calculations: 2nd row elements	210
B.1 Atomic Data	210
B.2 Molecular Data	221
B.3 Cl and Ar alternatives	226
C Supplementary Material: New generation of effective core potentials from correlated calculations: 3d transition metal series	233
C.1 Basis Sets	233
C.2 Atomic Spectra	234
C.2.1 Sc	235
C.2.2 Ti	235
C.2.3 V	237
C.2.4 Cr	240
C.2.5 Mn	242
C.2.6 Fe	244
C.2.7 Co	244
C.2.8 Ni	248
C.2.9 Cu	250
C.2.10 Zn	252

List of Tables

Table 3.1	The structural parameters are given. The bond lengths ($R/\text{\AA}$) and dihedral angles ($\angle/^\circ$) were obtained from DFT-TPSSh calculations. . .	29
Table 3.2	Bond distances ($R/\text{\AA}$) and dihedral angles ($\angle/^\circ$) of WBz.	31
Table 3.3	Bond distances ($R/\text{\AA}$) and dihedral angles ($\angle/^\circ$) of WBz ₂	31
Table 3.4	Binding energies [eV] of W-Bz and W-Bz ₂ and systems from DFT methods and from fixed-node DMC with two trial wave functions constructed with PBE and PBE0 orbitals.	32
Table 4.1	PbH bond length (r_e) and dissociation energy (D_e)	62
Table 4.2	Excitation energies for the Sn atom from the ³ P ₀ ground state. We include both LC and SC PPs. For completeness, we include COSCI and FCI to compare the FPSODMC and experiment[Nis].	63
Table 4.3	Electron Affinities for the 6 <i>p</i> elements. COSCI trial wave functions used throughout, with LC AREP/REPs for Pb[MS00], Bi[Sto02], Po[Sto02], and At[Sto02]. For Tl, no LC REP was found, so we utilize a SC AREP/REP [Met00].	66
Table 5.1	Atomic and ionic excitations and corresponding discrepancies for Boron. IP denotes the first ionization potential while EA is the electron affinity. Q is the ionization charge, 2S+1 the usual total spin multiplicity. AE denotes the calculated all-electron values while the rest of columns shows the discrepancies. UC means all-electron valence-only correlation with self-consistent but uncorrelated core, as explained in the text. All energies in eV. The MAD is the mean absolute difference over all of the discrepancies. Note that all gaps are calculated with reference to the ground state, namely Q=0 and 2S+1 = 2. The same notation applies to all the atomic/ionic data tables throughout the paper.	79
Table 5.2	ECP parameters for Constructed Boron. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.	81
Table 5.3	Atomic data for Carbon, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely Q=0 and 2S+1 = 3.	83
Table 5.4	ECP parameters for Constructed Carbon. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.	84

Table 5.5	Atomic data for Nitrogen, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely $Q=0$ and $2S+1 = 4$	84
Table 5.6	ECP parameters for Constructed Nitrogen. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.	86
Table 5.7	Atomic data for Oxygen, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely $Q=0$ and $2S+1 = 3$	88
Table 5.8	ECP parameters for Spectral Oxygen. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.	88
Table 5.9	Atomic data for Sulfur, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely $Q=0$ and $2S+1 = 3$	89
Table 5.10	ECP parameters for Constructed ECP for Sulfur. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.	92
Table 5.11	Mean absolute deviations of discrepancies of binding parameters at equilibrium (D_e , r_e and ω_e) and near the dissociation threshold (D_{diss}) at short bond lengths for our ECPs and previous constructions with respect to all-electron CCSD(T) calculations. The system sets correspond to Fig.5.14 except for the BH_3 molecule which was omitted from the MADs of the dissociation threshold energy.	93
Table 6.1	Parameter values for Ne-core ECPs. For all ECPs, the highest l value corresponds to the local channel.	110
Table 6.2	Parameter values for He-core ECPs. For all ECPs, the highest l value corresponds to the local channel.	111
Table 6.3	All-electron (AE) UCCSD(T) electron affinity and ionization potential of Na along with the errors from uncorrelated core (UC), ECPs and for information purposes also from experiment (Exp.). The uncontracted aug-cc-pCV5Z basis was used for all calculations. MAD is the mean absolute deviation of excitation energies, while MARE is the mean absolute relative error. All values in eV.	114
Table 6.4	All-electron UCCSD(T) ionization potentials for Mg along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. III for further description.	116

Table 6.5	All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Al along with the errors from the uncorrelated core (UC) and ECPs. Exp. gives experimental values for information purposes. The uncontracted aug-cc-pCV5Z basis was used for all calculations. AMAD is the mean absolute deviation for all excitation energies, LMAD is the mean absolute deviation for the first and second ionization potentials and electron affinity, while MARE is the mean absolute relative error for all states. All values in eV.	117
Table 6.6	All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Si along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.	119
Table 6.7	All-electron (AE) UCCSD(T) ionization potentials and electron affinity of P along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.	121
Table 6.8	All-electron (AE) UCCSD(T) valence ionization potentials and electron affinity of S along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.	123
Table 6.9	All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Cl along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.	125
Table 6.10	All-electron (AE) UCCSD(T) ionization potentials for Ar along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.	127
Table 6.11	Mean absolute deviations of discrepancies of binding parameters of all molecules considered in this work at equilibrium (D_e , r_e and ω_e) and near the all-electron dissociation threshold, $D_{\text{diss}}(\lesssim 0.05 \text{ \AA})$, at short bond lengths for our ECPs and previous constructions with respect to all-electron UCCSD(T) calculations. Errors in parenthesis such as “2.4(4)” denote “ 2.4 ± 0.4 ” and correspond to deviations of Morse potential fits to binding curves.	129

Table 6.12	Total energies of our ccECPs with Ne (LC) and He (SC) cores in their neutral ground states along with their core radii for each angular momentum channel – the values r_l are taken as the distance which the channel's full potential agrees with the bare Coloumb potential to within 10^{-5} Ha while $r_{l,nl}$ is the distance at which the channel's non-local potential drops below 10^{-5} Ha. Both RHF/ROHF and UCCSD(T) correlation energies are extrapolated to the CBS limit using the uncontracted aug-cc-pwCV{T,Q,5}Z bases. Energies given in Ha. Radii given in Å.	130
Table 7.1	A summary of the atomic state discrepancy data. For each atom, we provide mean absolute deviations (MADs). LMAD corresponds to the low-lying atomic spectrum, which includes the electron affinity, ionization potential, and the neutral and first ionized $s \leftrightarrow d$ transitions. The MAD includes the LMAD as well as ionizations down to the Ar core. A * indicates that the our final ccECP is equivalent to our spectral only ccECPs, as described in the atomic sections. N/A indicates that the particular ECP does not exist.	147
Table 7.2	Parameter values for hydrogen. The highest l value corresponds to the local channel. Each term takes the form $V_{\ell k}(r) = \beta_{\ell k} r^{n_{\ell k}-2} \exp[-\alpha_{\ell k} r^2]$	163
Table 7.3	Mean absolute deviations of binding parameters for various core approximations with respect to AE data for transition metal hydride and oxide molecules. All parameters were obtained using Morse potential fit. The parameters shown are dissociation energy D_e , equilibrium bond length r_e , vibrational frequency ω_e and binding energy discrepancy at dissociation bond length D_{diss}	165
Table 7.4	Parameter values for early transition metal ECPs. For all ECPs, the highest l value corresponds to the local channel. Each term takes the form $V_{\ell k}(r) = \beta_{\ell k} r^{n_{\ell k}-2} \exp[-\alpha_{\ell k} r^2]$	166
Table 7.5	Parameter values for late transtion metal ECPs. For all ECPs, the highest l value corresponds to the local channel. Each term takes the form $V_{\ell k}(r) = \beta_{\ell k} r^{n_{\ell k}-2} \exp[-\alpha_{\ell k} r^2]$	166
Table 8.1	Total energies for WO for the $^3\Sigma$ and $^5\Pi$ states at $r_e = 1.67$ Å. These results only utilize a scalar relativistic Hamiltonian and neglects spin-orbit.	179
Table 8.2	FNDMC (AREP) and FPSODMC (REP) dissociation energies of W_2 compared with other methods. In the fixed-node DMC calculations the trial nodes are from Slater determinants built from PBE0 single-particle orbitals.	182

Table A.1	The uncontracted aug-cc-pCV5Z basis set errors (a.u.) corresponding to nitrogen's CCSD(T) excitation energies calculated in the main text.	203
Table A.2	The uncontracted aug-cc-pCV5Z basis set errors (a.u.) corresponding to sulfur's CCSD(T) excitation energies calculated in the main text. . .	204
Table A.3	Correlation Consistent basis sets for the Boron ECP	205
Table A.4	Correlation Consistent basis sets for the Carbon ECP	206
Table A.5	Correlation Consistent basis sets for the Nitrogen ECP	207
Table A.6	Correlation consistent basis set for the Oxygen ECP.	208
Table A.7	Correlation Consistent basis sets for the Sulfur ECP	209
Table B.1	Total energy components of various states of the Na atom using our ccECP[Ne]	211
Table B.2	Total energy components of various states of the Na atom using our ccECP[He]	211
Table B.3	Total energy components of various states of the Mg atom using our ccECP[Ne]	211
Table B.4	Total energy components of various states of the Mg atom using our ccECP[He]	212
Table B.5	Total energy components of various states of the Al atom using our ccECP[Ne]	212
Table B.6	Total energy components of various states of the Al atom using our ccECP[He]	213
Table B.7	Total energy components of various states of the Si atom using our ccECP[Ne]	213
Table B.8	Total energy components of various states of the Si atom using our ccECP[He]	214
Table B.9	Total energy components of various states of the P atom using our ccECP[Ne]	214
Table B.10	Total energy components of various states of the P atom using our ccECP[He]	215
Table B.11	Total energy components of various states of the S atom using our ccECP[Ne]	215
Table B.12	Total energy components of various states of the S atom using our ccECP[He]	216
Table B.13	Total energy components of various states of the Cl atom using our ccECP[Ne]	217
Table B.14	Total energy components of various states of the Cl atom using our ccECP[He]	218
Table B.15	Total energy components of various states of the Ar atom using our ccECP[Ne]	219
Table B.16	Total energy components of various states of the Ar atom using our ccECP[He]	220

Table B.17	All-electron (AE) UCCSD(T) Na2 ground state ($^1\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	221
Table B.18	All-electron (AE) UCCSD(T) NaO ground state ($^2\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	222
Table B.19	All-electron (AE) UCCSD(T) MgO ground state ($^1\Sigma^+$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	222
Table B.20	All-electron (AE) UCCSD(T) Al2 ground state ($^3\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	224
Table B.21	All-electron (AE) UCCSD(T) AlO ground state ($^2\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	224
Table B.22	All-electron (AE) UCCSD(T) Si2 ground state ($^3\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	225
Table B.23	All-electron (AE) UCCSD(T) SiO ground state ($^1\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	225
Table B.24	All-electron (AE) UCCSD(T) P2 ground state ($^1\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	226
Table B.25	All-electron (AE) UCCSD(T) PO ground state ($^2\Pi$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	226
Table B.26	All-electron (AE) UCCSD(T) S2 ground state ($^3\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	227
Table B.27	All-electron (AE) UCCSD(T) SO ground state ($^3\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	227
Table B.28	All-electron (AE) UCCSD(T) Cl2 ground state ($^1\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPS. All energies in eV.	228
Table B.29	All-electron (AE) UCCSD(T) ClO ground state ($^2\Pi$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPS. All energies in eV.	228

Table B.30	All-electron (AE) UCCSD(T) ArH+ ground state ($^1\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.	229
Table B.31	Parameter values for Cl and Ar ccECP[Ne].S potentials. The highest l value corresponds to the local channel.	229
Table B.32	All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Cl along with the errors from ccECP[Ne].S. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V in main text for further description.	230
Table B.33	All-electron (AE) UCCSD(T) Cl ₂ ground state ($^1\Sigma_g$) binding parameters and potential energy surface along with the errors from ccECP[Ne].S. All energies in eV.	230
Table B.34	All-electron (AE) UCCSD(T) ClO ground state ($^2\Pi$) binding parameters and potential energy surface along with the errors from ccECP[Ne].S. All energies in eV.	231
Table B.35	All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Ar along with the errors from ccECP[Ne].S. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V in main text for further description.	231
Table B.36	All-electron (AE) UCCSD(T) ArH+ ground state ($^1\Sigma$) binding parameters and potential energy surface along with the errors from ccECP[Ne].S. All energies in eV.	232
Table C.1	Total energy components of the [Ar] $3d^1 4s^2 \ ^2D$ ground state of the Sc atom using our ccECP.	235
Table C.2	Sc AE gaps and relative errors for various core approximations. All gaps are relative to the [Ar] $3d^1 4s^2 \ ^2D$ state. All values in eV.	235
Table C.3	ScH AE molecular binding parameters and discrepancies for various core approximations. All parameters were obtained using Morse potential fit. The parameters shown are dissociation energy D_e , equilibrium bond length r_e , vibrational frequency ω_e and dissociation energy discrepancy at dissociation bond length D_{diss}	236
Table C.4	ScO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	236
Table C.5	Total energy components of the [Ar] $3d^2 4s^2 \ ^3F$ ground state of the Ti atom using our ccECP.	237
Table C.6	Ti AE gaps and relative errors for various core approximations. All in eV.	237
Table C.7	TiH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	238
Table C.8	TiO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	238

Table C.9	Total energy components of the [Ar] $3d^34s^2\ ^4F$ ground state of the V atom using our ccECP.	238
Table C.10	V AE gaps and relative errors for various core approximations. All data in eV.	239
Table C.11	VH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	239
Table C.12	VO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	240
Table C.13	Total energy components of the [Ar] $3d^54s^1\ ^7S$ ground state of the Cr atom for our ccECP.	240
Table C.14	Cr AE gaps and relative errors for various ECPs. All values in eV	241
Table C.15	CrH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	241
Table C.16	CrO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	242
Table C.17	Total energy components of the [Ar] $3d^54s^2\ ^6S$ ground state of the Mn atom for our ccECP.	242
Table C.18	Mn AE gaps and relative error for various core approximations. All in eV.	243
Table C.19	MnH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	243
Table C.20	MnO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	244
Table C.21	Total energy components of the [Ar] $3d^64s^2\ ^5D$ ground state of the Fe atom.	244
Table C.22	Fe AE gaps and relative errors for various ECPs. All data in eV	245
Table C.23	FeH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	245
Table C.24	FeO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	246
Table C.25	Total energy components of the [Ar] $3d^74s^2\ ^4F$ ground state of the Co atom using our ccECP.	246
Table C.26	Co AE gaps and relative errors in eV	246
Table C.27	CoH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	247
Table C.28	CoO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	247
Table C.29	Total energy components for the [Ar] $3d^84s^2\ ^3F$ state of the Ni atom for our ccECP.	248
Table C.30	Ni AE gaps and relative errors for various ECPs. All values in eV	248
Table C.31	NiH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	249

Table C.32	NiO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	249
Table C.33	Total energy components of the [Ar] $3d^{10}4s^1 2S$ ground state of the Cu atom using our ccECP.	250
Table C.34	Cu AE and relative errors for various core approximations. All values in eV	250
Table C.35	CuH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	251
Table C.36	CuO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	251
Table C.37	Total energy components for the [Ar] $3d^{10}4s^2 1S$ ground state of the Zn atom using our ccECP.	252
Table C.38	Zn AE gaps and relative errors for various ECPs. All values in eV	252
Table C.39	ZnH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	253
Table C.40	ZnO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3	253

List of Figures

Figure 2.1	A diffusion Monte Carlo simulation of the ground state of a one-dimensional harmonic oscillator. The walkers are initially distributed uniformly then evolved through 8 units of imaginary time and the birth/death re-weighting scheme is used during their evolution. After the simulation completes, the walkers are distributed according to ψ_0 as expected.	20
Figure 3.1	Molecular geometries of (a) bis(benzene)molybdenum (MoBz_2) and (b) mono(benzene)molybdenum (MoBz). The geometries of the corresponding W systems are similar.	27
Figure 3.2	The binding energies of the MoBz_2 and MoBz molecules calculated by DFT and FNDMC methods compared with experiment (a) and (b). FNDMC trial function has been constructed using TPSSh functional.	30
Figure 3.3	Radial components of $s(\ell = 0)$, $p(\ell = 1)$ and $d(\ell = 2)$ valence pseudorbitals plotted as $r^\ell \rho_\ell(r)$ for Mo (full, blue) and W (dashed, red) atoms. The maximum of s -orbital for W is higher than for Mo (relativity), while the opposite is true in the d -channel. Also note the significantly larger radius of the maximum (indicated by arrows) with the consequent smaller amount of charge in the core region in the d -channel of W.	33
Figure 4.1	Total Energy of the Pb ground state. The “No Drift” and “Drift” calculations are indistinguishable at this scale.	57
Figure 4.2	Total energies of the Pb ground state with COSCI trial wave functions. The calculations were performed with drift, $\mathbf{v}_D^S \neq 0$, and excluding drift, $\mathbf{v}_D^S = 0$	59
Figure 4.3	Total energy using COSCI and CISDT trial wave functions for the Pb ground state. FCI with cc-VQZ and a CBS extrapolation are included as a reference.	60
Figure 4.4	Binding Curve of the Sn_2 molecule using averaged spin-orbit AREP with FNDMC and spin-orbit REP with FPSODMC methods. The curves are offset to dissociation limit $2E_0(\text{Sn})$ within each method to enable comparison for the predicted binding energy of each method with experiment.	64
Figure 4.5	Binding curve of the Sn_2 molecule using large- and small-core REP. The large- and small-core systems have 8 and 44 valence electrons, respectively.	65

Figure 5.1	Boron dimer potential energy surface. UC represents an all-electron CCSD(T) calculation with a self-consistent but <i>uncorrelated</i> core, i.e., with no excitations from the core states. Spectral represents the optimization for the atomic spectrum alone, and Constructed indicates the ECP driven iteratively to minimize the dimer discrepancy while accepting a small increase in the spectrum discrepancy.	80
Figure 5.2	Boron dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve. The gray envelope represents a 0.05 eV window for the discrepancy. The vertical line indicates the equilibrium bond length as predicted by the all-electron CCSD(T) calculation.	81
Figure 5.3	Carbon dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.	83
Figure 5.4	Nitrogen dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.	85
Figure 5.5	Oxygen dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.	87
Figure 5.6	Potential energy surfaces of the S ₂ molecule from CCSD(T). We have plotted the predictions from various treatments of the sulfur cores. Shown are the all-electron core (AE), all-electron uncorrelated core (UC), Burkatzki-Filippi-Dolg (BFD), Dirac-Fock Trail-Needs (TN) and the CCSD(T) spectrum matched (Spectral) ECPs described in the text.	89
Figure 5.7	Sulfur dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.	90
Figure 5.8	Sulfur dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.	91
Figure 5.9	NH binding energy discrepancies for various ECPs	94
Figure 5.10	OH binding energy discrepancies for various ECPs. For Oxygen, we use our spectral ECP.	95
Figure 5.11	NO binding energy discrepancies for various ECPs. For Nitrogen, we use our constructed ECP and for Oxygen, we use our spectral ECP.	96
Figure 5.12	SH binding energy discrepancies compared to the all-electron CCSD(T) binding curve. For Sulfur, we use our constructed ECP and for Oxygen, we use our spectral ECP.	97
Figure 5.13	SO binding energy discrepancies compared to the all-electron CCSD(T) binding curve. For Sulfur, we use our constructed ECP.	98
Figure 5.14	Discrepancies of molecular binding parameters of our ECPs, UC and previous constructions with respect to all-electron CCSD(T) calculations. Parameters were obtained from Morse potential fits in all cases.	99

Figure 6.1	For the silicon atom, the spread of CCSD(T) valence-valence correlation errors (Δ_{cVV}) and spread of HF errors (Δ_{HF}) for various excitation energies from a variety of previously tabulated ECPs, in particular, BFD[Bur07], CRENBL[PC85], SBKJC[Ste84], STU[Dol87] and TN-DF[TN05].	105
Figure 6.2	For the phosphorus atom, the spread of CCSD(T) valence-valence correlation errors (Δ_{cVV}) and spread of HF errors (Δ_{HF}) for various excitation energies from a variety of previously tabulated ECPs, in particular, BFD[Bur07], CRENBL[PC85], SBKJC[Ste84], STU[Dol87] and TN-DF[TN05].	106
Figure 6.3	Binding energy discrepancies for (a) Na ₂ and (b) NaO molecules in their ground states $^1\Sigma_g$ and $^2\Sigma$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	115
Figure 6.4	Binding energy discrepancies for the MgO molecule in its ground state $^1\Sigma^+$. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	116
Figure 6.5	Binding energy discrepancies for (a) Al ₂ and (b) AlO molecules in their ground states $^3\Sigma_g$ and $^2\Sigma$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	118
Figure 6.6	Binding energy discrepancies for (a) Si ₂ and (b) SiO molecules in their ground states $^3\Sigma_g$ and $^1\Sigma$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	120
Figure 6.7	Binding energy discrepancies for (a) P ₂ and (b) PO molecules in their ground states $^1\Sigma_g$ and $^2\Pi$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	122
Figure 6.8	Binding energy discrepancies for (a) S ₂ and (b) SO molecules in their ground states $^3\Sigma_g$ and $^3\Sigma$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	124
Figure 6.9	Binding energy discrepancies for (a) Cl ₂ and (b) ClO molecules in their ground states $^1\Sigma_g$ and $^2\Pi$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	126

Figure 6.10	Binding energy discrepancies for the ArH ⁺ molecule in its ground state ¹ Σ. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	128
Figure 7.1	Spread of the contributions to the excitation energy for a variety of ECPs compared to all-electron for the Fe atom. ΔHF shows the variation in the HF errors, whereas ΔcVV shows the variation in the correlation energy error compared against the AE valence-valence correlation energy.	138
Figure 7.2	Discrepancies for the various core approximations compared to all-electron CCSD(T) for select states. Each state discrepancy uses the neutral <i>s</i> ² <i>d</i> ^{<i>n</i>} occupation as the reference, which is the neutral ground state for each transition metal except for Cr and Cu, which have <i>s</i> ¹ <i>d</i> ⁵ and <i>s</i> ¹ <i>d</i> ¹⁰ ground states correspondingly. (a) the neutral <i>s</i> ² <i>d</i> ^{<i>n</i>} → <i>s</i> ¹ <i>d</i> ^{<i>n</i>+1} excitation. (b) the neutral <i>s</i> ² <i>d</i> ^{<i>n</i>} → <i>d</i> ^{<i>n</i>+2} (c) the ionization from <i>s</i> ² <i>d</i> ^{<i>n</i>} → <i>s</i> ¹ <i>d</i> ^{<i>n</i>} (d) the ionization from <i>s</i> ² <i>d</i> ^{<i>n</i>} → [Ar]. The shaded gray window in each figure indicates a discrepancy of half of chemical accuracy in either direction from the all-electron reference. We note that for Sc and Ti, our final ccECP is equivalent to our spectral ccECP.S.	143
Figure 7.3	Mean Absolute Deviation, or MAD, for the TMs considering a large part of the spectrum from [Ar] up to low-lying neutral excited states and the anion. The shaded region of half of chemical accuracy is not visible on this scale. We note that for Sc and Ti, our final ccECP is equivalent to our spectral ccECP.S	144
Figure 7.4	Binding energy discrepancies for (a) ScH and (b) ScO molecules. The binding curves are relative to the AE CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	149
Figure 7.5	Binding energy discrepancies for (a) TiH and (b) TiO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	151
Figure 7.6	Binding energy discrepancies for (a) VH and (b) VO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	153
Figure 7.7	Binding energy discrepancies for (a) CrH and (b) CrO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	154
Figure 7.8	Binding energy discrepancies for (a) MnH and (b) MnO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	156

Figure 7.9	Binding energy discrepancies for (a) FeH and (b) FeO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	158
Figure 7.10	Binding energy discrepancies for (a) CoH and (b) CoO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	159
Figure 7.11	Binding energy discrepancies for (a) NiH and (b) NiO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	161
Figure 7.12	Binding energy discrepancies for (a) CuH and (b) CuO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	162
Figure 7.13	Binding energy discrepancies for (a) ZnH and (b) ZnO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	164
Figure 8.1	FPSODMC total energies of WO molecule as a function of spin time step τ_s /effective spin mass μ_s for the lowest states with $^3\Sigma$ and $^5\Pi$ symmetries. Note that the within the error bars the difference between the states remains very similar regardless of the time step. . . .	180
Figure 8.2	The difference in the charge density between DHF and PBE0 trial wave functions for the outermost spinors $\pi^4\sigma_{z^2}^2\sigma_s^2\delta^4$. In red, we show where PBE0 has a greater charge density and in blue where DHF shows a greater charge density. The black spheres indicate the location of the Tungsten atoms.	182
Figure B.1	Binding energy discrepancies for (a) Cl ₂ and (b) ClO molecules in their ground states $^1\Sigma_g$ and $^2\Pi$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	223
Figure B.2	Binding energy discrepancies for the ArH ⁺ molecule in its ground state $^1\Sigma$. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.	232

CHAPTER 1

Introduction

Quantum theory provides a remarkably accurate way to describe the properties of chemical systems and materials. In fact, it was immediately clear after its birth in the early 20th century that quantum theory underpins the entire field of chemistry [DF29] and as a consequence precisely describes any system composed of either atoms, molecules, or both. This was an astonishing success and the theory's discovery represented a significant advancement in humanity's understanding of the Universe – this is evident if one considers the rich variety of objects that exist above atomic scales and below scales relevant to the theory of gravity. As a practical matter, the human endeavour to improve technology drives the discovery of new materials and their properties; with quantum theory in place, a complete description of everyday material objects (both realized and unrealized) was available, for example, reaction pathways of chemical substances, phase diagrams of materials, the hardness, color, conductivity, and transparency of objects, and so on.

In the absence of time-varying external fields, the properties of these systems are governed primarily by the time-independent Schrödinger equation involving the constituent electrons and nuclei. Solutions to this equation provide access to zero temperature and finite-temperature properties and, in addition, time-independent and time-dependent phenomena. In practice, however, there is an impediment to utilizing the Schrödinger equation to predict an object's properties to arbitrary accuracy. This is because a general

polynomial-scaling solution to the equation so far does not exist. The difficulty stems from the fact that for a quantum system of any practical interest, the number of composite particles will be large – on the order of 10^{23} for objects of human-scale volumes – and they undergo complicated motions due to pairwise interactions that depend on inter-particle distances. This quantum many-body problem, as it is commonly referred, does not imply a complete impasse, however. There has been much effort over many decades to construct methodologies that provide solutions that are accurate enough for practical importance.

In the following section, an explicit description of the problem will be written down and described. The subsequent sections will detail numerous methodologies that have been formulated to yield approximate solutions to the quantum many-body problem involving electrons and nuclei – each with varying levels of sophistication and accuracy.

1.1 Electronic Structure

Throughout, atomic units (a.u.) will be used where $m_e = e = \hbar = 4\pi\epsilon_0 = 1$. In a.u., the time-independent Schrödinger for a general system of N electrons and M nuclei, is given by

$$\left[-\frac{1}{2} \sum_{i=1}^N \nabla_{\mathbf{r}_i}^2 - \frac{1}{2} \sum_{I=1}^M \frac{1}{M_I} \nabla_{\mathbf{R}_I}^2 - \sum_{i=1}^N \sum_{I=1}^M \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{i<j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J}^M \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \right] \Psi = E\Psi, \quad (1.1)$$

where $\{\mathbf{r}_i\}$ and $\{\mathbf{R}_I\}$ are the sets of electron and nuclear positions, respectively, and $\{M_I\}$ and $\{Z_I\}$ are the sets of nuclear masses and charges, respectively. Even for the lightest element, H, whose nucleus consists of just a single proton, the nuclear mass is nearly 2000 times larger than the mass of the electron. To good approximation, therefore, the nuclei can be assumed fixed in space and their respective kinetic energies can be dropped from equation 1.1 – this simplification is commonly known as the Born-Oppenheimer approximation. Additionally, given that the nuclei are assumed fixed, the last term on the left-hand side of equation 1.1 is just a constant which can be removed then added back later if needed, e.g., if a nuclear potential energy surface or cohesive energy, etc., is being calculated. Thus, under the Born-Oppenheimer approximation, we have the *electronic* time-independent

Schrödinger equation,

$$\left[-\frac{1}{2} \sum_{i=1}^N \nabla_{\mathbf{r}_i}^2 - \sum_{i=1}^N \sum_{I=1}^M \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{i < j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] \Psi = E \Psi, \quad (1.2)$$

which consists of the kinetic energy of the electrons, their interactions with the nuclei, and their interactions with each other.

The field of electronic structure is devoted to using equation 1.2 to determine properties of systems in atomic, molecular, or condensed phases. As a result, in fits and starts, numerous methodologies have been developed to approximate solutions to the equation in an attempt to reach levels of accuracy that are high enough such that quantitative predictions of these systems can be made.

1.2 Hartree-Fock

The simplest way to avoid the difficulty of solving the electronic Schrödinger equation in the presence of electron-electron interactions is to utilize the variational principle of quantum mechanics and assume a trial wave function that is a product state of single-body orbitals. This particular trial state is known as a Hartree product and neglects the interaction of the electrons as if the Hamiltonian were separable. The variational principle, for a trial state $|\Psi_T\rangle$, asserts that the energy expectation value of $|\Psi_T\rangle$ is no less than the exact ground state energy of the system, namely,

$$E_0 \leq \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} \quad (1.3)$$

where H is the system's Hamiltonian and E_0 is the ground-state eigenvalue of H . Since E_0 is the system's ground state energy, that is, it is the lowest energy the system can possibly have, it is then clear that equation 1.3 holds since in general $|\Psi_T\rangle$ is a superposition of eigenstates of H and its energy expectation value must be larger than E_0 due to contributions from higher energy states.

For a system of N electrons with space-spin coordinates $\{\mathbf{x}_i = (\mathbf{r}_i, \omega_i)\}$, the Hartree product wave function, $\langle \mathbf{X} | \Psi_H \rangle$, is given by

$$\langle \mathbf{X} | \Psi_H \rangle \equiv \Psi_H(\mathbf{X}) = \chi_1(\mathbf{x}_1) \chi_2(\mathbf{x}_2) \cdots \chi_N(\mathbf{x}_N), \quad (1.4)$$

where $\mathbf{X} \equiv (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ are the combined coordinates of all the particles and $\{\chi_i\}$ is a set of one-particle spin-orbitals. Assuming that the orbitals are orthonormal, the energy expectation value of the Hartree-product is

$$\begin{aligned} \langle \Psi_H | H | \Psi_H \rangle = & \sum_{i=1}^N \int d^4x \chi_i^*(\mathbf{x}) \left(-\frac{1}{2} \nabla^2 + V_{\text{ions}} \right) \chi_i(\mathbf{x}) \\ & + \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N \int d^4x \int d^4x' \frac{|\chi_i(\mathbf{x})|^2 |\chi_j(\mathbf{x}')|^2}{|\mathbf{r} - \mathbf{r}'|}. \end{aligned} \quad (1.5)$$

The optimum set of spin-orbitals can be derived by using the method of Lagrange multipliers to minimize $\langle \Psi_H | H | \Psi_H \rangle$ with the constraint that the orbitals remain normalized

$$\begin{aligned} & \delta \left[\langle \Psi_H | H | \Psi_H \rangle - \sum_{i=1}^N (\epsilon_i \langle \chi_i | \chi_i \rangle - 1) \right] \stackrel{!}{=} 0 \\ \Rightarrow & \sum_{i=1}^N \int d^4x \delta \chi_i^*(\mathbf{x}) \left\{ -\frac{1}{2} \nabla^2 + V_{\text{ions}} + \sum_{j \neq i}^N \int d^4x' \frac{|\chi_j(\mathbf{x}')|^2}{|\mathbf{r} - \mathbf{r}'|} - \epsilon_i \right\} \chi_i(\mathbf{x}) \\ & + \text{complex conjugate} = 0. \end{aligned} \quad (1.6)$$

Since the variations of $\delta \chi_i^*$ and $\delta \chi_i$ are independent and arbitrary, the expression in the curly brackets must be zero, which yields the following expression

$$\left[-\frac{1}{2} \nabla^2 + V_{\text{ions}} + \sum_{j \neq i}^N \int d^4x' \frac{|\chi_j(\mathbf{x}')|^2}{|\mathbf{r} - \mathbf{r}'|} \right] \chi_i(\mathbf{x}) = \epsilon_i \chi_i(\mathbf{x}). \quad (1.7)$$

The set of equations defined by 1.7 are known as the Hartree equations. These equations are non-linear given that the potential (or field) felt by an electron in spin-orbital χ_i depends on the spin-orbitals of the other electrons and therefore the equations need to be solved in an iterative fashion. The particular iterative process to solve the equations is known as the self-consistent field (SCF) method in which a set of spin-orbitals is initially assumed and used to define the field felt by each electron then the equations are solved for a new set of spin-orbitals which are subsequently plugged back into the equations and solved again; this process is iterated until the input and output spin-orbitals are identical to within some tolerance (typically measured by the density, energy, or both).

The Hartree product, unfortunately, misses essential physics by only accounting for the

electrostatics of the electrons and neglecting that they are indistinguishable particles which should obey the Pauli exclusion principle. For identical fermions, the exclusion principle is satisfied by wave functions that are antisymmetric under exchange of any pair of particle coordinates,

$$\Psi(\dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots) = -\Psi(\dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots). \quad (1.8)$$

An extension of the Hartree product that respects antisymmetry is the Slater determinant which is defined as

$$\Psi(\mathbf{X}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix}, \quad (1.9)$$

where $1/\sqrt{N!}$ is a normalization factor. The matrix here, with rows formed from particle coordinates and columns formed from spin-orbitals, is commonly referred to as the Slater matrix. The antisymmetry of the Slater determinant follows from the property that interchanging any two rows of a matrix changes the sign of its determinant and interchanging two rows within a Slater matrix is equivalent to interchanging the coordinates of a pair of particles.

If we minimize the energy expectation value of the Slater determinant with respect to its spin-orbitals, as we did for the Hartree product, we arrive at the following equations for the spin-orbitals

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{ions}} + \sum_{j=1}^N \int d^4x' \frac{|\chi_j(\mathbf{x}')|^2}{|\mathbf{r}-\mathbf{r}'|} - \sum_{j=1}^N \int d^4x' \frac{|\chi_j(\mathbf{x}')|^2}{|\mathbf{r}-\mathbf{r}'|} \left(\frac{\chi_j(\mathbf{x}) \chi_i(\mathbf{x}')}{\chi_j(\mathbf{x}') \chi_i(\mathbf{x})} \right) \right] \chi_i(\mathbf{x}) = \epsilon_i \chi_i(\mathbf{x}), \quad (1.10)$$

which are the well-known Hartree-Fock equations. Notice the form here is consistent with our assumption that the electrons occupy one-body spin-orbitals, that is, each electron experiences an average potential, or mean-field, generated from the other electrons rather than a (non-separable) pointwise interaction. This mean-field has two contributions; one coming from pure electrostatics of the one-body densities as can be seen in the 3rd term in the square brackets of equation 1.10 and the other contribution is the exchange interaction, namely, the average of the Coulomb interaction multiplied by an operator that interchanges

the coordinates of the two particles as seen in the 4th term in the square brackets of equation 1.10.

Relative to its simplicity, the accuracy of the Hartree-Fock wave function, – i.e., a Slater determinant of Hartree-Fock orbitals – is quite high. For a general system, the energy expectation value of this wave function is around 99% of the system’s exact ground state energy. However, before we tempt ourselves into thinking that this accuracy will be good enough for all practical purposes, consider how this picture can change when calculating an energy difference. As an example, Hartree-Fock underbinds the carbon dimer significantly and gives a binding energy, $E_b = 2E(C) - E(C_2)$, that is only 8%-10% of the experimental value for this quantity. With a potential for errors as large as these, it is clear that the missing energy of Hartree-Fock, known as the correlation energy, can be critical for achieving useful predictions for present (and future) experimental measurements.

1.3 Configuration Interaction

To capture the correlation energy missing from Hartree-Fock, a number of post-Hartree-Fock methods have been formulated. One such method, known as configuration interaction (CI), diagonalizes the Hamiltonian of the system in a many-body basis of Slater determinants. In *full* CI, the basis is constructed by taking a set of single-body orbitals (generally more than N) then forming a basis of determinants from all possible N -orbital combinations. Full CI gives an exact solution within the Hilbert space spanned by the set of determinants and the set spans the full Hilbert space if the single-body orbitals span the entire one-body Hilbert space. To see this, let us write down the expansion of an arbitrary function of the coordinates of a single particle into a set of spin-orbitals

$$f(\mathbf{x}) = \sum_i a_i \chi_i(\mathbf{x}). \quad (1.11)$$

To form an arbitrary function, g , of the coordinates of two particles, we could then write

$$g(\mathbf{x}, \mathbf{x}') = \sum_i a_i(\mathbf{x}') \chi_i(\mathbf{x}), \quad (1.12)$$

where we can then expand each function in the set $\{a_i\mathbf{x}'\}$ into spin-orbitals, yielding

$$g(\mathbf{x}, \mathbf{x}') = \sum_i \sum_j b_{ij} \chi_j(\mathbf{x}') \chi_i(\mathbf{x}). \quad (1.13)$$

We can further impose that g be antisymmetric and normalized,

$$g(\mathbf{x}, \mathbf{x}') = \frac{1}{\sqrt{2}} \sum_i \sum_j b_{ij} [\chi_j(\mathbf{x}') \chi_i(\mathbf{x}) - \chi_i(\mathbf{x}') \chi_j(\mathbf{x})]. \quad (1.14)$$

Therefore, within the space spanned by $\{\chi_i\}$, an arbitrary antisymmetric function of the coordinates of two particles is an expansion of Slater determinants formed from all combinations of spin-orbital pairs. For antisymmetric functions of the coordinates of N particles, the same holds except the Slater determinants are formed from all N spin-orbital combinations.

Full CI, is not a feasible approach in practice given that the number of determinants grows exponentially with the number of spin-orbitals. To combat this exponential blow-up, various strategies truncate the number of determinants in some fashion. For instance, CI with single and double excitations (CISD) uses all determinants formed from single- and double-orbital excitations out of a Hartree-Fock reference determinant. Similarly, CISDT, extends this to include triple-orbital excitations. One can push this as far as possible, i.e., CISDTQ, and so on, where each successive order obtains more and more of the correlation energy and therefore provides a systematically improvable approach.

Additionally, the multi-configurational self consistent field (MCSCF) approach goes a step further and optimizes the spin-orbitals along with the CI expansion coefficients for a given level of excitation. This can be achieved by first diagonalizing the Hamiltonian of the system with a basis of determinants set by the chosen excitation level then constructing *natural orbitals* from this wave function – orbitals that diagonalize its single-body reduced density matrix – then performing an additional diagonalization with the new set of orbitals. The truncation schemes mentioned so far can often lead to qualitatively incorrect results unless a high level of excitation can be chosen and therefore are not appropriate for larger systems. In addition, these methods suffer from a size consistency problem, that is when calculating the energy of a system of constituent parts where the parts do not interact with one another, the energy obtained is not equal to the sum of the energies obtained from calculating the parts independently. Furthermore, the methods also are not size extensive,

meaning they do not scale correctly for a system of interacting subsystems as the number of subsystems is increased and therefore are infeasible to use for systems in the solid state.

Another truncation scheme, known as selected-CI, begins with a Hartree-Fock reference and iteratively expands into larger numbers of determinants. At each round, the choice of whether a determinant is included in the expansion is based on some criteria. This approach attempts to build an expansion of the *most important* determinants and is often quite successful for small to medium sized systems.

1.4 Coupled Cluster

Similar in spirit to configuration interaction, is the coupled cluster method. In this approach, an exponential ansatz is introduced

$$|\Psi_{CC}\rangle = e^T |\Psi_0\rangle \quad (1.15)$$

where

$$T = \sum_i^N T_i \quad (1.16)$$

and each operator, T_i , generates a linear combination of all determinants which excite i electrons into the virtual (unoccupied) orbitals. In practice, a truncation scheme has to be used here as well. One particular approach that is often used is to truncate T at double excitations, namely $T = T_1 + T_2$, then treat the triple excitations in a perturbative manner – this is referred to as coupled-cluster with single and double excitations and triples treated perturbatively, or simply CCSD(T). CCSD(T) has the ability to give very accurate results and consequently is referred to as the golden standard in quantum chemistry. However, CCSD(T) scales as $\mathcal{O}(N^7)$ and therefore is not feasible for medium to large systems.

1.5 Density Functional Theory

In 1964, Hohenberg and Kohn realized that the dimensionality of many-body problem could be significantly reduced by working exclusively with the 3-dimensional electronic

density of the system, defined as

$$n(\mathbf{r}) \equiv N \int d^3r_2 \cdots \int d^3r_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (1.17)$$

This realization followed from two theorems. The first states:

Theorem 1.1. *The external potential, V_{ext} , uniquely determines the electronic density, n , of an interacting system of electrons.*

Proof. Consider two external potentials, V_{ext} and V'_{ext} – which differ from each other by more than just an irrelevant constant shift – give rise to the same density $n(\mathbf{r})$. To state this in another way, we are assuming that the ground states, Ψ and Ψ' , of the systems containing V_{ext} and V'_{ext} , respectively, give the same density $n(\mathbf{r})$ through equation 1.17. Let the Hamiltonians of the systems containing V_{ext} and V'_{ext} be denoted as H and H' , respectively. Furthermore, let the ground state eigenvalues of H and H' be denoted as E and E' , respectively. From the variational principle, we have

$$\begin{aligned} E + E' &< \langle \Psi' | H | \Psi' \rangle + \langle \Psi | H' | \Psi \rangle \\ &= \langle \Psi' | H' + V_{\text{ext}} - V'_{\text{ext}} | \Psi' \rangle + \langle \Psi | H + V'_{\text{ext}} - V_{\text{ext}} | \Psi \rangle \\ &= E' + E + \langle \Psi' | V_{\text{ext}} - V'_{\text{ext}} | \Psi' \rangle - \langle \Psi | V_{\text{ext}} - V'_{\text{ext}} | \Psi \rangle. \end{aligned} \quad (1.18)$$

Since, Ψ and Ψ' give the same density, the last two terms cancel, leaving

$$E + E' < E + E'. \quad (1.19)$$

Clearly, equation 1.19 is incorrect and therefore Ψ and Ψ' must yield different densities which are uniquely determined from the external potentials. \square

The second theorem states:

Theorem 1.2. *The energy of the system can be expressed as a functional of the density, $E[n(\mathbf{r})]$.*

Proof. From theorem 1.1, for N -electron systems, non-trivial changes in the external potential give different electronic densities. In principle, the wave function can be obtained from the density. Furthermore, from the wave function, the energy of the system that corresponds to a given external potential can be computed. \square

Let us express the energy functional in the following way,

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int d^3r V_{\text{ext}}(\mathbf{r})n(\mathbf{r}) \quad (1.20)$$

where

$$F[n(\mathbf{r})] = T[n(\mathbf{r})] + W[n(\mathbf{r})] \quad (1.21)$$

contains contributions from the kinetic energy of the electrons, T , and their interactions with each other, W , and thus is a universal functional of the density. Minimizing $E[n(\mathbf{r})]$ with respect to the density will lead to the ground state energy of the given system. Let us write W as a sum of a pure electrostatic interaction of the density with itself and a correction E_{xc} , known as the exchange-correlation functional, that adds in the missing correlation and exchange contributions,

$$W[n(\mathbf{r})] = \frac{1}{2} \int d^3r \int d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{xc}}[n(\mathbf{r})]. \quad (1.22)$$

A clear approximation of the kinetic energy operator became evident after Kohn and Sham [KS65] envisioned writing the density as a sum over single-particle densities of a fictitious non-interacting system of particles,

$$n(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2, \quad (1.23)$$

where the particles occupy the set of orbitals $\{\phi_i\}$. This allows us to write the energy as

$$E[n(\mathbf{r})] = -\frac{1}{2} \sum_i \int d^3r \phi_i^*(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}) + \int d^3r V_{\text{ext}}(\mathbf{r})n(\mathbf{r}) + \frac{1}{2} \int d^3r \int d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{xc}}[n(\mathbf{r})] \quad (1.24)$$

and the exchange-correlation functional contains the corrections to T and W ,

$$E_{\text{xc}} = \left(T + \frac{1}{2} \sum_i \int d^3r \phi_i^*(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}) \right) + \left(W - \frac{1}{2} \int d^3r \int d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \right). \quad (1.25)$$

Unfortunately, the exact exchange-correlation functional is unknown and therefore a precise mapping of the Schrödinger equation to a 3-dimensional problem is not available. There has been, however, much effort to construct approximations to E_{xc} with varying levels

of complexity. The simplest form is based on the uniform electron gas (UEG) and is known as the local density approximation (LDA). Here, E_{xc} , is built from an exact expression for the exchange energy of the UEG and an approximation to its correlation energy,

$$E_{xc}^{\text{LDA}} = \int d^3r (\epsilon_x^{\text{UEG}}(n(\mathbf{r})) + \epsilon_c^{\text{UEG}}(n(\mathbf{r}))) n(\mathbf{r}), \quad (1.26)$$

where $\epsilon_x^{\text{UEG}}(n(\mathbf{r}))$ and $\epsilon_c^{\text{UEG}}(n(\mathbf{r}))$ are the exchange and correlation energies per electron, respectively. Parameterized forms of the correlation energy have been fit to high-accuracy quantum Monte Carlo calculations of the UEG [CA80] and are in wide use today. Another type of exchange-correlation functional, known as the generalized gradient approximation (GGA), writes the exchange-correlation energy per electron as a function of both the density and its gradient,

$$E_{xc}^{\text{GGA}} = \int d^3r \epsilon_{xc}^{\text{GGA}}(n(\mathbf{r}), \nabla n(\mathbf{r})) n(\mathbf{r}). \quad (1.27)$$

This approximation attempts to correct the fact that the density is not uniform everywhere as in the electron gas. Functionals known as meta-GGA attempt to push this further and include a dependence on the second derivative of the density. To correct the exchange energy, a type of functional known as a hybrid mixes in the exact exchange from Hartree-Fock.

Though there is a plethora of approximate functionals, there is no clear hierarchy of them based on their accuracy. Therefore, one is not able to determine, blindly, whether one functional will perform better for a given system and property than another. Furthermore, since there doesn't exist a well-defined litmus test for functionals, one ultimately has to compare to experiment to determine the accuracy of a calculated property; making it unreliable as a predictive theory.

Quantum Monte Carlo (QMC) methods are a powerful set of tools that use stochastic approaches to approximate solutions to the Schrödinger equation. A key advantage of using stochastic techniques is the favorable scaling with electron number. In addition, the methods are able to work directly with the exact Hamiltonian of the system rather than relying on an approximate model. In the following sections a description of Monte Carlo integration and the Metropolis algorithm is given, we then cover the variational and diffusion Monte Carlo methods in the final two sections – these two flavors of real-space QMC are used extensively in subsequent chapters of this work.

2.1 Monte Carlo Integration

Monte Carlo integration is a powerful tool that allows one to estimate integrals of high-dimensional functions. This is done by bringing the integral into a form such that it represents an expectation value of a *different* function over an appropriate probability distribution. For instance, consider the D -dimensional integral

$$I = \int d^D X f(\mathbf{X}). \quad (2.1)$$

Given a probability distribution, $\rho(\mathbf{X})$, which by definition must satisfies $\rho(\mathbf{X}) \geq 0$ and $\int d^D X \rho(\mathbf{X}) = 1$, we can re-write I as

$$I = \int d^D X \rho(\mathbf{X}) \left[\frac{f(\mathbf{X})}{\rho(\mathbf{X})} \right], \quad (2.2)$$

which clearly appears as an expectation value of $f(\mathbf{X})/\rho(\mathbf{X})$ over the distribution $\rho(\mathbf{X})$. If one were able to generate samples from $\rho(\mathbf{X})$, i.e., generate a set of values, $\{\mathbf{X}_i\}$, distributed according to $\rho(\mathbf{X})$, then an obvious way to approximate the integral presents itself, namely,

$$I \approx \frac{1}{M} \sum_{i=1}^M \left[\frac{f(\mathbf{X}_i)}{\rho(\mathbf{X}_i)} \right]. \quad (2.3)$$

The approximation given in equation 2.3 becomes exact in the limit that the number of samples approaches infinity. If we assume that the samples $\{\mathbf{X}_i\}$ are uncorrelated then by the central limit theorem multiple estimates of I (generated from equation 2.3) will be normally distributed with average

$$\mu_{f/\rho} = I \quad (2.4)$$

and variance

$$\sigma_{f/\rho}^2 = \frac{\sigma_I^2}{M} \quad (2.5)$$

where

$$\sigma_I^2 \equiv \int d^D X \rho(\mathbf{X}) \left[\frac{f(\mathbf{X})}{\rho(\mathbf{X})} - I \right]^2. \quad (2.6)$$

Consequently, the magnitude of the error bar associated with a given estimate of I is $\sigma_{f/\rho} = \sigma_I/\sqrt{M}$ and therefore the estimate will agree with the exact value of I with absolute error less than $\sigma_{f/\rho}$ with 68.2% probability. Generally, $\sigma_{f/\rho}^2$ is not known exactly, but it can also be estimated in a similar fashion as I in the following way

$$\sigma_{f/\rho}^2 \approx \frac{1}{M(M-1)} \sum_{i=1}^M \left[\frac{f(\mathbf{X}_i)}{\rho(\mathbf{X}_i)} - \frac{1}{M} \sum_{j=1}^M \frac{f(\mathbf{X}_j)}{\rho(\mathbf{X}_j)} \right]^2, \quad (2.7)$$

where the prefactor $1/(M-1)$ is there to ensure an unbiased estimate of the variance, σ_I^2 , given that there are only $M-1$ degrees of freedom since the residuals sum to zero. The key benefit of Monte Carlo integration over grid-based quadratures such as Simpson's rule is

that the error of the integral's estimate is completely independent of the dimensionality of the problem and therefore much more efficient for high-dimensional integrands.

2.2 Metropolis Algorithm

In cases where the samples $\{\mathbf{X}_i\}$ can not be generated directly from the distribution, $\rho(\mathbf{X})$, the Metropolis algorithm provides an alternative way to generate them by way of a Markov chain. A Markov chain is a stochastic sequence in which a *walker* moves iteratively along a trajectory $\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$ and the probability that the walker moves from \mathbf{X}_i to \mathbf{X}_{i+1} depends only on the walker's current location and not on its previous history.

The Metropolis algorithm begins with a walker located at a randomly chosen position \mathbf{X} . Then, based on a probability distribution, $T(\mathbf{X}' \leftarrow \mathbf{X})$, a trial move to \mathbf{X}' is proposed. The trial move is then accepted with the following probability

$$A(\mathbf{X}' \leftarrow \mathbf{X}) = \text{Min} \left(1, \frac{T(\mathbf{X} \leftarrow \mathbf{X}')\rho(\mathbf{X}')}{T(\mathbf{X}' \leftarrow \mathbf{X})\rho(\mathbf{X})} \right). \quad (2.8)$$

If accepted, the walker moves to location \mathbf{X}' , otherwise the walker remains at position \mathbf{X} . Then from the walker's current location, another trial move is made and the process repeats. After many iterations, the locations that the walker traverses will be distributed according to $\rho(\mathbf{X})$.

Let us prove that this procedure indeed gives the right distribution, $\rho(\mathbf{X})$. First, instead of imagining a single walker traversing a very long path, let's imagine a different scenario where multiple walkers traverse shorter paths simultaneously. This scenario will give the same distribution of visited locations as long as we allow the ensemble of walkers to first equilibrate in order to remove any bias introduced by initially distributing the walkers *incorrectly*—since the walkers' starting positions are chosen randomly, the distribution of the initial positions will very likely be wrong. After some time, we expect the ensemble to reach equilibrium and settle into some distribution $n(\mathbf{X})$ and consequently the number of walkers moving from volume element $d^{3N}X$ to $d^{3N}X'$ in a single timestep will be the same as the number moving from volume element $d^{3N}X'$ to $d^{3N}X$. The probability that a single walker in volume element $d^{3N}X$ steps to volume element $d^{3N}X'$ is equal to the probability that a move to $d^{3N}X'$ is proposed multiplied by the probability that the move is accepted.

Thus, the total number of walkers that move from $d^{3N}X$ to $d^{3N}X'$ in a single timestep is

$$d^{3N}X' A(\mathbf{X}' \leftarrow \mathbf{X}) T(\mathbf{X}' \leftarrow \mathbf{X}) [n(\mathbf{X}) d^{3N}X] \quad (2.9)$$

where $n(\mathbf{X}) d^{3N}X$ is the number of walkers currently in volume element $d^{3N}X$. Since this equals the number of walkers moving from $d^{3N}X'$ to $d^{3N}X$, we have

$$A(\mathbf{X}' \leftarrow \mathbf{X}) T(\mathbf{X}' \leftarrow \mathbf{X}) n(\mathbf{X}) = A(\mathbf{X} \leftarrow \mathbf{X}') T(\mathbf{X} \leftarrow \mathbf{X}') n(\mathbf{X}') \quad (2.10)$$

and therefore

$$\frac{n(\mathbf{X})}{n(\mathbf{X}')} = \frac{A(\mathbf{X} \leftarrow \mathbf{X}') T(\mathbf{X} \leftarrow \mathbf{X}')}{A(\mathbf{X}' \leftarrow \mathbf{X}) T(\mathbf{X}' \leftarrow \mathbf{X})}. \quad (2.11)$$

Without loss of generality, let's assume $A(\mathbf{X} \leftarrow \mathbf{X}') < 1$ and thus $A(\mathbf{X}' \leftarrow \mathbf{X}) = 1$ from equation 2.8. Their ratio is then

$$\frac{A(\mathbf{X} \leftarrow \mathbf{X}')}{A(\mathbf{X}' \leftarrow \mathbf{X})} = \frac{T(\mathbf{X}' \leftarrow \mathbf{X}) \rho(\mathbf{X})}{T(\mathbf{X} \leftarrow \mathbf{X}') \rho(\mathbf{X}')} \quad (2.12)$$

and we have

$$\frac{n(\mathbf{X})}{n(\mathbf{X}')} = \frac{\rho(\mathbf{X})}{\rho(\mathbf{X}')}. \quad (2.13)$$

Therefore, we have shown that the density of the ensemble of walkers is proportional to the desired distribution $\rho(\mathbf{X})$.

2.3 Variational Monte Carlo

The simplest flavor of quantum Monte Carlo is variational Monte Carlo (VMC). This method begins with a trial many-body wave function, $\psi_T(\mathbf{R}; \alpha)$, where \mathbf{R} is the collection of all spatial coordinates of the electrons in the system and α is a vector of variational parameters, then optimizes the set of parameters with respect to a chosen objective function in order to approximate a desired eigenstate of the system.

Often one is interested in approximating the ground state of the system and therefore the variational principle can be used to minimize the energy expectation value of ψ_T . The expression of ψ_T 's energy expectation value can be placed into a form such that Monte

Carlo integration can be readily performed, shown here

$$\begin{aligned}
E[\boldsymbol{\alpha}] &= \langle \psi_T | \psi_T \rangle^{-1} \int d^{3N} R \psi_T(\mathbf{R}; \boldsymbol{\alpha}) H \psi_T(\mathbf{R}; \boldsymbol{\alpha}) \\
&= \langle \psi_T | \psi_T \rangle^{-1} \int d^{3N} R |\psi_T(\mathbf{R}; \boldsymbol{\alpha})|^2 \left(\frac{H \psi_T(\mathbf{R}; \boldsymbol{\alpha})}{\psi_T(\mathbf{R}; \boldsymbol{\alpha})} \right) \\
&\equiv \langle \psi_T | \psi_T \rangle^{-1} \int d^{3N} R |\psi_T(\mathbf{R}; \boldsymbol{\alpha})|^2 E_{\text{Loc}}(\mathbf{R}; \boldsymbol{\alpha})
\end{aligned} \tag{2.14}$$

where E_{Loc} is referred to as the *local energy*. The final line in equation 2.14 can be identified as the expectation value of E_{Loc} over the distribution $|\psi_T(\mathbf{R}; \boldsymbol{\alpha})|^2 / \langle \psi_T | \psi_T \rangle$. Clearly then, we can generate M samples $\{\mathbf{R}_i\}$ from $|\psi_T(\mathbf{R}; \boldsymbol{\alpha})|^2$ with the Metropolis algorithm then approximate $E[\boldsymbol{\alpha}]$ as

$$E[\boldsymbol{\alpha}] \approx \frac{1}{M} \sum_{i=1}^M E_{\text{Loc}}(\mathbf{R}_i; \boldsymbol{\alpha}). \tag{2.15}$$

Note that since the normalization of ψ_T drops out of the expression for E_{Loc} , we can thankfully ignore it all together and use unnormalized trial forms. Since the integrals required to compute the energy of ψ_T are approximated with Monte Carlo integration, this provides a key advantage over many other wave function based methods since one is not restricted to trial forms whose energy can be analytically determined but rather almost any arbitrary form can be utilized.

Another optimization strategy in VMC takes advantage of the fact that E_{Loc} becomes a constant when ψ_T is equal to an eigenstate of the system and consequently the variance of E_{Loc} will be zero. For a given ψ_T , the corresponding variance of E_{Loc} will be bounded from below since it is a positive definite quantity and can therefore be used as an objective function to improve ψ_T . The variance of E_{Loc} is approximated by Monte Carlo integration in the following way

$$\begin{aligned}
\sigma^2[\boldsymbol{\alpha}] &= \langle E_{\text{Loc}}^2 \rangle_{|\Psi_T|^2} - \langle E_{\text{Loc}} \rangle_{|\Psi_T|^2}^2 \\
&\approx \frac{1}{M-1} \sum_{i=1}^M \left[E_{\text{Loc}}(\mathbf{R}_i; \boldsymbol{\alpha}) - \frac{1}{M} \sum_{j=1}^M E_{\text{Loc}}(\mathbf{R}_j; \boldsymbol{\alpha}) \right]^2
\end{aligned} \tag{2.16}$$

where $\langle Q \rangle_{|\Psi_T|^2}$ is the expectation value of quantity Q over the distribution $|\Psi_T|^2$. In practice, an initial optimization of the local energy's variance can help to improve the efficiency a

subsequent energy minimization and therefore this is often done.

2.4 Diffusion Monte Carlo

A more sophisticated flavor of quantum Monte Carlo is the diffusion Monte Carlo (DMC) method. In this method, the ground state of the system is obtained by evolving an initial state, $|\psi(0)\rangle$, through imaginary time by applying the time evolution operator

$$|\psi(\tau)\rangle = e^{-(H-E_T)\tau} |\psi(0)\rangle \quad (2.17)$$

where E_T is an energy offset introduced for convenience and $\tau \equiv it$ is chosen to be real and thus the time t is purely imaginary. Expanding $|\psi(0)\rangle$ into the eigenstates, $\{|\phi_i\rangle\}$, of H then taking the limit $\tau \rightarrow \infty$ gives

$$\begin{aligned} \lim_{\tau \rightarrow \infty} |\psi(\tau)\rangle &= \lim_{\tau \rightarrow \infty} e^{-(H-E_T)\tau} |\psi(0)\rangle \\ &= \lim_{\tau \rightarrow \infty} e^{-(H-E_T)\tau} \sum_i c_i |\phi_i\rangle \\ &= \lim_{\tau \rightarrow \infty} \sum_i c_i e^{-(E_i-E_T)\tau} |\phi_i\rangle. \end{aligned} \quad (2.18)$$

If we assume $E_T = E_0$ then all excited states are exponentially damped since their eigenvalues are greater than E_0 yielding

$$\lim_{\tau \rightarrow \infty} |\psi(\tau)\rangle = c_0 |\phi_0\rangle, \quad (2.19)$$

and only the ground state component of $|\psi(0)\rangle$ remains – note that an implicit assumption has been made that the initial state has a non-zero overlap with the ground state.

In the real-space basis, the time-evolved state is given by

$$\begin{aligned} \langle \mathbf{R} | \psi(\tau) \rangle \equiv \psi(\mathbf{R}, \tau) &= \int d^3N_{R'} \langle \mathbf{R} | e^{-(H-E_T)\tau} | \mathbf{R}' \rangle \langle \mathbf{R}' | \psi(0) \rangle \\ &= \int d^3N_{R'} G(\mathbf{R} \leftarrow \mathbf{R}', \tau) \psi(\mathbf{R}', 0) \end{aligned} \quad (2.20)$$

where $G(\mathbf{R} \leftarrow \mathbf{R}', \tau) \equiv \langle \mathbf{R} | \exp(-(H-E_T)\tau) | \mathbf{R}' \rangle$ is the Green function for the many-body Schrödinger equation. A general expression for the Green function is not known, but we can

approximate it in terms of functions that we know by first re-writing it over N time-slices as

$$\begin{aligned} & \langle \mathbf{R} | e^{-(H-E_T)\tau} | \mathbf{R}' \rangle = \\ & \int d^{3N}R_1 \cdots d^{3N}R_{N-1} \langle \mathbf{R} | e^{-(H-E_T)\delta\tau} | \mathbf{R}_{N-1} \rangle \langle \mathbf{R}_{N-1} | e^{-(H-E_T)\delta\tau} | \mathbf{R}_{N-2} \rangle \cdots \langle \mathbf{R}_1 | e^{-(H-E_T)\delta\tau} | \mathbf{R}' \rangle \end{aligned} \quad (2.21)$$

where $\delta\tau = \tau/N$ and N is a large positive integer. For small $\delta\tau$, we can use the Trotter-Suzuki formula on each matrix element in the following way

$$\begin{aligned} \langle \mathbf{R} | e^{-(H-E_T)\delta\tau} | \mathbf{R}' \rangle & \approx \langle \mathbf{R} | e^{-(V-E_T)\delta\tau/2} e^{-T\delta\tau} e^{-(V-E_T)\delta\tau/2} | \mathbf{R}' \rangle \\ & = (2\pi\delta\tau)^{-3N/2} \exp\left[-\frac{(\mathbf{R}-\mathbf{R}')^2}{2\delta\tau}\right] \exp[-(V(\mathbf{R}) + V(\mathbf{R}') - 2E_T)\delta\tau/2] \end{aligned} \quad (2.22)$$

where the Hamiltonian has been expressed as a sum of kinetic and potential operators, namely, $H = T + V$. The error from factorizing the exponential is small and proportional to $\delta\tau^3$. From the last line of equation 2.22, we see that the dynamics are controlled by two processes: a diffusion process which is a result of the kinetic energy operator and a re-weighting process that is a result of the potential energy operator.

In order to realize the wave function, ψ , let us first map it to a probability distribution and represent it as a discrete collection of walkers that are distributed accordingly. We'll first represent $\psi(\mathbf{R}, 0)$ in the way just described

$$\psi(\mathbf{R}, 0) \rightarrow \sum_i \delta(\mathbf{R} - \mathbf{R}_i) \quad (2.23)$$

then we can evolve the walkers one time-step as follows

$$\begin{aligned} \psi(\mathbf{R}, \delta\tau) & = \int d^{3N}R' \langle \mathbf{R} | e^{-(V-E_T)\delta\tau/2} e^{-T\delta\tau} e^{-(V-E_T)\delta\tau/2} | \mathbf{R}' \rangle \sum_i \delta(\mathbf{R}' - \mathbf{R}_i) \\ & = (2\pi\delta\tau)^{-3N/2} \sum_i \exp\left[-\frac{(\mathbf{R}-\mathbf{R}_i)^2}{2\delta\tau}\right] \exp[-(V(\mathbf{R}) + V(\mathbf{R}_i) - 2E_T)\delta\tau/2]. \end{aligned} \quad (2.24)$$

To discretize the evolved distribution, we can use a re-weighting scheme whereby we first generate new walker positions by sampling the Gaussian distributions then include the

potential energy exponents as weights for each walker, which gives

$$\psi(\mathbf{R}, \delta\tau) \rightarrow \sum_i \delta(\mathbf{R} - \mathbf{R}_i) \exp[-(V(\mathbf{R}_i) + V(\mathbf{R}'_i) - 2E_T) \delta\tau/2] \quad (2.25)$$

where $\{\mathbf{R}_i\}$ and $\{\mathbf{R}'_i\}$ are the new and old positions of the walkers, respectively. After repeating this procedure over many time-slices, the resulting distribution will begin to mimic the ground state wave function of the system, except there will be a small bias given we assumed a non-zero $\delta\tau$.

The re-weighting scheme, however, is generally not efficient since over time the weight of one walker will tend to be exponentially larger than the others and all others will have negligible contribution. A way to avoid this is by using the branching algorithm (also known as the birth/death algorithm). At each time-step, the algorithm effectively increases or decreases the weights of each walker by allowing individual walkers to either proliferate or exit the simulation, respectively. The likelihood that a walker is multiplied or removed is based on the re-weighting factor

$$P = \exp[-(V(\mathbf{R}_i) + V(\mathbf{R}'_i) - 2E_T) \delta\tau/2] \quad (2.26)$$

and if $P < 1$ then the walker remains in the simulation with probability P , otherwise if $P \geq 1$ then the walker remains in the simulation and in addition a new walker is added to the simulation at the same position of the original walker with probability $P - 1$. The offset E_T can be adjusted on the fly in order to maintain a desired number of walkers on average. As an example of the DMC method, figure 2.1 shows a simulation of the ground state of the one-dimensional harmonic oscillator. After completion, the walkers clearly converge to the correct ground state ψ_0 .

So far, the DMC method has been described in its most simple form and unfortunately is still rather inefficient. The reason for this is that the re-weighting factor depends on the potential V which can cause it to undergo large fluctuations or even diverge. To generate a more efficient scheme, importance sampling is often used. Assuming a trial (or guiding) wave function $\psi_T(\mathbf{R})$, let's define a new function $f(\mathbf{R}, \tau) \equiv \psi_T(\mathbf{R})\Phi(\mathbf{R}, \tau)$ where $\Phi(\mathbf{R}, \tau)$ satisfies the imaginary-time Schrödinger equation

$$-\partial_\tau \Phi(\mathbf{R}, \tau) = (H - E)\Phi(\mathbf{R}, \tau). \quad (2.27)$$

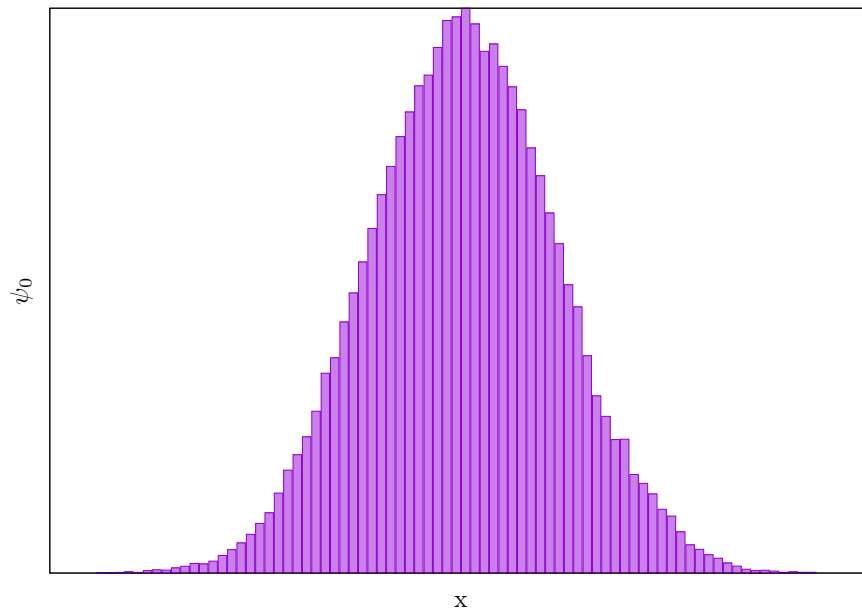


Figure 2.1 A diffusion Monte Carlo simulation of the ground state of a one-dimensional harmonic oscillator. The walkers are initially distributed uniformly then evolved through 8 units of imaginary time and the birth/death re-weighting scheme is used during their evolution. After the simulation completes, the walkers are distributed according to ψ_0 as expected.

We can re-write the Schrödinger equation in terms of $f(\mathbf{R}, \tau)$ which yields

$$-\partial_\tau f(\mathbf{R}, \tau) = -\frac{1}{2}\nabla^2 f(\mathbf{R}, \tau) + \nabla \cdot [\mathbf{v}_D(\mathbf{R})f(\mathbf{R}, \tau)] + [E_L(\mathbf{R}) - E_T], \quad (2.28)$$

where $E_L(\mathbf{R})$ is the same local energy we defined above and $\mathbf{v}_D(\mathbf{R})$ is known as the drift velocity and is defined as

$$\mathbf{v}_D(\mathbf{R}) \equiv \nabla \ln |\psi_T(\mathbf{R})| = \psi_T^{-1}(\mathbf{R}) \nabla \psi_T(\mathbf{R}). \quad (2.29)$$

From an initial function $f(\mathbf{R}, 0)$, the time-evolved solution then takes the form

$$f(\mathbf{R}, \tau) = \int d^{3N}R' \tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) f(\mathbf{R}', 0) \quad (2.30)$$

where $\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \psi_T(\mathbf{R})G(\mathbf{R} \leftarrow \mathbf{R}', \tau)\psi_T^{-1}(\mathbf{R})$ is known as the importance sampling Green function and $G(\mathbf{R} \leftarrow \mathbf{R}', \tau)$ is the Green function we defined above. For small time-steps, the importance sampling Green function can be approximated as

$$\begin{aligned} \tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) &\approx (2\pi\delta\tau)^{-3N/2} \exp\left[-\frac{(\mathbf{R} - \mathbf{R}' - \delta\tau\mathbf{v}_D(\mathbf{R}'))^2}{2\delta\tau}\right] \exp[-(E_L(\mathbf{R}) + E_L(\mathbf{R}') - 2E_T)\delta\tau/2] \\ &\equiv G_d(\mathbf{R} \leftarrow \mathbf{R}', \delta\tau)G_b(\mathbf{R} \leftarrow \mathbf{R}', \delta\tau) \end{aligned} \quad (2.31)$$

where we define the birth and re-weighting factors as $G_b(\mathbf{R} \leftarrow \mathbf{R}', \delta\tau)$ and $G_d(\mathbf{R} \leftarrow \mathbf{R}', \delta\tau)$, respectively. With this modified form, we see that the re-weighting factor now depends on the local energy rather than the potential which should fluctuate much less since the local energy should be roughly constant if a ψ_T is chosen such that is close to the system's ground state.

Once the walkers are distributed according to $f(\mathbf{R}, \tau)$, a *mixed estimator* of the ground

state energy can be made in the following way

$$\begin{aligned}
E_0 &= \lim_{\tau \rightarrow \infty} \frac{\langle \psi_T | e^{-\tau H/2} H e^{-\tau H/2} | \psi_T \rangle}{\langle \psi_T | e^{-\tau H/2} e^{-\tau H/2} | \psi_T \rangle} \\
&= \lim_{\tau \rightarrow \infty} \frac{\langle \psi_T | e^{-\tau H} H | \psi_T \rangle}{\langle \psi_T | e^{-\tau H} | \psi_T \rangle} \\
&= \frac{\langle \psi_0 | H | \psi_T \rangle}{\langle \psi_0 | \psi_T \rangle} \\
&\approx \frac{1}{M} \sum_{i=1}^M E_{\text{loc}}(\mathbf{R}_i)
\end{aligned} \tag{2.32}$$

where $E_{\text{loc}} = \psi_T^{-1}(\mathbf{R}) H \psi_T(\mathbf{R})$ is the local energy and the walkers $\{\mathbf{R}_i\}$ are distributed according to $f(\mathbf{R}, \tau \rightarrow \infty)$

A key limitation still remains in the DMC algorithm as it has been described. The wave function of a many-body fermionic system is not positive definite and therefore the algorithm must be extended in some way if the solution of a system of fermions is desired. Often the extension that is used is the fixed-node approximation if the desired wave function is real – for complex wave functions, the fixed-phase approximation can be used which is discussed in chapter 4. In fixed-node DMC, a trial wave function is introduced and used to define the nodal surface of the evolving DMC wave function. Recall that our distribution from importance sampling had the form

$$f(\mathbf{R}, \tau) = \psi_T(\mathbf{R}) \Phi(\mathbf{R}, \tau) \tag{2.33}$$

and therefore, if the nodes and sign of $\Phi(\mathbf{R}, \tau)$ are fixed to those of $\psi_T(\mathbf{R})$ then clearly

$$f(\mathbf{R}, \tau) \geq 0. \tag{2.34}$$

We can implement this method by evolving a set of walkers as we did previously, however, we now keep track of whether a walker crosses a node of $\psi_T(\mathbf{R})$ by checking if a given step causes $\psi_T(\mathbf{R})$ to change sign. If a sign change is observed, then we can either remove the walker from the simulation entirely or we can reject the move and move it back to its position before the step – both methods are equivalent in the limit of $\delta\tau \rightarrow 0$. Note also that the drift velocity, defined in equation 2.29, assists in keeping walkers away from the

nodes of $\psi_T(\mathbf{R})$ since it drives walkers away from regions where the function is small.

The nodal constraint on the evolving wave function introduces a bias, of course, however a large amount of correlation energy is generally obtained – often on the order of 95%. This level of accuracy places fixed-node DMC in a class of the most reliably predictive theories available within electronic structure, and in addition, given its favorable scaling with electron number $\mathcal{O}(N^3)$, strikes an unprecedented balance of accuracy and efficiency.

CHAPTER 3

Quantum Monte Carlo Study of mono(benzene)TM and bis(benzene)TM

Portions of this chapter also appeared in:

A Quantum Monte Carlo Study of mono(benzene)TM and bis(benzene)TM Systems.

Chem. Phys. Lett., **667**, 74 (2017)

Contributions:

In the work described in this chapter, the author performed all geometry relaxations and fixed-node DMC calculations of the mono(benzene)W and bis(benzene)W systems and played a significant role in the analysis of those data and in the manuscript's preparation.

3.1 Introduction

The fixed-node bias is generally the dominant source of error in DMC calculations though accuracy in many cases can be quite high even for the simplest trial nodes, e.g., those from a single Slater determinant. This nodal problem has been studied in various ways (see, for example, Refs. [Baj05; BM09]) and several methods have been proposed to reduce the bias through some form of optimization of the trial nodes (see Refs. [Umr07; TU07; Lüc07]). Due to the stochastic nature of QMC, however, such optimizations are rather demanding on several fronts, in particular, the efficiency of the optimizer is vital and systematic improvements to the nodes can be difficult to obtain, especially for large systems. To generate optimization methods that improve on these difficulties, a viable strategy would be to first obtain a better understanding of sources of nodal errors and of nodal properties, in general. Some progress has been achieved in this respect in several studies [RM12; Ras14; Kul14], where it was shown for 1st and 2nd row elements that nodal errors grow with node nonlinearities (a property related to the multiplicity of bonds) and with increasing electronic density. In this study, we add to this investigation by looking at nodal properties of benzene complexes containing molybdenum (Mo) and tungsten (W). In particular, we consider the half-sandwich mono(benzene)Mo (MoBz) and the full-sandwich bis(benzene)Mo (MoBz₂) systems, depicted in Figure 3.1, as well as the structurally and chemically similar WBz and WBz₂ systems. The choice of these molecular systems has been motivated by several considerations.

Mo and W are 4*d* and 5*d* transition metal (TM) elements, respectively, and are stacked vertically within the periodic table's VIB column. Given their partially filled *d*-shells, both elements have the potential to form many types of bonds and, therefore, can be found in variety of molecular systems that exhibit interesting properties with examples including catalysts and bioenzymes. As a consequence of these attributes, a comparison between Mo- and W-containing systems enables us to further examine the dependencies of fixed-node errors to electronic density (or, similarly, state localization) and bond multiplicity. To achieve a sound comparison, we look at the same electronic states and similar geometries so that can we contrast the results between systems containing the 4th period metal to those containing the 5th period metal.

Another point of interest regarding these systems is in the fundamental chemistry of metals bonded to organic molecules. As an example, CrBz₂ was first synthesized only a

couple of decades ago and was the first stable synthetic structure of its kind, namely, where a TM was bound to two benzene rings in a full-sandwich geometry. It was a breakthrough in the field of organometallic chemistry and was followed by many novel sandwich-type complexes such as VBz and CoBz. Some of these systems are well-known as catalysts since transition metals have the ability to form multiple bonds with low barriers and the structural flexibility of the sandwich geometries enable easy conformational changes that are crucial for efficient reaction paths. For example, the MoBz₂ system that we include in this study, is a well-known catalyst. Furthermore, the benzene-based organometallics later proved to be of high interest in the field of spintronics. In particular, benzene multi-decker nanowires with sandwiched transition metal atoms, such as V and Co, have become some of the most studied systems recently due to differences in conductivities in the two spin channels and their possible functioning as spin filters [Hor12]. Yet another area of research that involves these constituents is graphene doped with transition metal adatoms where benzene can serve as the simplest cluster model that can describe the local chemistry at the graphene hollow bonding site [NI11]. The Mo and W systems we consider here provide interesting examples of organometallic bonding which is often quite challenging to describe by either the correlated or density functional theory (DFT) approaches.

It is well-known that systems containing TM elements often pose significant challenges for experiments and simulations. These difficulties can be attributed to (i) strong many-body effects arising from the compact nature of the *d*-states resulting in significant electronic correlation [And94; Dac99; HM12; Mic82; Mül09; Bal02; BZ02; Bor08; Efr78; GG82; AH82; Bay84; Zha04] and (ii) near degeneracy of the *d*-shell and outer *s*, *p*-shells which can generate a large number of low-lying hybridized states in molecular environments. In order to probe the electronic correlations and many-body effects we employ the FNDMC method that, apart from the fixed-node approximation, can provide very accurate results. In addition, relativistic effects impact the valence electronic structure and must be accounted for at some level if reasonable agreement with experiment is expected. To account for relativity at some level, we have utilized scalar-relativistic effective core potentials, throughout.

3.2 Computational Details

After examining orbitals generated by different techniques such as Hartree-Fock (HF), natural orbitals from configuration interaction with single and double excitations (CISD)

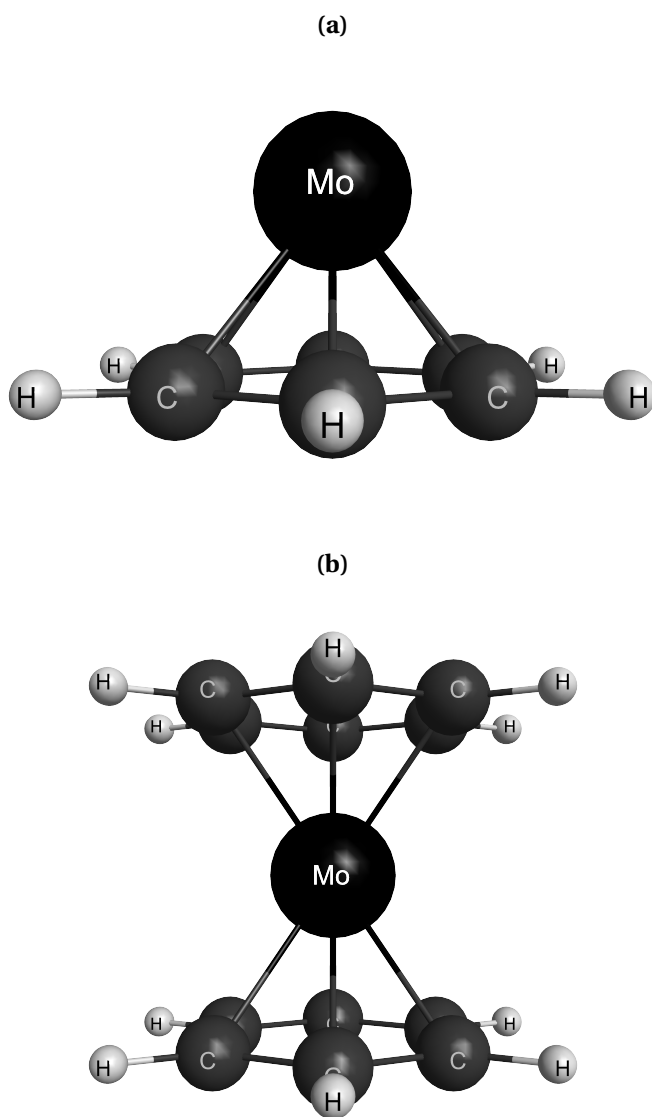


Figure 3.1 Molecular geometries of (a) bis(benzene)molybdenum (MoBz_2) and (b) mono(benzene)molybdenum (MoBz). The geometries of the corresponding W systems are similar.

with varying size of virtual space and DFT orbitals, we found that the DFT orbitals provided lower energies, in general, and therefore were used in the FNDMC calculations.

We replace the atomic cores with effective core potentials (ECPs) since the QMC computational cost grows as $Z_{\text{eff}}^{5.5-6.5}$, where Z_{eff} is the effective nuclear charge. ECPs therefore provide significant boost in efficiency. In addition, impact of relativity on valence states is built into the ECPs by construction so that relativistic effects at one-particle level are included. We note that the ECPs we use involve averaged spin-orbit interaction (often denoted as AREP, averaged relativistic effective potentials).

ECP nonlocal operators were treated by the locality approximation [Mit91]. The corresponding bias vanishes quadratically in the trial function error so that in many cases it gets folded into the fixed-node bias since it has the same type of scaling. The FNDMC calculations were done with T-moves algorithm [Cas06] that preserve the variational bound for the total energy regardless of the localization approximation treatment of ECPs. The QMC calculations were performed using the software package QWALK [Wag09a].

3.3 Results and Discussion

3.3.1 MoBz and MoBz₂ systems

Several functionals were examined for each system and the DFT calculations were carried out with the GAUSSIAN09 [Fri] and GAMESS [Sch93a] codes. The geometry optimizations were carried out within C_{6v} and D_{6h} for Mo-Bz and Mo-Bz₂, respectively. For Mo we used pseudopotentials with $4s4p4d5s$ valence space [Pet07] while for the rest of atoms we used ECPs from Ref. [Bur07]. The basis sets were of aug-cc-pVTZ quality.

The spin multiplicity of the ground states of MoBz and MoBz₂ were examined in DFT, each yielding singlet ground states. The trial wave functions used were the single-reference Slater-Jastrow as described above. The Slater determinant was constructed from single-particle orbitals generated by the DFT-TPSSH method that produced the lowest overall total energies, i.e., it led to the lowest fixed-node biases. The Jastrow function in the trial wave functions contained up to three-body correlation terms, namely, nucleus-electron, electron-electron and nucleus-electron-electron terms. The Jastrow function was optimized within the framework of the VMC method.

The relaxed structural parameters used in the MoBz₂ and MoBz DMC calculations are

Table 3.1 The structural parameters are given. The bond lengths ($R/\text{\AA}$) and dihedral angles ($\angle/^\circ$) were obtained from DFT-TPSSh calculations.

	$R(\text{Mo-Bz})$	$R(\text{C-C})$	$R(\text{C-H})$	$\angle \text{CCCH}$
MoBz ₂	1.782	1.419	1.083	0.378
MoBz	1.617	1.438	1.093	3.463

given in Table 3.1. The binding energies calculated by several DFT methods and the fixed-node DMC together with experiment are presented in Figure 3.2. The binding energy is defined as follows:

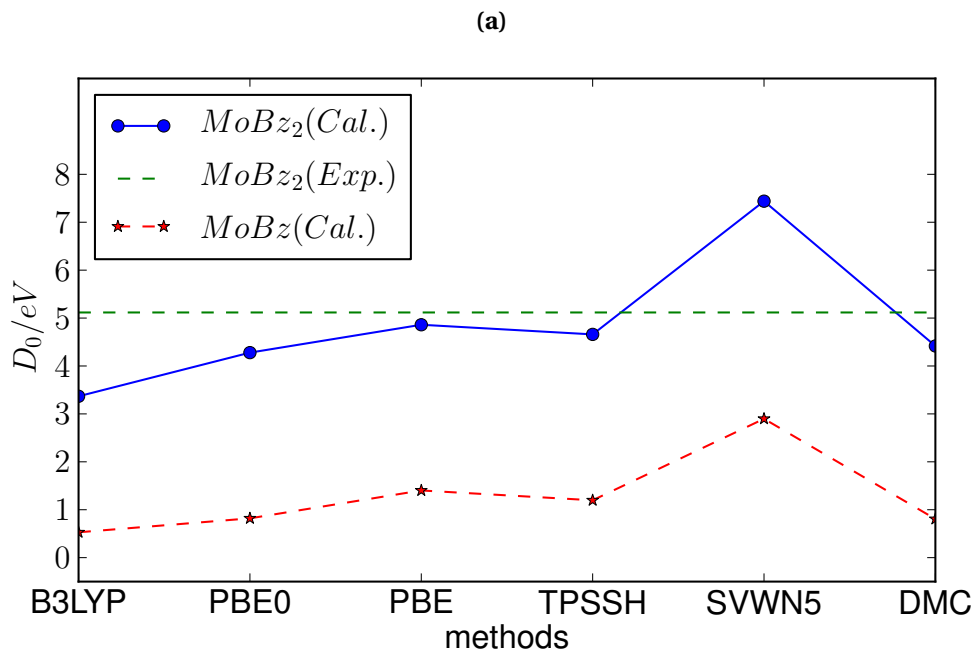
$$D_0(\text{MoBz}_2) = 2E(\text{Bz}) + E(\text{Mo}) - E(\text{MoBz}_2) \quad (3.1)$$

$$D_0(\text{MoBz}) = E(\text{Bz}) + E(\text{Mo}) - E(\text{MoBz}) \quad (3.2)$$

Interestingly, for MoBz₂, all the DFT functionals except LDA/SVWN5 underbind with regard to the experimental value. An underbinding of about 0.66(3) eV is present also in the FNDMC result. We attribute most of this error to fixed-node error suggesting that the wave function has a significant multi-reference character. This is not surprising considering the high multiplicity of bonds in that system. A qualitatively similar pattern appears also for MoBz, however, in this case we were not able to find the experimental data. Our best guess is that the actual bonding energy lies around 1 eV and that the FNDMC shows underbinding by ≈ 0.3 eV, again due to the need of a multi-reference wave function to describe multiple bonds. The size of the fixed-node bias is similar to that in a recent study of corresponding dimers, in particular, the impact of the fixed-node errors on binding of the Mo dimer turns out to be about ≈ 0.7 eV.

3.3.2 WBz and WBz₂ systems

For the benzene complexes containing W, we also carried out DFT calculations with various functionals using the GAMESS [Sch93a] software package. We again considered basis sets of aug-cc-pVTZ quality. An ECP was utilized for the W atom with a $5s5p5d6s$ valence space [Pet07] and again ECPs from Ref. [Bur07] were employed for the benzene constituents. We imposed C_{6v} and D_{6h} symmetries for WBz and WBz₂, respectively, and examined each system in a singlet spin state. The geometries of both systems were optimized with the



(b)

Method	D_0/eV ($MoBz_2$)	D_0/eV ($MoBz$)
B3LYP	3.37	0.53
PBE0	4.28	0.82
PBE	4.86	1.40
TPSSh	4.66	1.20
SVWN5	7.43	2.90
DMC/TPSSh	4.42(4)	0.81(3)
Exp.	5.12 [SY06]	

Figure 3.2 The binding energies of the $MoBz_2$ and $MoBz$ molecules calculated by DFT and FNDMC methods compared with experiment (a) and (b). FNDMC trial function has been constructed using TPSSh functional.

quadratic approximation to the augmented Hessian technique until the maximum component of the gradient was less than 10^{-4} in magnitude. The equilibrated structural parameters are listed in tables 3.2 and 3.3, where we define the distance $R(\text{W-Bz})$ to be the distance from the W atom to the plane formed by the carbon atoms.

Table 3.2 Bond distances ($R/\text{\AA}$) and dihedral angles ($\angle/^\circ$) of WBz.

	$R(\text{W-Bz})$	$R(\text{C-C})$	$R(\text{C-H})$	$\angle \text{CCCH}$
PBE	1.589	1.442	1.092	3.717
PBE0	1.590	1.431	1.084	3.216
SVWN5	1.567	1.441	1.101	4.421

Table 3.3 Bond distances ($R/\text{\AA}$) and dihedral angles ($\angle/^\circ$) of WBz_2 .

	$R(\text{W-Bz})$	$R(\text{C-C})$	$R(\text{C-H})$	$\angle \text{CCCH}$
PBE	1.781	1.423	1.090	0.773
PBE0	1.773	1.413	1.082	0.760
SVWN5	1.748	1.422	1.099	1.218

For the FN-DMC calculations, single-reference Slater-Jastrow trial wave functions were used. We considered Slater determinants built from PBE and PBE0 single-particle orbitals and up to three-body interaction terms in the Jastrow factors. The systems were placed in their respective optimized geometries. The motivation for considering orbitals generated from the PBE functional in addition to the hybrid functional, PBE0, was to take the opportunity to investigate whether exact Hartree-Fock mixing in these systems would have any impact, as was the case in $3d$ systems, see for example, [Kol10]. Table 3.4 shows the calculated binding energies for various DFT methods as well as fixed-node DMC; the latter we present as a benchmark result given the absence of experimental data for these systems. The binding energies are defined similarly to those given in expressions (6) and (7), in addition to the following definition

$$D_0(\text{WBz+Bz}) = E(\text{WBz}) + E(\text{Bz}) - E(\text{WBz}_2). \quad (3.3)$$

Provided that we only accounted for averaged spin-orbit interactions, we took the tungsten atom to be in its septet spin state which we determined to be the ground state of the Hamiltonian with spin-orbit averaged ECPs.

Table 3.4 Binding energies [eV] of W-Bz and W-Bz₂ and systems from DFT methods and from fixed-node DMC with two trial wave functions constructed with PBE and PBE0 orbitals.

	W+Bz	W+Bz+Bz	WBz+Bz
PBE	1.62	5.67	4.05
PBE0	1.25	5.36	4.11
SVWN5	3.20	8.36	5.16
DMC/PBE	1.42(3)	6.00(3)	4.58(3)
DMC/PBE0	1.47(3)	6.08(3)	4.61(3)

It is interesting that the results for WBz and WBz₂ look rather different. Let us analyze the differences point by point. First, we suspect that the part of the fixed-node bias that affects energy differences is significantly lower due to the lower electronic density and *5d* states that are less localized when compared to *4d*-states. This can be explained by inspection of the one particle atomic states plotted in Figure 3.3. Note that the orbital maxima of the semicore *s*-states, namely, *4s* for Mo and *5s* for W, are almost the same. This contrasts with the *5d* (Mo) and *6d* (W) states that differ significantly. Not only is the maximum of the W *6d* state at a larger radius but also the density inside the core region is much smaller. This is reminiscent of the contrast between the *2p* states in the first-row and *3p* states in the second-row where a much smaller fixed-node error was found in the latter case (both in absolute and in differences) [Ras14].

Second, the sensitivity of both FNDMC energy differences and total energies to changes in the DFT functionals used to generate the orbitals is very minimal in the W systems, showing a reduced impact on the nodes. This conclusion is further supported by our study of W₂ that will be described elsewhere.

Third, FNDMC binding energies are actually larger than predicted by DFT functionals (except LDA, that significantly overbinds, as we commented above). Indeed, this is also in line with our calculations of W₂ that agrees with experiment rather well [Mel18]. Therefore, we conjecture that our estimation of binding energies in these systems is the most accurate to date. Note that we were not able to find reliable experimental data for these systems

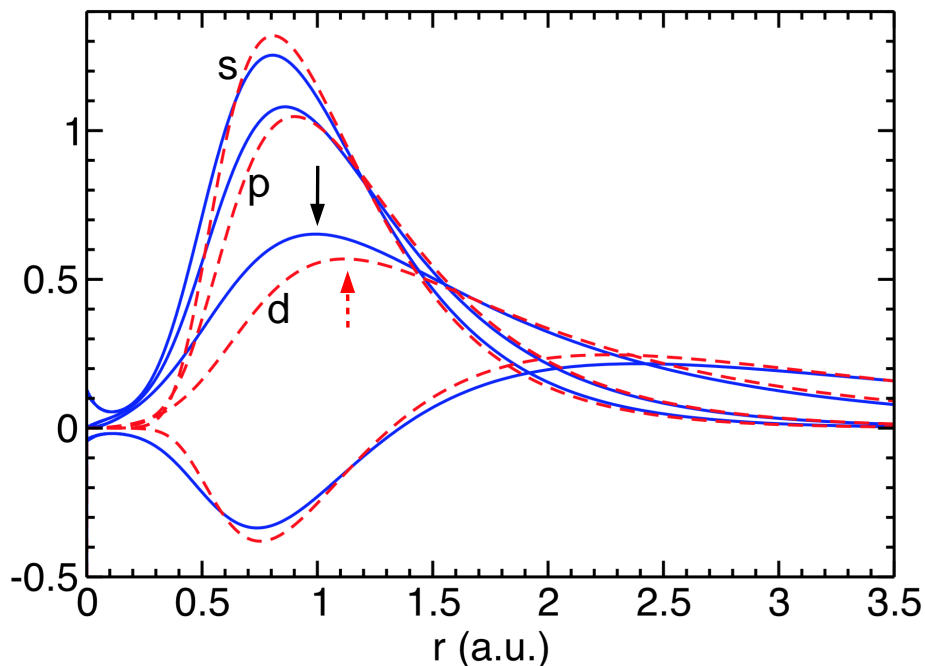


Figure 3.3 Radial components of $s(\ell = 0)$, $p(\ell = 1)$ and $d(\ell = 2)$ valence pseudorbitals plotted as $r^\ell \rho_\ell(r)$ for Mo (full, blue) and W (dashed, red) atoms. The maximum of s -orbital for W is higher than for Mo (relativity), while the opposite is true in the d -channel. Also note the significantly larger radius of the maximum (indicated by arrows) with the consequent smaller amount of charge in the core region in the d -channel of W.

in the literature and therefore we provide genuine predictions for these binding energies. The fixed-node errors in Mo systems deserve further study as well as the impact of the spin-orbit interactions that have been taken into consideration in an averaged manner in the present study. Overall, the FNDMC methods show very high accuracy for the 5th period systems in general and thus provide very accurate methodology for the correlated wave function calculations.

3.4 Conclusions

We present calculations of single and double benzene systems with molybdenum and tungsten atoms in half- and full-sandwich geometries. We used several DFT functionals and FNDMC method to estimate the binding energies in these systems. For MoBz₂ we found that single-reference wave function appears to underestimate the binding by about

0.5 eV when compared with experiment. Interestingly, except LDA that is known to overbind by significant amount almost universally (20-30%), the DFT functionals underbind these systems as well with significant spread 0.4-1.8 eV, depending on the functional. For MoBz we were not able to find the experimental value and therefore we estimate the binding to be around 1 eV. Since the overall behavior of the total ad binding energies is qualitatively similar to MoBz₂ we also estimate the fixed-node bias to cause underestimation of the FNDMC binding by about $\approx 0.4-0.6$ eV with the uncertainty being assigned to the averaging of the spin-orbit effects. The results for tungsten systems appear to be significantly less impacted by the fixed-node errors due to the lower electronic density and more favorable spatial ordering of the *s*, *p* and *d* levels. This is a very encouraging finding since it implies that accurate calculations are feasible for materials with 5*d* elements at the single reference trial function level. In the absence of experimental data, our calculations provide the most accurate values of binding energies in these tungsten molecular systems to date.

CHAPTER 4

Quantum Monte Carlo with Variable Spins

Portions of this chapter also appeared in:

Quantum Monte Carlo with Variable Spins.

Phys. Rev. E, **96**, 043305 (2017)

Contributions:

In the work described in this chapter, the author performed the COSCI and fixed-node DMC calculations of the AREP-containing systems.

4.1 Introduction

Despite its successes, FNDMC actually solves only the spatial part of the eigenstate problem since in typical calculations the electronic spins are treated statically rather than as true quantum variables. This is perfectly adequate in systems where the Hamiltonian does not include spin, and then the nontrivial part of the solution is reduced to spatial dependencies only. The particular spin configuration of individual electrons is conserved and therefore is imposed as a symmetry of the system, e.g., a singlet or triplet state. However, many interesting systems exhibit interactions between the spin and spatial degrees of freedom such as the spin-orbit interaction. For nuclear systems, the quantum nature of spins have been realized in variational[CK85] and auxiliary field Monte Carlo[Sar03; Ped04; Gan06; Gan07] methods. In condensed matter applications, a DMC method was implemented for the 2D homogeneous electron gas with Rashba interaction [Amb09] as well as its modification to the VMC method applied to atoms[Amb12]. In these two approaches, the spinor states are stochastically sampled as opposed to sampling the particle coordinate space that underlies the DMC methods. Recently, we introduced a DMC method which keeps the trial spinors intact during the imaginary time evolution [Mel16a]. This method is particularly useful in that it has the zero-variance property, namely that for arbitrary configurations (spin and spatial coordinates) the bias in energy is proportional to the square of the trial wave function error. Since spin-orbit is nonlocal in the particle coordinate space, the method deals with spin-orbit terms in a manner similar to nonlocal pseudopotentials [Mit91; Cas06] and, as we show below, many of the developed techniques then carry over.

In this paper, we elaborate and expand upon the details and technical issues of the method introduced previously [Mel16a]. In §4.2, we give a short discussion of the fixed-phase method in a configuration space without the spin degrees of freedom. In later sections, we generalize the fixed-phase algorithm to incorporate varying spins and spin-dependent Hamiltonians. In §4.3, we introduce one particular choice of spin-dependent Hamiltonian, namely the spin-orbit interaction via a pseudopotential. We note that any spin-dependent Hamiltonian could be included, however we choose to focus on the spin-orbit interaction. In §4.4, we discuss the inclusion of the spin variables and our choice for the spin representation. In §4.5, we discuss corresponding timestep dependencies in evolutions of both spatial and spin degrees of freedom. We present applications of the method to several atoms and molecules in §4.6. We conclude in §4.7.

4.2 Fixed-Phase Diffusion Monte Carlo

In order to include spin-dependent Hamiltonians within DMC, we must work with inherently complex wave functions. FNDMC nominally treats real-valued wave functions, so we must resort to a generalization of the method. Before dealing with spin terms, we present a short review of the fixed-phase method (FPDMC) [Ort93] and its relation to the fixed-node flavor of the DMC.

For an N -electron system, we work in a configuration space $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \in \mathbb{R}^{dN}$, where d is dimensionality and here we assume $d = 3$. For now, we consider particle spin to be a label rather than a variable. We assume the Born-Oppenheimer approximation, so that we have the Hamiltonian of the form $H = -(1/2)\nabla^2 + V(\mathbf{R})$, where $\nabla = (\nabla_1, \nabla_2, \dots, \nabla_N)$ and V denotes the electron-ion and electron-electron Coulomb interactions. Since the wave function $\Psi(\mathbf{R}, \tau)$ is complex, we write $\Psi(\mathbf{R}, \tau) = \rho(\mathbf{R}, \tau)e^{i\Phi(\mathbf{R}, \tau)}$ and substitute into the imaginary-time Schrödinger equation. This yields two coupled differential equations for the amplitude, $\rho(\mathbf{R}, \tau)$, and phase, $\Phi(\mathbf{R}, \tau)$, as

$$-\frac{\partial \rho(\mathbf{R}, \tau)}{\partial \tau} = \left[T_{kin} + V(\mathbf{R}) + \frac{1}{2} |\nabla \Phi(\mathbf{R}, \tau)|^2 \right] \rho(\mathbf{R}, \tau) \quad (4.1)$$

$$-\frac{\partial \Phi(\mathbf{R}, \tau)}{\partial \tau} = \left[T_{kin} + \frac{\nabla \rho(\mathbf{R}, \tau) \cdot \nabla}{\rho(\mathbf{R}, \tau)} \right] \Phi(\mathbf{R}, \tau) \quad (4.2)$$

where we abbreviate $T_{kin} = -(1/2)\nabla^2$. In order to obtain an approximate solution, we invoke the fixed phase approximation by taking $\partial_\tau \Phi(\mathbf{R}, \tau) = 0$ with $\Phi(\mathbf{R}, 0) = \Phi_T(\mathbf{R})$, where $\Phi_T(\mathbf{R})$ is the phase of a trial or guiding wave function. Writing the trial wave function as $\Psi_T(\mathbf{R}) = \alpha(\mathbf{R}) + i\beta(\mathbf{R})$, we obtain one expression for the fixed trial amplitude and phase

$$\rho_T(\mathbf{R}) = \sqrt{\alpha^2(\mathbf{R}) + \beta^2(\mathbf{R})} \quad (4.3)$$

$$\Phi_T(\mathbf{R}) = \tan^{-1} \frac{\beta(\mathbf{R})}{\alpha(\mathbf{R})} \quad (4.4)$$

so that $-\pi/2 \leq \Phi_T(\mathbf{R}) \leq \pi/2$. Since the overall constant phase is irrelevant we can alternatively define

$$\Phi_T(\mathbf{R}) = \cot^{-1} \frac{\beta(\mathbf{R})}{\alpha(\mathbf{R})} \quad (4.5)$$

so that we get $0 \leq \Phi_T(\mathbf{R}) \leq \pi$. The stationary phase condition makes the equation (4.2) moot,

while the equation for the non-negative amplitude determines the energy eigenvalue.

4.2.1 Fixed-phase upper bound property

The fixed-phase approximation is variational since the repulsive potential $1/2|\nabla\Phi_T(\mathbf{R})|^2$ can only raise the energy for an approximate phase [Ort93]. This is easy to see from the energy expectation with $\rho \exp(i\Phi_T)$ that must be an upper bound to the exact energy for an arbitrary symmetric $\rho \geq 0$.

4.2.2 Fixed-phase as a special case of the fixed-node

We note that the fixed-phase approximation is a generalization of the more familiar fixed-node approximation for a real-valued $\Psi_T(\mathbf{R})$. Let us show that explicitly. We denote the nodes of Ψ_T as

$$\Gamma = \{\mathbf{R}; \Psi_T(\mathbf{R}) = 0\} \quad (4.6)$$

Consider another wave function $\Psi_B(\mathbf{R})$ that is symmetric, normalizable and real. It should be also nonvanishing and positive in the domain of $\Psi_T(\mathbf{R})$. An appropriate prototype for Ψ_B can be, for example, an approximation to the bosonic ground state of H . We construct a new complex trial function

$$\tilde{\Psi}_T = \Psi_T + i\varepsilon\Psi_B \quad (4.7)$$

The potential that is generated by the phase of $\tilde{\Psi}_T(\mathbf{R})$ is given by

$$V_{ph} = \frac{1}{2}|\nabla\tilde{\Psi}_T|^2 = \frac{1}{2}\left|\frac{\varepsilon\mathbf{h}}{\Psi_T^2 + \varepsilon^2\Psi_B^2}\right|^2 \quad (4.8)$$

where

$$\mathbf{h} = \Psi_T\nabla\Psi_B - \Psi_B\nabla\Psi_T. \quad (4.9)$$

Away from the node Γ the limit $\varepsilon \rightarrow 0$ produces $V_{ph} = 0$ since then $\Psi_T^2 > 0$. At the node Γ the situation is a little bit more subtle. The key point is that the function $|\nabla\Psi_T|^2 \geq 0$ is generically nonzero at the node (exceptions might possibly be non-analytical points of Ψ_T due to interaction singularities, which are, however, of zero measure). Therefore taking the limit $\varepsilon \rightarrow 0$ we get

$$V_{ph}(\mathbf{R}) = V_\infty\delta(\mathbf{R} - \mathbf{R}_\Gamma) \quad (4.10)$$

where $\mathbf{R}_\Gamma \in \Gamma$ and V_∞ diverges as $\propto 1/\varepsilon^2$, therefore V_{ph} enforces vanishing of any wave function at the node Γ . In this limit V_{ph} become the fixed-node ‘‘potential’’ that is more naturally understood as a boundary condition. The fixed-phase approximation is therefore more general than the fixed-node approximation. However, the accuracy of the method depends on the choice of the phase that nominally varies in the full configuration space unlike in the fixed-node condition that applies only on the nodal subspace that is $(3N - 1)$ -dimensional for N fermions in 3D space (ie, its codimension is 1).

4.2.3 Importance sampling

If we try to solve equation (4.1) without modification, fluctuations in the weights due to the potentials will make the DMC implementation inefficient. We therefore apply an importance sampling transformation with a trial amplitude [RC82]. If we denote $g(\mathbf{R}, \tau) = \rho_T(\mathbf{R})\rho(\mathbf{R}, \tau)$, equation (4.1) becomes

$$-\frac{\partial g(\mathbf{R}, \tau)}{\partial \tau} = -\frac{1}{2}\nabla^2 g(\mathbf{R}, \tau) + \nabla \cdot [\mathbf{v}_D(\mathbf{R})g(\mathbf{R}, \tau)] + [E_L(\mathbf{R}) - E_T]g(\mathbf{R}, \tau) \quad (4.11)$$

where we have included also an energy offset E_T . The importance sampling introduces two new terms, namely a drift velocity

$$\mathbf{v}_D(\mathbf{R}) = \nabla \ln \rho_T(\mathbf{R}) = \rho_T^{-1}(\mathbf{R})\nabla \rho_T(\mathbf{R}) \quad (4.12)$$

and the local energy

$$E_L(\mathbf{R}) = \rho_T^{-1}(\mathbf{R}) \left[-\frac{1}{2}\nabla^2 + V + \frac{1}{2}|\nabla \Phi_T(\mathbf{R})|^2 \right] \rho_T(\mathbf{R}) \quad (4.13)$$

For later purposes we can simplify the evolution equation by denoting the dynamical part of the operator acting on the function g as $H_{\mathbf{R}}^{\text{drift}}$ so that we can write

$$-\frac{\partial g(\mathbf{R}, \tau)}{\partial \tau} = [H_{\mathbf{R}}^{\text{drift}} + E_L(\mathbf{R}) - E_T]g(\mathbf{R}, \tau) \quad (4.14)$$

One can find straightforward formulas for the drift, potential generated by the phase and local energy by using the gradient and laplacian of Ψ_T . Clearly, we have

$$\nabla \rho_T e^{i\Phi_T} = e^{i\Phi_T} \nabla \rho_T + \Psi_T (i\nabla \Phi_T) \quad (4.15)$$

which implies

$$\nabla\Phi_T = \text{Im}(\Psi_T^* \nabla\Psi_T) / \rho_T^2 \quad (4.16)$$

$$\nabla \ln \rho_T = \text{Re}(\Psi_T^* \nabla\Psi_T) / \rho_T^2 \quad (4.17)$$

Similarly for the laplacian we write

$$\begin{aligned} \nabla^2\Psi_T = \nabla^2[\rho_T e^{i\Phi_T}] &= e^{i\Phi_T} \nabla^2\rho_T - \rho_T e^{i\Phi_T} (\nabla\Phi_T)^2 \\ &+ 2ie^{i\Phi_T} (\nabla\rho_T \cdot \nabla\Phi_T) + i\rho_T e^{i\Phi_T} (\nabla^2\Phi_T) \end{aligned} \quad (4.18)$$

so that the real contribution can be further arranged as

$$\text{Re}[\Psi_T^* T_{\text{kin}}\Psi_T] = \rho_T(-1/2)\nabla^2\rho_T + \rho_T^2(1/2)(\nabla\Phi_T)^2 \quad (4.19)$$

where $T_{kin} = -(1/2)\sum_i \nabla_i^2$. Therefore we can write

$$\text{Re}[\Psi_T^* T_{\text{kin}}\Psi_T] / \rho_T^2 = \rho_T^{-1} T_{\text{kin}}\rho_T + (1/2)(\nabla\Phi_T)^2 \quad (4.20)$$

and the local energy is then

$$E_L(\mathbf{R}) = \text{Re}[\Psi_T^* T_{kin}\Psi_T] / \rho_T^2 + V \quad (4.21)$$

Since gradient and laplacian of Ψ_T are routinely calculated in DMC, by using the above formulas the needed quantities can be evaluated straightforwardly.

Rewriting equation (4.11) in integral form yields

$$g(\mathbf{R}', t + \tau) = \int d\mathbf{R} \rho_T(\mathbf{R}') G(\mathbf{R}' \leftarrow \mathbf{R}, \tau) \rho_T^{-1}(\mathbf{R}) g(\mathbf{R}, t) \quad (4.22)$$

The Green's function for this process is of the exact same form as the in FNDMC,

$$\begin{aligned} G(\mathbf{R}' \leftarrow \mathbf{R}; \tau) &\simeq (2\pi\tau)^{-3N/2} \exp\left[\frac{-|\mathbf{R}' - \mathbf{R} - \tau\mathbf{v}_D(\mathbf{R})|^2}{2\tau}\right] \\ &\times \exp\left[-\frac{\tau}{2}(E_L(\mathbf{R}') + E_L(\mathbf{R}) - 2E_T)\right] \end{aligned} \quad (4.23)$$

At this point, it is clear that the implementation of FPDMC proceeds in the exact same manner as FNDMC. The main difference is that the mixed-distribution is made from the

trial amplitude rather than the trial wave function. Additionally, the local energy has an additional term from the trial phase. Since both amplitudes in $g(\mathbf{R}, \tau)$ are positive-definite everywhere, any proposed move in the imaginary time evolution is accessible since there is no nodal surface (any incidental zeros of ρ_T are at most codimension 2, i.e., of zero measure, similarly to points in 3D space). The accuracy of this method clearly depends on the accuracy of the trial phase. If the trial phase happens to be the exact phase, then the projected solution will be $g(\mathbf{R}, \infty) \propto \rho_T(\mathbf{R})\rho_0(\mathbf{R}, \infty)$ producing the ground state energy while the convergence towards the exact value scales with the square of the difference between the exact and approximate trial function.

4.3 Spin Orbit Interactions

In this section, we introduce the relativistic part of the Hamiltonian for use in FPMC. In §4.3.1, we give a quick discussion of relativistic quantum mechanics in a 4-component formalism. We then discuss the reduction to a 2-component formalism with an effective Hamiltonian. This effective Hamiltonian approach uses pseudopotentials to replace the relativistic core-electrons with a suitable effective field for the valence electrons. We discuss the similarities and differences between the standard effective core potentials used in DMC. In §4.3.2, we show that an upper bound can be obtained within the fixed-phase approximation for complex nonlocal operators like our spin-orbit Hamiltonian.

4.3.1 AREP and SO Operators

For heavy atomic and molecular systems, bonding and spectral properties cannot be accurately predicted without the inclusion of scalar-relativistic and spin-orbit effects [Des73; Lu71]. In a relativistic treatment, one must begin with the approximation for the relativistic Hamiltonian [Gra70; Des75], known as the Dirac-Coulomb Hamiltonian, given as

$$H = \sum_{i=1}^{N_e} [-ic(\boldsymbol{\alpha} \cdot \nabla)_i + \beta c^2] - \sum_{i=1}^{N_e} \sum_{I=1}^{N_{ion}} \frac{Z_I}{r_{iI}} + \sum_{i>j} \frac{1}{r_{ij}} \quad (4.24)$$

where $\boldsymbol{\alpha}$ and β are defined through the Pauli matrices $\boldsymbol{\sigma}$ and identity I_2 ,

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} \quad (4.25)$$

Here we have ignored the Breit interactions[Bre29], which gives rise to higher order retardation effects such as spin-other orbit, dipole interactions between two spins, and Fermi-contact interactions. This Hamiltonian, to order $1/c^2$, contains the dominant relativistic effects of the mass-velocity correction, the Darwin contribution to the $\ell = 0$ atomic level, and the spin-orbit interaction. The eigenfunctions of the Hamiltonian in equation (4.24) will be 4-component Dirac spinors. These spinors can be decomposed into large and small 2-component spinors, ψ_L and ψ_S respectively. Analysis of the individual ψ_L and ψ_S for all-electron systems indicate that ψ_S is negligible in the valence region where chemical bonding is important [DK75]. From this, to a reasonable approximation the valence electrons can be accurately described by 2-component spinors. Since the relativistic effects are strongest in the core region, valence electrons can be treated nonrelativistically subject to an effective field that mimics the repulsion of the core electrons [Lee77a]. This points to representing the relativistic effects on the valence electrons through an operator W^{REP} that leads to the following Hamiltonian

$$H = T_{kin} + V + W^{REP} \quad (4.26)$$

which contains only the valence electrons. The effective potential W^{REP} for an electron i from a given ion is typically expanded in the form

$$W_i^{REP} = \sum_{\ell} \sum_{j=|\ell-1/2|}^{\ell+1/2} \sum_{m_j=-j}^j W_{\ell j}^{REP}(r_i) |\ell j m_j\rangle \langle \ell j m_j| \quad (4.27)$$

where r_i is the electron-ion distance. The effective operator W^{REP} contains all of the relativistic effects from the core region and allows one to only consider the valence electrons, as is typically done in nonrelativistic calculations [Kah76]. This is very important for the application in DMC, where for all-electron systems the computational demands scale with the atomic number Z as $\approx Z^6$, but scales more favorably as $N_{valence}^{2-3}$ when effective potentials remove the core electrons. Therefore, relativistic QMC calculations can be done using the effective non-local pseudopotentials.

The relativistic effective potential can be written in a different form that separates the relativistic effects into scalar relativistic and spin-orbit[Erm81], namely $W^{REP} = W^{AREP} + W^{SO}$. The operator is semi-local; i.e. local in the relative distance to the nearest nucleus, but non-local in the solid angle for a given radius. The first term is spin-averaged core

potential, which includes the effect of the mass-velocity, Darwin, averaged spin-orbit, and the effective field under which the valence electrons respond. The AREP term takes the form

$$W_i^{AREP} = W_L^{AREP}(r_i) + \sum_{\ell}^{L-1} \sum_{m=-\ell}^{\ell} [W_{\ell}^{AREP}(r_i) - W_L^{AREP}(r_i)] |\ell m\rangle \langle \ell m| \quad (4.28)$$

where W_L^{AREP} is the local part of the potential, and $W_{\ell}^{AREP}(r_i)$ is weighted average over the j terms in equation (4.27),

$$W_{\ell}^{AREP}(r_i) = \frac{1}{2\ell + 1} [\ell W_{\ell, \ell-1/2}^{REP}(r_i) + (\ell + 1) W_{\ell, \ell+1/2}^{REP}(r_i)] \quad (4.29)$$

The spin-orbit interaction is included in the W^{SO} operator and takes the form

$$W_i^{SO} = s \cdot \sum_{\ell=1}^L \frac{2}{2\ell + 1} \Delta W_{\ell}^{SO}(r_i) \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} |\ell m\rangle \langle \ell m| \ell |\ell m'\rangle \langle \ell m'| \quad (4.30)$$

with the definition $\Delta W_{\ell}^{SO}(r_i) = W_{\ell, \ell+1/2}^{REP}(r_i) - W_{\ell, \ell-1/2}^{REP}(r_i)$. The radial functions are expanded in gaussians in the same form as traditional nonrelativistic pseudopotentials, namely

$$W_Y^X(r_i) = \frac{1}{r_i^2} \sum_{\alpha} A_{\ell\alpha} r_i^{n_{\ell\alpha}} e^{-B_{\ell\alpha} r_i^2} \quad (4.31)$$

where $X \in \{AREP, SO, REP\}$ and $Y \in \{\ell, L\}$ and s is the spin. The parameters $A_{\ell,\alpha}$, $n_{\ell,\alpha}$ and $B_{\ell,\alpha}$ for the *AREP* and *SO* terms have been developed by various groups including the Stuttgart-Cologne group [Stu] and Clarkson University group [Cla].

In order to include relativistic effects into DMC, we must consider the action of the pseudopotential. The action of the pseudopotential on the wave function will have real and imaginary parts, and thus the amplitude and phase equations are transformed to

$$-\frac{\partial \rho}{\partial \tau} = \left[-\frac{1}{2} \nabla^2 + V + \frac{1}{2} |\nabla \Phi|^2 + W^{Re} \right] \rho \quad (4.32)$$

$$-\frac{\partial \Phi}{\partial \tau} = \left[-\frac{1}{2} \nabla^2 + \frac{\nabla \rho \cdot \nabla}{\rho} + W^{Im} \right] \Phi \quad (4.33)$$

with

$$W^{Re/Im} = \text{Re/Im} \left[\frac{W^{REP} \Psi}{\Psi} \right] \quad (4.34)$$

The imaginary part describes the phase flux determined both by ρ and W^{Im} . The real part

is the eigenvalue equation that provides the total energy eigenvalue of the system. Since we do not know the exact phase or the exact wave function to determine $W^{Re/Im}$, we invoke the fixed-phase approximation as discussed earlier as well as the localization approximation used in many conventional DMC calculations [Mit91]. This projects the pseudopotential onto the trial wave function Ψ_T as

$$W^{Re} \rightarrow W_T^{Re} = \text{Re} \left[\frac{W^{REP} \Psi_T}{\Psi_T} \right] \quad (4.35)$$

Note that the localization approximation eliminates the fundamental difficulty of the nonlocal operator that, in general, leads to introduction of another type of fermion sign problem, even for a single electron. This is easy to see by considering the matrix elements of the pseudopotential. If we denote the configuration space of space \mathbf{r}_i and spin s_i coordinates

$$\mathbf{X} = (\mathbf{R}, \mathbf{S}) = (\mathbf{r}_1, \dots, \mathbf{r}_N, s_1, \dots, s_N) \quad (4.36)$$

the matrix element that enters the Green's function after a Trotter expansion can be written as

$$\begin{aligned} \langle \mathbf{X}' | \exp(-\tau W) | \mathbf{X} \rangle &= \delta(\mathbf{X}' - \mathbf{X}) \\ &\quad -\tau \langle \mathbf{X}' | W | \mathbf{X} \rangle + \mathcal{O}(\tau^2) \end{aligned} \quad (4.37)$$

The key problem lies with the matrix elements $\langle \mathbf{X}' | W | \mathbf{X} \rangle$ that, in general, can have a complicated sign structure and thus generate negative or complex values; obviously, this is also true without the spin-orbit terms. The locality approximation eliminates this problem. However, it generates a bias that vanishes quadratically with the error in the trial wave function and it also does not guarantee the variational property with regard to the original Hamiltonian[Mit91]. For real valued wave functions, the variational property can be recovered [Haa95] using the T-moves algorithm [Cas06]. In the next section we show that this upper bound can also be obtained for the complex pseudopotentials and wave functions.

4.3.2 Variational property of the fixed-phase method for nonlocal, complex, Hermitian operators

In the following, we present a generalization of the proof of the upper-bound for nonlocal operators and real wave functions [Haa95] with combined sampling and localization projection given above that enables to recover the upper bound property. Here we will show it for more general nonlocal Hermitian operators and complex wave functions. This proof enables us to be able to apply the T-moves technique [Cas06] in this more general setting. In the proof we will follow rather closely the original arguments [Haa95] that will be generalized at a few important points.

For a nonlocal operator W sign changes arise when the following condition is fulfilled

$$\text{Re}[\Psi_T^*(\mathbf{X})\Psi_T(\mathbf{X}'')\langle\mathbf{X}|W|\mathbf{X}''\rangle] > 0 \quad (4.38)$$

as is clear from considerations of Eq. 4.37 in the importance sampling Green's function. Note that we can express the matrix elements of arbitrary Hermitian operator as $\langle\mathbf{X}|W|\mathbf{X}''\rangle = w(\mathbf{X},\mathbf{X}'')e^{i\gamma(\mathbf{X},\mathbf{X}'')}$ where $w(\mathbf{X},\mathbf{X}'')$ is symmetric in $\mathbf{X} \leftrightarrow \mathbf{X}''$ and positive-definite. Then $\gamma(\mathbf{X},\mathbf{X}'') = -\gamma(\mathbf{X}'',\mathbf{X})$ as must be the case when W is Hermitian, i.e., $W = W^\dagger$. We write the trial wave function as $\Psi_T(\mathbf{X}) = \rho_T(\mathbf{X})e^{i\Phi_T(\mathbf{X})}$. Denoting

$$\alpha(\mathbf{X},\mathbf{X}'') = \Phi_T(\mathbf{X}'') - \Phi_T(\mathbf{X}), \quad (4.39)$$

condition (4.38) becomes

$$\text{Re}[w(\mathbf{X},\mathbf{X}'')e^{i\gamma(\mathbf{X},\mathbf{X}'')} \rho_T(\mathbf{X})\rho_T(\mathbf{X}'')e^{i\alpha(\mathbf{X},\mathbf{X}'')}] > 0 \quad (4.40)$$

which reduces to

$$\cos(\alpha(\mathbf{X},\mathbf{X}'') + \gamma(\mathbf{X},\mathbf{X}'')) > 0 \quad (4.41)$$

since $w(\mathbf{X},\mathbf{X}'')$, $\rho_T(\mathbf{X})$, and $\rho_T(\mathbf{X}'')$ are positive-definite. Although α and γ are anti-symmetric, $\cos(\alpha + \gamma)$ is a symmetric function of \mathbf{X} and \mathbf{X}'' . Following [Haa95; Cas06], we construct an effective Hamiltonian

$$\langle\mathbf{X}|H_{eff}|\mathbf{X}'\rangle = \begin{cases} \langle\mathbf{X}|H|\mathbf{X}'\rangle & , \mathbf{X} \neq \mathbf{X}' \text{ and } \cos(\alpha + \gamma) < 0 \\ 0 & , \mathbf{X} \neq \mathbf{X}' \text{ and } \cos(\alpha + \gamma) > 0 \\ \langle\mathbf{X}|H + V_{sf}|\mathbf{X}'\rangle & , \mathbf{X} = \mathbf{X}' \end{cases} \quad (4.42)$$

where V_{sf} is the *sign-flip* potential defined as

$$\langle \mathbf{X} | V_{sf} | \mathbf{X} \rangle = \int_{\cos(\alpha+\gamma)>0} d\mathbf{X}' \langle \mathbf{X} | W | \mathbf{X}' \rangle \frac{\Psi_T(\mathbf{X}')}{\Psi_T(\mathbf{X})} \quad (4.43)$$

We want to show that H_{eff} produces an upper bound for the original Hamiltonian. We begin with any state with the *same phase* as the trial wave function, namely

$$|\Psi\rangle = \int d\mathbf{X} \Psi(\mathbf{X}) |\mathbf{X}\rangle = \int d\mathbf{X} \rho(\mathbf{X}) e^{i\Phi_T(\mathbf{X})} |\mathbf{X}\rangle \quad (4.44)$$

The discrepancy of H_{eff} and H with this state is

$$\begin{aligned} \Delta E &= \langle \Psi | H_{eff} - H | \Psi \rangle \\ &= \langle \Psi | V_{sf} - H_{sf} | \Psi \rangle \end{aligned} \quad (4.45)$$

Rewriting this in configuration space, we obtain

$$\begin{aligned} \Delta E &= \int d\mathbf{X} \Psi^*(\mathbf{X}) \left[\langle \mathbf{X} | V_{sf} | \mathbf{X} \rangle \Psi(\mathbf{X}) \right. \\ &\quad \left. - \int d\mathbf{X}' \langle \mathbf{X} | H_{sf} | \mathbf{X}' \rangle \Psi(\mathbf{X}') \right] \end{aligned} \quad (4.46)$$

Rewriting this over the terms that generate sign-flips, we have

$$\begin{aligned} \Delta E &= \int d\mathbf{X} \int_{sf} d\mathbf{X}' |\Psi(\mathbf{X})|^2 \langle \mathbf{X} | W | \mathbf{X}' \rangle \frac{\Psi_T(\mathbf{X}')}{\Psi_T(\mathbf{X})} \\ &\quad - \langle \mathbf{X} | W | \mathbf{X}' \rangle \Psi^*(\mathbf{X}) \Psi(\mathbf{X}') \end{aligned} \quad (4.47)$$

Denoting $h = \rho_T(\mathbf{X}')/\rho_T(\mathbf{X})$, we see that

$$\begin{aligned} \Delta E &= \iint_{\Omega} d\mathbf{X} d\mathbf{X}' w(\mathbf{X}, \mathbf{X}') \cos(\alpha(\mathbf{X}, \mathbf{X}') + \gamma(\mathbf{X}, \mathbf{X}')) \\ &\quad \times [h\rho^2(\mathbf{X}) + h^{-1}\rho^2(\mathbf{X}') - 2\rho(\mathbf{X})\rho(\mathbf{X}')] \end{aligned} \quad (4.48)$$

where $\Omega = \{\mathbf{X}, \mathbf{X}'; \cos(\alpha + \gamma) > 0\}$. If we simplify this expression once more, we see that ΔE becomes

$$\Delta E = \iint_{\cos(\alpha+\gamma)>0} d\mathbf{X}d\mathbf{X}' w(\mathbf{X}, \mathbf{X}') \times \cos(\alpha(\mathbf{X}, \mathbf{X}') + \gamma(\mathbf{X}, \mathbf{X}')) \frac{[h\rho(\mathbf{X}) - \rho(\mathbf{X}')]^2}{h} \geq 0 \quad (4.49)$$

which is clearly positive everywhere, since the integration is over the region where $\cos(\alpha(\mathbf{X}, \mathbf{X}') + \gamma(\mathbf{X}, \mathbf{X}'))$ is positive. Thus, the effective Hamiltonian produces an upper bound for H and recovers the variational property. Note that this point - that the approximation makes sense only in the fixed-node framework - has been emphasized in the original paper on the localization approximation [Mit91]. The upper bound can be recovered in this framework by implementation of the so-called T -moves algorithm [Cas06]. Note that even in this algorithm the fixed-node/phase condition is of key importance, the T -moves can recover only the best possible energy within the given constraint.

4.4 Spin Representation and Sampling

We discuss how spins can be treated as a quantum variable. In §4.4.1, we discuss our choice for a continuous and overcomplete representation of the spin variable. Once we have a representation for the spin variable, we discuss the form of the one-particle spinors and trial wave functions which couple the spin and spatial degrees of freedom in §4.4.2. We discuss the evaluation of the pseudopotential with this spin representation in §4.4.3. Lastly, we discuss our choice of sampling spin degrees of freedom and how this modifies the Green's function in FPDMC in §4.4.4.

4.4.1 Spin Representations

Let us denote one-particle spinors as

$$\chi(\mathbf{r}, s) = \alpha \varphi^\uparrow(\mathbf{r}) \chi^\uparrow(s) + \beta \varphi^\downarrow(\mathbf{r}) \chi^\downarrow(s) \quad (4.50)$$

where s is the spin projection coordinate on the z -axis. In its usual minimal representation the spin variable have discrete values $s = \pm 1/2$ so that for S_z eigenstates $\chi^\uparrow(1/2) = \chi^\downarrow(-1/2) = 1$, $\chi^\downarrow(1/2) = \chi^\uparrow(-1/2) = 0$. The evaluation of any expectation $\langle \mathcal{B} \rangle$ for a variational wave function $\Psi_{var}(\mathbf{R}, \mathbf{S}) = \Psi_{var}(\mathbf{X})$ includes spatial integrals as well as summation over 2^N spin configurations space of $(-1/2, 1/2)^N$

$$\langle \mathcal{B} \rangle_{var} = \frac{\int d\mathbf{R} \sum_{\mathbf{S}} \Psi_{var}^* \mathcal{B} \Psi_{var}}{\int d\mathbf{R} \sum_{\mathbf{S}} \Psi_{var}^* \Psi_{var}} = \frac{\int d\mathbf{X} \Psi_{var}^* \mathcal{B} \Psi_{var}}{\int d\mathbf{X} \Psi_{var}^* \Psi_{var}} \quad (4.51)$$

assuming we have N fermions. This can be recast as sampling according to the positive density $w(\mathbf{R}, \mathbf{S}) = |\Psi_{var}(\mathbf{R}, \mathbf{S})|^2$

$$\langle \mathcal{B} \rangle_{var} = \frac{\int d\mathbf{X} w(\mathbf{X}) (\Psi_{var}^*(\mathbf{X}))^{-1} \mathcal{B} \Psi_{var}(\mathbf{X})}{\int d\mathbf{X} w(\mathbf{X})} \quad (4.52)$$

This expression can be implemented in the variational Monte Carlo (VMC) as one can simply add the sampling of the spin configurations to the discrete sampling of the spatial coordinates.

However, generalization to projection methods is more complicated. Note that any change of the discrete spin coordinate(s) will lead to “jumps” in the stochastic path. These jumps can cause the local energy fluctuations to increase substantially and that could possibly compromise the utility and efficiency of the method. Since we are employing a diffusion-drift sampling process in imaginary time with weights that include local energy in the exponential, any large fluctuations would make a reliable estimate of the expectations difficult to obtain, especially if we increase the system size. Another strategy would be to sum over all of the spin configurations for every spatial step. However, this has an exponential scaling so that for large systems this is intractable.

One possibility how to address this obstacle is to make the spin configuration space compact and continuous, which allows for smooth sampling [Sch01]. We can choose an *overcomplete* spin representation through the utilization of a 1D ring (or S^1) lowest pair of degenerate eigenstates as follows:

$$\langle s_j | \chi^\uparrow \rangle = e^{is_j}, \quad \langle s_j | \chi^\downarrow \rangle = e^{-is_j} \quad (4.53)$$

where the spin variable $s_j \in [0, 2\pi)$. This implies the normalization condition for two arbi-

trary spin states

$$\langle \chi^\alpha | \chi^\beta \rangle = \int_0^{2\pi} \frac{ds}{2\pi} \langle \chi^\alpha | s \rangle \langle s | \chi^\beta \rangle = \delta_{\alpha\beta} \quad (4.54)$$

As the simplest illustration, consider an arbitrary one-electron spinor of the form $|\chi\rangle = a|\chi^\uparrow\rangle + b|\chi^\downarrow\rangle$. The expectation value of the $S_x = \frac{1}{2}[|\chi^\uparrow\rangle\langle\chi^\downarrow| + |\chi^\downarrow\rangle\langle\chi^\uparrow|]$ operator with the spinor is clearly $\langle \chi | S_x | \chi \rangle = ab$. If we now consider this expectation value in a VMC formulation, we have the following expectation value

$$\langle \chi | S_x | \chi \rangle = \int ds |\chi(s)|^2 E_L(s) \quad (4.55)$$

where $\chi(s) = ae^{is} + be^{-is}$ and the local energy is $E_L(s) = \chi^{-1}(s)S_x\chi(s) = 1/2(ae^{-is} + be^{is})/(ae^{is} + be^{-is})$. Plugging in, this yields the expectation value ab as expected, where we sample the distribution $|\chi(s)|^2$ and evaluate the average of the local energy.

The introduced representation has several important consequences. First, it enables to define a continuous path for the evolving sampling points (walkers) and therefore all the associated quantities along the path are smooth by definition. Second, the spin coordinate space that is introduced has some desirable properties, namely it is compact and the interval $(0, 2\pi)$ can be sampled rapidly. The harmonic functions have minimal curvature and are complex so that no additional (artificial) node is created, i.e., so that the formulation fits the fixed-phase formulation. Third, note that unlike discrete coordinates that switch on and off the up and down components of the spinor, the spin functions are always somewhere "in between" due to the fact that they are weighted by complex values with unit modulus. In effect, they introduce a complex weighted spinors that for many-spins hedge the average effect of the spin summations. This will prove important at sampling the spin coordinates as explained below.

4.4.2 Trial Wave Functions

In FNDMC calculations without spin terms in the Hamiltonian, the electrons can be labeled as N_\uparrow spin-up and N_\downarrow spin-down ones and these labels remain static. This is due to the fact that spins commute with the Hamiltonian, implying that both the total and individual spins are conserved. It can be shown [Fou01] that expectation values can be then calculated using spatial only averaging with configuration space of $\mathbf{R} \in \mathbb{R}^{3N}$. Consequently, the trial wave functions are typically constructed as bipartitioned spin-up and -down Slater deter-

minant(s) built from one-particle orbitals obtained from Hartree-Fock, post-Hartree-Fock or DFT methods

$$\Psi_T(\mathbf{R}) = e^{U(\mathbf{R})} \sum_m c_m \det_m^\uparrow[\phi_i(\mathbf{r}_k)] \det_m^\downarrow[\phi_j(\mathbf{r}_l)] \quad (4.56)$$

The particle correlations are explicitly approximated by the Jastrow factor given as

$$U(\mathbf{R}) = \sum_{iI} U_1(r_{iI}) + \sum_{i \neq j} U_2(r_{ij}) + \sum_{I, i \neq j} U_3(r_{iI}, r_{jI}, r_{ij}) \quad (4.57)$$

where we have one-, two-, and three-body terms ($U_1(r_{iI})$, $U_2(r_{ij})$, and $U_3(r_{iI}, r_{jI}, r_{ij})$ respectively), that describe electron-ion, electron-electron, etc, correlations.

With our choice of spin representation in §4.4.1, we have a wave function that lives in a configuration space $\mathbf{X} = \{(\mathbf{r}_1, s_1), \dots, (\mathbf{r}_N, s_N)\} \in \mathbb{R}^{3N} \times [0, 2\pi)^N$. We write the trial wave function as

$$\Psi_T(\mathbf{X}) = e^{U(\mathbf{R})} \sum_\alpha c_\alpha \det_\alpha[\dots, \chi_i(\mathbf{r}_k, s_k), \dots] \quad (4.58)$$

where $\{\chi_i(\mathbf{r}, s)\}$ are one-particle spinors. In general, each spinor has different spatial dependence for the up and down spin components, namely

$$\chi(\mathbf{r}, s) = a\varphi^\uparrow(\mathbf{r})e^{is} + b\varphi^\downarrow(\mathbf{r})e^{-is} \quad (4.59)$$

Each spatial function $\varphi^{\uparrow(\downarrow)}$ is expanded in appropriate basis functions (for example, gaussian type orbitals or plane waves). With regards to the Jastrow factor, we use the same form as described above. Seemingly, every electron should be treated as having the “same” spin since there is only *one* determinant of spinors, rather than the spin-like and spin-unlike distinction in Jastrow forms employed in conventional calculations. This would imply that the cusp should correspond to the like-spin value since the determinant vanishes at the two-electron coincidence point. However, spatial coincidence configurations are of zero measure with regard to coincidence at the full space-spin configurations space. Clearly, the differences in spin coordinates make the determinant, in general, nonvanishing even when the spatial coordinates of two electrons coincide. Therefore the more appropriate is the unlike spins cusp condition [Cep78; CA80]

$$\left. \frac{dU_2(r_{ij})}{dr_{ij}} \right|_{r_{ij}=0} = 1/2 \quad (4.60)$$

We also note that the precise cusp value has only a marginal impact on the results since it really affects only a very small part of the configuration space. Much more substantial effect comes from the shape of the Jastrow correlations for $r_{ij} > 0.1 - 0.2$ Bohr, namely, at medium- and long-range distances. These are the ranges of distances where correlations affect one- and two-electron pair densities very significantly over a sizable part of the configuration space. The accuracy in these regions has then important consequences for both minimization of energy fluctuations as well as for accurate projections and minimization of the localization bias.

In the limit of vanishing spin-orbit interaction the single-reference spinor determinant (regardless of the chosen representation) should simplify to the product of spin-up and -down determinants. This is true also for our trial wave function. Let us consider N occupied spinors that can be grouped as $N/2$ Kramer's pairs (for simplicity assuming N to be even). We can write the Kramer's pair as

$$\chi^+ = (\varphi + \Delta\varphi)\chi^\uparrow + (\varphi - \Delta\varphi)\chi^\downarrow \quad (4.61)$$

$$\chi^- = (\varphi - \Delta\varphi)\chi^\uparrow - (\varphi + \Delta\varphi)\chi^\downarrow \quad (4.62)$$

where the $\Delta\varphi$ is the spin-orbit induced splitting of the spatial orbital φ . We sketch a block of the first four rows from the corresponding Slater determinant as given by

$$\det \begin{bmatrix} \chi_1^+(1), \chi_1^+(2), \chi_1^+(3), \chi_1^+(4), \dots \\ \chi_1^-(1), \chi_1^-(2), \chi_1^-(3), \chi_1^-(4), \dots \\ \chi_2^+(1), \chi_2^+(2), \chi_2^+(3), \chi_2^+(4), \dots \\ \chi_2^-(1), \chi_2^-(2), \chi_2^-(3), \chi_2^-(4), \dots \\ \dots \end{bmatrix}. \quad (4.63)$$

Let all the spin variables $\{s_i\}$ have distinct values and the spin orbit splitting $\Delta\varphi \rightarrow 0$. After some linear rearrangements we can write the matrix as sketched for the first four rows

$$\det \begin{bmatrix} \varphi_1(1), 0, \varphi_1(3), 0, \dots \\ 0, \varphi_1(2), 0, \varphi_1(4), \dots \\ \varphi_2(1), 0, \varphi_2(3), 0, \dots \\ 0, \varphi_2(2), 0, \varphi_2(4), \dots \\ \dots \end{bmatrix} \quad (4.64)$$

up to a common complex prefactor. After reshuffling rows and columns, the single determinant of spinors factorizes into the product of two determinants with spin-up and -down particles. Generalization to odd N with an unpaired spinor is straightforward.

For the sake of completeness we note that the simplest trial function based on a *pair spinor* orbital $\chi_{pair}(\mathbf{r}_i, s_i, \mathbf{r}_j, s_j)$ written as an antisymmetrized product of distinct pairs of particles results in a pfaffian

$$\Psi_T(\mathbf{R}, \mathbf{S}) = \text{pf}[\chi_{pair}(\mathbf{r}_i, s_i, \mathbf{r}_j, s_j)] \exp[U(\mathbf{R})]. \quad (4.65)$$

Obviously, the pair orbital itself is antisymmetric since the pfaffian is defined for a skew-symmetric matrix so that $\chi_{pair}(\mathbf{r}_i, s_i, \mathbf{r}_j, s_j) = -\chi_{pair}(\mathbf{r}_j, s_j, \mathbf{r}_i, s_i)$, as explained previously [Baj06]. Note also that for odd number of electrons the skew symmetric matrix can be expanded by an unpaired row and column with an unpaired spinor so that the resulting matrix is of even dimension. Therefore systems with odd number of electrons can be described by the corresponding pfaffian [BM09] as well (without boosting the matrix by another row and column the pfaffian of matrix with odd dimensions would vanish identically).

4.4.3 Evaluation of the Pseudopotential and Importance Sampling

The evaluation of the term $\text{Re}[\Psi_T^{-1} W_T^{REP} \Psi_T]$ is similar to the evaluation done in standard QMC calculations. Consider electron i and nucleus I whose relative distance is r_{iI} . We will need to calculate the contribution

$$\begin{aligned} \frac{W_{T,(iI)}^{REP} \Psi_T}{\Psi_T} &= \sum_{\ell, j} W_{\ell, j}^{REP}(r_{iI}) \int d\Omega'_{iI} \int ds'_i \\ &\quad \times \sum_m \langle \Omega_{iI} s_i | \ell j m \rangle \langle \ell j m | \Omega'_{iI} s'_i \rangle \\ &\quad \times \frac{\Psi_T((\mathbf{r}_1, s_1), \dots, (\mathbf{r}'_i, s'_i), \dots, (\mathbf{r}_N, s_N))}{\Psi_T((\mathbf{r}_1, s_1), \dots, (\mathbf{r}_i, s_i), \dots, (\mathbf{r}_N, s_N))} \end{aligned} \quad (4.66)$$

where $\mathbf{r}_{iI} = \mathbf{r}_i - \mathbf{r}_I$, Ω, Ω' are corresponding solid angles, and $\mathbf{r}'_i = (r_i, \Omega'_i)$, while the integral in s' is over the spin degree of freedom in a given representation. One advantage of the projection used in localization approximation is that the integration over the spins can be done explicitly and exactly for each determinant in the trial function expansion. Of course, this is true only for the case of the Jastrow factor being spin independent as is our

choice here.

To illustrate this, let us consider a trial wave function which is built from a single determinant. Focusing on an individual electron i near a nucleus I as we have in equation (4.66), we will need the ratio of the wave function evaluated at (\mathbf{r}'_i, s'_i) to the original wave function. This can be written as

$$\frac{\det[\dots, \chi_\alpha(\mathbf{r}'_i, s'_i), \dots]}{\det[\dots, \chi_\alpha(\mathbf{r}_i, s_i), \dots]} = \sum_\alpha C_{\alpha,i}(\mathbf{r}_i, s_i) \chi_\alpha(\mathbf{r}'_i, s'_i) \quad (4.67)$$

where $C_{\alpha,i}(\mathbf{r}_i, s_i)$ are the matrix elements of the inverse transpose of the Slater matrix, where α labels the spinors and i labels the electron. Plugging this into equation (4.66), we obtain

$$\begin{aligned} \frac{W_{T,(iI)}^{REP} \Psi_T}{\Psi_T} &= \sum_{\ell,j} W_{\ell,j}^{REP}(r_{iI}) \sum_\alpha C_{\alpha,i}(\mathbf{r}_i, s_i) \\ &\int d\Omega'_{iI} \int ds'_i \sum_m \langle \Omega_{iI} s_i | \ell j m \rangle \langle \ell j m | \Omega'_{iI} s'_i \rangle \chi_\alpha(\mathbf{r}'_i, s'_i) \end{aligned} \quad (4.68)$$

Focusing on an individual ℓ, j element in the summation, we see

$$\begin{aligned} W_{\ell,j}^{REP}(r_{iI}) \sum_\alpha C_{\alpha,i}(\mathbf{r}_i, s_i) \int d\Omega'_{iI} \int ds'_i \sum_m \\ \times \langle \Omega_{iI} s_i | \ell j m \rangle \langle \ell j m | \Omega'_{iI} s'_i \rangle [a \phi_\alpha^\uparrow(\mathbf{r}') \chi^\uparrow(s') + b \phi_\alpha^\downarrow(\mathbf{r}') \chi^\downarrow(s')] \end{aligned} \quad (4.69)$$

where we have expanded the individual spinor χ_α into its spin and spatial functions. We can simplify the previous expression by defining two functions $A_{\ell,j}$ and $B_{\ell,j}$ as

$$\begin{aligned} A_{\ell,j}(\Omega_{iI}, \Omega'_{iI}) &= \int ds'_i \sum_m \langle \Omega_{iI} s_i | \ell j m \rangle \langle \ell j m | \Omega'_{iI} s'_i \rangle \chi^\uparrow(s'_i) \\ B_{\ell,j}(\Omega_{iI}, \Omega'_{iI}) &= \int ds'_i \sum_m \langle \Omega_{iI} s_i | \ell j m \rangle \langle \ell j m | \Omega'_{iI} s'_i \rangle \chi^\downarrow(s'_i) \end{aligned} \quad (4.70)$$

such that the individual ℓ, j element can be written as

$$W_{\ell,j}^{REP}(r_{iI}) \sum_{\alpha,i} C_{\alpha,i}(\mathbf{r}_i, s_i) \times \int d\Omega'_{iI} [a\phi^\uparrow(\mathbf{r}_i)A_{\ell,j}(\Omega_{iI}, \Omega'_{iI}) + b\phi^\downarrow(\mathbf{r}_i)B_{\ell,j}(\Omega_{iI}, \Omega'_{iI})] \quad (4.71)$$

Note that the spin integration has been eliminated, and we are only left with a integral over the solid angle Ω'_{iI} , which is carried out numerically using standard techniques as is the case of spatial-only nonlocality [Mit91]. The terms $\langle \Omega s | \ell j m \rangle$ are spin-spherical harmonics with $j = \ell \pm 1/2$, and can be written in our spin representation as

$$\begin{aligned} \mathcal{Y}_{\ell+1/2,m}^\ell(\Omega, s) &= \langle \Omega s | \ell, \ell + 1/2, m \rangle \\ &= \sqrt{\frac{\ell + m + 1/2}{2\ell + 1}} Y_{\ell, m-1/2}(\Omega) e^{is} \\ &\quad + \sqrt{\frac{\ell - m + 1/2}{2\ell + 1}} Y_{\ell, m+1/2}(\Omega) e^{-is} \end{aligned} \quad (4.72)$$

$$\begin{aligned} \mathcal{Y}_{\ell-1/2,m}^\ell(\Omega, s) &= \langle \Omega s | \ell, \ell - 1/2, m \rangle \\ &= -\sqrt{\frac{\ell - m + 1/2}{2\ell + 1}} Y_{\ell, m-1/2}(\Omega) e^{is} \\ &\quad + \sqrt{\frac{\ell + m + 1/2}{2\ell + 1}} Y_{\ell, m+1/2}(\Omega) e^{-is} \end{aligned} \quad (4.73)$$

Generalization to the inclusion of a spin-free Jastrow and/or multiple Slater determinant wave functions is straightforward.

In order to obtain the entire contribution of the pseudopotential, we simply sum over all of the electrons and ions and add the local contribution W_L^{REP} . This yields a total pseudopotential contribution of

$$W_T^{Re} = \text{Re} \left\{ \sum_{i=1}^{N_e} \sum_{I=1}^{N_I} \left[W_L(r_{iI}) + \frac{W_{T,(iI)}^{REP} \Psi_T}{\Psi_T} \right] \right\} \quad (4.74)$$

which is added to the local energy in equation (8.5).

Once we are able to evaluate the nonlocal potential contribution in the localization

approximation as $W_T^{Re} = Re[\Psi_T^{-1} W^{REP} \Psi_T]$, so that it becomes a multiplicative *many-body* and Ψ_T -dependent potential, we can apply the importance sampling transformation to the Eq. 4.32. Note that due to the continuous spin values the corresponding local energy is a continuous and smooth function almost everywhere (the exceptions might be, possibly, zero measure configurations for which both the amplitude ρ and the phase Φ vanish simultaneously). The evolution equation is therefore now solved for the product $g = \rho_T \rho_{FP}$ where FP denotes the fixed-phase solution.

4.4.4 Spin Sampling

In order to sample the spin variables smoothly within this spin representation, we add a spin “kinetic” energy and corresponding offset into the Hamiltonian for all s_i , $i \in \{1, 2, \dots, N\}$ of the form

$$T_i^s = -\frac{1}{2\mu_s} \left[\frac{\partial^2}{\partial s_i^2} + 1 \right] \quad (4.75)$$

such that $H \rightarrow H' = H + \sum_{i=1}^N T_i^s$. Consider the action of T_i^s on an arbitrary one-particle spinor $\psi(\mathbf{r}_i, s_i) = \alpha \varphi^\uparrow(\mathbf{r}_i) e^{is_i} + \beta \varphi^\downarrow(\mathbf{r}_i) e^{-is_i}$, where $\varphi^{\uparrow(\downarrow)}(\mathbf{r})$ are different different spatial orbitals for the spin-up and -down components. Clearly, $T_i^s \psi(\mathbf{r}_i, s_i) = 0$ due to the introduced offset so that there is no contribution from the spin laplacian.

At this point one has to consider whether to apply the importance sampling transformation before or after adding this spin variables Hamiltonian. Let us carry out the transformation with the spin Hamiltonian included, assuming the fixed-phase approximation. Using the notation from Eq. (4.14) it can be written as

$$\begin{aligned} -\frac{\partial g(\mathbf{X})}{\partial \tau} &= [H_{\mathbf{R}}^{\text{drift}} + E_L(\mathbf{X}) + W_T^{Re} - E_T] g(\mathbf{X}) \\ &\quad - \frac{1}{2\mu_s} \nabla_{\mathbf{s}}^2 g(\mathbf{X}) + \frac{1}{\mu_s} \nabla_{\mathbf{s}} \cdot [\mathbf{v}_D^{\mathbf{s}}(\mathbf{X}) g(\mathbf{X})] \\ &\quad + \frac{1}{2\mu_s} |\nabla_{\mathbf{s}} \phi_T(\mathbf{X})|^2 g(\mathbf{X}) \end{aligned} \quad (4.76)$$

where we have written $\nabla_{\mathbf{s}} = (\frac{\partial}{\partial s_1}, \dots, \frac{\partial}{\partial s_N})$ and for the sake of completeness we write down also the expression for the local energy

$$E_L(\mathbf{X}) = Re[\Psi_T^*(\mathbf{X}) T_{kin} \Psi_T(\mathbf{X})] / \rho_T^2 + V \quad (4.77)$$

where T_{kin} denotes the spatial kinetic energy. Note that there are formally three new terms generated by spin degrees of freedom: diffusion, drift term with velocity

$$\mathbf{v}_D^{\mathbf{S}}(\mathbf{X}) = \nabla_{\mathbf{S}} \ln \rho_T(\mathbf{X}) = \rho_T^{-1}(\mathbf{X}) \nabla_{\mathbf{S}} \rho_T(\mathbf{X}) \quad (4.78)$$

as well as contribution to the local energy from the spin gradient of the trial phase. We have introduced H' only to be able to sample the spin variables \mathbf{S} since in reality spin does not have any kinetic energy. Therefore we drop the contribution to the energy from the spin gradient of the trial phase as it creates an artificial contribution. This will be further discussed in §4.5.

For completeness, the inclusion of the spin kinetic energy and offset modifies importance-sampling Green's function in equation (8.7) to

$$\tilde{G}(\mathbf{X}' \leftarrow \mathbf{X}; \tau) \simeq T_{\mathbf{X}', \mathbf{X}} e^{-\tau[E_L(\mathbf{X}) + E_L(\mathbf{X}') - 2E_T]/2} \quad (4.79)$$

with

$$T_{\mathbf{X}', \mathbf{X}} \propto \exp \left[\frac{-|\mathbf{R}' - \mathbf{R} - \tau \mathbf{v}_D^{\mathbf{R}}(\mathbf{R})|^2}{2\tau} \right] \times \exp \left[\frac{-|\mathbf{S}' - \mathbf{S} - \tau_s \mathbf{v}_D^{\mathbf{S}}(\mathbf{S})|^2}{2\tau_s} \right] \quad (4.80)$$

where we have introduced a spin time-step $\tau_s = \tau/\mu_s$.

4.5 Time-step Errors and Approximations

A potential source of error comes from the choice of spin time-step τ_s (or equivalently, the spin mass μ_s) due to the complex representation on S^1 . In its minimal representation, the expectation value for the energy is given by (4.51), i.e. the entire 2^N spin configurations must be summed over at every step in the imaginary time evolution or one must sample the discrete spins causing “jumps.” In our continuous representation, the speed of the spin sampling can be chosen at our disposal due to the effective spin mass μ_s . If we consider

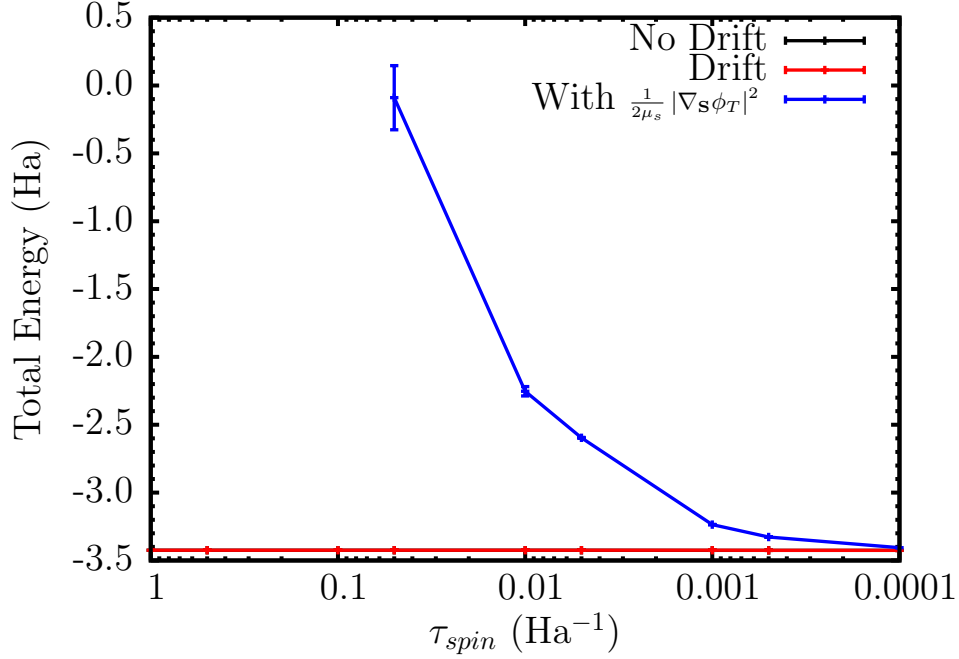


Figure 4.1 Total Energy of the Pb ground state. The “No Drift” and “Drift” calculations are indistinguishable at this scale.

the i th walker, the expected root-mean-square displacement (rms) in coordinate space is

$$r_i^{rms}(t) \propto \sqrt{t} \quad (4.81)$$

neglecting the drift velocity. In spin space, however, the rms goes as

$$s_i^{rms}(t) \propto \sqrt{\frac{t}{\mu_s}} \quad (4.82)$$

By taking the limit $\mu_s \rightarrow 0$, the spin rms $s_i^{rms} \rightarrow \infty$, i.e. the spin has sampled its entire space. At $\mu_s = 0$, this is equivalent to summing over all spin configurations.

In order to determine the dependence on the spin mass/time-step, we studied the FPSODMC energy as a function of spin mass/time-step for a fixed spin time-step of the Pb atom ground state. In order to minimize the spatial time-step error, we choose a spatial time-step of $\tau = 0.001$ Ha⁻¹ throughout. Our trial wave functions are complete open-shell configuration interaction (COSCI), with one-particle spinors obtained from the DIRAC relativistic quantum chemistry code [Dir]. Additionally, we performed calculations that

include and exclude the spin drift velocity. This amounts to a choice of when the importance sampling transformation is invoked in the algorithm. With the inclusion of the spin drift, the spin variables are updated with

$$s'_i = s_i + \sqrt{\tau_s} \eta + \tau_s v_D^{s_i} \quad (4.83)$$

where η is a normally distributed random variable and the greens function in equation (4.79) is used. Excluding the spin drift amounts to the condition $\mathbf{v}_D^S = 0$.

We first justify dropping the term $1/2\mu_s |\nabla_{\mathbf{s}} \phi_T(\mathbf{X})|^2$ from the local energy. This term arises from the importance-sampling transformation when we include the spin kinetic energy, which was introduced to allow for efficient sampling of the spin degrees of freedom. Clearly, this term is spurious since spins do not have kinetic energy. In order to illustrate its effect, a plot of the total energy of Pb ground state is shown in Figure 4.1. As τ_s gets large, μ_s approaches zero and fully integrates the spin degrees of freedom. However, since the spin phase gradient is proportional to μ_s^{-1} , it adds a positive term to the local energy with rapidly increasing value for larger spin time steps. From here on, all calculations are performed without this term.

The dependence on the spin drift velocity is shown in Figure ???. For large spin timesteps and small spin masses, the total energies for both methods saturate at ~ -3.4244 Ha. For spin timesteps $\tau_s \geq 1.0$ Ha $^{-1}$, the acceptance ratios drop below 0.7. In this regime, the method is no longer purely DMC. For spin timesteps $\tau_s < 1.0$ Ha $^{-1}$, the energies agree between the with/without spin drift calculations to within the error bars. The inclusion of the spin drift acts to increase the acceptance ratio; for spin timesteps larger than $\tau_s = 0.1$ Ha $^{-1}$, the acceptance ratio is greater than 0.99, which is desired for a DMC calculation. An interesting feature is a small decrease in total energy near $\tau_s = 0.001$ Ha $^{-1}$. For very small spin timesteps, the energy saturates at ~ -3.4265 Ha.

In order to determine how the energies compare to the exact eigenvalue for this Hamiltonian, we performed full configuration interaction (FCI) calculations with cc-*VnZ* basis sets and extrapolate to the complete basis set limit (CBS), using the same effective Hamiltonian

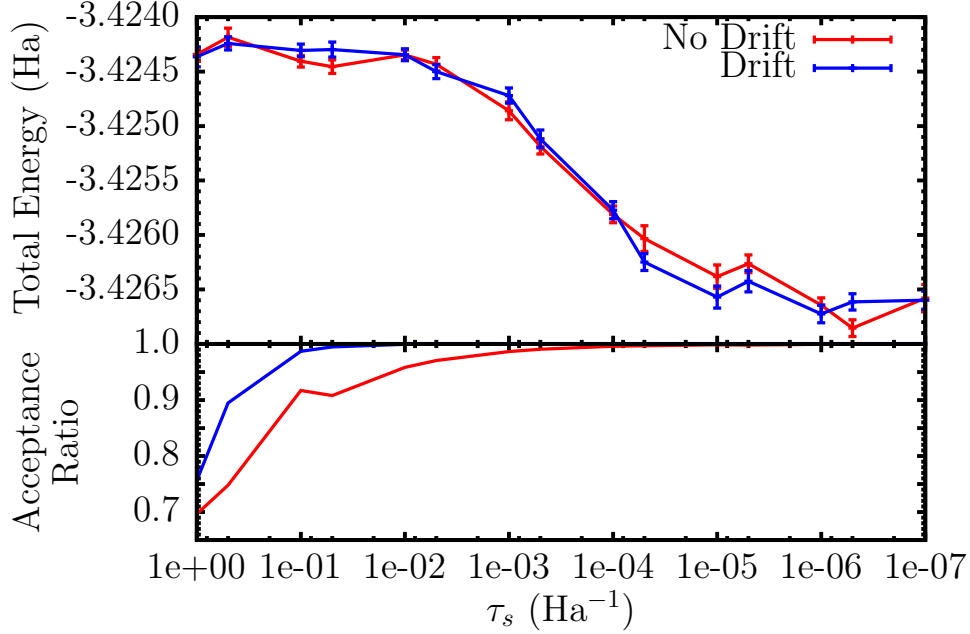


Figure 4.2 Total energies of the Pb ground state with COSCI trial wave functions. The calculations were performed with drift, $\mathbf{v}_D^S \neq 0$, and excluding drift, $\mathbf{v}_D^S = 0$.

within the two-component spinor formalism. For extrapolation, we use fits of the form

$$E_{CBS} = E_{CBS}^{f(\text{COSCI})} + E_{CBS}^{g(\text{FCI-COSCI})} \quad (4.84)$$

$$f(x) = E_{CBS}^{f(x)} + \alpha e^{-\beta x} \quad (4.85)$$

$$g(x) = E_{CBS}^{g(x)} + \frac{\gamma}{(n-3/8)^3} + \frac{\delta}{(n-3/8)^5} \quad (4.86)$$

where n refers to the size of the basis set and α , β , γ , and δ are fitting parameters. In addition to a COSCI trial wave function, we also calculated the total energy using a CISDT trial wave function for the ground state. The results are shown in Figure 4.3. For $\tau_s < 0.0001$ Ha⁻¹, both ψ_T used for the FPSODMC calculations are below the FCI calculations with a cc-VQZ basis set. The FCI with a CBS extrapolation is the best estimate for the exact ground state of this effective Hamiltonian, and the FPSODMC method lies above in both cases, due to the fixed-phase bias.

For a large spin time-step and small spin-mass, the spin configuration space is sampled faster than the spatial degrees of freedom. This corresponds to the plateau at -3.4244 Ha as seen in Figure ???. However if τ_s gets too large relative to τ , the spin steps become very

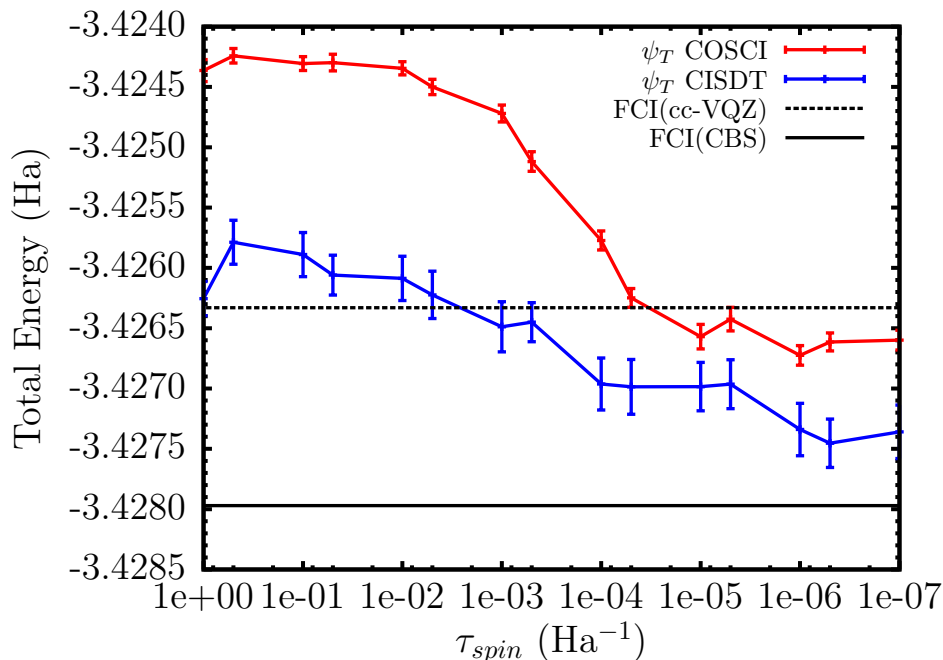


Figure 4.3 Total energy using COSCI and CISDT trial wave functions for the Pb ground state. FCI with cc-VQZ and a CBS extrapolation are included as a reference.

large and lowers the acceptance ratio significantly. In the other regime, namely where $\tau_s \ll \tau$ and μ_s is large, the spin degrees of freedom diffuse slowly relative to the spatial degrees of freedom and are effectively slowly moving barriers. Therefore spatial degrees of freedom have time to find and favor the minima subject to the fixed-phase given by the slowly moving spins. This drives the energy down (Figures ?? & 4.3), however it is still variationally bound with respect to the exact eigenvalue given from the FCI(CBS) energy. This is interesting to observe since due to overcomplete representation for the spins one cannot rule out that the algorithm in the limit of slow spin evolution might find energies that would not be variational. However, all the indication are are that the fixed-phase approximation dominates over the full range of the spin timesteps.

For the sake of completeness, we also performed FPSODMC calculations using the minimal spin representation, where the spin variables are randomly sampled from $s_i \in \{-1/2, 1/2\}$. For the Pb atom with a LC REP, there are only $2^4 = 16$ spin configurations. For larger systems, the spin space grows as 2^N . The spin variables were updated at each spatial step with uniform sampling, and we found an energy of -3.4239(3) Ha that is comparable to results that we have obtained with our S^1 representation. However, the DMC acceptance

ratio is only ≈ 0.92 and clearly the acceptance can get significantly lower for larger systems (in fact, we expect exponential decrease of the acceptance in large systems). It appears to be challenging to counter this trend since discrete representation does not provide any drift that could boost the acceptance. Therefore our expanded spin representation is significantly more efficient and provides several other advantages as elaborated upon above. In addition, there might be further gains in improving our method that will be explored in subsequent work.

4.6 Applications

We present applications of the FPSODMC method to several examples of electronic structure problems with significant impact of the spin-orbit terms. We apply the method to several electronic properties of atoms and molecules. In §4.6.1, we present results for the lead hydride PbH, where we present results for the binding energy and bond length. We perform both averaged scalar-relativistic and spin-orbit treatments of this molecule using FPSODMC to demonstrate the importance of spin-orbit on the electronic properties. In §4.6.2, we present a study of the Sn atom and dimer as an example of the 4th row element that shows significant impact of the spin-orbit. For the atom, we calculate the electronic structure of the first few excited states with remarkable agreement with experiment. For the dimer, we present scalar and spin-orbit treatments to predict the binding energy and bond length. Lastly in §4.6.3, we calculate the electron affinities of the $6p$ elements. All trial wave functions were constructed from one-particle spinors obtained from the DIRAC [Dir] code.

4.6.1 PbH

We present calculations of bond lengths and dissociation energies of the linear molecule PbH as a testing cases which can be compared with previous high-accuracy calculations. Previous theoretical studies have focused on two aspects of Pb molecules: treatment of the correlation that requires multi-reference wave functions and relativistic effects including both scalar and beyond. Both of these are also addressed in our DMC calculations.

For PbH molecule, a large core (LC) REP[MS00] was found to be sufficiently accurate and the results are presented in Table 4.1. We contrast two types of DMC calculations with regard to treatment of the spin degrees of freedom. One is the usual static-spin (1-component) calculation with weighted average of the spin-orbit terms for the same l in the PP, denoted

Table 4.1 PbH bond length (r_e) and dissociation energy (D_e)

Method	r_e (Å)	D_e (eV)
spin-free CCSD(T) ^a	1.836	2.66
MRCIS-spss(CCSD(T)) ^b	1.834	1.61-1.71
DMC/1-component AREP	1.834	2.582(3)
DMC/2-component REP	1.838	1.67(2)
Exp[HH79]	1.839	≤ 1.69

^a 1-component CCSD(T) with large-core AREP [MS00]
^b 2-component MRCIS with spin-free-state shift (spss) evaluated at the 1-component CCSD(T) level [MS00]

as AREP defined above. The second type employs the present method with spinors and genuine 2-component relativistic PP that includes SO interaction terms explicitly (REP). The result of REP calculations show excellent agreement with the experiment and previous SO-CI [MS00; HH79] study. We point out the large discrepancy (≈ 0.9 eV) between the results of AREP and REP for the dissociation energy. It is clear that the averaged SO treatment is grossly inadequate for producing a reliable result for this system. The dominant correction for SO effects in this case comes from the atom, about 1 eV. The 2-component result is clearly more consistent with the experiment value. Note that the molecular spin-orbit correction $\Delta_{\text{SO}}^{\text{M}} = E_{\text{DMC}}^{\text{AREP}} - E_{\text{DMC}}^{\text{REP}}$ for PbH molecule is estimated to be actually rather small, about 0.11 eV. However, in this case the impact of the SO terms for Pb atom is dominant, about 1.1 eV. Thus, the atomic SO correction contributes most of the large discrepancy between the AREP and REP results.

4.6.2 Sn and Sn₂

We investigate the effect of the spin-orbit interaction on Sn systems, namely the Sn atom and dimer. We investigate both large-[Sto02] and small-core[MS00] PPs with 46 and 28 electrons removed respectively. In order to isolate the effect of spin-orbit, we perform static-spin calculations using only the PP via FNDMC as well as full dynamical spin calculations with the full REP via FPSODMC using COSCI trial wave functions. By comparing the AREP and REP calculations, we calculate an atomic spin-orbit correction $\Delta_{\text{SO}} = E_{\text{DMC}}^{\text{AREP}} - E_{\text{DMC}}^{\text{REP}}$ between the ground states. For the LC, we find a correction of $\Delta_{\text{SO}}^{\text{LC}} = 0.1689(2)$ eV and $\Delta_{\text{SO}}^{\text{SC}} = 0.27(2)$ eV for the SC. From Table 4.2, we see that the SC REP performs significantly better than

the LC REP and agrees remarkably well with the experimental excitation energies. We also performed full configuration interaction (FCI) calculations using the LC REP in order to compare with our FPSODMC excitation energies.

For the dimer, we aim to find the equilibrium geometry as well as the binding energy. To determine the effect of spin-orbit on the dimer, we calculate the dimer both with the AREP and REP with COSCI trial wave functions. We perform calculations at various bond lengths and fit the binding curve to the Morse potential of the form

$$V(r) = D_e \left(e^{-2a(r-r_e)} - 2e^{-a(r-r_e)} \right) \quad (4.87)$$

Results are shown in Figure 4.4. As was the case for PbH in §4.6.1, the scalar relativistic calculation without spin-orbit (AREP) significantly overbinds the dimer. The scalar relativistic method predicts an equilibrium bond length of $r_e = 2.74(2)$ Å and a dissociation energy of $D_e = 2.32(3)$ eV. We note that these results are comparable to other DMC studies of the Sn₂ molecule with scalar relativistic pseudopotentials [Naz16] For the dimer including the SO interaction, we find an equilibrium bond length of $r_e = 2.742(9)$ Å and a dissociation energy of $D_e = 1.80(2)$ eV. We calculate a molecular spin-orbit correction $\Delta_{SO}^M = E_{\text{DMC}}^{\text{AREP}} - E_{\text{DMC}}^{\text{REP}}$ at the equilibrium bond length of $\Delta_{SO}^M = 0.12(3)$ eV. We note that a COSCI-MVOQ scalar relativistic treatment predicts $r_e = 2.860$ Å and a $D_e = 1.704$ eV whereas COSCI with spin orbit treatment predicts a bond length of $r_e = 2.91$ Å and a dissociation energy $D_e = 0.858$ eV. We also calculated the binding curve for the LC REP within FPSODMC with a COSCI trial wave function, shown in Figure 4.5. The LC REP produces an overall shift in the binding curve of ~ 0.1 eV.

Table 4.2 Excitation energies for the Sn atom from the 3P_0 ground state. We include both LC and SC PPs. For completeness, we include COSCI and FCI to compare the FPSODMC and experiment[Nis].

State	COSCI	DMC	COSCI	DMC	FCI [†]	CISD [†]	Expt. [Nis]
	LC	LC	SC	SC	LC	SC	
3P_1	0.168	0.145(2)	0.180	0.23(2)	0.175	0.196	0.210
3P_2	0.392	0.367(3)	0.412	0.43(2)	0.375	0.416	0.425
1D_2	1.311	0.967(3)	1.308	1.08(2)	1.035	1.146	1.068
1S_0	2.783	2.119(3)	2.742	2.17(2)	2.214	2.279	2.128

[†] cc-pVTZ [Pet03]

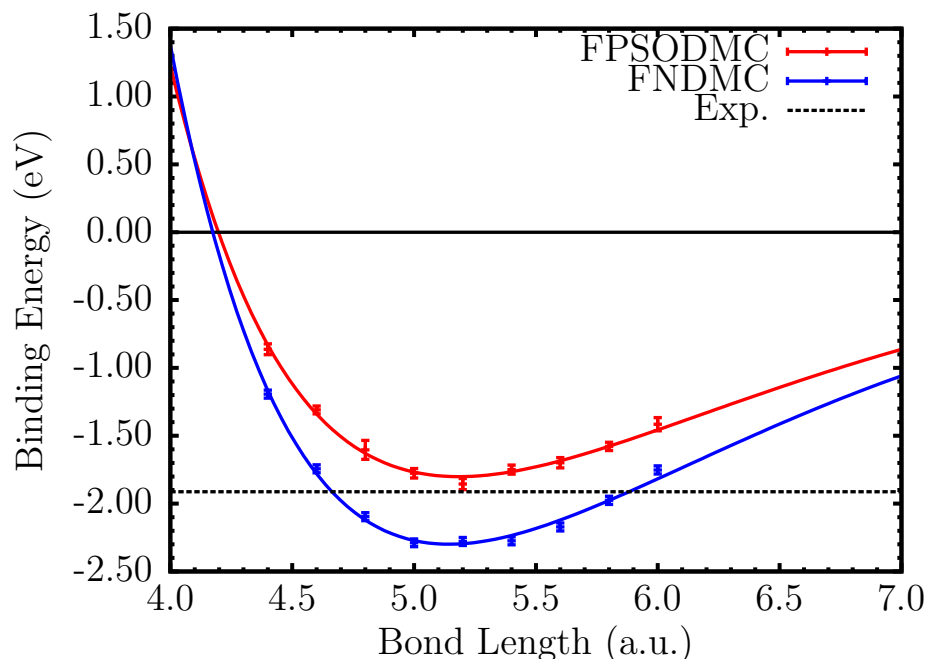


Figure 4.4 Binding Curve of the Sn_2 molecule using averaged spin-orbit AREP with FNDMC and spin-orbit REP with FPSODMC methods. The curves are offset to dissociation limit $2E_0(\text{Sn})$ within each method to enable comparison for the predicted binding energy of each method with experiment.

Experimental data for the dimer agrees remarkably well with the FPSODMC calculations. The experimental bond length was found to be $r_e^{\text{exp}} = 2.748 \text{ \AA}$ [Pak88] and a dissociation energy of $D_e^{\text{exp}} = 1.9125(31) \text{ eV}$ [Pak88], which agree to our FPSODMC numbers to approximately 0.1 eV. We note that all DMC calculations (with and without SO) predict an equilibrium geometry that agrees with experiment to approximately 0.1 \AA . Due to the fact that Sn is from the fourth row, the spin-orbit correction to the atom and dimer energies are intermediate size (0.27(2) eV and 0.12(3) eV respectively), these contributions are not negligible if one desires predictions to chemical accuracy.

4.6.3 Electron Affinities

As a last illustration of the method, we present results for the electron affinities (EA) of various atoms, namely the $6p$ elements Tl-At. These elements have rather strong spin orbit coupling compared to their isovalent counterparts. For example, the respective

EAs of C, Si, Ge and Sn are 1.2629 eV, 1.385 eV, 1.2 eV, and 1.2 eV[HL85]. However, due to the strong spin-orbit effect on the 3P_0 ground state of Pb, the EA is significantly reduced which was experimentally determined to be 0.365(8) eV [Fei81]. Theoretical calculations that did not include spin-orbit predict an EA of 1.284 eV, which is comparable to Ge and Sn where the spin-orbit interaction is significantly weaker[Pet03]. The inclusion of spin-orbit has a significant effect; a full four-component relativistic treatment predicts an EA of 0.403(39) eV[Tat09] whereas FPSODMC using REPs predicts 0.417(7) eV[Mel16a].

The electron affinity is calculated from the ground state energies of the neutral, E_0 , and anionic E_0^- species, namely

$$EA = E_0 - E_0^- \quad (4.88)$$

For each species, we calculate the EA using an AREP and REP to isolate the effect of spin-orbit. The results are presented in Table 4.3. Electron affinity grows as proton number increases. The scalar relativistic (AREP) calculation for Bi does not follow this trend, and is significantly lower. For Pb, Po, and At, the EA is higher than experiment or other quantitative estimates. The inclusion of spin-orbit significantly improves these results to be accurate

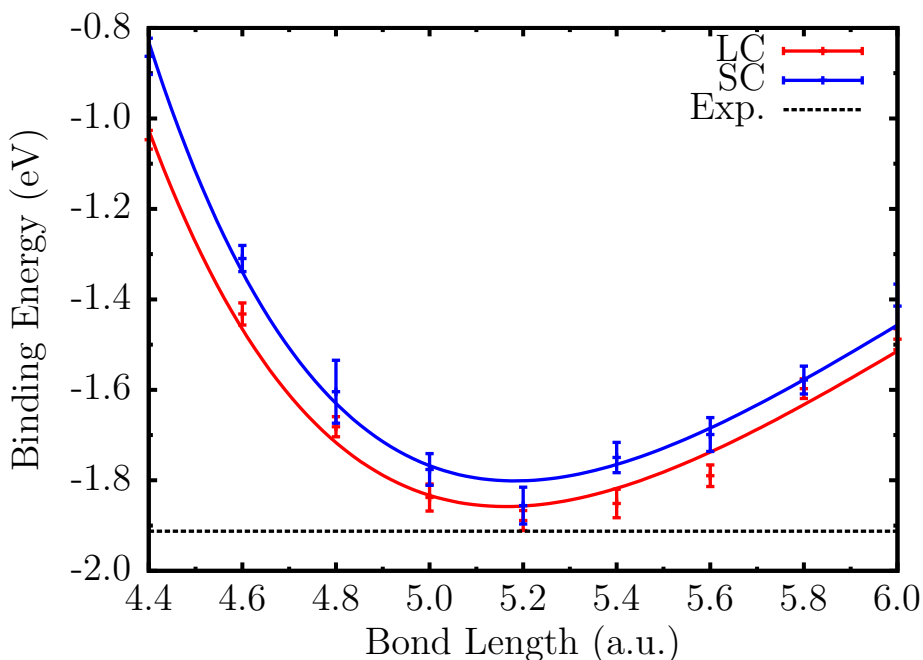


Figure 4.5 Binding curve of the Sn_2 molecule using large- and small-core REP. The large- and small-core systems have 8 and 44 valence electrons, respectively.

Table 4.3 Electron Affinities for the $6p$ elements. COSCI trial wave functions used throughout, with LC AREP/REPs for Pb[MS00], Bi[Sto02], Po[Sto02], and At[Sto02]. For Tl, no LC REP was found, so we utilize a SC AREP/REP [Met00].

Species	COSCI AREP	COSCI REP	FNDMC AREP	FPSODMC REP	Expt
Tl	-0.015	-0.195	0.29(2)	0.17(3)	0.377(13) [Eat]
Pb	0.951	-0.085	1.35(1)	0.417(7)[Mel16a]	0.365(8)[Fei81]
Bi	-0.144	0.080	0.82(1)	1.04(2)	0.942362(13) [BH01]
Po	0.981	0.556	1.94(1)	1.32(6)	1.32 [†]
At	2.280	1.545	3.22(1)	2.83(8)	2.80(2) [†]

[†] No experimental data. Quantitative estimates [Zol69]

to within 0.1 eV of the experimental values for Pb and Bi and in complete agreement with other quantitative estimates for Po and At where no experimental data exists to our knowledge.

4.7 Conclusions

In this paper we elaborate in detail on the fixed-phase spin-orbit DMC (FPSODMC) method that we have introduced recently [Mel16a]. We provide detailed derivations for several aspects of the method. One important point is the proof of upper bound property when dealing with complex nonlocal operators that enables us to employ techniques based on T-moves, i.e., using combination of projection on the trial function and nonlocal sampling. The next point we demonstrate is more practical and has to do with the time step biases with regard to spin sampling together with the overcompleteness of the representation. We provide calculations of several systems that illustrate the capabilities of the method as well as show the impact of the spin-orbit interactions on energy differences. The method opens new perspectives for many-body electronic structure calculations spinor formalism that take into account variable nature of the spin degrees of freedom.

CHAPTER 5

A New Generation of Effective Core Potentials for Correlated Calculations

Portions of this chapter also appeared in:

A New Generation of Effective Core Potentials for Correlated Calculations

J. Chem. Phys. **147**, 224106 (2017)

Contributions:

In the work described in this chapter, the author played a significant role in the investigation of the various objective functions used to generate the final high-accuracy ECPs and the preparation of the manuscript. The author generated all data corresponding to the second-row S ECP.

5.1 Introduction

In recent decades, many-body electronic structure methods have enjoyed steady progression in accuracy and efficiency; this has largely been driven by algorithmic improvements coupled with the increasing power of computational resources. These advances can be readily seen, for instance, in methods that rely on wave function expansions, such as coupled cluster (CC) methods, as well as stochastic approaches that sample either particle configurations or the Fock space of antisymmetric wave functions. For example, quantum Monte Carlo (QMC) methods that utilize explicitly correlated many-body wave functions often reach accuracies of 1-2 kcal/mol for systems containing elements from several rows of the periodic table [Foulkes:2001rmp; KM11]. For practical reasons these methods often approximate properties of atomic cores, especially when calculating systems with heavier elements. In particular, the cores are either left uncorrelated, fully frozen or even eliminated entirely using pseudopotentials or effective core potentials (ECPs). Note, however, that ECPs are generated within approximate theories as they are often tuned to match Density Functional Theory (DFT) or Hartree-Fock (HF) all-electron atomic properties in the valence space.

In correlated treatments of very heavy elements, the use of ECPs is essentially unavoidable since explicit correlation of atomic cores would become another bottleneck for such calculations. For instance, in QMC the corresponding computational cost grows as $\mathcal{O}(Z^{5-6})$ where Z is the nuclear charge. Although ECPs provide an important efficiency boost, they inevitably introduce errors that can compromise the overall quality of results. Furthermore, the increasing accuracy of the many-body methods mentioned above reveals that ECP imperfections can become a dominating source of systematic bias [SM13; Naz16].

In recent decades a number of ECP tables as well as on-demand computational packages have been established using several types of electronic structure approaches. This includes DFT constructions such as the comprehensive table of Bachelet et al [Bac82] based on relativistic DFT and ECPs generated by the OPIUM package [Wal]. We also mention recent advances in DFT constructions with multiple projectors [Ham13]. Several sets of tables suitable for many-body calculations have been developed using Hartree-Fock/Dirac-Fock (HF/DF) methods [Dol87; DC12; Ste84; PC85; Ros90; HW85; Bur07; TN05].

Most of the existing ECP constructions have relied on reproducing one-particle properties such as norm/charge-conservation [Bac82] or closely related shape consistency [PC85;

Ste84; Ovc01], HF/Dirac-Fock (DF) excitation energies[Dol87] or their extensions [Ham13] including a recent composite scheme designed to accurately build-in many-body effects [TN13; TN15; TN17]. Comprehensive reviews [Pic89; DC12] contain details as well as many relevant references which we skip here. To the best of our knowledge, the incorporation of many-body approaches into ECP constructions have been rather limited so far; first by the QMC work of Acioli and Ceperley [AC94], also the works of Maron, et al.[MT98] and Fromager et al.[Fro04] and recent works of Trail and Needs [TN13; TN15; TN17]. Overall, ECPs have arguably been some of the most successful ways to eliminate unnecessary degrees of freedom through effective Hamiltonians. They have helped to save huge amounts of computational time and effort over the past few decades, and are largely responsible for enabling well-converged many-body solid state calculations.

As indicated above, the current generation of ECPs is becoming a significant hurdle in achieving a level of accuracy consistent with high-accuracy all-electron correlated calculations. Presently, valence-only correlated approaches often require additional ECP testing, redeveloping, adapting, modifying, and revalidation in order to get the systematic errors originating in ECPs below a desired threshold. Such practice is highly undesirable and indeed hampers a sizable fraction of QMC and other correlated wave function calculations. This situation is similar to that of past DFT calculations, where pseudopotentials were largely constructed by hand. Optimization approaches [SG15] and systematic testing[Lej14; Lej16] versus all-electron approaches and solid-state calculations have recently been very successful and show that there is a significant room for improvements. However, the substantially higher cost of many-body calculations makes a direct use of these approaches for ECP development quite challenging.

In this work, we envision and assess practical approaches for constructing a new generation of ECPs. We have several goals in mind. First, an effective Hamiltonian that includes an ECP is much more useful if it is reasonably simple and constructed for broad use in basically all theories, i.e., that its universality is similar to the original all-electron Hamiltonian. One goal, therefore, is to probe for effective and simple forms that are appropriate for use in both mean-field and/or explicitly correlated methods. In this way, the effective ECP Hamiltonian provides a well-defined reference so that systematic errors from subsequent approximations and approaches are on full display and can be clearly benchmarked.

Another key goal is to increase the accuracy of the new ECPs beyond previous constructions, using measures that we define later. We explore a few strategies for constructing ECPs

using correlated wave function methods from the outset. Ideally, these effective Hamiltonians should reproduce, as closely as possible, the behavior of the valence electrons with the original Hamiltonian in a vast range of chemical environments; regardless of whether one calculates a molecule or a condensed system, in an equilibrium or non-equilibrium conformation, and within weak or strong bonding settings. Therefore, our goal is to construct an effective Hamiltonian that mimics the many-body valence spectrum, the spatial structure of eigenstates and overall scattering properties of the original, relativistic, all-electron atom's Hamiltonian. Clearly, some compromises will have to be made and in this work, we investigate the accuracy limits of ECPs of a simple semi-local form with almost a minimal number of parameters and we derive such ECPs for a small set of testing elements. As a guiding principle we have in mind *isospectrality for a subset of states*, i.e., we demand that the all-electron and ECP spectra are as close as possible to a set of valence states. The isospectrality is a very general property and applies even in cases when the Hilbert spaces for the two isospectral operators are different, e.g., due to different boundary conditions, different spatial domains, etc.

Note that when we refer to the spectrum, we have in mind the spectrum for not only the neutral atom, but states of cations, anions and binding curves for molecular systems containing this species. In addition to spectra, we explore the use of many-body spatial information such as correlated single-body density matrices and natural orbitals as well as constructions that are matched to a combination of spatial and spectral information and further iterated constructions. These explorations help to identify trade-offs between the accuracy for spectra and spatial many-body properties versus demands on the constructions useful for a variety of electronic structure packages. Additional demands such as systematic record of benchmarks from correlated calculations, appropriate updates, etc., are further specified below.

In Section 5.2, we outline a list of properties that we argue are necessary for any ECP that leads to optimal results within high-accuracy valence-only many-body calculations. Section 5.3 outlines the inverse-problem approach that we have utilized throughout this work as well as the form of our ECPs. Section 5.4 covers general optimization methods and constructions for generating ECPs consistent with our desired set of properties. Results for the first- and second-row atoms, namely, boron, carbon, nitrogen, oxygen, and sulfur are shared in Section 5.5. An analysis of each pseudoatom's CCSD(T) dimer properties as compared to all-electron calculations is also given in Section 5.5. An extensive test of

transferability is applied to our most accurate ECP constructions in Section 5.6 in order to further ascertain their performance within various molecular environments. We conclude in Section 5.7.

5.2 Desired properties

For this new generation of ECPs, we envision including the following set of properties.

a) **Many-body construction** is the key and perhaps the most salient point in this list. Some ideas in this direction have been tried before, for example, Acioli and Ceperley [AC94] explored correlated wave functions to generate ECPs and required that the pseudoatom's density matrix matched that of the full atom beyond a given core radius. Extending on this methodology, Trail and Needs have generated tables of ECPs from MCHF wave functions [TN13; TN15] and CCSD(T) excitation energies [TN17] to improve their previously generated table [TN05]. Additionally, attempts were made to build core effects into ECPs in the works of Maron, et al.[MT98] and Fromager et al.[Fro04]. We believe that a more systematic and accurate approach here can still be achieved, especially when considering properties outside of equilibrium (see below). In our construction, we use Coupled Cluster (CC) and Configuration Interaction (CI) methods that are very effective for atomic and small molecular systems, and moreover, we analyze how our generated ECPs perform in non-equilibrium geometries.

b) **The simplicity of the ECP operator** has significant benefits since it allows use in many methods (QMC, quantum chemistry based on expansions in basis sets, and DFT, using established codes). We, therefore, build upon a simple semi-local form that has been in use for some time. Per need, a number of extensions could prove to be useful and could be incorporated in future. This might involve different choices of core-valence partitions, several types of representations such as numerical on radial meshes besides usual Gaussian expansions, fixed radial cutoffs, etc. More desired properties could be added as the project develops, e.g., how to optimally deal with "fat" cores of heavy elements, possible inclusion of core polarization and core relaxation terms, minimizing the fixed-node errors generated by the core part of the wave function without compromising the accuracy, minimizing the locality approximation in QMC applications following [KK17], etc. For plane wave applications, the suitability of recasting the ECP into the Kleinman-Bylander form [KB82] could be included as well.

c) **Testing** on a set of systems in order to delineate the accuracy limits for energy differences, equilibrium bond lengths and other properties. Potentially, some of these systems could be included into the retuning set, if important or necessary. For use and further improvements, it will be very useful to have documented systematic errors of generated ECPs documented upfront.

d) **Systematic updates**, i.e., keeping the data, developments, and history on a website [Web] that can be eventually updated by interested contributors at large.

5.3 Effective ECP Hamiltonian: Isospectrality on a subspace of valence states

One way to approach the ECP construction is to formulate it as an inverse problem, i.e., finding an effective operator that produces a subset of valence atomic states such that outside the core they reproduce the all-electron properties as closely as possible. We assume that for this subset both the all-electron and pseudized atoms can be solved in the same systematic and consistent framework, ideally exactly, or, in practice, nearly exactly, using the state-of-the-art many-body methods.

What makes the two operators, the all-electron Hamiltonian and ECP Hamiltonian, close? Assuming we are interested in the valence subset of states and properties there are the following two key aspects:

- i) the two spectra should be the same/very close; and,
- ii) the spatial characteristics (one- and multi-particle many-body density matrices) of the corresponding two sets of eigenstates outside the core region should be the same/very close.

Note that, in general, the two Hamiltonians differ substantially, in the number of particles/degrees of freedom, presence/absence of ionic Coulomb singularities and relativity. Since in the reference calculations all correlations (including cores) are assumed to be present, i) and ii) implicitly demand that the ECPs capture, as much as possible, also the *impact of core-core and core-valence correlations on the valence properties*. Our results below show that to a certain extent, this is indeed the case.

Finding the desired solution in this setting essentially defines an inverse problem. We can expect that this problem will be ill-conditioned with many nearly optimal or non-unique solutions. Why is this the case? The inclusion of many states into the optimization set

often leads to a frustrated optimization problem, e.g., improvement for one state increases penalties for other states. With sufficient number of such frustrating couplings one ends up with a problem that is qualitatively similar, say, to finding a ground state of a spin glass (a well-known problem in statistical mechanics).

One way how to deal with such ill-conditioned problems is to impose appropriate constraints that limit the space of possible solutions. That makes the problem solvable but could result in compromises on the resulting accuracy or incurred biases. Often, it is quite difficult to find the right set of constraints. If the objective function is over-constrained the search for a minimum is fast but the solution might be too biased by the constraints. On the other hand, in an under-constrained formulation, the optimization could be very inefficient. Therefore, the goal is to find the best trade-off(s) between these two limits.

5.3.1 ECP Form

The valence-only electronic Born-Oppenheimer Hamiltonian we consider has the following form

$$H_{\text{val}} = \sum_i [T_i^{\text{kin}} + V_i^{\text{pp}}] + \sum_{i < j} 1/r_{ij} \quad (5.1)$$

For this work, we use a semi-local ECP form with a minimal number of parameters [Bur07]

$$V_i^{\text{pp}} = V_{\text{loc}}(r_i) + \sum_{l=0}^{l_{\text{max}}} V_l(r_i) \sum_m |lm\rangle \langle lm|, \quad (5.2)$$

where r_i is the radial distance of electron i from the core's origin. The non-local terms contain the projectors on lm -th angular momentum state. The local term, V_{loc} , is chosen as

$$V_{\text{loc}}(r) = -\frac{Z_{\text{eff}}}{r} (1 - e^{-\alpha r^2}) + \alpha Z_{\text{eff}} r e^{-\beta r^2} + \gamma e^{-\delta r^2}, \quad (5.3)$$

where Z_{eff} is the effective core charge, $Z_{\text{eff}} = Z - Z_{\text{core}}$. The V_l potential was chosen to consist of a single gaussian term

$$V_l(r) = \beta_l e^{-\alpha_l r^2}. \quad (5.4)$$

All variables labeled by Greek letters are treated as optimization parameters in the minimization of a chosen objective function. Additionally, a constraint that forces a concave

shape at the origin is imposed [Bur07]

$$\gamma\delta + \alpha_l\beta_l > 0, \quad \forall l. \quad (5.5)$$

In the case(s) of combined constructions below, the number of Gaussians in each channel is extended.

Note that the chosen form imposes a very significant restriction on the variational freedom of the effective operator. We have used this simple form not only to simplify the optimization problem but also to investigate the true many-body accuracy limit for this “minimal model” version. The results are actually very encouraging. As presented in the following sections, even with such restricted variational freedom we were able to construct more accurate effective potentials than the existing ones as well as to illustrate the presented concepts above.

5.4 Optimization methods and constructions

In this section, we give a summary of a few atomic and molecular properties predicted from our constructed ECPs for several atoms from the first and second rows. For comparison, along with our results, we juxtapose the predictions of these properties from correlated all-electron, uncorrelated core/correlated valence for a handful of other ECPs that have been used in many-body calculations in recent years. Furthermore, we present results that utilize different strategies to construct the ECPs, namely, ECPs built from all-electron spectral data only, from spatial data only as well as combined and iterated constructions.

For correlated calculations, a number of codes can be used in a non-relativistic setting. For relativistic calculations, however, the choices are more limited. Nevertheless, it is interesting to see how far one could push the accuracy limits with currently existing methods and codes.

5.4.1 Objective function with atomic spectral discrepancies only

For the all-electron spectral-only references, we used the MOLPRO quantum chemistry package [Wer12] to calculate a subset of states from each atom’s spectrum using the CCSD(T) method. To account for scalar relativistic effects, in the all-electron reference we used the Douglas-Kroll-Hess Hamiltonian throughout. For each atom, the uncontracted aug-cc-

pCV5Z basis set (which includes core state correlation functions) was chosen in order to minimize finite basis set errors as much as possible. We used this same basis for both pseudo-atom and all-electron cases so that the basis set errors would largely cancel when gap discrepancies between the two cases were calculated. We carried out limited tests with even larger basis (6Z) and we found out that the gained accuracy had a marginal effect on the spectral differences when compared with other discrepancies such as binding curves for molecular systems as explained below. In the supplementary material, we show again marginal level of errors from the extrapolation to complete basis set limit for both BFD and all-electron spectral gaps for the first and second row atoms N and S, respectively.

To generate the ECPs, we followed an energy consistent scheme, as described in reference [Bur07] and references therein, and minimized the differences in the all-electron and ECP excitation energies in a least-squares way, with our objective function defined as

$$f = \sum_s (\Delta E_s^{AE} - \Delta E_s^{PP})^2, \quad (5.6)$$

where s labels a given excited state and ΔE_s is the energy gap between the excited state and the neutral atomic ground state.

We chose to include as many ionizations as possible in the atomic references; from the most deeply ionized single-valence state up to (at least) the M^{th} anionic state where M is the predicted number of bound anions from the all-electron CCSD(T) spectrum. The spectrum here therefore includes not only neutral atomic states, but also numerous states for the anions and cations. In our investigation, we observed that when the most deeply ionized states were not included in the reference, their corresponding excitations tended to have significant discrepancies from the all-electron excitations and consequently the pseudoatom's transferability could be negatively affected. Additionally, for each ionization, we expanded the reference to include the bound ground states of all possible total spin channels of that ion, e.g., for an ion with four electrons in the valence space, we included the ground states of the quintet, triplet and singlet spin channels into the reference, provided that each state was predicted to be bound from the all-electron calculation. The latter choice was motivated by a desire to minimize any possible contamination from energetically lower states within the same symmetry channel.

5.4.2 Objective function with spectral and spatial density matrix discrepancies

In order to incorporate spatial information into the pseudopotential, we explored the utilization of the single-body density matrix following the previous work of Acioli and Ceperley. For the all-electron reference and pseudoatom, we generated the single-body density matrices from a CISD wave function. We expect that if the pseudoatom density matrix and the all-electron density matrix agree beyond a chosen core region defined by cut-off radius r_c , then the pseudoatom valence will mimic that of the all-electron atom. Given an AE atom and a pseudoatom, we can measure the difference between the density matrices outside of r_c as

$$\Delta\rho_{r_c} = \iint_{r,r'>r_c} d\mathbf{r}d\mathbf{r}' |\rho_{AE}(\mathbf{r},\mathbf{r}') - \rho_{ECP}(\mathbf{r},\mathbf{r}')|^2. \quad (5.7)$$

If we express each density matrix through the natural orbitals $\{\phi_i\}$ and occupation numbers n_i of a correlated calculation such as CISD, we can rewrite the discrepancy between the density matrices as

$$\begin{aligned} \Delta\rho_{r_c} = & \sum_{ij}^N n_i^{AE} n_j^{AE} \left| \langle \phi_i^{AE} | \phi_j^{AE} \rangle_{r_c} \right|^2 \\ & + \sum_{i,j}^M n_i^{ECP} n_j^{ECP} \left| \langle \phi_i^{ECP} | \phi_j^{ECP} \rangle_{r_c} \right|^2 \\ & - 2 \sum_i^N \sum_j^M n_i^{AE} n_j^{ECP} \left| \langle \phi_i^{AE} | \phi_j^{ECP} \rangle_{r_c} \right|^2 \end{aligned} \quad (5.8)$$

where the expression $\langle \phi_i | \phi_j \rangle_{r_c}$ is the overlap of two natural orbitals evaluated from r_c to infinity.

An important point worth noting here is that one should not expect to rigorously match the all-electron and pseudoatom density matrices. The reason is that the density matrix of the original atom generally contains a contribution from the tails of the core states beyond the cutoff radius. Since it is difficult to disentangle the tails in the many-body setting this "contamination" will be present. Note that in traditional one-particle formulations this is a non-issue since HF or Kohn-Sham self-consistent orbitals decompose the core and

valence states into distinct sets. The closest analogy to the one-particle decomposition comes from the natural orbitals, however, even there the occupation numbers and the natural orbitals will slightly differ (reflecting thus the presence or the absence of the core states). This becomes particularly relevant whenever smaller core radii are imposed. The fits can be further elaborated to approximately take this into account, however, here we opted for a simple agreement assuming that this effect will not dominate the achieved discrepancies.

We further constructed ECPs where we include both the spectral and spatial information into the objective function. Using the integration of the density matrix beyond a cutoff radius as defined above, we define our spatial objective function as

$$g = \sum_s \Delta\rho_{r_c}^s \quad (5.9)$$

where we sum over all states considered. We construct a new objective function where we include both spatial and spectral information as

$$h = \frac{f}{f_0} + \frac{g}{g_0} \quad (5.10)$$

where f_0 and g_0 are the optimal values of the respective objective functions. In the case of sulfur below, the two objective functions were left unweighted given the similarity between f_0 and g_0 in that case.

5.4.3 Optimization methods

Throughout, we have utilized the nonlinear DONLP2 optimization code of Spellucci [Spe] for generating the final parameter values of our ECPs with respect to the various objective functions which we have considered. The specifics of the method implemented in DONLP2 are outlined in two papers [Spe98b; Spe98a], where generally, the solver extends the sequential quadratic programming method, an iterative method that relies on the second derivatives of both the objective function and the constraints, such that it can be applied to nonlinear problems. We considered this choice for the chosen solver to be appropriate given the nonlinear “smoothness” constraint that has been imposed on the parameter sets of our ECPs.

5.4.4 Constructed, Combined and Iterated Schemes

The final procedure that we pursued took molecular data into account. Having generated molecular binding curve predictions from an assortment of ECP constructions, we noticed that there was a potential to improve the ECPs further. In particular, the optimization could be guided to fulfill additional criteria such as reproducing the dimer binding curve in a few iterations. This type of optimization loop produced very small discrepancies from the all-electron CCSD(T) curves along the desired range of geometries even at the steep repulsive side of the binding curves. In particular, for the corresponding dimers, we found that this strategy is able to produce discrepancies as small as 0.05 eV or lower along the binding curve all the way up to the dissociation limit at the repulsive side. Due to time demands of correlated calculations and rather slow coupling between different codes we guided this level of optimization by interventions to speed up the search. That was also very useful to understand the qualitative relationships between shapes/amplitudes of the ECP versus its properties and therefore we leave the automation of this part for future. We comment on these constructions at each presented atomic case.

In all of the systems presented below, we compare our ECPs to all-electron CCSD(T), all-electron CCSD(T) with an uncorrelated core (UC), i.e., no excitations from the core, and various ECPs often used in QMC calculations, namely Burkatzki-Filippi-Dolg (BFD) [Bur07], Trail-Needs (TN-DF) [TN05] DF ECPs, and where applicable, the shape-consistent correlated Trail-Needs (TN-CEPP) ECPs [TN13] as well as the recent shape and energy consistent correlated Trail-Needs (TN-eCEPP) ECPs [TN17]. Note that when utilizing the TN-CEPP and TN-eCEPP, we only test the effective core potential and do not include the semi-empirical core polarization potentials utilized by Trail and Needs, given that we were interested in comparing all effective cores within a consistent level of approximation. A key question that we chose to pursue was how well an ECP alone would capture core-valence effects without the need for additional adjustments or approximations. By not including CPPs along with the ECPs allowed us to isolate this effect. Note that, Trail and Needs have studied the effect of the CPP in Figure 6 of [TN17], illustrating that CPPs do have only minor impact on the dissociation energies over a wide variety of molecules. As explained in results, we found that much larger discrepancies almost invariably appear at shorter molecular bond lengths and therefore we have focused on addressing this aspect in our constructions.

5.5 Results

Here we present our initial results for a number of explicitly correlated ECPs using the various objective functions described above. In the first row, we present optimized ECPs for Boron, Carbon, Nitrogen, and Oxygen. For the second row elements, we show our results for Sulfur.

5.5.1 Boron

Table 5.1 Atomic and ionic excitations and corresponding discrepancies for Boron. IP denotes the first ionization potential while EA is the electron affinity. Q is the ionization charge, $2S+1$ the usual total spin multiplicity. AE denotes the calculated all-electron values while the rest of columns shows the discrepancies. UC means all-electron valence-only correlation with self-consistent but uncorrelated core, as explained in the text. All energies in eV. The MAD is the mean absolute difference over all of the discrepancies. Note that all gaps are calculated with reference to the ground state, namely $Q=0$ and $2S+1 = 2$. The same notation applies to all the atomic/ionic data tables throughout the paper.

	Q	2S+1	AE Gap	UC	BFD	TN-DF	TN-CEPP	TN-eCEPP	Spectral	Constructed
	+2	2	33.4290	-0.0726	0.0048	-0.0316	0.1186	0.0224	0.0027	0.0109
(IP)+1	1	1	8.2771	-0.0379	-0.0670	-0.0711	-0.0030	-0.0280	-0.0077	-0.0301
+1	3	3	12.9083	-0.0085	0.0249	0.0121	0.0326	0.0032	-0.0042	-0.0129
0	4	4	3.5752	0.0187	0.0202	0.0523	0.0184	0.0107	0.0018	-0.0079
-1	1	1	0.2854	-0.0026	0.0067	0.0092	0.0079	0.0046	-0.0034	0.0031
(-EA)-1	3	3	-0.2482	0.0043	0.0142	0.0220	0.0144	0.0117	0.0006	0.0098
-1	5	5	2.5144	0.0207	0.0183	0.0683	0.0301	0.0193	-0.0004	-0.0027
MAD				0.0236	0.0223	0.0381	0.0321	0.0143	0.0030	0.0110

Boron is perhaps the lightest element where time savings from an appropriate ECP become significant while the valence space is sufficiently large to represent the atom in chemical settings. Table 5.1 shows several constructions of the ECP obtained by minimizing the atomic spectral error only (Spectral) and another that has been adjusted to reproduce the binding curve of the ground state of B_2 in the range of bond lengths (Constructed). Note the very high accuracy that was obtained with minimizing only the spectral discrepancy. A significantly improved dimer solution has been found by minimizing a compromise between reproducing the spectrum and the binding curve. Note that the shortest bond length

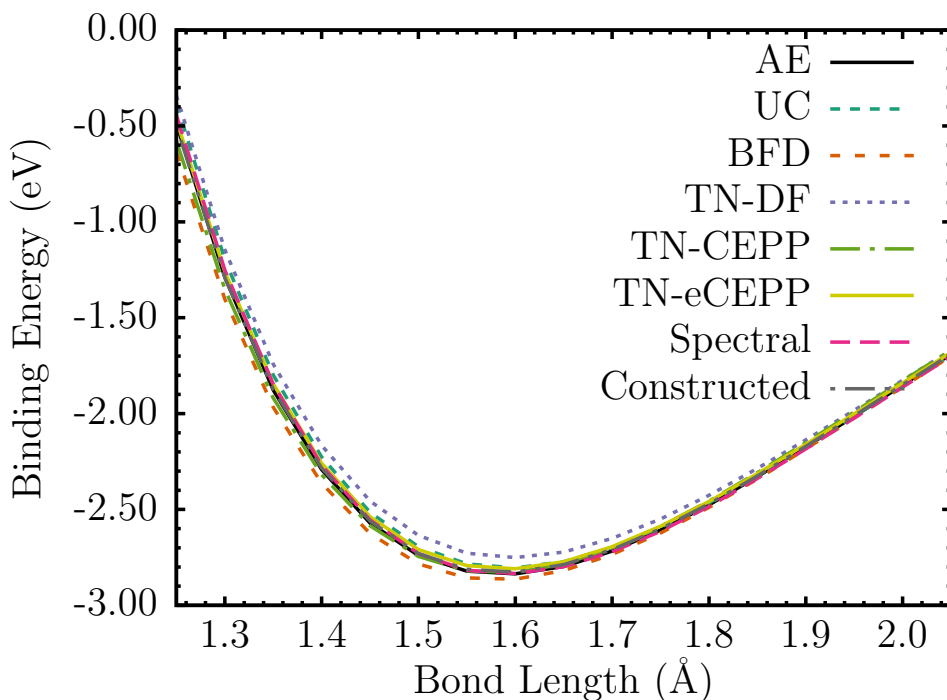


Figure 5.1 Boron dimer potential energy surface. UC represents an all-electron CCSD(T) calculation with a self-consistent but *uncorrelated* core, i.e., with no excitations from the core states. Spectral represents the optimization for the atomic spectrum alone, and Constructed indicates the ECP driven iteratively to minimize the dimer discrepancy while accepting a small increase in the spectrum discrepancy.

corresponds approximately to the dissociation point due to nucleus-nucleus repulsion. In solids, such distance between the atoms would correspond to very high pressures roughly beyond 500 GPa. Figure 5.1 shows the B_2 potential energy surface for the $^3\Sigma_g$ state and 5.2 provides a set of discrepancies from the all-electron CCSD(T) for our constructed ECPs compared against previous ECPs: BFD, TN-DF, the recently correlated constructions by Trail and Needs (TN-CEPP and TN-eCEPP) as well as the all-electron uncorrelated core result.

Right away we can observe the following:

a) The constructed ECPs provide a significantly more accurate picture of the molecular system, both for the valence spectrum optimization (Spectral) as well as more accurate binding curve optimization (Constructed), than the existing examples. Remarkably, our ECPs show *smaller errors* for both the spectrum and for binding curve than the all-electron

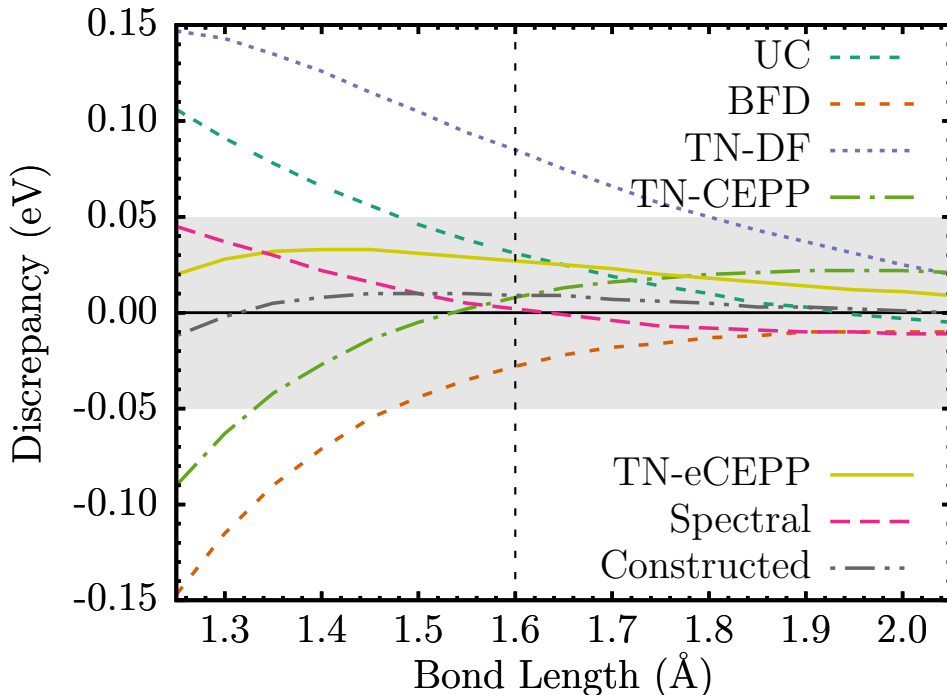


Figure 5.2 Boron dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve. The gray envelope represents a 0.05 eV window for the discrepancy. The vertical line indicates the equilibrium bond length as predicted by the all-electron CCSD(T) calculation.

Table 5.2 ECP parameters for Constructed Boron. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.

Channel	n_{lk}	α_{lk}	β_{lk}
p	1	31.49298	3.00000
	3	22.56509	94.47895
	2	8.64669	-9.74800
$s-p$	2	4.06246	20.74800

correlated valence-valence calculation with an *uncorrelated core* (UC). In UC all one-particle states are solved self-consistently including the cores (i.e., it is not the atomic frozen core in that sense), however, only valence-valence correlations are invoked in the CCSD(T) method while core-valence and core-core correlations are neglected. The fact that our effective Hamiltonians provide a better description than UC for the dimer and atomic spectra is quite surprising. We, therefore, observe that the inverse problem formulation enables to

mimic, within a certain level of accuracy, effects of core excitations on the valence space.

b) The simple form of the ECP described above is able to accommodate higher accuracy demands.

Note that overall accuracy in reproducing correlated properties is roughly at the level of 0.02 eV, significantly higher than in previous constructions. Both of these observations are quite unexpected and will be further elaborated in the cases of other elements. Since these solutions were satisfactory we have not pursued further improvements by including spatial correlations explicitly into the objective function. Note that the spectrum and refinement alone were sufficient to determine the high accuracy solution within the minimal and significantly restricted representation. Our best ECP parameters (Constructed) are given in Table 5.2.

Labeling. For the sake of clear identification and recording of progress we introduce here the label for this and subsequent ECPs as “correlation consistent ECP”, version 0.1, or **ccECP0.1**, in short. The same label is assigned to all tabulated ECPs in this paper.

Note that the notation can be further expanded, for example, as ttECP.xx.yy.ENSOCO. Here the theory would be labeled as tt=cc,hf,df,..., while major.minor release is represented as above by xx.yy, respectively. In addition, fitted quantities could include ENSOCO=excitations-norms/shapes-spatialDM-combined/iterated-other approaches. As far as systems used for fit is concerned one could possibly affix a label ADHOCCG=atom-dimer-hydride-oxide-cluster-general/other system. As an example, the full label for our Boron case would be ccECP0.1.EC.

5.5.2 Carbon

Since the spectral only optimization worked quite well for B, only needing a slight readjustment with our constructed ECP, we test the same for the C atom. The atomic spectrum for various ECPs is given in Table 5.3 and dimer discrepancies in the $^1\Sigma_g$ state are given in Figure 5.3. The spectral only construction does remarkably well on minimizing the energy differences for the atomic spectra and produces an ECP that is improved over other ECPs. While there is slight overbinding, the Spectral curve is relatively flat indicating accurate vibrational frequencies compared to the all-electron. The electron affinity (EA) for the spectral differs with the all electron CCSD(T) value by only -1 meV. The ionization potential (IP) is an order of magnitude better and only differs by $+0.1$ meV. In addition, a minor compromise in reproducing the atomic spectra produces a much better dimer binding that

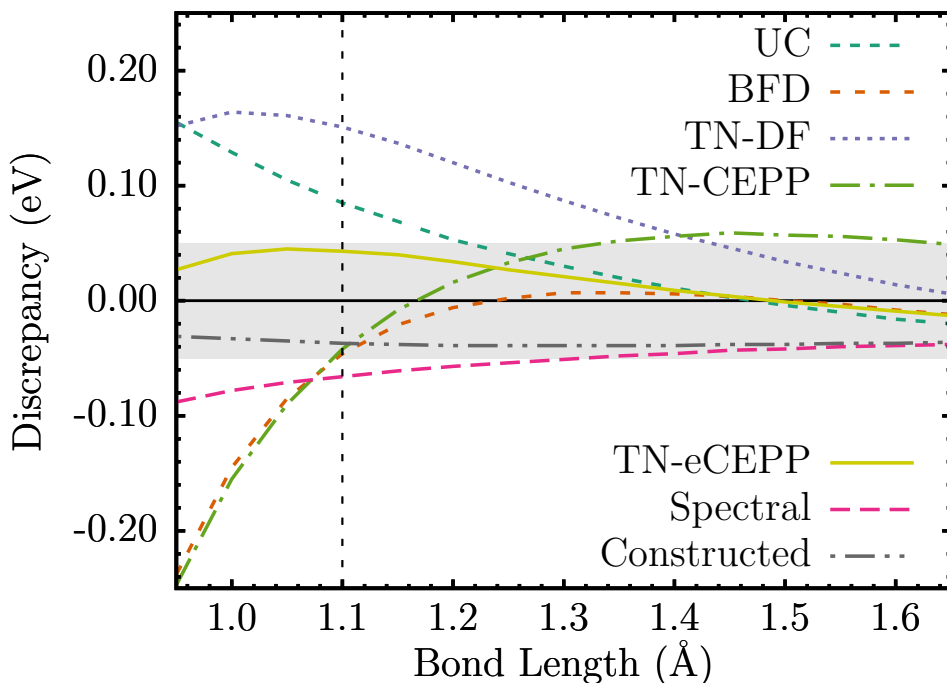


Figure 5.3 Carbon dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.

is within 0.05 eV across the entire curve and has a flat discrepancy, i.e., the same vibrational frequencies as the all-electron curve.

Table 5.3 Atomic data for Carbon, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely $Q=0$ and $2S+1 = 3$.

Q	2S+1	AE Gap	UC	BFD	TN-DF	TN-CEPP	TN-eCEPP	Spectral	Constructed
+3	2	83.4895	-0.1469	-0.1090	-0.1561	-0.0824	0.2544	0.0005	-0.0024
+2	1	35.6041	-0.1020	-0.2208	-0.1723	-0.0953	0.0326	-0.0007	0.0110
+2	3	42.1035	-0.0561	-0.1083	-0.1080	-0.0368	0.0751	-0.0009	-0.0061
(IP)+1	2	11.2452	-0.0334	-0.0725	-0.0631	-0.0277	-0.0073	0.0016	0.0027
+1	4	16.5590	-0.0022	-0.0955	-0.0552	-0.0173	0.0075	0.0001	0.0019
0	1	1.3950	-0.0143	0.0013	0.0000	0.0045	0.0095	0.0006	-0.0009
0	5	4.1491	0.0231	-0.0743	-0.0085	-0.0030	0.0126	0.0005	0.0084
(-EA)-1	4	-1.2421	0.0072	0.0259	0.0249	0.0098	0.0270	-0.0010	-0.0006
MAD			0.0481	0.0884	0.0735	0.0532	0.0346	0.0008	0.0046

The parameters for our best Carbon ECP (Constructed) is given in Table 5.4.

Table 5.4 ECP parameters for Constructed Carbon. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.

Channel	n_{lk}	α_{lk}	β_{lk}
p	1	14.43502	4.00000
	3	8.39889	57.74008
	2	7.38188	-25.81955
$s-p$	2	7.76079	52.13345

5.5.3 Nitrogen

Table 5.5 Atomic data for Nitrogen, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely Q=0 and 2S+1 = 4.

Q	2S+1	AE Gap	UC	BFD	TN	TN-CEPP	TN-eCEPP	Spectral	Spatial	Spec/Space	Constructed
+4	2	169.0094	-0.2308	-0.2139	-0.3301	0.5136	-0.3450	0.0024	-0.2259	0.0022	-0.2758
+3	1	91.5371	-0.1784	-0.3903	-0.2882	0.2012	-0.1921	-0.0043	-0.2368	-0.0046	-0.0494
+3	3	99.8869	-0.1205	-0.2582	-0.2868	0.2291	-0.2538	-0.0011	-0.0157	-0.0009	-0.2531
+2	2	44.1212	-0.0866	-0.1827	-0.1429	0.0792	-0.0862	0.0018	-0.0635	0.0017	0.0014
+2	4	51.1908	-0.0455	-0.2695	-0.2339	0.0645	-0.1732	-0.0023	0.0548	-0.0019	-0.2317
+1	1	16.5790	-0.0448	-0.0472	-0.0370	0.0380	-0.0168	0.0105	0.0028	0.0105	0.0232
(IP)+1	3	14.5319	-0.0297	-0.0646	-0.0547	0.0125	-0.0291	-0.0013	-0.0070	-0.0013	0.0100
+1	5	20.5319	0.0009	-0.2524	-0.1729	-0.0041	-0.1077	0.0011	0.0559	0.0015	-0.2154
0	2	2.6789	-0.0295	-0.0037	-0.0056	0.0046	-0.0104	-0.0084	-0.0080	-0.0085	-0.0068
MAD			0.0852	0.1870	0.1725	0.1274	0.1349	0.0037	0.0745	0.0037	0.1185

In constructions of nitrogen, we explored several objective functions and refinement strategies, as shown in Figure 5.4 for the molecule in the $^1\Sigma_g$ state and Table 5.5 for the atomic properties. The spectrum only optimization performs rather well and is a significant improvement over many of the previous constructions. The results show that using the spatial information only such as one-particle density matrices appears less favorable overall. We conjecture that this is caused due to very weak correlation “signal” in the corresponding density matrices. The difference between the mean-field and correlated density matrices

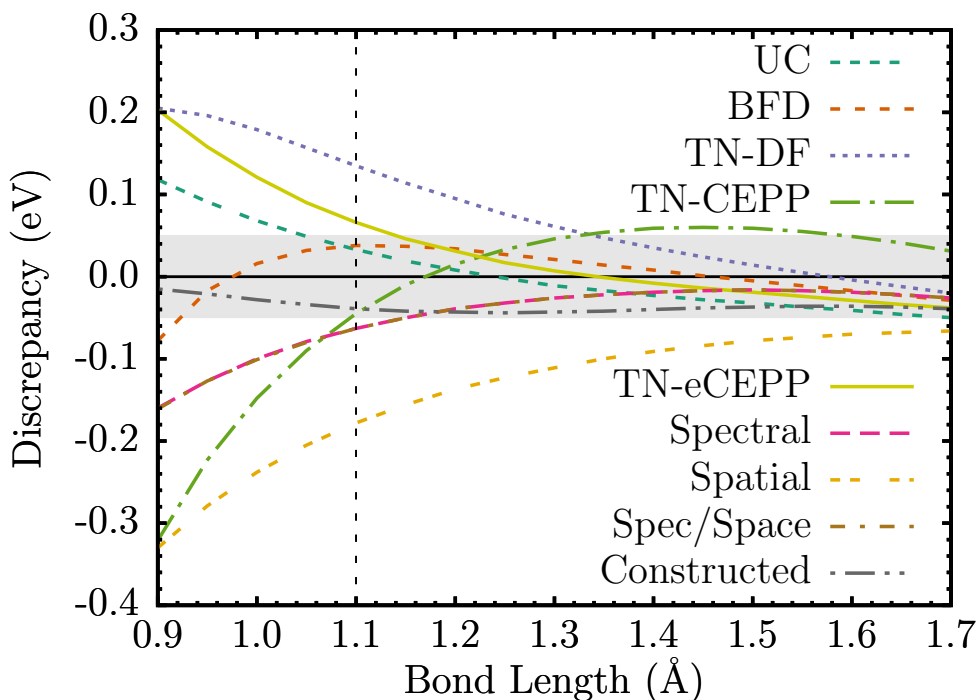


Figure 5.4 Nitrogen dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.

is very tiny so that the correlation effect is overwhelmed by the one-particle character that is dominant. On the other hand, the spectrum alone provides a much stronger signal since the eigenvalues basically determine the wave function tails essentially exactly. When combining the spatial and spectral information, the spatial signal was not sufficient to significantly change the parameters, and we obtain essentially the same ECP for both constructions. Although the spectral ECP has excellent atomic properties and a reasonable dimer, we sought to construct a better ECP overall. The constructed ECP was generated by adding additional constraints that were not included in the original objective function in order to alter systematic trends we observed in our various constructions to minimize the dimer discrepancy. In particular, we observed that pairs of ECPs with dimer discrepancies of opposite trends could be produced based on different constraints and subsequently the ECPs could be combined as a linear sum leading to increased agreement with the molecular properties from all-electron CCSD(T) while also preserving a high level of accuracy on the atomic spectrum. This procedure could certainly be used to define a new objective function where the dimer fit is directly included, however, this proved to be computationally

inefficient and we therefore chose to improve the molecular properties by hand as described. The constructed ECP has a discrepancy of below 0.05 eV across the entire curve, although we compromise the spectrum when compared to our spectral construction, particularly for high-spin atomic states. However, if the high-spin atomic states are not considered, the atomic spectrum is better than all previous generations of ECPs. Note that in molecular systems the nitrogen almost invariably appears in low/lowest spin configurations.

Another observation is that ECPs with a combination of minimizing the dimer binding and spectrum discrepancies could be constructed without generating an overly large negative impact to the atomic spectra, provided an appropriate constraint is utilized. Though some accuracy of the atomic properties would diminish as the dimer binding is improved, as can be seen in Table 5.5, we found that a reasonable balance could be obtained between the two when compared to other ECP constructions.

The parameters for our best Nitrogen ECP (Constructed) is given in Table 5.6.

Table 5.6 ECP parameters for Constructed Nitrogen. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.

Channel	n_{lk}	α_{lk}	β_{lk}
p	1	12.91881	3.25000
	1	9.22825	1.75000
	3	12.96581	41.98612
	3	8.05477	16.14945
	2	12.54876	-26.09522
	2	7.53360	-10.32626
$s-p$	2	9.41609	34.77692
	2	8.16694	15.20330

5.5.4 Oxygen

As in the previous case, we test a variety of objective functions when constructing an ECP for O. We present the results in Figure 5.5 for the molecule in the ${}^3\Sigma_g$ state and in Table 5.7 for the atomic properties. Using a spectral-only construction, we are able to construct an ECP that provides a binding curve discrepancy to the all-electron CCSD(T) curve within

0.05 eV along the entire curve. Since the Spectral ECPs are reasonably flat compared to other ECP constructions and the all-electron uncorrelated core (UC), the vibrational frequencies agree quite well to the all-electron CCSD(T) calculations. As in N, the spatial information alone appears insufficient in generating a high-quality ECP. When combining both spatial and spectral information (Spec/Space), we actually obtain an ECP that is essentially flat across a wide range of bond lengths and shows the best dissociation energy. This also results in a slight compromise on the atomic properties. However, this ECP begins to overbind significantly in the shorter bond length regime which would correspond to high pressure in solids. Based on this and the quality of our spectral only ECP, we believe that the spectral only optimization for Oxygen produces the best ECP overall and we do not pursue iterated constructions. The parameters for Spectral Oxygen are given in Table 5.8

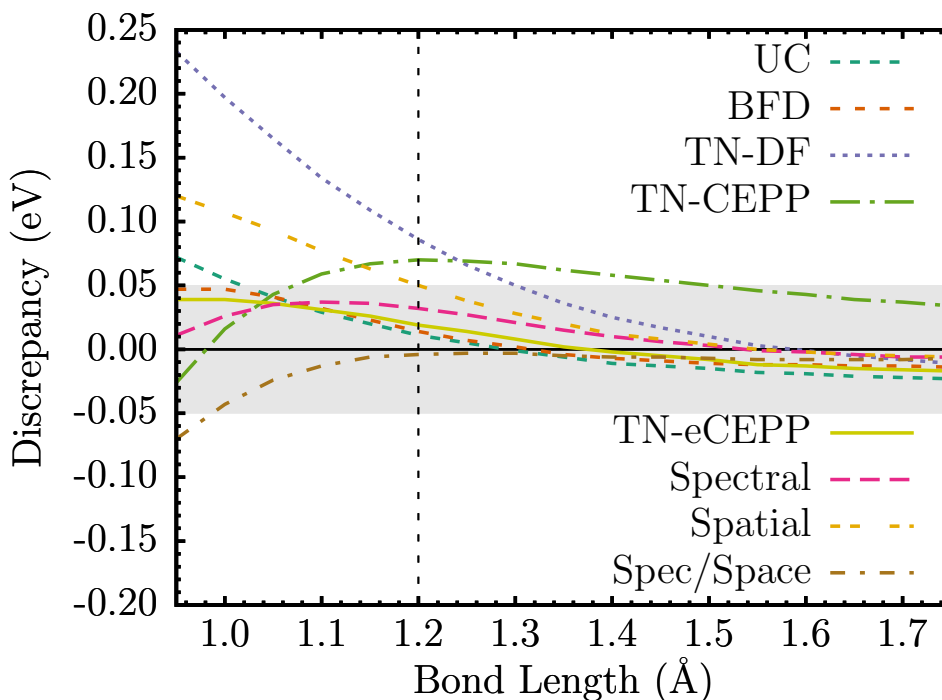


Figure 5.5 Oxygen dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.

Table 5.7 Atomic data for Oxygen, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely Q=0 and 2S+1 = 3.

Q	2S+1	AE Gap	UC	BFD	TN-DF	TN-CEPP	TN-eCEPP	Spectral	Spatial	Spec/Space
+5	2	294.8903	-0.3059	-0.4334	-0.6186	0.7414	-0.5643	-0.0021	-0.4531	-0.0228
+4	1	180.9900	-0.2478	-0.6864	-0.4477	0.3505	-0.1826	0.0017	-0.2573	-0.0328
+4	3	191.1861	-0.1808	-0.4825	-0.5444	0.3467	-0.4327	-0.0037	-0.2651	0.0252
+3	2	103.6268	-0.1383	-0.3375	-0.2391	0.1617	-0.0780	0.0060	-0.0801	0.0282
+3	4	112.4585	-0.0894	-0.5304	-0.4781	0.0894	-0.3264	-0.0021	-0.2007	0.0209
+2	1	51.3705	-0.0760	-0.1056	-0.0811	0.0815	-0.0026	0.0200	0.0153	0.0541
+2	3	48.7002	-0.0627	-0.1445	-0.1185	0.0349	-0.0297	-0.0035	-0.0223	0.0264
+2	5	56.1201	-0.0257	-0.5550	-0.4073	-0.0454	-0.2398	0.0061	-0.1885	-0.0026
+1	2	17.1953	-0.0621	-0.0367	-0.0448	-0.0097	-0.0183	-0.0206	-0.0102	-0.0005
(IP)+1	4	13.5653	-0.0142	-0.0438	-0.0436	-0.0192	-0.0053	-0.0058	-0.0118	0.0083
+1	6	43.9706	-0.0290	-0.7264	-0.4557	-0.1001	-0.2281	-0.0075	-0.2131	-0.0299
0	1	2.1816	-0.0210	0.0057	0.0007	0.0069	-0.0052	-0.0055	0.0039	-0.0017
0	5	9.4860	-0.0182	-0.1259	-0.0750	-0.0545	-0.0110	-0.0250	-0.0350	-0.0164
0	7	39.4276	-0.0247	-0.6879	-0.4390	-0.0948	-0.2207	0.0025	-0.1962	-0.0152
(-EA)-1	2	-1.4209	0.0017	0.0105	0.0092	0.0259	-0.0083	0.0044	-0.0012	-0.0036
MAD			0.0865	0.3275	0.2669	0.1442	0.1434	0.0078	0.1303	0.0192

Table 5.8 ECP parameters for Spectral Oxygen. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.

Channel	n_{lk}	α_{lk}	β_{lk}
p	1	12.30997	6.00000
	3	14.76962	73.85984
	2	13.71419	-47.87600
$s-p$	2	13.65512	85.86406

5.5.5 Sulfur

As a last illustration, we include our progress from the second-row atom, Sulfur. Construction of the ECP for sulfur has followed similar steps as in previous cases. In Fig.5.6 we show the impact on the S₂ dimer its ground state ($^3\Sigma_g^{(-)}$) of leaving the core uncorrelated on the overall accuracy and we see that near equilibrium its agreement with the all-electron prediction is at the level of ≈ 0.03 eV.

For this case, we utilized a Ne core ECP and attempted a set of optimization strategies similar to the investigation of N and O. In this case, we again consider constructing ECPs to mimic all-electron many-body spectral, spatial and combined properties, in turn. Figures 5.7 and 5.8 show CCSD(T) binding energy discrepancies from the all-electron S₂ molecule ($^3\Sigma_g^{(-)}$) for various approximations to the sulfur cores including our generated ECPs.

Table 5.9 Atomic data for Sulfur, similar to Table 5.1. Energies in eV. Note that all gaps are calculated with reference to the ground state, namely $Q=0$ and $2S+1 = 3$.

Q	2S+1	AE Gap	UC	BFD	TN-DF	Spectral	Spatial	Spec/Space	Constructed
+5	2	188.3560	-0.4975	-0.3201	-0.6243	-0.0469	-0.3620	-0.9587	-0.5141
+4	1	115.7750	-0.2143	-0.0922	-0.1412	0.0123	0.0693	-0.4847	-0.1838
+4	3	126.1610	-0.3461	-0.1881	-0.3452	0.0580	0.2173	-0.3349	-0.2871
+3	2	68.5115	-0.1152	-0.0295	-0.0381	0.0290	0.2974	-0.1429	-0.0926
+3	4	77.3129	-0.2375	-0.1931	-0.2359	0.0318	0.3573	-0.0642	-0.1816
+2	1	35.1649	-0.1012	-0.0399	-0.0338	-0.0130	0.2559	-0.0326	-0.0717
+2	3	33.6900	-0.0515	-0.0332	-0.0177	-0.0076	0.2565	-0.0246	-0.0582
+2	5	40.8647	-0.1675	-0.2658	-0.2325	-0.0577	0.2273	-0.0197	-0.1568
+1	2	12.4609	-0.0920	-0.0732	-0.0636	-0.0673	0.0794	-0.0484	-0.0830
(IP)+1	4	10.2976	-0.0159	-0.0581	-0.0377	-0.0536	0.0826	-0.0320	-0.0559
+1	6	30.7228	-0.1557	-0.2981	-0.3143	-0.0025	0.2499	0.1068	-0.0134
+0	1	1.3319	-0.0399	-0.0029	-0.0084	-0.0015	0.0090	-0.0011	-0.0094
+0	5	8.9838	-0.0223	-0.0563	-0.0648	-0.0244	0.0749	-0.0076	-0.0136
+0	7	27.8996	-0.1643	-0.3088	-0.3504	0.0094	0.2134	0.1109	0.0170
(-EA)-1	2	-2.0494	-0.0018	0.0277	0.0064	0.0302	-0.0741	-0.0124	0.0149
MAD			0.1482	0.1325	0.1676	0.0297	0.1884	0.1588	0.1169

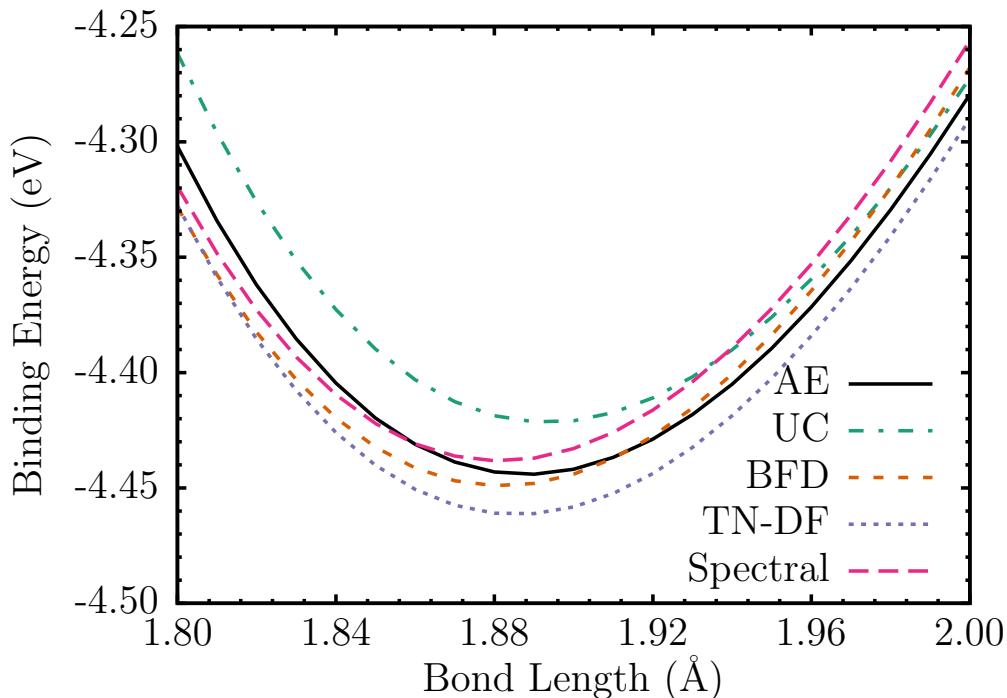


Figure 5.6 Potential energy surfaces of the S_2 molecule from CCSD(T). We have plotted the predictions from various treatments of the sulfur cores. Shown are the all-electron core (AE), all-electron uncorrelated core (UC), Burkatzki-Filippi-Dolg (BFD), Dirac-Fock Trail-Needs (TN) and the CCSD(T) spectrum matched (Spectral) ECPs described in the text.

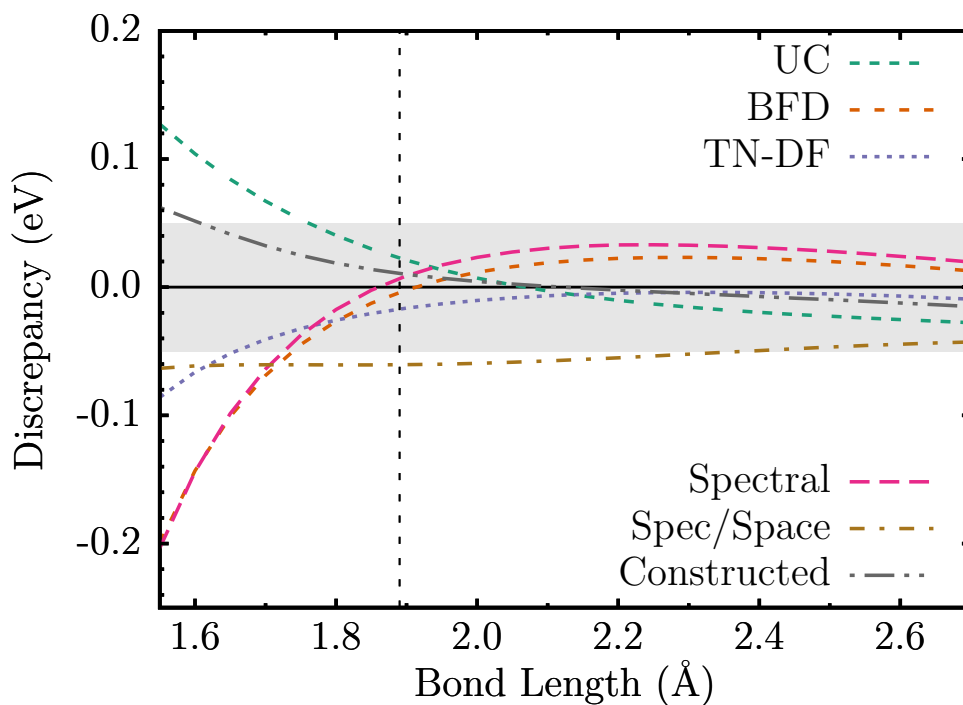


Figure 5.7 Sulfur dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.

For the spectral case, we again have used an all-electron reference formed from bound excitations and included all possible valence ground states of total spin and charge whereby we minimized the spectral discrepancies exclusively with respect to ECP’s parameters. In Table 5.9, it is shown that the MAD from the all-electron excitation energies that is no more than 0.03 eV in this case. This agreement is nearly an order of magnitude improvement over the other approximations to the core. Additionally, with the spectral objective function, we observe that the agreement with the all-electron binding energy at equilibrium is very good with an error of no more than 0.01 eV. For shorter separations of the dimer, however, the spectral ECP undershoots the all-electron binding energy by tenths of eV and as a result of this large change from equilibrium, we see a non-negligible slope and curvature in the discrepancy at equilibrium which negatively impacts the agreement with the all-electron prediction of the ground state vibrational frequency.

Using the spatial information for the same set of states, we generated the single-body density matrices from all-electron CISD wave functions and subsequently imposed that the single-body density matrix of the pseudoatom’s CISD wave function match beyond

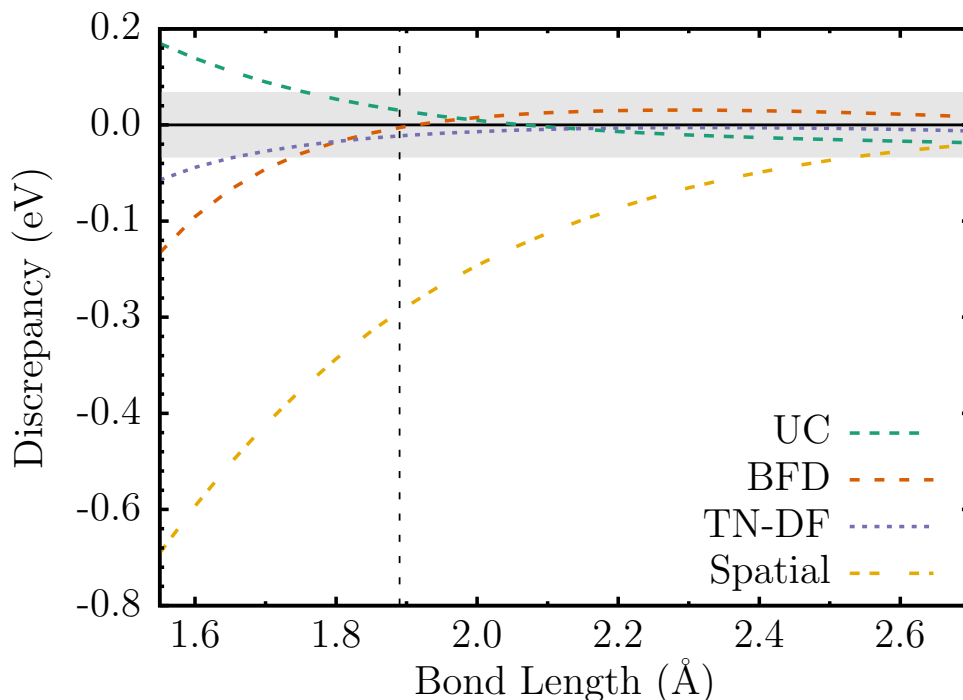


Figure 5.8 Sulfur dimer binding energy discrepancies compared to the all-electron CCSD(T) binding curve.

a core radius. The resulting ECPs constructed from this procedure were not adequately transferable which we attribute again to marginal electronic correlations signal in the natural orbitals as we argued for the cases of N and O. It can also be observed from Table 5.9 that the atomic properties are generally negatively affected by optimizing spatial data alone.

For sulfur, we also attempted to match both the all-electron excitation energies and single-body density matrices, simultaneously. For this case, the binding energy discrepancies from the all-electron atom are shown in Fig. 5.7. Here we see that the error is quite uniform over the entire region plotted, and moreover, it is no more than about 0.06 eV.

Lastly, we again considered additional molecular constraints on the objective function in order to further improve the ECP's dimer properties as described for the nitrogen atom. The constraints were constructed in such a way as to reach a balance between the spectral and molecular properties. We show its discrepancies in Fig. 5.7. For this ECP, we see that the dimer's error is mostly within 0.05 eV throughout the plotted region and the spectral properties are an improvement over both UC and previously generated ECPs. The parameters

for sulfur’s constructed ECP are shared in Table 5.10.

Table 5.10 ECP parameters for Constructed ECP for Sulfur. The parametrization for each channel is given by $V_l(r) = \sum_k \beta_{lk} r^{n_{lk}-2} e^{-\alpha_{lk} r^2}$. The corresponding correlation consistent basis sets are included in the Supplementary Material.

Channel	n_{lk}	α_{lk}	β_{lk}
<i>d</i>	1	4.23812	3.06000
	1	2.19773	2.94000
	3	1.71348	12.96866
	3	10.20072	6.46132
	2	3.41487	-10.45671
	2	1.40439	-9.79751
<i>s - d</i>	2	3.91958	23.19840
	2	3.91388	22.28866
<i>p - d</i>	2	2.71232	8.39601
	2	3.20078	11.15610

5.6 Transferability

One of the desired properties for our ECPs is transferability, i.e., high-quality performance in systems that were not used in our optimization procedure. In the case of our spectral ECPs, the optimization of the parameters only involved atomic properties; for these ECPs, the dimer discrepancies compared to the all-electron results illustrate the transferability of our ECPs to a degree. However, for our constructed ECPs, the parameters were tuned in order to produce improved dimer properties while not sacrificing the atomic properties and thus the transferability should be verified also on independent examples. Therefore, it is desirable to test those cases in other bonding environments in order to illustrate their transferability beyond the dimers. For this purpose, we have calculated potential energy surface discrepancies for hydrides, oxides and a handful of additional molecules BH_3 , BN , BS and CN and in order to ascertain their level of transferability. All molecular calculations were performed with either the MOLPRO quantum chemistry package [Wer12] or the GAUSSIAN09 code [Fri].

We show binding energy discrepancies with respect to all-electron CCSD(T) for the NH ,

OH, NO, SH and SO molecules in Figs. 5.9,5.10,5.11,5.12 and 5.13, respectively. Furthermore, we summarize the discrepancies in the binding parameters, D_e , r_e , and ω_e of all molecules considered in this work with respect to all-electron CCSD(T) values in Fig. 5.14, where D_e is the dissociation energy, r_e is the equilibrium bond length, and ω_e is the vibrational frequency. For each case, the parameters and their errors were obtained from fitting the potential energy surface to the Morse potential near equilibrium, where the potential is given as

$$V(r) = D_e(e^{-2a(r-r_e)} - 2e^{-a(r-r_e)}) \quad (5.11)$$

where a is related to the vibrational frequency by

$$\omega_e = \sqrt{\frac{2a^2 D_e}{\mu}} \quad (5.12)$$

and μ is the reduced mass of the molecule.

To make the comparison of the pseudoatom transferabilities easier, we also share the MADs of these parameters, both at equilibrium and at the dissociation limit at short bond lengths corresponding to high pressures, in Table 6.11. From the table, it is shown that our ECPs perform the best overall, where the mean-absolute deviation from the all-electron dissociation energy is smaller than all considered core approximations; furthermore, this improvement and better overall balance is achieved with very limited variational freedom as given by the choice of the ECP form. The only exception is a marginally better MAD for the vibrational frequencies of TN-CEPP with 5(3) versus ours 7(2) cm^{-1} .

Table 5.11 Mean absolute deviations of discrepancies of binding parameters at equilibrium (D_e , r_e and ω_e) and near the dissociation threshold (D_{diss}) at short bond lengths for our ECPs and previous constructions with respect to all-electron CCSD(T) calculations. The system sets correspond to Fig.5.14 except for the BH_3 molecule which was omitted from the MADs of the dissociation threshold energy.

	UC	BFD	TN-DF	TN-CEPP	TN-eCEPP	This Work
D_e (eV/ 10^2)	2.9(1)	2.3(1)	7.8(1)	1.9(2)	2.7(2)	1.8(1)
r_e (Å/ 10^3)	2.9(2)	2.9(2)	3.6(2)	3.2(3)	1.1(2)	1.0(2)
ω_e (cm^{-1})	9(2)	12(2)	23(2)	5(3)	12(3)	7(2)
D_{diss} (eV/ 10^2)	11.94	23.76	9.78	22.64	7.10	6.13

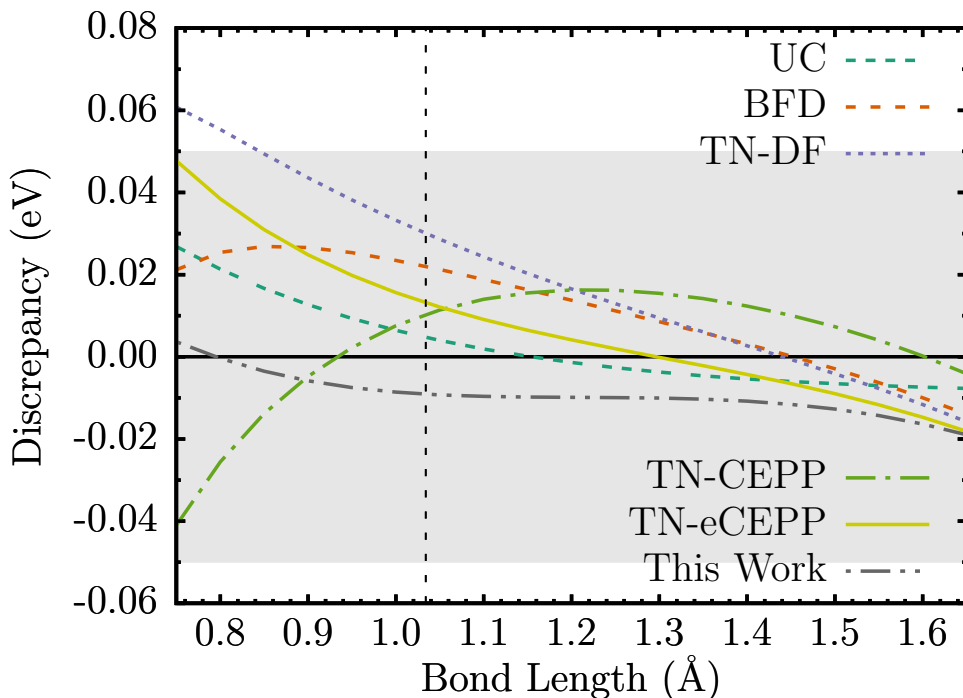


Figure 5.9 NH binding energy discrepancies for various ECPs

5.7 Conclusions

Our paper presents an advancement in the construction of effective core potentials for accurate, correlated valence-only calculations. A key difference from previous constructions being the consistent use of nearly exact many-body approaches to build the ECPs and balancing this with refinement by molecular data that improves the transferability. We introduce isospectrality of all-electron and ECP Hamiltonians on the subspace of valence states as a foundation in formulating the objective functions. Additional criteria that were explored included many-body spatial information from density matrices. This was followed by analyzing the accuracy on dimers and if significant differences were observed we included the dimer information to boost the accuracy and transferability. This was done in an iterative manner and combined constructions were used for efficiency reasons. We were especially careful in obtaining consistent binding curves within a large portion of the bond, i.e., from separations near the dissociation limit to at least equilibrium. We further ascertained the transferability in those cases on hydride, oxide molecules and selected

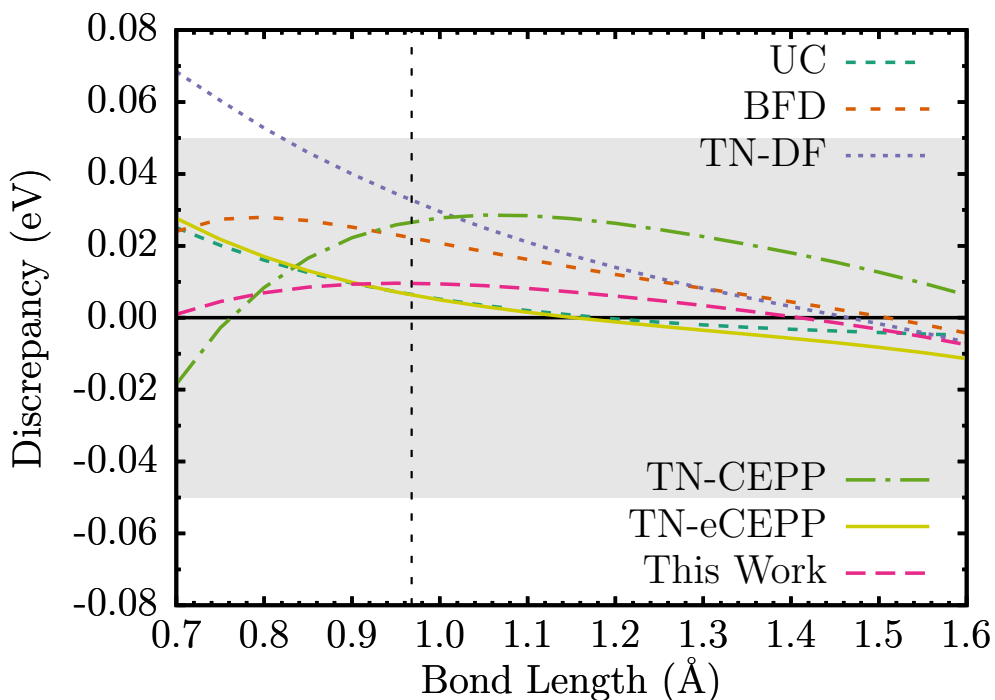


Figure 5.10 OH binding energy discrepancies for various ECPs. For Oxygen, we use our spectral ECP.

other molecules that involved the considered element.

The calculations are done at the relativistic level, therefore, energy differences are very close to the actual experimental values for the relevant quantities.

We were able to decrease the atomic spectral errors for B, C, N, O, and S atoms to significantly smaller values when compared to previous constructions.

Interestingly, the spectral discrepancies are also *smaller* than in uncorrelated core calculations. Therefore the constructed operators are able to effectively accommodate at least some of the effects from core-core and core-valence correlations as it is clear from comparisons with uncorrelated core calculations. Similar argument applies to the improved description of simple molecular systems for a range of bond lengths including very short bonds that are relevant for calculations of materials at high pressures. Rather surprisingly, this implies that for a range of valence properties these ECPs will provide the same or *even more accurate* results than all-electron treatments with uncorrelated cores. Note the caveat “for a range of valence properties” in the previous statement since this is true only for states/properties that are not too far from the one used in optimization. Clearly, inclusion

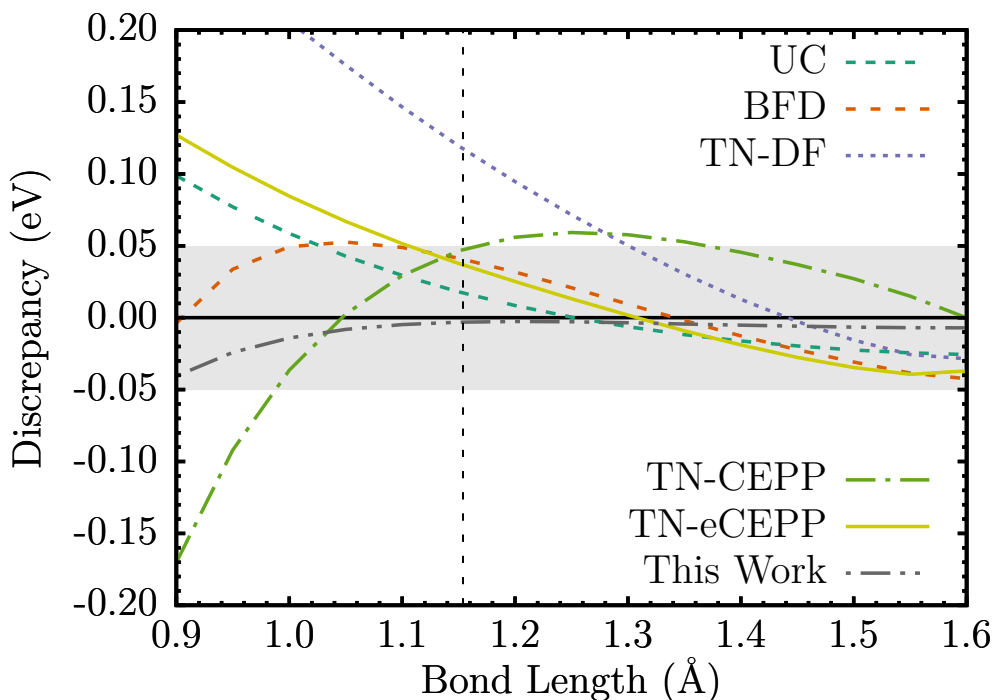


Figure 5.11 NO binding energy discrepancies for various ECPs. For Nitrogen, we use our constructed ECP and for Oxygen, we use our spectral ECP.

of larger range of energies and states that can be described would require either additional terms in the effective ECP Hamiltonian or, in general, including more subshells into the valence space (as is common for heavier elements).

Our experience from these constructions shows:

a) the accuracy of even the simplest forms of semi-local ECPs can be boosted very significantly by appropriate many-body constructions;

b) optimization is quite involved since minor changes in the inputs and the objective function (e.g., increasing the size of the basis set, boosting the spectral accuracy from CCSD to CCSD(T), etc) complicate the optimizations with many local minima solutions, i.e., we see a clear hallmark of an ill-conditioned problem;

c) there are differences from element to element influenced by the extent and choice of excitations, occupations and other details showing thus further complications from the ill-conditioned character of the task.

The obtained results and experience offers a route for constructing comparable quality ECPs for more involved elements such as $3d$ transition series and beyond. Some adjust-

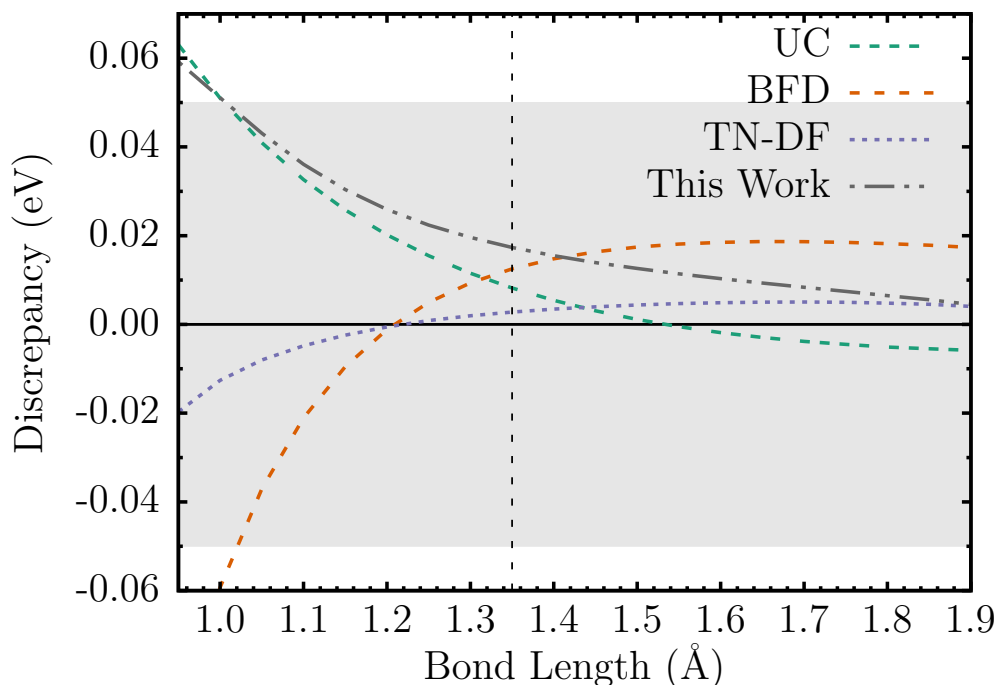


Figure 5.12 SH binding energy discrepancies compared to the all-electron CCSD(T) binding curve. For Sulfur, we use our constructed ECP and for Oxygen, we use our spectral ECP.

ments in the constructions and expanding the parameter sets can be expected. Overall prospects for heavier elements would be limited basically only by the availability of codes that can do accurate correlated all-electron relativistic calculations for atomic and small molecular systems.

In this work we have pursued the construction of ECPs from a many-body framework as opposed to one-particle schemes. In addition, we have included further considerations into the construction such as molecular systems at and away of the equilibrium for better overall transferability.

One-particle constructions have generally involved norm/shape and charge conservations, one-particle eigenvalues, logarithmic derivatives, differences between approximate total energies, etc. Unfortunately, all of these were subject to the biases of the (approximate) method used to solve the underlying atomic problem. The expectation and hope were that in subsequent calculations the systematic cancellations of the underlying biases would be the same as in all-electron setting. In many cases, this worked reasonably well. However, this was not true systematically and for some elements, states and bonding environments

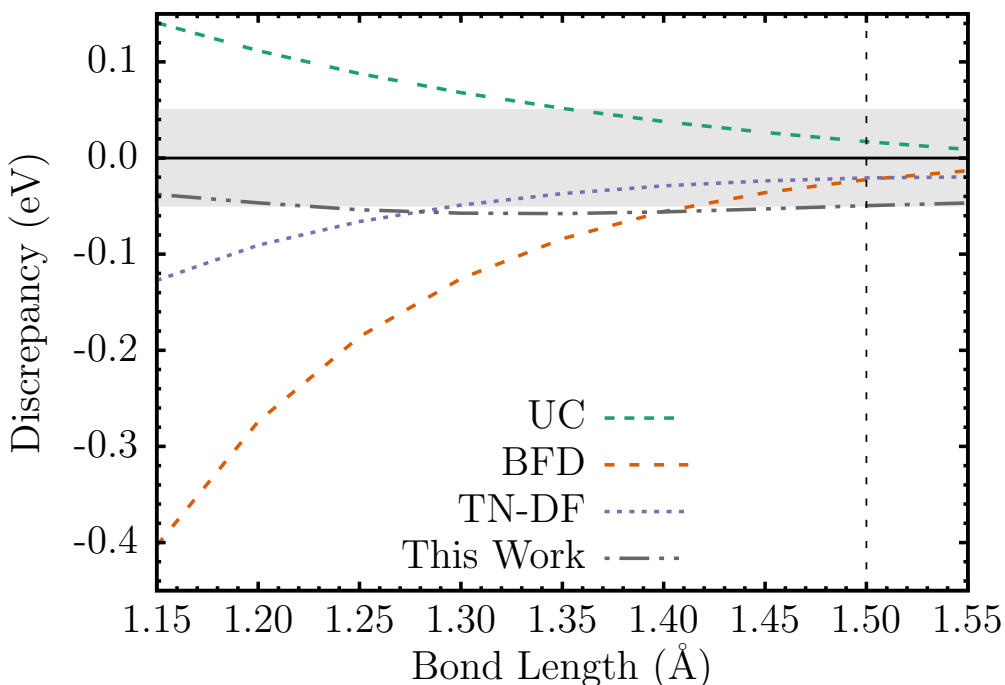


Figure 5.13 SO binding energy discrepancies compared to the all-electron CCSD(T) binding curve. For Sulfur, we use our constructed ECP.

the errors are significant. For example, in DFT all-electron and ECP calculations for just a single atom could lead to large differences. In addition, special adjustments might be necessary such as the nonlinear core corrections for transition metal atoms [Lou82], etc. These complications, however, defy the universal use of such ECPs in other methods, what is one of our stated goals. In addition, there has always been a lingering (and valid) question why an ECP constructed in DFT or HF methods should be any good in a correlated many-body approach. Clearly, that was one of our motivations to use a many-body framework consistently throughout the construction, in testing and also in further developments.

The validity and limits of the constructed ECPs also become much more transparent if the *information about systematic errors built into ECPs is provided upfront* so as to clear the stage for subsequent calculations. The simplicity of the basic ECP form is also highly desirable since that enables wide use in many settings and approaches.

Note that presented ECPs are of the simplest semilocal type with minimal form and size, with only one or two gaussians per channel. It is realistic to expect that more careful and more elaborate fits could further tune the properties beyond accuracies we have currently

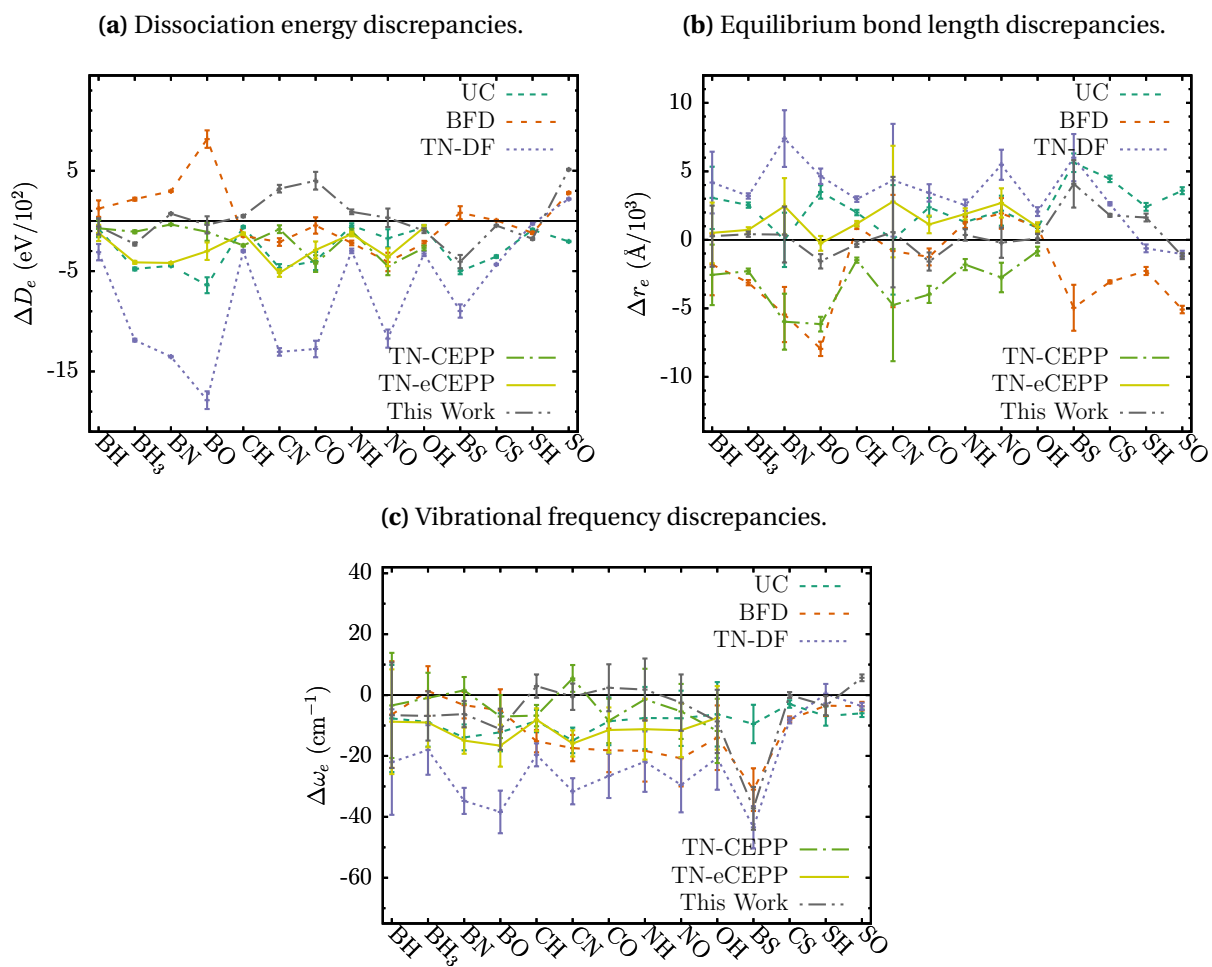


Figure 5.14 Discrepancies of molecular binding parameters of our ECPs, UC and previous constructions with respect to all-electron CCSD(T) calculations. Parameters were obtained from Morse potential fits in all cases.

obtained. The corresponding optimizations are still challenging and are hampered by complicated couplings between the atomic solvers and optimization methods and require further refinement, higher efficiency, and more robustness. At the same time, the presented results offer encouraging examples of the accuracy that can be obtained with present day capabilities and show that there is significant room for further improvements and expansions.

A New Generation of Effective Core Potentials from Correlated Calculations: 2nd row elements

Portions of this chapter also appeared in:

A New Generation of Effective Core Potentials from Correlated Calculations: 2nd row elements

J. Chem. Phys. **149**, 104108 (2017)

Contributions:

In the work described in this chapter, the author played a primary role in the investigation of the objective function used to generate the final high-accuracy 2nd-row ECPs and a significant role in the preparation of the manuscript. The author generated all data in this work – except for the CCSD(T) total energies provided in the appendix which were performed by Cody Melton – and played a key role in the analysis of those data. Authors Guangming Wang and Abdulgani Annaberdiyev also contributed to the preparation of the manuscript.

6.1 Introduction

Remarkably, for the first row, we found that it was possible to derive ECPs that show significant improvements in various valence properties at the CCSD(T) level of theory when compared with existing ECP tables or the all-electron uncorrelated-core (UC) approximation where excitations out of the core are suppressed. Quite surprisingly this was achieved with simple, “minimal”, non-divergent ECP parameterizations, i.e., with very few gaussians per nonlocal channel.

In order to maintain the simplicity and direct connection of the ECP to the original Hamiltonian, the solution was formulated as an inverse problem. Namely, we searched for an effective Hamiltonian operator that reproduces a given set of many-body eigenenergies and eigenstates, outside the core region, as closely as possible. Not surprisingly, this can lead to an ill-conditioned problem (high sensitivity to inputs) and/or a complicated optimization landscape with many near-degenerate minima. Consequently, such optimizations can be rather difficult. Such difficulties can be understood conceptually, for instance, if one considers the well-known example from the spectral theory that two operators with different domains can have the same spectrum, i.e., they can be exactly isospectral. This non-uniqueness suggests that the problem requires adequate and proper constraining in order to arrive at a desirable solution with a reasonable amount of effort. For the B, C, N, O and S elements, we were able to accomplish just that and produced ECPs for these atoms that exceed the accuracy of existing tabulated ECPs and that have led to significant improvements in transferability.

We also aimed to construct ECPs that are simple and can be employed in a variety of electronic structure packages and methods, i.e., as widely applicable as possible. To this end, we chose a well-established form [Bur07; Ovc01] that is non-singular and behaves quadratically at the origin and is parameterized by a short sum of radial powers multiplied by gaussians. This form leads to one-particle orbitals that also behave quadratically at the origin which makes their expansion in gaussians more efficient. Additionally, the smooth form provides computational savings for quantum Monte Carlo by reducing fluctuations.

Another aspect of our constructed ECPs was to probe for the transferability in bonds both near and out of the equilibrium configurations, having in mind, for example, high-pressure applications in future. Therefore we monitored discrepancies of dimer binding curves including short bond lengths up to the dissociation limit, where the binding energy

goes to zero for the compressed bond. Note that this limit is particularly challenging since it directly probes whether the ECP correctly mimics the (missing) core charge at short lengths. In cases where larger errors were present, we included the dimer into the construction as an additional constraint which improved the quality of the resulting ECPs. Indeed, we found that the key molecular properties were more accurately reproduced in this way. As an independent probe of transferability, we then tested hydride and oxide molecules. Further benchmarks and testing will come from future applications for systems where all-electron correlated calculations can be carried out.

In the present work, we extend our previous collection of correlation-consistent ECPs to include the entire 2nd row of elements Na through Ar (following the convention that H and He comprise the zeroth row). For these elements, we have further improved and modified the methodology used in generating the ECPs which we describe in more detail in what follows. The constructed ECPs that we present here are not meant to be definitive. It is clear that with more elaborate forms, say, with a larger number of free parameters, the accuracy might still improved. Another consideration is that different applications might require further refinement or even new constructions. For example, alkaline elements have nominally large cores and one or two valence electrons. Clearly, there are applications where this would not provide an accurate replacement of the all-electron Hamiltonian (high pressures) and the inclusion of the outermost shell from the core to the valence space might be needed. This implies that one needs more options to guarantee the accuracy in various physical or chemical applications. Therefore keeping track of updates and any new constructions, as well as maintaining benchmark sets is an important part of providing clear choices and validations for future use of ECPs.

The paper is organized as follows. In section II, we discuss the different components of a number of atomic energy gaps of a variety of tabulated 2nd row ECPs and how this affects their agreement with all-electron gaps at the many-body level as well as how this motivated the particular construction we use for the ECPs presented in this work. We outline this construction in section III. In section IV, we describe the particular parameterization we use for our ECPs. Section V outlines the scheme we followed to construct the ECPs' corresponding valence basis sets. In section VI, we share our results and present our ECPs together with several existing ones for comparison and highlight the shortcomings of Ne-core ECPs in particular settings (short polar bonds) and therefore, in addition, present He-core ECPs for the entire 2nd row that significantly improve the accuracy of ECPs with

Ne-core partitioning and serve also as benchmarks for comparisons. In section VII, we give our concluding remarks.

6.2 ECP Atomic Correlation Energies

We have looked in a more quantitative manner at the contributions from valence-valence correlation energies as well as Hartree-Fock (HF) energies for a variety of atomic excitations among a testbed of tabulated ECPs. This provides additional insights into how various components of the energy are modified and respond to ECP differences (form, parameterization, etc.) which ultimately enabled us to improve our ECP construction scheme and helped to improve their overall performance. Specifically, for a number of tabulated ECPs, we have analyzed the errors (relative to relativistic all-electron results) of the HF and correlation contributions to the CCSD(T) spectral energies for the Si and P atoms and we have looked at how the errors for these energy components change between different ECPs. Relativity in all-electron calculations was included through the 10th-order Douglas-Kroll-Hess Hamiltonian for both HF and CCSD(T) approaches.

In Figs. 6.1 and 6.2, for Si and P, we show examples of the spread of HF and CCSD(T) valence-valence correlation errors across a number of previously tabulated ECPs, which employ a wide variety of parameterizations, for a number of excitation energies. In particular, we plot the spread of the following quantities over the set of ECPs:

$$\Delta\text{HF}_s = \Delta E_s^{\text{ECP}} - \Delta E_s^{\text{AE}}, \quad (6.1)$$

where ΔE_s represents the energy gap between a given state and the ground state, and

$$\Delta\text{cVV}_s = \left| \text{VV}_{\text{corr},s}^{\text{ECP}} \right| - \left| \text{VV}_{\text{corr},s}^{\text{AE}} \right|, \quad (6.2)$$

where VV_{corr} is the valence-valence correlation energy. In these figures, the spread (represented by the shaded areas) are bounded by the maximum discrepancy (upper bound) and minimum discrepancy (lower bound) among the various ECPs. For the ECP cases, the valence-valence correlation is taken to be the fully correlated CCSD(T) energy of a particular state. For the all-electron case, the valence-valence correlation is taken to be the UC CCSD(T) correlation energy of a particular state.

There are several observations that can be gleaned from these plots. The first is that

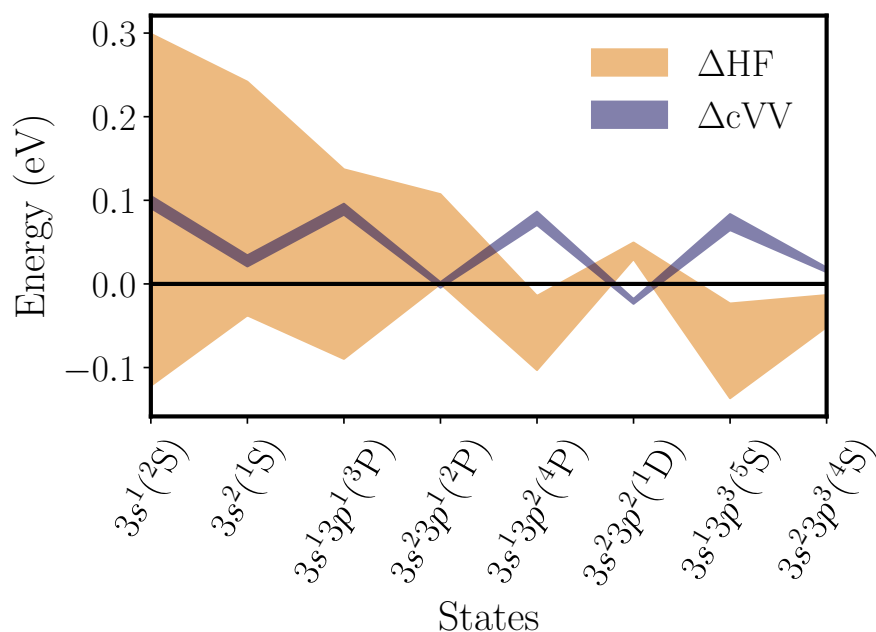


Figure 6.1 For the silicon atom, the spread of CCSD(T) valence-valence correlation errors (ΔcVV) and spread of HF errors (ΔHF) for various excitation energies from a variety of previously tabulated ECPs, in particular, BFD[Bur07], CRENL[PC85], SBKJC[Ste84], STU[Dol87] and TN-DF[TN05].

the ECPs' correlation energies are on average larger than the corresponding all-electron valence-only correlated values with the increase varying between 0 and 0.15 eV. This has been known for some time, see the analysis by Dolg [Dol96]. The reason is the absence of the radial node(s) and increased smoothness of ECP (pseudo)orbitals since this increases the probability for unlike spin pairs of electrons to encounter each other, thus increasing the correlation energy. However, the next striking feature is the rigidity of the correlated energy differences across various ECPs, regardless of construction, whether it shows singular or bounded/smooth behavior of local and nonlocal terms at the origin or other details. Essentially the ECP correlation energies are almost invariant to the the form or the construction of ECP, typically within 0.02 eV or less. Consequently, the largest differences between ECPs for atomic spectra come from fluctuations that originate in the HF component of energy differences. Finally, the third point is that aiming for atomic spectra accuracy better than 0.1 eV requires the HF component to "compensate" for the rigid correlation contribution

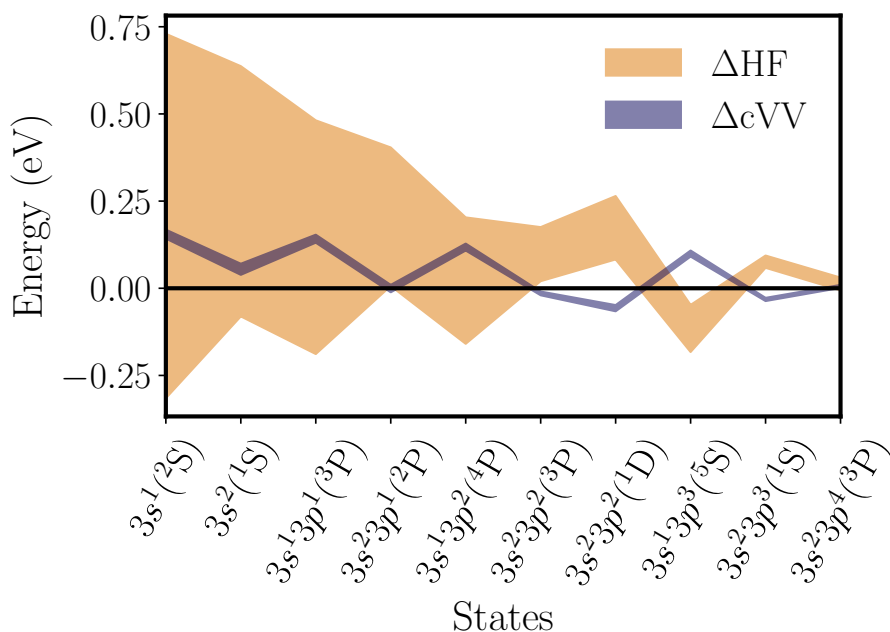


Figure 6.2 For the phosphorus atom, the spread of CCSD(T) valence-valence correlation errors (ΔcVV) and spread of HF errors (ΔHF) for various excitation energies from a variety of previously tabulated ECPs, in particular, BFD[Bur07], CRENL[PC85], SBKJC[Ste84], STU[Dol87] and TN-DF[TN05].

component for a number of states. Given this rigidity of ECP correlations, we incorporate mean-field data into the construction of the 2nd row ECPs along with many-body data in order to avoid large mean-field biases which we found helps to improve transferability (shown below in section V). In the following section we give the specifics of the procedure we use to construct the ECPs.

6.3 Construction

The strategy that we employ to generate the 2nd row ECPs is a combination of many-body energy consistency and single-body norm conservation. We provide the details of each and how they are combined to form our full objective function in this section.

6.3.1 Many-body Energy Consistency

In order to account for correlation when generating our ECPs, we have employed a spin-unrestricted CCSD(T) energy consistency scheme as part of our full objective function. The details of this scheme are as follows: For a given atom, we first generate a large number of all-electron reference atomic states which includes various ionizations and spin-multiplicities. The states are calculated with the fully correlated (both core and valence electrons) spin-unrestricted CCSD(T) method. We account for relativity by using the 10th order Douglas-Kroll-Hess Hamiltonian [RW04]. We formulate this part of the objective function by calculating the same atomic states with a parameterized ECP at the same level of theory, and then vary the parameters until the gaps between these states and the neutral ground state agree with that from the all-electron to a satisfying level. The discrepancies of the ECP gaps relative to the all-electron gaps are minimized in the least squares way which defines the first component of our objective function, \mathcal{E}^2 , given by

$$\mathcal{E}^2 = \sum_s (\Delta E_s^{\text{ECP}} - \Delta E_s^{\text{AE}})^2, \quad (6.3)$$

where s labels a given atomic state and ΔE_s is the energy gap between the state and the neutral ground state and "ECP" and "AE" label the ECP and all-electron gaps, respectively.

The energetics from all-electron and ECP were calculated with the MOLPRO quantum chemistry package [Wer12]. For all atoms, the uncontracted aug-cc-pCV5Z basis set [THD89], where core state correlation functions are present, was utilized. If linear dependencies developed as a result of uncontracting the basis, the problematic terms with nearly identical exponents were removed. This same basis was used for calculating both the all-electron and ECP gaps in order to construct similar basis set errors between the two which would largely cancel when calculating their difference in Eq. 6.3. Quantitative estimates of the residual finite basis set errors were given in our previous work [Ben17a] for a test set of atoms from the 1st and 2nd row.

We chose not to use the \mathcal{E}^2 objective function exclusively when training our ECPs given that in doing so we observed an overall large increase in error of the HF gaps and subsequently poorer transferability when calculating potential energy surfaces of diatomic molecules (dimers and oxides) at the CCSD(T) level of theory. We suspect this is due to the rigidity in the ECP correlation energies as outlined in section 6.2, namely, in order to reduce the total error of the CCSD(T) gap, the magnitude of the HF error increases to

compensate the error from the more rigid correlations. We observed that reaching small gap errors at the many-body level at the full expense of the HF gap errors showed tendencies to increase transferability errors. In fact, the best results came from a balance between the errors in correlation and HF components. In addition, significant variational freedom inside the core is proved to be helpful and we have therefore included a single-body norm conservation into our objective function, as outlined in the next subsection, in order to strike this balance.

6.3.2 Single-body Norm Conservation

In order to maintain accuracy at the HF level of theory, we employed a norm conservation scheme as another component of our full objective function. We have taken a strategy similar to what has been used in previous works [Ste84; Ovc01; Chr79; TN05], however, the specifics of our approach are as follows. For the Ne-core ECPs, we first generated all-electron scalar relativistic HF radial orbitals, ϕ_l^{AE} , where the 10th order DKH hamiltonian was employed. For each angular momentum channel, we chose a cutoff radius, R_l , corresponding to the outer most extremum of $r^{l+1}\phi_l^{\text{AE}}$. At this radius, we then obtained the orbital's norm inside this point, $N_l^{\text{AE}} = \int_0^{R_l} (r^{l+1}\phi_l^{\text{AE}}(r))^2 dr$, its value, $V_l^{\text{AE}} = \phi_l^{\text{AE}}(R_l)$ and its slope, $S_l^{\text{AE}} = \left. \frac{d}{dr}\phi_l^{\text{AE}}(r) \right|_{R_l}$. The norm-conserving objective function is then formed as

$$\mathcal{N}^2 = \sum_l (N_l^{\text{ECP}} - N_l^{\text{AE}})^2 + (V_l^{\text{ECP}} - V_l^{\text{AE}})^2 + (S_l^{\text{ECP}} - S_l^{\text{AE}})^2 + (\epsilon_l^{\text{ECP}} - \epsilon_l^{\text{AE}})^2, \quad (6.4)$$

where we also include the errors in the orbital eigenvalues ϵ_l . This procedure is similar to what is generally done in norm-conserving schemes, however, we allow the core more freedom by not matching it to a particular functional form but rather focus on matching the quantities mentioned. For the elements Al-Ar we took reference s and p orbitals from the atom's neutral ground state and took the reference d orbital from the $p \rightarrow d$ excitation. For Na and Mg, we took the single-valence doublet states as the reference in each channel.

6.3.3 Weighted Combination and Core-Valence Partitioning

To form our full objective function, O^2 , we take \mathcal{E}^2 and \mathcal{N}^2 as a linear combination,

$$O^2 = w_0 \mathcal{E}^2 + \mathcal{N}^2, \quad (6.5)$$

where w_0 is a chosen weight, typically taken as 5×10^{-2} . To construct our ECP, we vary its parameters until O^2 is minimized. The value of w_0 is chosen so as to strike a balance between the accuracy of the many-body and single-body properties of the ECP – or viewed in another way, we are allowing a violation of single-body norm conservation in order to achieve higher accuracies on the many-body spectral properties.

When using the conventional Ne-core and $3s3p$ valence, the transferability of the ECPs in particular environments was observed to be less than satisfying. This is particularly true for compressed polar bonds (as will be shown). In order to reach high many-body and single-body accuracies in these regimes as well, we have also generated He-core ECPs for all of the second row atoms. Though, more expensive than their Ne-core counterparts the He-core ECPs still provide a non-negligible computational savings when considering their use in quantum Monte Carlo which scales with atomic number as $\mathcal{O}(Z^{5.5-6.5})$ and within planewave codes where calculations with the full Coulomb potential are not feasible. For these small-core ECPs, the optimization procedure was nearly identical, except that we took the neutral ground states for all references in the norm-conserving component and we generated the cutoff radii from the semi-core orbitals. In this case, for each angular momentum channel, the cutoff radius was taken as the inner most finite extremum of $r^{l+1}\phi_l^{\text{AE}}$. We then used these radii as the match points of the norm, value and slope of the valence orbitals while also matching their eigenvalues. Furthermore, we found that a much larger weight could be used on the UCCSD(T) spectral component (two or three orders of magnitude) than the weight used for the Ne-core ECPs without introducing a significant negative impact in transferability. This is due to a much smaller discrepancy in the correlation energies (valence and semi-core electrons) between the He-core and all-electron cases.

We have applied the construction outlined in this section to all the second row atoms, Na-Ar. For all optimizations, we have used the DONLP2 solver of Spellucci [Spe] to generate the optimal parameter sets for all ECPs. The methodologies used in DONLP2 are outlined in [Spe98b] and [Spe98a]. Though finding the global minimum in these types of non-linear optimizations can be difficult, our incorporation of the many-body spectral component and mean-field spatial component (ie, additional constraints) should alleviate issues related to the ill-conditioned nature of this problem. Further justification of our chosen objective function came after analyzing the accuracy of our ECPs in settings that were not included as part of their training, which we cover in section VI.

The parameter values of our ccECPs for all atoms and all core sizes are shared in Tables I and II and are also available at the website in Ref. [Web].

Table 6.1 Parameter values for Ne-core ECPs. For all ECPs, the highest l value corresponds to the local channel.

Atom	Z_{eff}	l	$n_{l,k}$	$\alpha_{l,k}$	$\beta_{l,k}$
Na	1	0	2	5.377666	6.234064
		0	2	1.408414	9.075931
		1	2	1.379949	3.232724
		1	2	0.862453	2.494079
		2	1	4.311678	1.000000
		2	3	1.925689	4.311678
		2	2	1.549498	-2.083137
Mg	2	0	2	5.936017	6.428631
		0	2	1.592891	14.195491
		1	2	1.583969	3.315069
		1	2	1.077297	4.403025
		2	1	6.048538	2.000000
		2	3	2.796989	12.097075
		2	2	2.547408	-17.108313
Al	3	0	2	7.863954	14.879513
		0	2	2.061358	20.746863
		1	2	3.125175	7.786227
		1	2	1.414930	7.109015
		2	1	5.073893	3.000000
		2	3	8.607001	15.221680
		2	2	3.027490	-11.165685
Si	4	0	2	9.447023	14.832760
		0	2	2.553812	26.349664
		1	2	3.660001	7.621400
		1	2	1.903653	10.331583
		2	1	5.168316	4.000000
		2	3	8.861690	20.673264
		2	2	3.933474	-14.818174
P	5	0	2	12.091334	15.259383
		0	2	3.044535	31.707918
		1	2	4.310884	7.747190
		1	2	2.426903	13.932528
		2	1	5.872694	5.000000
		2	3	9.891298	29.363469
		2	2	4.692469	-17.011136
S	6	0	2	16.117687	15.925748
		0	2	3.608629	38.515895
		1	2	6.228956	8.062221
		1	2	2.978074	18.737525
		2	1	6.151144	6.000000
		2	3	11.561575	36.906864
		2	2	5.390961	-19.819533
Cl	7	0	2	17.908432	15.839234
		0	2	4.159880	44.469504
		1	2	7.931763	8.321946
		1	2	3.610412	24.044745
		2	1	7.944352	7.000000
		2	3	12.801261	55.610463
		2	2	6.296744	-22.860784
Ar	8	0	2	27.068139	18.910152
		0	2	4.801263	53.040012
		1	2	11.135735	8.015534
		1	2	4.126631	28.220208
		2	1	8.317181	8.000000
		2	3	13.124648	66.537451
		2	2	6.503132	-24.100393

6.4 ECP Form

The ECPs we have constructed have the following form

$$V_i^{\text{ECP}} = V_{\text{loc}}(r_i) + \sum_{l=0}^{l_{\text{max}}} V_l(r_i) \sum_m |lm\rangle \langle lm|, \quad (6.6)$$

Table 6.2 Parameter values for He-core ECPs. For all ECPs, the highest l value corresponds to the local channel.

Atom	Z_{eff}	l	$n_{l,k}$	$\alpha_{l,k}$	$\beta_{l,k}$	Atom	Z_{eff}	l	$n_{l,k}$	$\alpha_{l,k}$	$\beta_{l,k}$
Na	9	0	2	61.004364	25.194598	P	13	0	2	101.982019	25.197230
		0	2	14.829519	60.933576			0	2	37.485881	189.426261
		1	1	8.997115	9.000000			1	1	15.073300	13.000000
		1	3	8.958240	80.974036			1	3	18.113176	195.952906
		1	2	8.691287	-57.117991			1	2	17.371539	-117.611086
Mg	10	0	2	72.718763	25.152097	S	14	0	2	111.936344	25.243283
		0	2	19.375648	84.969927			0	2	43.941844	227.060768
		1	1	10.136825	10.000000			1	1	17.977612	14.000000
		1	3	10.457653	101.368253			1	3	20.435964	251.686565
		1	2	10.109595	-68.596401			1	2	19.796579	-135.538891
Al	11	0	2	81.815564	25.157259	Cl	15	0	2	124.640433	26.837357
		0	2	24.522883	113.067525			0	2	52.205433	277.296696
		1	1	11.062056	11.000000			1	1	22.196266	15.000000
		1	3	12.369778	121.682619			1	3	26.145117	332.943994
		1	2	11.965444	-82.624567			1	2	25.015118	-161.999982
Si	12	0	2	92.046246	25.228329	Ar	16	0	2	135.620522	25.069215
		0	2	30.895726	150.483122			0	2	60.471053	332.151842
		1	1	12.066048	12.000000			1	1	23.431337	16.000000
		1	3	15.276621	144.792575			1	3	26.735872	374.901386
		1	2	14.506273	-99.229393			1	2	26.003325	-178.039517

where r_i is the radial distance of electron i from the core's origin and l_{max} is chosen to be the maximum occupied angular momentum channel from the ground state core. The non-local terms include $l m$ -th angular momentum projectors, $|l m\rangle\langle l m|$.

The local potential, V_{loc} , is taken to be

$$\begin{aligned}
 V_{\text{loc}}(r) &= -\frac{Z_{\text{eff}}}{r} + \sum_{k=1} \beta_{\text{loc},k} r^{n_{\text{loc},k}-2} e^{-\alpha_{\text{loc},k} r^2} \\
 &= -\frac{Z_{\text{eff}}}{r} + \frac{\beta_{\text{loc},1}}{r} e^{-\alpha_{\text{loc},1} r^2} + \beta_{\text{loc},2} r e^{-\alpha_{\text{loc},2} r^2} + \sum_{k=3} \beta_{\text{loc},k} e^{-\alpha_{\text{loc},k} r^2},
 \end{aligned} \tag{6.7}$$

where $Z_{\text{eff}} = Z - Z_{\text{core}}$ is the pseudo-atom's total core charge. We fix $\beta_{\text{loc},1} = Z_{\text{eff}}$, $\beta_{\text{loc},2} = Z_{\text{eff}}\alpha_{\text{loc},1}$, while $n_{\text{loc},1} = 1$, $n_{\text{loc},2} = 3$, so that $V_{\text{loc}}(r)$ is finite everywhere and its first derivative vanishes at the origin. Similarly, the set of non-local potentials, $\{V_l\}$, are parameterized by

$$V_l(r) = \sum_{k=1} \beta_{l,k} r^{n_{l,k}-2} e^{-\alpha_{l,k} r^2}, \tag{6.8}$$

where the powers $\{n_{l,k}\}$, are all equal to 2. Except for $\beta_{\text{loc},1}$ and $\beta_{\text{loc},2}$, which are fixed as

given above, the Greek letter parameters are all varied in the minimization of our objective function. For an initial round of optimization, for V_{loc} and all V_l 's, we set the number of gaussian terms to one. For a given Ne-core ECP, we included 1-2 more gaussians in these channels for additional flexibility if further optimization rounds were required to reach agreement within roughly 0.05-0.1 eV of the low-lying all-electron many-body spectrum. Pushing the accuracies further, generally led to significant mean-field errors for reasons discussed in section II. In the case of the He-core ECPs, we again only included one gaussian term in these functions initially. However, for these ECPs further terms were added until we observed agreement with all-electron data within roughly the chemical accuracy window. This simple form keeps the volume of the parameter space as small as practically possible and therefore helps to simplify the optimization problem.

6.5 Valence basis sets

For each atom and each core size, we have generated a corresponding basis set. The general recipe that we followed to generate the basis sets (similar to Ref. [Bur07]) was the same for both core sizes and we outline it in this section.

To generate the basis sets for the He-core ECPs, we have minimized the atomic ground-state CCSD(T) energy for all atoms. The basis sets for the Ne-core ECPs minimize the atomic ground-state CCSD(T) energy for the atoms Al through Ar and the dimer ground-state CCSD(T) energy for the atoms Na and Mg. For a given basis set, we start by generating contractions, one for each occupied HF ground-state orbital, by starting from a set of even-tempered primitives and optimizing the largest exponent until the ground-state CCSD(T) energy is minimized. The resulting exponents are then taken as the contraction primitives whose coefficients are taken to be the ground-state HF expansion coefficients. We take the same contractions for each VDZ, VTZ, VQZ and V5Z basis. For a given basis, we append additional even-tempered uncontracted primitives following the correlation-consistent scheme [Dun02] and determine the largest exponent among these primitives by minimizing the ground-state CCSD(T) energy.

In addition to these basis sets, we have also generated augmented varieties which follow the same scheme, but include an additional even-tempered diffuse primitive into each symmetry channel. All of our basis sets are shared at the website given in Ref. [Web].

6.6 Results

For all 2nd row atoms, we share in Tables 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9 and 6.10 the all-electron CCSD(T) valence ionization potentials and (if bound) the electron affinity along with the discrepancies of the same quantities from various core approximations which includes all-electron UC approximation, a number of previously tabulated ECPs [Ste84; Bur07; Dol87; TN05] and our Ne-core and He-core correlation consistent ECPs, ccECP[He] and ccECP[Ne], respectively. We also share the discrepancies from all-electron UCCSD(T) potential energy surfaces for one or two diatomic molecules for all the atoms in Figs. 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9 and 6.10. For the cases of oxides, we utilized our previously published oxygen ECP [Ben17a] when calculating potential energy surfaces for our ECPs and for all other ECPs we used that table's corresponding oxygen ECP. For the molecular discrepancies, we span a large range of geometries in all cases, from near the dissociation threshold ($\lesssim 0.05 \text{ \AA}$) on the left to just past equilibrium. For argonium, we used a well-tuned unpublished hydrogen ECP for all cases other than uncorrelated core. Our ECP parameter sets are shared in Tables 6.1 and 6.2.

In the following subsections, we provide a discussion for each pseudoatom's atomic and molecular accuracies, in turn, and summarize the results. In the final subsection, we share a comparison of the mean absolute deviations of the binding parameters of the molecules presented in this work between our ECPs and the all-electron UC approximation as well as the other tabulated ECPs we've benchmarked against in this work to provide a more global picture of the accuracy of our potentials. In this section we also present total atomic energy components of our ECPs along with their core radii for future reference.

6.6.1 Sodium

In the case of sodium, the atomic and molecular results of our Ne-core and He-core ECPs are shared in Table 6.3 and Fig. 6.3, respectively. For the case of the Ne-core, our construction led to a significant improvement of the atom's first ionization potential, when compared to the all-electron UCCSD(T) result, relative to the previously tabulated ECPs as well as the uncorrelated core (UC) result – the error being roughly a factor of 3 smaller than the other ECPs and roughly a factor of 2 smaller than UC. The electron affinity of our Ne-core ECP is also in good agreement with the all-electron result with an error that is less than 0.01 eV. In the case of our He-core ECP, the agreement with the all-electron atom is better still and for

each quantity, the errors are at sub meV scales.

For the Na₂ and NaO molecules, the accuracies of both of our core partitions are also quite high with our Ne-core and He-core ECPs being within 0.05 eV and 0.01 eV of the all-electron UCCSD(T) result over the full range of geometries, respectively.

Table 6.3 All-electron (AE) UCCSD(T) electron affinity and ionization potential of Na along with the errors from uncorrelated core (UC), ECPs and for information purposes also from experiment (Exp.). The uncontracted aug-cc-pCV5Z basis was used for all calculations. MAD is the mean absolute deviation of excitation energies, while MARE is the mean absolute relative error. All values in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	5.1391 ^a	5.1334	-0.1215	-0.1847	-0.1770	-0.1738	-0.1774	-0.0665	-0.0020
EA	0.5479 ^b	0.5470	-0.0017	0.0005	0.0030	0.0025	0.0026	0.0077	-0.0014
MAD			0.0616	0.0926	0.0900	0.0882	0.0900	0.0371	0.0017
MARE			0.0134	0.0184	0.0200	0.0192	0.0197	0.0135	0.0015

^a Reference [Kra18]

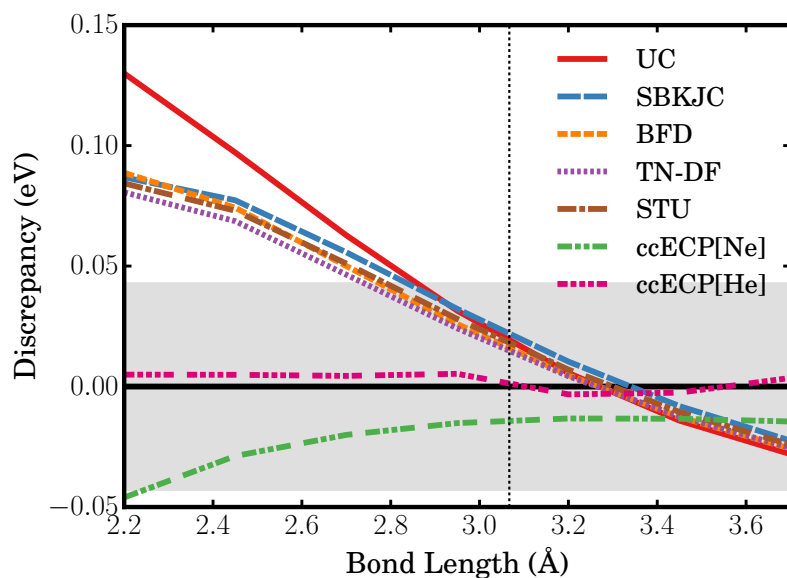
^b Reference [And99]

6.6.2 Magnesium

For magnesium, the atomic and molecular results of our Ne-core and He-core ECPs are shared in Table 6.4 and Fig. 6.4, respectively. The Ne-core partition shows an increased accuracy over UC and the previously tabulated ECPs with respect to the first and second ionization potentials. The He-core partition further increases the accuracy by more than an order of magnitude, with errors less than a meV for these quantities.

In the case of the MgO molecule, our Ne-core ECP shows better agreement with all-electron UCCSD(T) at shorter atomic separations than both UC and previously tabulated ECPs and at equilibrium, the agreement is within chemical accuracy. Our He-core ECP improves upon this further, with errors less than 0.01 eV across all geometries plotted.

(a) Na₂ binding curve discrepancies.



(b) NaO binding curve discrepancies.

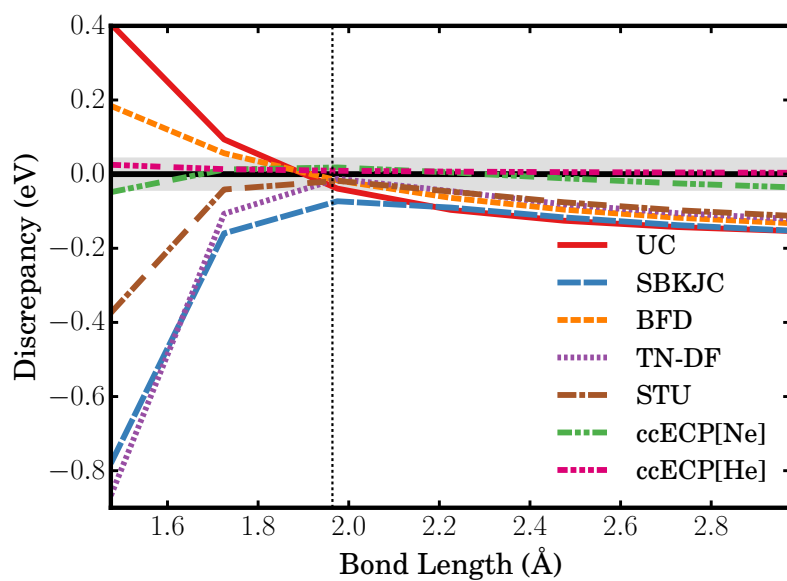


Figure 6.3 Binding energy discrepancies for (a) Na₂ and (b) NaO molecules in their ground states $^1\Sigma_g$ and $^2\Sigma$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

Table 6.4 All-electron UCCSD(T) ionization potentials for Mg along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. III for further description.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	7.6462 ^a	7.6400	-0.0980	-0.0884	-0.0727	-0.0693	-0.0617	-0.0578	0.0059
IP(II)	15.0353 ^a	15.0287	-0.2858	-0.3003	-0.2898	-0.2709	-0.2757	-0.2050	-0.0050
MAD			0.1919	0.1944	0.1812	0.1701	0.1687	0.1314	0.0055
MARE			0.0159	0.0158	0.0144	0.0135	0.0132	0.0106	0.0006

^a Reference [Kra18]

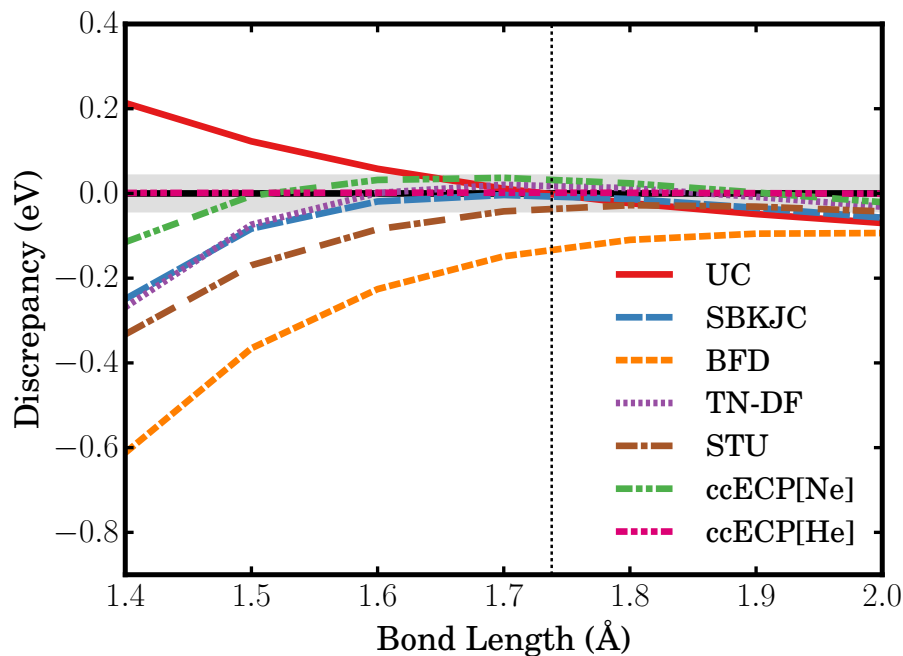


Figure 6.4 Binding energy discrepancies for the MgO molecule in its ground state $^1\Sigma^+$. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

6.6.3 Aluminum

For aluminum, the atomic and molecular results of our Ne-core and He-core ECPs are shared in Table 6.5 and Fig. 6.5, respectively. In the Ne-core case, the errors over the ionization potentials and the electron affinity were not improved over the other core approximations

but instead remained comparable. We note the lower variational freedom for our case due to bounded forms and smaller radial range when compared with some of the constructions. However, in the case of the He-core, we observed errors less than a hundredth of an eV.

The Al₂ molecule is described well by both core partitions, both being within chemical accuracy scales. For the AlO molecule, similar to previously tabulated ECPs, our Ne-core ECP is overbound by tenths of eV as the molecule is compressed towards dissociation. This potentially signifies a limitation of this particular choice of core for this type of environment, namely, compressed polar bonds. We suspect that key contributors to the error in this regime are a breakdown of the assumed point-charge interactions of the cores and also the neglect core-valence overlaps [DC12; Cha77]. For our smaller He-core ECP, we see that these difficulties evaporate and we observe errors well within chemical accuracy.

Table 6.5 All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Al along with the errors from the uncorrelated core (UC) and ECPs. Exp. gives experimental values for information purposes. The uncontracted aug-cc-pCV5Z basis was used for all calculations. AMAD is the mean absolute deviation for all excitation energies, LMAD is the mean absolute deviation for the first and second ionization potentials and electron affinity, while MARE is the mean absolute relative error for all states. All values in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	5.9858 ^a	5.9606	0.0045	0.0214	0.0135	0.0167	0.0667	0.0157	-0.0027
IP(II)	18.8286 ^a	18.8216	-0.1698	-0.1720	-0.1588	-0.1408	-0.0862	-0.1491	-0.0051
IP(III)	28.4476 ^a	28.4447	-0.3588	-0.4331	-0.3339	-0.3640	-0.3384	-0.3915	0.0052
EA	0.4328 ^b	0.4183	-0.0165	0.0215	0.0225	0.0182	0.0294	0.0171	0.0010
AMAD			0.1374	0.1620	0.1322	0.1349	0.1302	0.1434	0.0035
LMAD			0.0636	0.0716	0.0649	0.0586	0.0608	0.0606	0.0029
MARE			0.0155	0.0198	0.0191	0.0166	0.0245	0.0163	0.0008

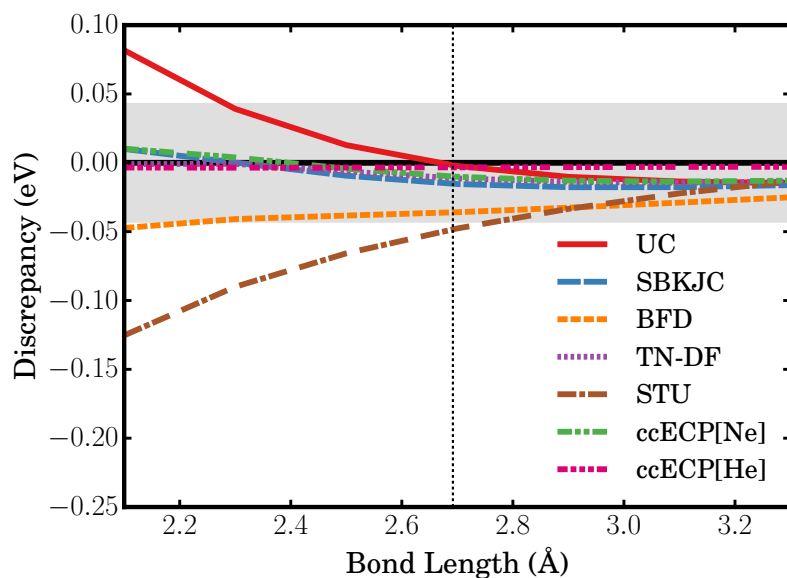
^a Reference [Kra18]

^b Reference [And99]

6.6.4 Silicon

For silicon, the atomic and molecular results of our Ne-core and He-core ECPs are shared in Table 6.6 and Fig. 6.6, respectively. In the Ne-core case, like Al, the accuracies among the ionization potentials and electron affinity were not improved over the other core approxi-

(a) Al₂ binding curve discrepancies.



(b) AlO binding curve discrepancies.

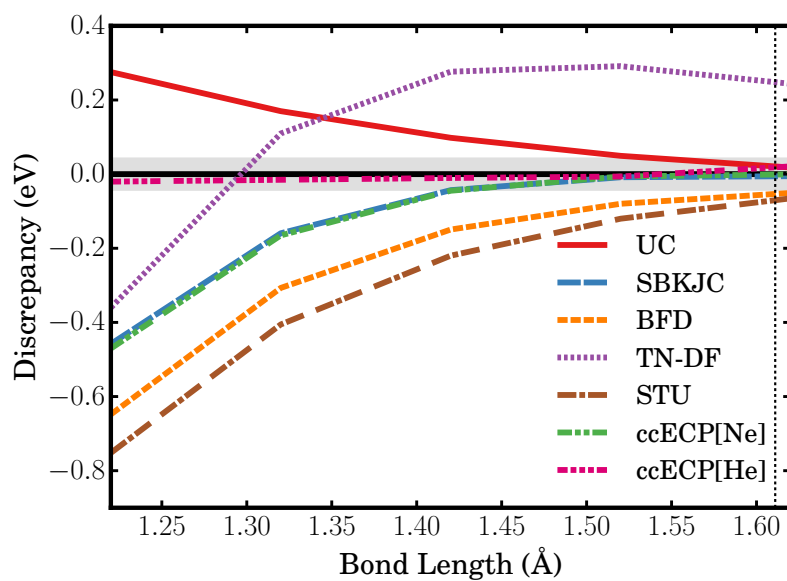


Figure 6.5 Binding energy discrepancies for (a) Al₂ and (b) AlO molecules in their ground states $^3\Sigma_g$ and $^2\Sigma$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

mations, though the He-core ECP shows errors smaller than chemical accuracy for all of these quantities. The results from the Ne-core ECP potentially reveal the limitations of this particular core partition.

In the case of the silicon dimer, both core partitions perform very well and their errors at the UCCSD(T) level are no more than about 0.02 eV. The SiO molecule shows similar and profound discrepancies for the Ne-core ECP as was seen in AlO. At short atomic separations, the pseudo-molecule is severely overbound for all tested ECPs including our Ne-core partition. Interestingly, at equilibrium, the agreement for our ECP and some others is much better and is within chemical accuracy. For our He-core ECP, the errors are significantly reduced and we observe chemical accuracy agreement across all geometries plotted. On this element, it is clearly visible that atomic spectrum is not sufficient to assess the overall quality of the ECPs. Seemingly, BFD, TN, and STU have somewhat smaller energy discrepancies than our Ne-core ECP. However, molecular calculations reveal lower or even significantly lower transferability. In fact, our Ne-core ECP is on par with SBKJC that however is unbounded ($1/r^2$ terms) so that our variational freedom is then significantly lower in comparison.

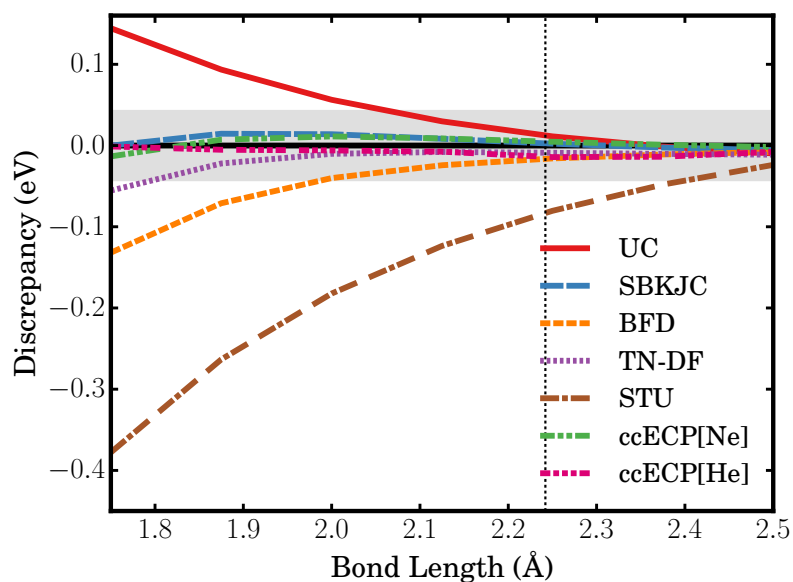
Table 6.6 All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Si along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	8.1517 ^a	8.1392	-0.0053	0.0190	0.0152	0.0183	0.0977	0.0166	-0.0011
IP(II)	16.3459 ^a	16.3014	-0.0320	-0.0096	-0.0131	-0.0072	0.1280	-0.0096	-0.0139
IP(III)	33.4930 ^a	33.4791	-0.2143	-0.2827	-0.1909	-0.2143	-0.0744	-0.2175	-0.0173
IP(IV)	45.1418 ^a	45.1325	-0.4016	-0.5920	-0.3528	-0.4683	-0.1474	-0.5123	0.0242
EA	1.3895 ^b	1.3928	0.0097	0.0210	0.0144	0.0193	0.0417	0.0175	0.0029
AMAD			0.1326	0.1849	0.1173	0.1455	0.0978	0.1547	0.0119
LMAD			0.0157	0.0165	0.0142	0.0149	0.0891	0.0146	0.0060
MARE			0.0050	0.0079	0.0053	0.0067	0.0111	0.0066	0.0008

^a Reference [Kra18]

^b Reference [And99]

(a) Si₂ binding curve discrepancies.



(b) SiO binding curve discrepancies.

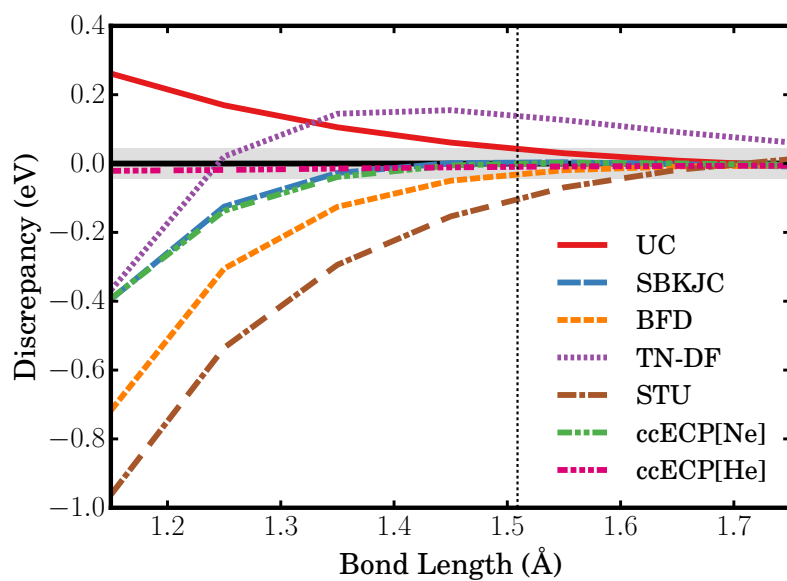


Figure 6.6 Binding energy discrepancies for (a) Si₂ and (b) SiO molecules in their ground states $^3\Sigma_g$ and $^1\Sigma$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

6.6.5 Phosphorus

For phosphorous, the atomic and molecular results of our Ne-core and He-core ECPs are shared in Table 6.7 and Fig. 6.7, respectively. For the case of our Ne-core partition, we observed that comparable accuracies of the atomic properties were only achievable with respect to the other core approximations. However, significant improvement was again achieved for the case of our He-core partition.

For the case of P_2 , both core partitions perform well with the Ne-core ECP obtaining maximum errors that are only marginally larger than chemical accuracy and the He-core ECP's errors being a factor of 2-3 smaller still. The results from PO are also quite good for both partitions. At equilibrium, both He- and Ne-core ECPs show errors no more than about 0.03 eV. For our He-core ECP, this level of accuracy is maintained as the bond is compressed, however, the Ne-core ECP at these shorter separations begins to overbind. Note that our constructions compete in overall optimality with SBKJC while having smaller variational freedom in bounded potentials. In addition, the molecular results show rather nonsystematic behavior for STU ECPs that are different (underbinding) from the rest of the row (typically, overbinding).

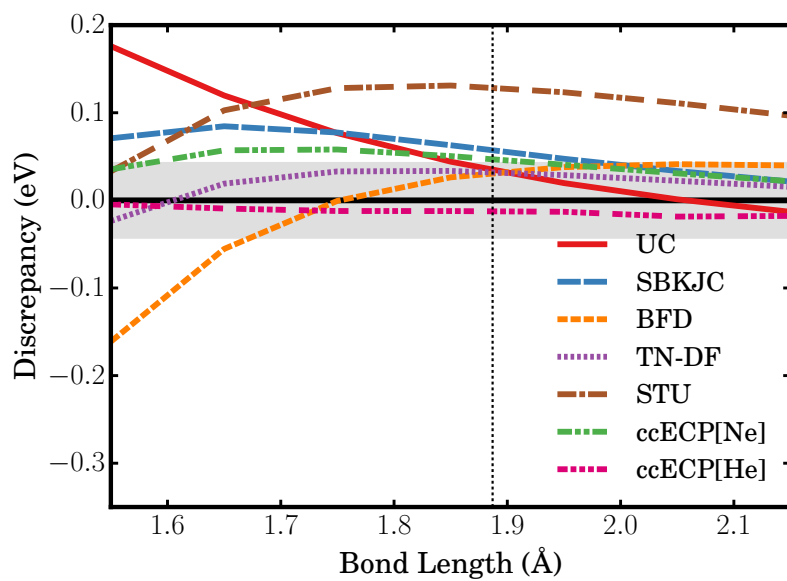
Table 6.7 All-electron (AE) UCCSD(T) ionization potentials and electron affinity of P along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	10.4867 ^a	10.5104	-0.0114	0.0171	0.0138	0.0221	-0.0054	0.0252	0.0011
IP(II)	19.7695 ^a	19.7562	-0.0355	-0.0116	-0.0003	0.0019	0.0050	-0.0018	-0.0109
IP(III)	30.2026 ^a	30.1328	-0.0677	-0.0798	-0.0535	-0.0522	0.0121	-0.0729	-0.0347
IP(IV)	51.4439 ^a	51.4304	-0.2527	-0.4559	-0.2187	-0.3254	-0.2051	-0.3577	-0.0241
IP(V)	65.0251 ^a	65.0181	-0.4402	-0.9174	-0.3825	-0.6667	-0.4504	-0.7180	0.0348
EA	0.7465 ^b	0.7003	0.0055	-0.0103	-0.0260	-0.0102	-0.0350	-0.0114	0.0020
AMAD			0.1355	0.2487	0.1158	0.1797	0.1188	0.1978	0.0179
LMAD			0.0175	0.0130	0.0134	0.0114	0.0151	0.0128	0.0047
MARE			0.0041	0.0071	0.0084	0.0058	0.0103	0.0065	0.0009

^a Reference [Kra18]

^b Reference [And99]

(a) P₂ binding curve discrepancies.



(b) PO binding curve discrepancies.

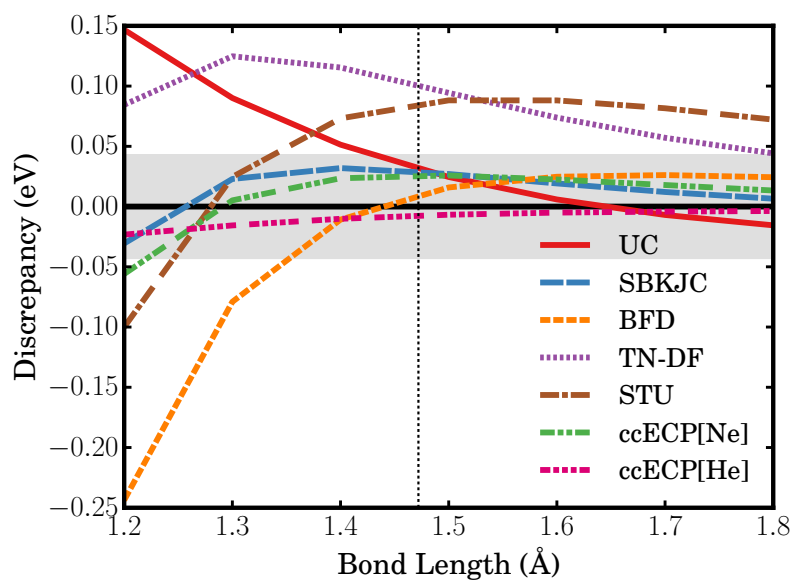


Figure 6.7 Binding energy discrepancies for (a) P₂ and (b) PO molecules in their ground states ¹Σ_g and ²Π, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

6.6.6 Sulfur

For sulfur, the atomic and molecular results of our Ne-core and He-core ECPs are shared in Table 6.8 and Fig. 6.8, respectively. In this case, the accuracies of the Ne-core ECP at the UCCSD(T) level are rather good with errors within chemical accuracies throughout all bond lengths plotted for both S_2 and SO molecules. And for our He-core ECP, we again see the benefit of this particular choice of core-valence partitioning where the atomic accuracies show large improvements over the other Ne-core approximations and ≈ 0.02 - 0.03 eV accuracy is achieved for both the sulfur dimer and oxide across all geometries plotted.

Table 6.8 All-electron (AE) UCCSD(T) valence ionization potentials and electron affinity of S along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	10.3600 ^a	10.2999	-0.0166	-0.0338	-0.0583	-0.0335	0.0851	-0.0397	0.0028
IP(II)	23.3378 ^a	23.3950	-0.0375	0.0231	0.0234	0.0098	0.2906	0.0194	0.0001
IP(III)	34.86 ^a	34.8258	-0.0670	-0.0301	0.0010	-0.0431	0.3369	-0.0388	-0.0202
IP(IV)	47.222 ^a	47.2693	-0.1042	-0.1361	-0.0673	-0.1393	0.3210	-0.1532	-0.0556
IP(V)	72.5945 ^a	72.5882	-0.2887	-0.6879	-0.2331	-0.4965	-0.0172	-0.5524	-0.0360
IP(VI)	88.0529 ^a	88.0550	-0.4769	-1.1502	-0.3611	-0.8586	-0.4356	-0.9956	0.0532
EA	2.0771 ^b	2.0500	0.0019	0.0020	-0.0276	0.0004	0.0326	-0.0026	0.0056
AMAD			0.1418	0.2947	0.1103	0.2259	0.2170	0.2574	0.0248
LMAD			0.0187	0.0196	0.0364	0.0146	0.1361	0.0206	0.0028
MARE			0.0025	0.0045	0.0041	0.0035	0.0083	0.0042	0.0008

^a Reference [Kra18]

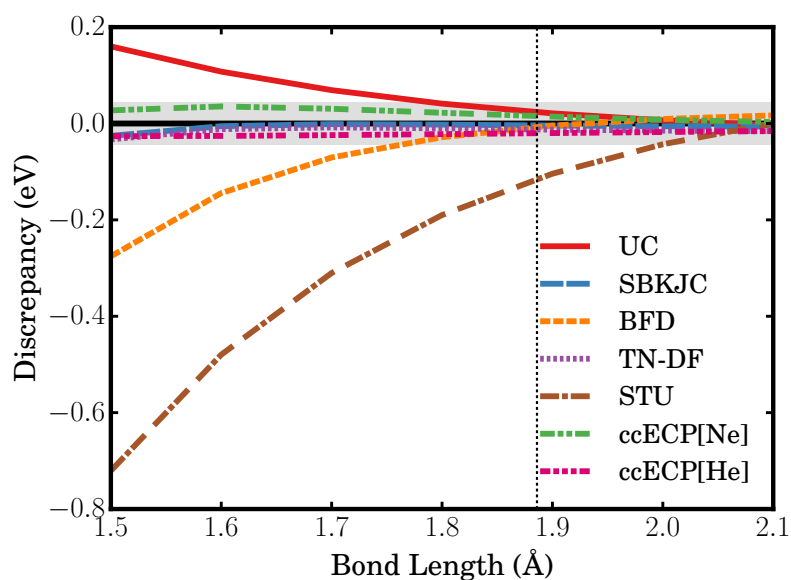
^b Reference [And99]

6.6.7 Chlorine

For chlorine, the atomic and molecular results of our Ne-core and He-core ECPs are shared in table 6.9 and figure 6.9, respectively. For this case, we only observed a significant improvement of the ionization potentials and electron affinity for our He-core ECP – its errors reaching no more than about 0.1 eV even for very deeply ionized cases.

Our construction, however, led to He- and Ne-core ECPs that reproduce the all-electron

(a) S₂ binding curve discrepancies.



(b) SO binding curve discrepancies.

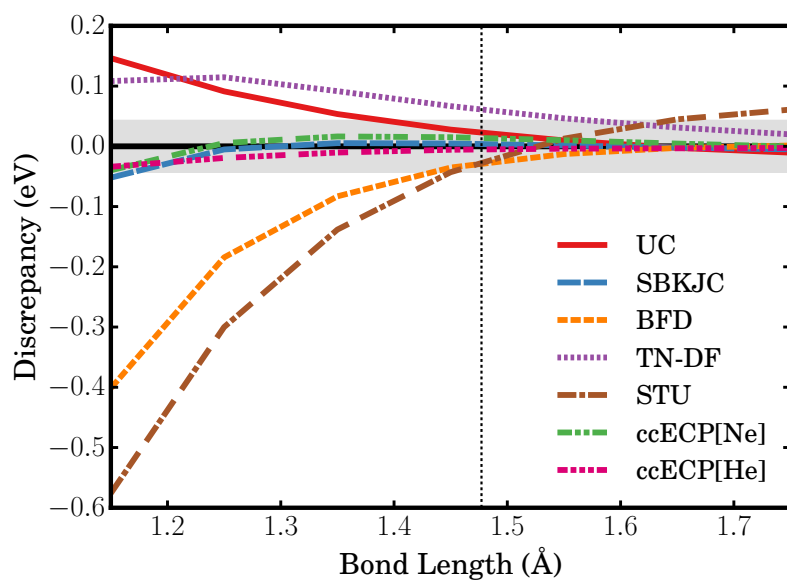


Figure 6.8 Binding energy discrepancies for (a) S₂ and (b) SO molecules in their ground states ³Σ_g and ³Σ, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

Cl₂ and ClO potential energy surfaces quite well – where both large and small core ECPs remain well within chemical accuracy for the full range of geometries. To provide another option with lower spectral errors, given the larger errors of our Ne-core ECPs in deeply ionized cases, we also generated a Ne-core ECP that weighted the energy component of the objective function more heavily to reduce the spectral errors further at the cost of slightly larger errors in the accuracy of Cl₂ and ClO. We share this alternative ECP in the supplemental material.

Table 6.9 All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Cl along with the errors from uncorrelated core (UC) and ECPS. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	12.9676 ^a	12.9388	-0.0175	-0.0082	-0.0396	-0.0210	0.1591	-0.0121	0.0015
IP(II)	23.8136 ^a	23.7309	-0.0401	-0.0649	-0.0860	-0.0743	0.2194	-0.0637	-0.0103
IP(III)	39.80 ^a	39.6980	-0.0614	0.0090	0.0091	-0.0126	0.4779	0.0234	-0.0238
IP(IV)	53.24 ^a	53.1811	-0.0939	-0.0807	-0.0816	-0.0952	0.5020	-0.0655	-0.0572
IP(V)	67.68 ^a	67.6102	-0.1328	-0.2335	-0.2471	-0.2280	0.4487	-0.2217	-0.1069
IP(VI)	96.94 ^a	96.8961	-0.3139	-0.9935	-0.0873	-0.6785	0.0301	-0.7001	-0.0441
IP(VII)	114.2013 ^a	114.2079	-0.5023	-1.5507	-0.2635	-1.0900	-0.4557	-1.1858	0.0646
EA	3.6127 ^b	3.6210	-0.0007	0.0193	-0.0153	0.0079	0.0781	-0.0123	0.0064
AMAD			0.1453	0.3700	0.1037	0.2759	0.2964	0.2856	0.0394
LMAD			0.0194	0.0308	0.0470	0.0344	0.1522	0.0294	0.0061
MARE			0.0020	0.0047	0.0024	0.0036	0.0094	0.0037	0.0008

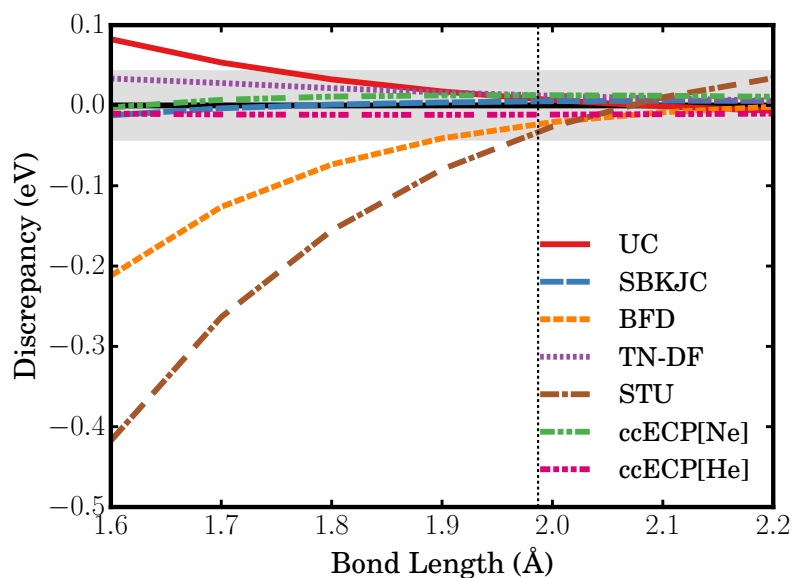
^a Reference [Kra18]

^b Reference [And99]

6.6.8 Argon

For argon, the atomic and molecular results of our Ne-core and He-core ECPs are shared in Table 6.10 and Fig. 6.10, respectively. With respect to the atomic properties, the Ne-core ECP performs well in this case with a total mean absolute deviation that is comparable to the all-electron uncorrelated core results. The He-core again performs even better where the total mean absolute deviation is improved over our Ne-core ECP by another 0.4 eV. To test the ECPs' performances in a molecular setting we calculated the errors from the all-electron UCCSD(T) ArH⁺ binding curve. Both core partitions show high accuracies for this case

(a) Cl₂ binding curve discrepancies.



(b) ClO binding curve discrepancies.

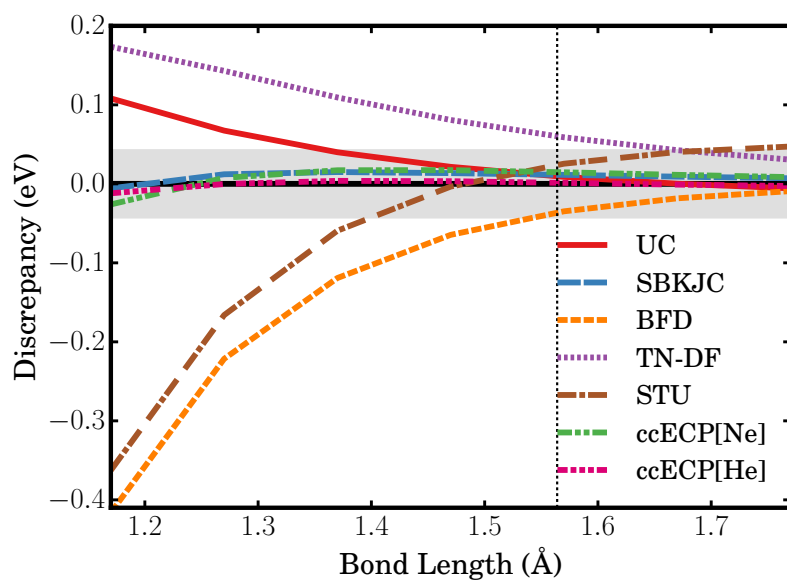


Figure 6.9 Binding energy discrepancies for (a) Cl₂ and (b) ClO molecules in their ground states $^1\Sigma_g$ and $^2\Pi$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

where the Ne-core and He-core errors are no larger than ≈ 0.01 eV in magnitude. Similar to chlorine, we also generated a Ne-core ECP that weighted the energy component of the objective function more heavily to reduce the spectral errors further at the cost of larger errors in the accuracy in the molecule. We share this alternative ECP in the supplemental material.

Table 6.10 All-electron (AE) UCCSD(T) ionization potentials for Ar along with the errors from uncorrelated core (UC) and ECPs. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V for further description.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
IP(I)	15.7596 ^a	15.7829	-0.0172	0.0146	-0.0156	-0.0016	-0.0116	-0.00327	0.0201
IP(II)	27.6297 ^a	27.6005	-0.0369	-0.0386	-0.0510	-0.0567	-0.0071	-0.05349	0.0135
IP(III)	40.735 ^a	40.6121	-0.0619	-0.1305	-0.1243	-0.1451	-0.0132	-0.13884	-0.0024
IP(IV)	59.58 ^a	59.2804	-0.0835	-0.0581	-0.0243	-0.0929	0.2038	-0.05261	-0.0183
IP(V)	74.84 ^a	74.7437	-0.1178	-0.2015	-0.1742	-0.2220	0.2207	-0.18799	-0.0622
IP(VI)	91.290 ^a	91.1085	-0.1580	-0.4181	-0.4141	-0.4046	0.2016	-0.39684	-0.1260
IP(VII)	124.41 ^a	124.3356	-0.3353	-1.3918	0.0652	-0.9451	-0.4787	-1.13569	-0.1348
IP(VIII)	143.4567 ^b	143.4706	-0.5238	-2.0213	-0.0730	-1.3990	-0.9534	-1.7643	0.1849
AMAD			0.1668	0.5343	0.1177	0.4084	0.2613	0.4666	0.0703
LMAD			0.0271	0.0266	0.0333	0.0291	0.0094	0.0284	0.0168
MARE			0.0019	0.0049	0.0018	0.0040	0.0026	0.0043	0.0008

^a Reference [Kra18]

^b Reference [And99]

6.6.9 Molecular binding parameters, total energies and core radii

We compiled all the results for molecular equilibrium parameters into the Tab. 6.11. The statistics provide useful information for an overall performance of the presented constructions as well as comparison with previously published tables. We have achieved rather significant improvements in the molecular properties due to the more flat discrepancy curves than other ECPs. In general, the ECPs have a tendency to overbind for short bonds and the cause is the missing core charge and the related core-valence repulsion. Interestingly enough, except for Al and Si, we were able to keep these biases typically below 0.05 eV with mildly increased deviations (0.1 eV) at very small bond length for Mg. Clearly, minor improvements might still be possible although we expect that one would encounter

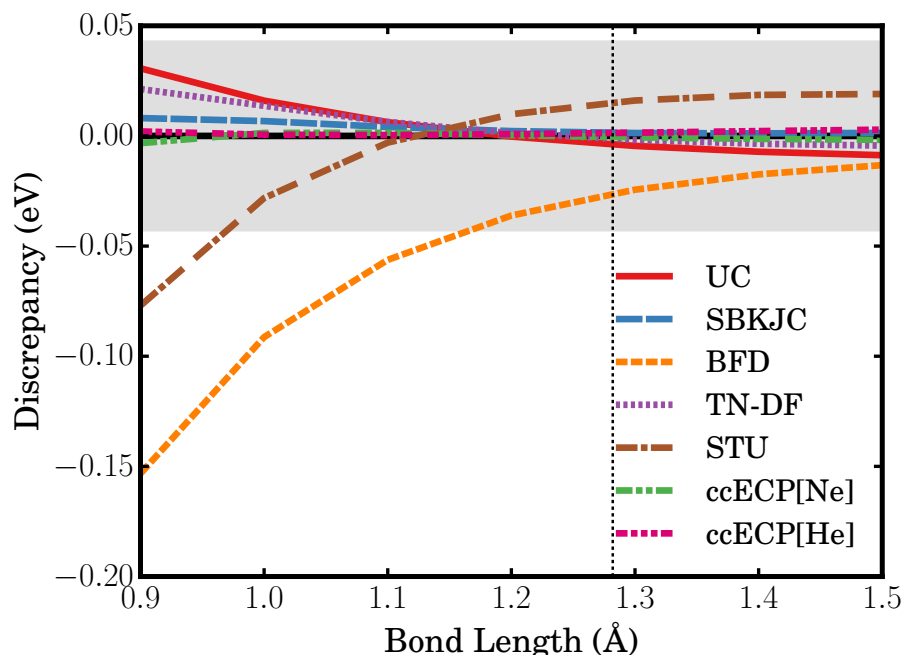


Figure 6.10 Binding energy discrepancies for the ArH⁺ molecule in its ground state $^1\Sigma$. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

diminishing returns unless the ECP form is further generalized. On the other hand, He-core framework clearly eliminates these deficiencies and provide very high accuracy across the whole row. Finally, we list total HF and correlation energies from UCCSD(T) as extrapolated to the complete basis set limit in Table 6.12 along with the core radii of our ccECPs. While HF energies are saturated to about 0.1 mHa or better in all cases, for Ne-core we estimate extrapolation errors on correlation energies to be within about 1 mHa. For He-core correlation energies we guess underestimation by about 3 mHa for Na and up to about 6 mHa for Ar, showing limitations of 5Z basis set and excitations up to perturbational triples to fully capture the many-body effects of 16 electrons. This estimate comes from an approximate analysis of core-core, core-valence and valence-valence correlation energy components in all-electron calculations [Ran15] with an added correction from Dolg [Dol96] that reflects increase in correlations in the ECP setting. The core radii, r_l , of each ccECP is taken as the radial distance from the origin at which the channel's full potential agrees to within 10^{-5} Ha of the bare Coulomb potential, $-Z_{\text{eff}}/r$. Additionally, we share the cutoffs, $r_{l,\text{nl}}$, of each

channel’s non-local potential which we define as the point where the non-local potential drops below 10^{-5} Ha. The non-local radii have implications primarily related to efficiency of QMC calculations as further explained elsewhere [CM96]. Additional data can be found also in Supplementary Information.

Table 6.11 Mean absolute deviations of discrepancies of binding parameters of all molecules considered in this work at equilibrium (D_e , r_e and ω_e) and near the all-electron dissociation threshold, $D_{\text{diss}}(\approx 0.05 \text{ \AA})$, at short bond lengths for our ECPs and previous constructions with respect to all-electron UCCSD(T) calculations. Errors in parenthesis such as “2.4(4)” denote “ 2.4 ± 0.4 ” and correspond to deviations of Morse potential fits to binding curves.

	UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
D_e (eV/ 10^2)	1.9(4)	2.4(3)	3.2(4)	5.1(3)	5.8(4)	2.0(4)	0.8(4)
ω_e (cm^{-1})	6(3)	14(3)	9(3)	23(3)	12(3)	11(3)	1(3)
r_e (mÅ)	17(2)	10(2)	15(2)	12(2)	18(2)	3(2)	1(2)
D_{diss} (eV/ 10^2)	16.9	16.9	31.9	16.5	38.9	9.2	1.4

6.7 Conclusions

In this work, we offer several new advances and insights into the construction of correlation consistent ECPs and apply this to the construction of potentials for the 2nd row elements.

In certain aspects, the 2nd row elements appear to be somewhat more complicated than the 1st row even though the corresponding valence spaces are very similar and the total and kinetic energies are lower (by about 20% when compared with the 1st row) indicating thus smoother electronic densities. Intuitively, therefore, accurately solving systems of 2nd row elements might be considered “easier” than, say, describing the deeper lying and more localized p-subshells in systems containing 1st row elements. However, the small number of valence electrons combined with larger core sizes generate larger biases than in the 1st row. In particular, the core-valence effects in correlation (important for spectra) compete with the absence of electrostatic core-valence interactions (important for short bonds) and therefore these tendencies are more difficult to reconcile. We found that the fits with many-body spectra only had limited accuracy but when we included more of the mean-field spatial information we were able to reach the most acceptable compromises.

Table 6.12 Total energies of our ccECPs with Ne (LC) and He (SC) cores in their neutral ground states along with their core radii for each angular momentum channel – the values r_l are taken as the distance which the channel’s full potential agrees with the bare Coloumb potential to within 10^{-5} Ha while $r_{l,nl}$ is the distance at which the channel’s non-local potential drops below 10^{-5} Ha. Both RHF/ROHF and UCCSD(T) correlation energies are extrapolated to the CBS limit using the uncontracted aug-cc-pwCV{T,Q,5}Z bases. Energies given in Ha. Radii given in Å.

Atom	r_s	r_p	r_d	$r_{s,nl}$	$r_{p,nl}$	HF	Corr.
Na(LC)	1.648	2.009	1.464	1.652	2.009	-0.186203	
Na(SC)	0.675	0.675		0.543		-47.357015	-0.325551
Mg(LC)	1.578	1.838	1.232	1.578	1.838	-0.788370	-0.035077
Mg(SC)	0.625	0.625		0.480		-62.927451	-0.363681
Al(LC)	1.406	1.633	1.135	1.406	1.633	-1.876999	-0.059869
Al(SC)	0.591	0.591		0.431		-80.993054	-0.395813
Si(LC)	1.273	1.427	1.006	1.273	1.427	-3.672423	-0.088666
Si(SC)	0.564	0.564		0.387		-101.625857	-0.428232
P(LC)	1.173	1.278	0.925	1.173	1.278	-6.340917	-0.117781
P(SC)	0.508	0.508		0.354		-125.258405	-0.461752
S(LC)	1.085	1.165	0.867	1.085	1.165	-9.918156	-0.177994
S(SC)	0.471	0.471		0.329		-151.916385	-0.525628
Cl(LC)	1.015	1.068	0.807	1.015	1.068	-14.689386	-0.235794
Cl(SC)	0.422	0.422		0.303		-181.612402	-0.587434
Ar(LC)	0.950	1.004	0.795	0.950	1.004	-20.779601	-0.289756
Ar(SC)	0.418	0.418		0.283		-214.891481	-0.642806

We were lead to the particular construction scheme introduced in this work after analyzing the behavior of the one-particle HF energies together with the correlation energies. Not too surprisingly, we found that the correlation energies for particular states and for a broad variety of existing ECPs with different parameterizations were essentially constant – the changes were on the order of 0.01-0.02 eV. This appears consistent with the key tenet of effective core potential theory that implies that the correlation will faithfully “follow” the constructed effective valence operators. However, we found this to be correct essentially only up to about 0.1 eV accuracy for atomic spectra. As soon as higher accuracy was sought and molecules were considered, the interplay between various energy components became more complicated and required refinements.

The testing of transferability proved to be very revealing. In particular, for the aluminum oxide and silicon oxide molecules at short bond lengths, we see a significant departure from the assumed accuracy standards of ≈ 0.05 eV for both ours and existing ECPs. We traced the deficiency to the missing electron core tails and likely core-valence effects which are absent in the ECP and would need to be included to reach higher accuracies. Similar problems have been noticed before, see for example, Ref.[OD03]. Here, we present a systematic view on this issue for Ne-core ECPs within polar bonds at short separations and, in addition, we point out that elimination of this issue might require more elaborated solutions. We have tested the inclusion of classical effective repulsive charges that would correspond to the electronic core-core interactions that could reasonably restore the desired accuracy. In order to introduce this in a more thought-through fashion, we leave this aspect for future studies. This is a notable point that can affect many oxide calculations[OD03] such as Al and Si perovskites at high pressures (for example, these are important components of the Earth’s mantle). Due to the importance of oxides, it is highly desirable to probe for these particular aspects in calculations of solids and other systems.

The atomic spectra and Tab. 6.11 show that apart from the inherent bias for a few short polar bonds our constructions show consistent accuracy in both atomic and molecular calculations and offer close to an optimal balance overall. Our results, somewhat unexpectedly, are on par with (or in a few cases significantly improve upon) the ECPs from the SBKJC table. Except for the Na atom (which is a bit special having only one valence electron), SBKJC ECPs appear to represent perhaps the most systematic consistency among the tested previous constructions. Note that the SBKJC ECPs are not bound at the nucleus, diverging as $1/r^2$ in the repulsive channels and as $-Z_{\text{eff}}/r$ in the local channel. Due to our choice

of the ccECP form with bounded potentials and minimized radial extent of the nonlocal terms (although not explicitly enforced) for several elements we hit a comparably optimal point as has been achieved in SBKJC table before. In some cases, we were able to actually find significantly better compromises. As mentioned before, we note that Ne-core ECPs for atoms such as Na and Mg with one or two valence electrons might prove to have only limited applicability in some bonding environments where shallow core states contribute to the chemical effects (for example, in compounds at high pressures).

It is also clear that core polarization and relaxation potentials [Shi91; SM93; Mul84; MM84] can alleviate some of the mentioned deficiencies. In this study, however, we wanted to see the limits of the simplest ECP form before employing new terms. In addition, oxide molecules for elements in the middle might be difficult to “fix” since the root of the problem appears to be the missing repulsive (essentially classical) charge from core electrons and this might require a more general and more extensive modification of the ECP form.

On the other hand, since we report here on the existing limits of the Ne-core ECPs in their current forms, we have decided to provide another option for high accuracy calculations that offer a significant boost in accuracy and eliminate many compromises mentioned above. To this end, we constructed high accuracy He-core ECPs that significantly decrease the discrepancies with all-electron relativistic calculations. The gain is by an order of magnitude or more, with typical errors of 0.001-0.005 eV on the atomic side and ≈ 0.01 eV for molecular systems. Obviously, these operators imply additional cost that is most pronounced in stochastic methods due to the increase of total energies and fluctuations. Note that some of that is compensated from much smaller effective core radius of nonlocal terms [CM96].

Overall, we believe that the constructed ECPs provide both new insights as well as practical new options for high accuracy valence-only calculations. Supplementary material contains further information including very accurate CCSD(T) total energies extrapolated to the complete basis set limit (better than 1 mHa accuracy) for atomic states and relativistic CCSD(T) molecular potential energy surfaces and binding parameters from all-electron calculations as well as the discrepancies from all core approximations considered in this work – experimental binding parameters are also included for comparison. All ccECPs and basis sets are accessible at the website [Web]. Besides this data, the presented analysis offers a clear comparison between existing ECPs and should open new perspectives for further advances in this important research area.

CHAPTER 7

A New Generation of Effective Core Potentials from Correlated Calculations: $3d$ Transition Metal Series

Portions of this chapter also appeared in:

A New Generation of Effective Core Potentials from Correlated Calculations: $3d$ Transition Metal Series

J. Chem. Phys. **149**, 134108 (2017)

Contributions:

In the work described in this chapter, the author helped to generate CCSD(T) total energies of the high-accuracy ECPs for the $3d$ transition metal series.

7.1 Introduction

As detailed in chapters 5 and 6, we have advanced the use of correlated methods in ECP constructions in a systematic manner and we have proposed to increase the accuracy of ECP operators to a significantly higher level for first and second row elements [Ben17a; Ben18]. This was motivated by the needs of accurate correlated methods such as Coupled Cluster (CC), selected Configurations Interaction (CI) combined with quantum Monte Carlo (QMC) [Caf16] and full CIQMC (FCIQMC) [Boo13] that allow for correlated calculations of larger systems. Indeed, the accuracy of these calculations started to be hampered by the biases in the existing sets of ECPs [Foy14; SM13; KK17] and therefore much more reliable, accurate and tested ECPs became highly desirable.

Let us recall the key principles of our correlation consistent ECP (ccECP) constructions:

i) reproducing many-body spectrum of all-electron, relativistic, nearly-exact calculations for a subset of atomic states;

ii) adjusting the ECPs to capture the behavior of the original Hamiltonian in bonded situations for both equilibrium and non-equilibrium atomic geometries;

iii) using a simple semi-local form so that the potentials are given by a minimal or small set of gaussian functions; in addition, the potentials are bounded and smooth at the origin/nucleus; we also try to keep the tails in nonlocal channels short so as to diminish their actions in bonding regions;

iv) and, finally, establishing a database and systematic labeling together with data on obtained accuracy benchmarks; in addition, we keep the table open to further updates, benchmarks and refinements.

In this work, we apply these ideas to the 3d transition metal series. These atoms present their own set of challenges when compared with the first- and second-rows and therefore several considerations have to be taken into account. First, although the nominal valence levels are only 3d and 4s, accurate ECPs require the semicore 3s and 3p states to be included in the valence space as has been shown repeatedly in the past [Dol87; CM96]. This is straightforward to understand considering that 3s, 3p, and 3d states occupy the same electronic shell and therefore bonding and hybridization of 3d levels have a significant impact also on the semicore levels. Second, the elements in the middle of the 3d series are key constituents in many magnetic materials and therefore require rather a precise description of related electronic structure properties. Often the Hund's rule high-spin

atomic ground state is significantly modified in the bonded environment. This can involve several bonding mechanisms such as charge transfer from 4s states and/or d-s, d-p or d-d hybridizations with resulting partial or full quench of the local magnetic moment. Similarly, spin flip energies and large number of states/multiplicities of the open d shell are important in many catalytic chemicals and materials. Particular importance for these changes in the electronic structure is related $s \leftrightarrow d$ promotion energies as well as higher occupations of the d subshells in both neutral and ionized cases.

Several tables of ECPs for 3d transition metal atoms and beyond have been produced over the past three decades, for example, see Refs. [PC85; Dol87; Bur08; TN17; Kro16]. In particular, the full periodic table developed over a period of time by the Stuttgart-Bonn-Koeln collaboration [Dol87; Dol05; DC12] has been used quite extensively.

We expand upon these advances here and construct high accuracy ccECPs for the elements Sc to Zn. We use a very compact form with a few gaussians per channel and adjust the terms so that the potentials are finite and smooth in the origin as established previously [Ovc01; Bur08]. This form can, therefore, be used in many packages as well as with a variety of basis sets. In particular, we are interested in its usefulness also for periodic systems such as extended 2D materials or 3D solids calculations based on plane waves.

The all-electron and ECP calculations we use are based on scalar relativistic CC method with large basis sets and basis extrapolations. Although the calculations are very accurate, obviously they have limitations. The important factors are extrapolations to the infinite basis set limits, level of correlations in the CC method and scalar relativistic treatment. We estimate that these biases could sum up to ≈ 0.05 eV for almost all of the valence energy differences we consider. Therefore we deem ECPs with discrepancies within this threshold as being of comparable accuracy.

An additional point for future use is that we provide rather accurate values of atomic total energies for the ground and selected excited states. The high accuracy CC method with extrapolations provide estimations of the exact eigenvalues within systematic errors that vary from a few to about ≈ 10 mHa for the heaviest atoms. This data offers a useful checkpoint for many methods even for those that rely on sizable error cancellations in differences such as DFT. Since practical versions of DFT employ approximate functionals this also provides a valuable data for analysis of the performance in all-electron vs. ECP formulations since they can lead to non-negligible differences.

We point out that the presented ECPs include also core-core and core-valence correla-

tions and therefore they represent effective Hamiltonians that have almost as broad use as the original all-electron one, ie, for many purposes they aspire to be almost *universal*. That should be true essentially for any method and any valence property that does not require explicit core presence. We believe that this is important because it offers a well-defined platform to develop a systematic understanding of biases in a variety of approaches. For the transition metal series this is particularly relevant since fully correlated relativistic all-electron calculations of such systems are out of reach for just a few atom systems. The ECPs, therefore, enable to expand the system sizes that can be treated by many-body methods and at the same time provide reference framework to study systematic biases.

In what follows we first present the methods and objective functions used in the optimization process. The results section show the atomic and molecular properties of the constructed operators compared with all-electron and other existing tables. We comment and discuss several aspects of these new constructions in the conclusions section.

7.2 Methods

7.2.1 ECP Parametrization

The ECP electronic Born-Oppenheimer Hamiltonian (a.u) has the following form

$$H_{\text{val}} = \sum_i [T_i^{\text{kin}} + V_i^{\text{ECP}}] + \sum_{i < j} 1/r_{ij} \quad (7.1)$$

For this work, we use a semi-local ECP form with a minimal number of parameters [Bur07]

$$V_i^{\text{ECP}} = V_{\text{loc}}(r_i) + \sum_{\ell=0}^{\ell_{\text{max}}} V_{\ell}(r_i) \sum_m |\ell m\rangle \langle \ell m|, \quad (7.2)$$

where r_i is the radial distance of the i^{th} electron from the core's origin, and ℓ_{max} is the maximum angular momentum for the non-local projectors. In this work, we choose $\ell_{\text{max}} = 1$. The non-local terms contain the projectors on the angular momentum states with ℓm quantum numbers. The local term, V_{loc} , is chosen to cancel the Coulomb singularity at the

origin, i.e.

$$\begin{aligned}
V_{\text{loc}}(r) &= -\frac{Z_{\text{eff}}}{r} + \sum_{k=1} \beta_{\text{loc},k} r^{n_{\text{loc},k}-2} e^{-\alpha_{\text{loc},k} r^2} \\
&= -\frac{Z_{\text{eff}}}{r} + \frac{\beta_{\text{loc},1}}{r} e^{-\alpha_{\text{loc},1} r^2} + \beta_{\text{loc},2} r e^{-\alpha_{\text{loc},2} r^2} + \sum_{k=3}^4 \beta_{\text{loc},k} e^{-\alpha_{\text{loc},k} r^2}
\end{aligned} \tag{7.3}$$

where Z_{eff} is the effective core charge, $Z_{\text{eff}} = Z - Z_{\text{core}}$. To cancel the Coulomb singularity at the origin and force the first derivative to be zero at the origin, we fix $\beta_{\text{loc},1} = Z_{\text{eff}}$, $\beta_{\text{loc},2} = Z_{\text{eff}}\alpha_{\text{loc},1}$ with $n_{\text{loc},1} = 1$ and $n_{\text{loc},2} = 3$ and all other $n_{\text{loc},k} = 2$. The $V_{\ell}(r)$ potentials are expanded as follows

$$V_{\ell}(r) = \sum_{k=1}^2 \beta_{\ell,k} r^{n_{\ell,k}-2} e^{-\alpha_{\ell,k} r^2}, \tag{7.4}$$

where all the $n_{\ell,k}$ are fixed to $n_{\ell,k} = 2$. All variables are treated as optimization parameters in the minimization of a chosen objective function, except for the terms we have fixed as described above. In addition, a constraint that imposes a concave shape of the potential at the origin is imposed [Bur07]

$$\sum_{k=3}^4 \beta_{\text{loc},k} \alpha_{\text{loc},k} + \sum_{k=1}^2 \alpha_{\ell,k} \beta_{\ell,k} > 0, \quad \forall \ell. \tag{7.5}$$

7.2.2 Objective Function and Optimization Protocols

The fully correlated calculations of transition metals and heavier elements with very large basis sets and high accuracy methods such as CCSD(T) are very costly and, eventually, impractical. Therefore, we constructed an optimization strategy that introduces correlation into the ECP without explicitly calculating the correlation energies at each step of the optimization. This approach relies on the fact that the correlation energy for a given atomic state is similar across different ECPs, and the variation for the correlation energy error is significantly smaller than the variation in the HF error, as shown in Figure ???. For a particular excitation energy, the all-electron (AE) contribution can be written as $\Delta_g^{\text{AE}} = \Delta_{\text{HF}}^{\text{AE}} + \Delta_{\text{CC}}^{\text{AE}} + \Delta_{\text{CV}}^{\text{AE}} + \Delta_{\text{VV}}^{\text{AE}}$, where CC, CV, and VV represent the core-core, core-valence, and valence-valence correlation contribution to the excitation energy. When using an ECP, there are only valence electrons, so the ECP excitation energy can be written as $\Delta_g^{\text{ECP}} = \Delta_{\text{HF}}^{\text{ECP}} + \Delta_{\text{corr}}^{\text{ECP}}$. In order to match the AE excitation energy, the ECP can either adjust $\Delta_{\text{HF}}^{\text{ECP}}$ or $\Delta_{\text{corr}}^{\text{ECP}}$. By

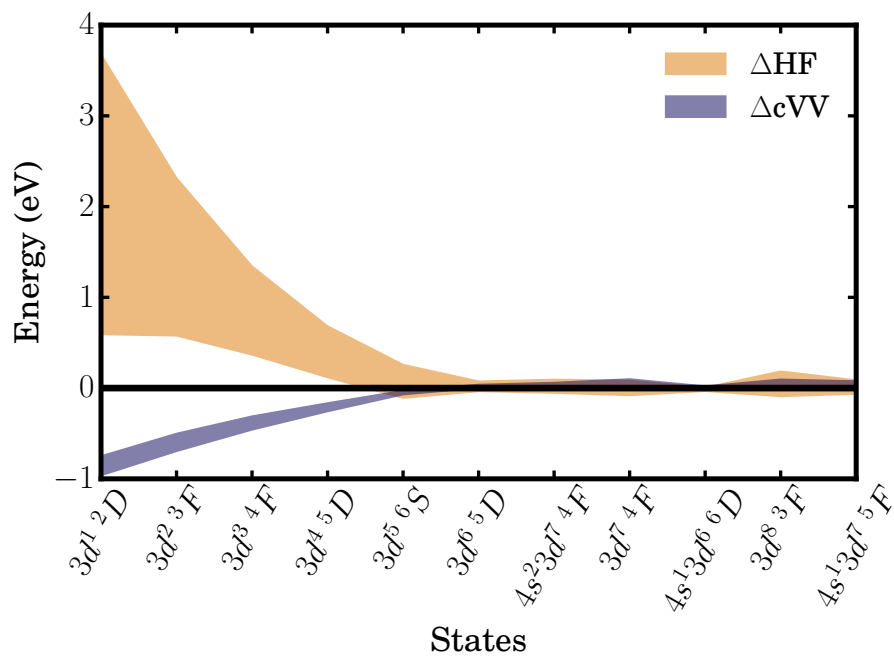


Figure 7.1 Spread of the contributions to the excitation energy for a variety of ECPs compared to all-electron for the Fe atom. ΔHF shows the variation in the HF errors, whereas ΔcVV shows the variation in the correlation energy error compared against the AE valence-valence correlation energy.

considering a variety of ECPs, we plot the spread of HF errors and the difference between the ECP correlation energy and the valence-valence correlation energy of the AE in Figure ?? for a variety of ECPs, including BFD [Bur07], STU [Dol87], eCEPP [TN17], and a variety of our constructed ECPs at both the HF and correlated levels. The spread of HF errors is larger than the correlation energies, which indicates that the HF contribution is more flexible across ECP parameterizations whereas the correlation energy varies much less. This is true whether the ECP form shows smooth and finite amplitudes at the nucleus or whether it contains $-Z_{\text{eff}}/r$ or $1/r^2$ divergences. This suggests that the $\Delta_{\text{HF}}^{\text{ECP}}$ can be adjusted while treating $\Delta_{\text{corr}}^{\text{ECP}}$ as a constant during an optimization procedure, which will result in $\Delta_g^{\text{ECP}} = \Delta_g^{\text{AE}}$ as is desired. This fact has been used in the optimization strategy below.

1. For a given atom, a set of all-electron (AE) atomic states is calculated using CCSD(T) method with uncontracted aug-cc-pwCVnZ-DK [BP05] basis set and extrapolated to the complete basis set limit (CBS) using data for $n=T, Q, 5$. We utilize the DKH Hamiltonian to incorporate scalar relativistic effects for all atomic states to generate our AE reference data. For the HF, we extrapolate to the CBS limit with

$$E_n^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + a \exp[-bn] \quad (7.6)$$

where n labels the basis set size, $E_{\text{CBS}}^{\text{HF}}$ is the CBS limit, and a and b are other fitting parameters and

$$E_n^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + \frac{c}{(n+3/8)^3} + \frac{d}{(n+3/8)^5} \quad (7.7)$$

where the correlation energy is defined as $E_n^{\text{corr}} = E_n^{\text{CCSD(T)}} - E_n^{\text{HF}}$. It is known that the atomic states of some transition metals such as the ground state of V are incorrectly described by a single reference using real spherical harmonics. In order to obtain the correct atomic symmetries, we generate ‘‘symmetry equivalenced’’ orbitals by state-averaging the irreducible representations (irreps) of D_{2h} for a particular atomic state [BT88]. We use the resulting natural orbitals to calculate the correlation energies with CCSD(T). This way, we are able to obtain reference AE energy gaps from a consistent description of the atomic state.

2. The initial construction of the ECP is launched by reproducing the scalar relativistic Dirac-Hartree-Fock (DHF) spectrum. We generate a large set of random ECPs and optimize each using the nonlinear DONLP2 code by Spelluci [Spe]. The objective

function, Σ , to be minimized over the spectrum S is given by

$$\Sigma(S) = \sum_{s \in S} w_s (\Delta E_{\text{ECP}}^{(s)} - \Delta E_{\text{AE}}^{(s)})^2, \quad (7.8)$$

where $\Delta E^{(s)}$ is the excitation energy for state s relative to the neutral ground state and w_s is the weight assigned to that excitation. The spectrum is optimized using numerical DHF code [Shi] to avoid basis set errors and to speed up the calculations. The set of states considered includes the neutral and singly ionized $s \leftarrow d$ excitations, as well as further ionizations and the anion ground state. These states are chosen due to the fact that the low-lying $s \leftarrow d$ transitions and various oxidation states play a crucial role in transition metal chemistry [Dol87]. The resulting ECPs are further refined as follows.

3. The DHF spectrum optimized ECP with the lowest objective function in step 2 is used to calculate the correlation energies using CCSD(T) for the states mentioned. This defines $\Delta_{\text{corr}}^{\text{ECP}}$, i.e. the correlation energy contribution to the gaps. For the ECPs, we utilize the same uncontracted aug-cc-pwCVnZ-DK basis set. It should be noted that for the ECP calculations, no DKH Hamiltonian is used since we intend to build the scalar relativistic effects directly into the effective potential.
4. It is clear that DHF spectrum optimized ECP atomic excitation energies $\Delta E_{\text{ECP}}^{(s)}$ are not going to match the corresponding AE excitation energies $\Delta E_{\text{AE}}^{(s)}$ when a post-HF method such as CCSD(T) or CI is used. However, it is possible to obtain $\Delta_g^{\text{ECP}} = \Delta_g^{\text{AE}}$ by only matching $\Delta_{\text{HF}}^{\text{ECP}}$ to *shifted* $\Delta_{\text{HF}}^{\text{AE}}$:

$$\Delta_{\text{HF}}^{\text{ECP}} = \Delta_{\text{HF}}^{\text{AE}} + \epsilon \quad (7.9)$$

where ϵ is given as:

$$\begin{aligned} \epsilon &= \Delta_{\text{corr}}^{\text{AE}} - \Delta_{\text{corr}}^{\text{ECP}} \\ &= (E_{\text{corr,excited}}^{\text{AE}} - E_{\text{corr,ground}}^{\text{AE}}) - \\ &\quad (E_{\text{corr,excited}}^{\text{ECP}} - E_{\text{corr,ground}}^{\text{ECP}}) \end{aligned} \quad (7.10)$$

Here AE correlation energy refers to the total correlation energy (CC+CV+VV) of the state. The new shifted gaps are inserted into the objective function in equation 7.8 to

be minimized.

5. The steps 3-4 are iterated until a self-consistent ECP is obtained. For each step ϵ is re-evaluated using the ECP from the last iteration. In every iteration, the ECP parameters are randomly perturbed within ≈ 1 -2% of parameter values to ensure better scanning of possible local minimas around the current values of the parameters. Usually, a set of self-consistent parameters are obtained within a few iterations.

In cases where pure energy consistency does not result in an acceptably transferable ECP (as described in section 7.3), we reduce the spectrum by removing very highly ionized states and add additional constraints to improve the transferability. Note that the full spectral fits include also core-core and core-valence correlations that become quite significant for the highly ionized states and therefore they have tendency to steer the ECP from the optimal valence-valence description. This is especially true for our ECP form with small number of variational parameters. Although the bias is typically small (say, 0.1 eV) we opted for further refinements. This is accomplished by including one-particle eigenvalue discrepancies into the objective function. It directs the highly ionized states to be less influenced by the corresponding correlations so as to be increase the weight of the HF character for these states. In these cases, we utilize a new objective function, Γ , given as

$$\Gamma(S) = \Sigma(S) + \gamma \sum_{\ell} (\epsilon_{\ell}^{\text{ECP}} - \epsilon_{\ell}^{\text{AE}})^2 \quad (7.11)$$

where the first term is the initial spectral objective function and ϵ_{ℓ} is one particle eigenvalue. The eigenvalues are weighted by γ to allow the spectrum to be minimized while keeping the one-particle eigenvalues reasonably close to the corresponding AE ones. With few iterations one can find a compromise that reproduces a large part of the spectrum as well achieves transferability in molecular calculations. Clearly, further improvements are possible but in this work we are mostly concerned with the demonstration of principle and with providing simple ECPs that fulfill the accuracy criteria.

We note that the objective function utilized here for the $3d$ transition metals is different than what was used for the first and second row elements [Ben17a] & [Ben18]. As we have seen, energy consistency alone (our ccECPs) does not in general result in an ECP with accurate atomic properties as well being transferable to molecular systems.

In order to improve transferability, we include HF properties into our objective function as we did in the second row. Due to the large number of valence electrons for the transition metals, we found that our objective functions only needed a modest level of HF information to improve overall transferability as opposed to the second row, where the smaller number of electrons and large cores required a more complicated construction.

7.3 Results

We present results comparing our correlation consistent ECPs, labeled ccECP, to various other core approximations relative to the all-electron calculation for both the atomic spectra as well as monohydrides and monoxides. The all-electron reference in all following calculations is spin-restricted CCSD(T) where we have correlated all electrons in the system, i.e. no core orbitals have been frozen. To include relativity, we utilize a 10^{th} order Douglas-Kroll-Hess Hamiltonian [RW04]. To fully correlate the core, we use an uncontracted aug-cc-pwCV n Z-dk basis [BP05] and extrapolate to the complete basis set limit to minimize basis set error. For the ECP calculations, we perform CCSD(T) calculations without DKH (since the effects are built into the effective potential) with the same aug-cc-pwCV n Z-dk basis sets and extrapolate to the complete basis set limit. We compare our ECPs against an all-electron uncorrelated core, which does not allow for any excitations from the Neon core in the CCSD(T), as well as other ECPs. The uncorrelated core approximation, labeled UC, only accounts for valence-valence correlation, and any core-core or core-valence correlation is absent despite it being an all-electron calculation. The other ECPs include Burkatzki-Filippi-Dolg ECPs (BFD) [Bur07] which are energy-consistent DHF ECPs designed for use in QMC, Stuttgart ECPs (STU) [Dol87] which are energy-consistent DHF ECPs, and the recently constructed Trail-Needs ECPs (eCEPP) [TN17] which use a shape and energy consistent scheme to construct correlated ECPs.

Obviously, the methods we use have their accuracy limits. In particular, we rely on scalar relativistic approach with averaged spin-orbit effects. This introduces additional bias of ≈ 0.025 eV as a representative value that we estimated from accurate spin-orbit correlated calculations of atomic excitations[Liu15]. Further important sources of bias are the sizes of basis sets (alleviated by extrapolations) and level of correlation captured by the CCSD(T) method. On smaller systems where we were able to push the basis set limits

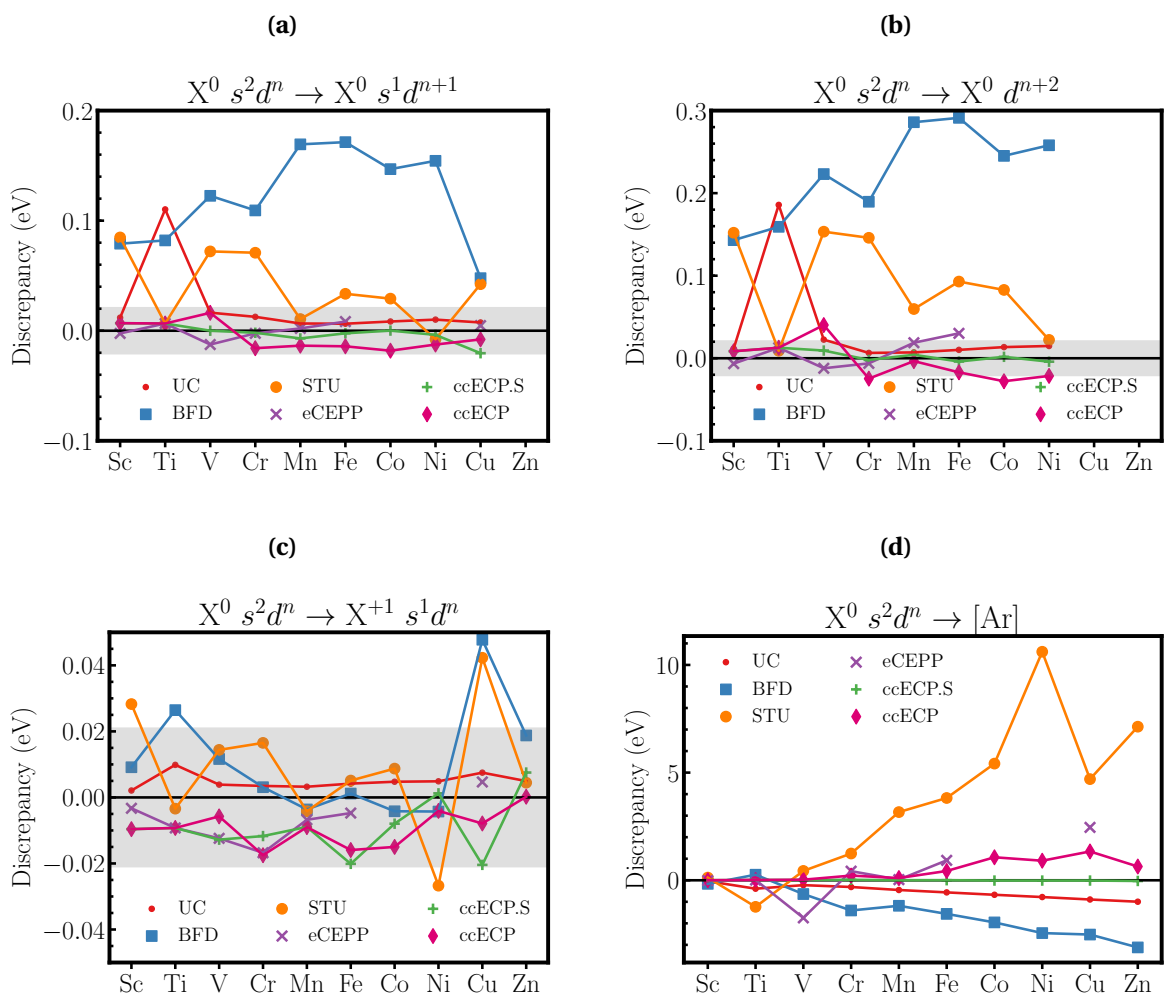


Figure 7.2 Discrepancies for the various core approximations compared to all-electron CCSD(T) for select states. Each state discrepancy uses the neutral $s^2 d^n$ occupation as the reference, which is the neutral ground state for each transition metal except for Cr and Cu, which have $s^1 d^5$ and $s^1 d^{10}$ ground states correspondingly. (a) the neutral $s^2 d^n \rightarrow s^1 d^{n+1}$ excitation. (b) the neutral $s^2 d^n \rightarrow d^{n+2}$ (c) the ionization from $s^2 d^n \rightarrow s^1 d^n$ (d) the ionization from $s^2 d^n \rightarrow [\text{Ar}]$. The shaded gray window in each figure indicates a discrepancy of half of chemical accuracy in either direction from the all-electron reference. We note that for Sc and Ti, our final ccECP is equivalent to our spectral ccECP.S.

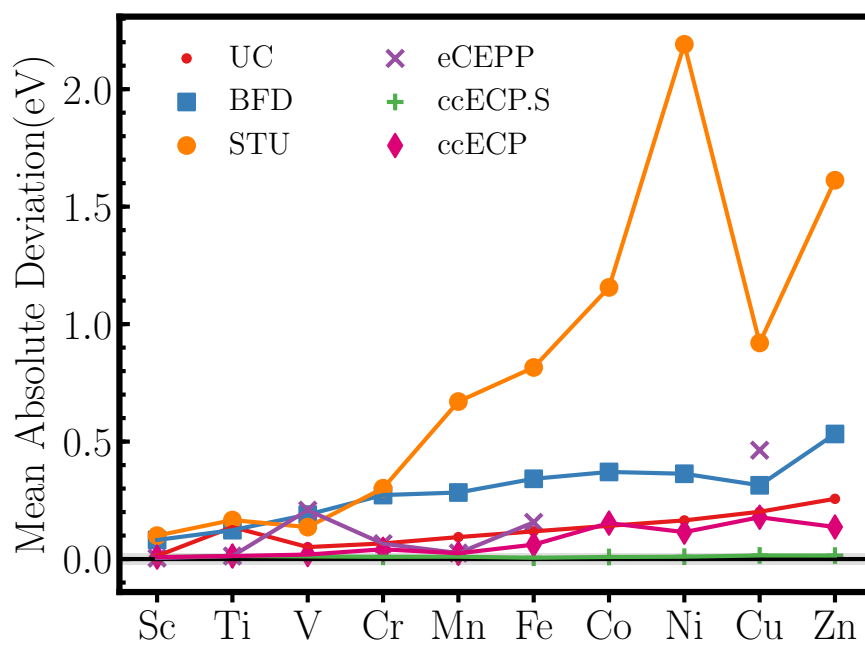


Figure 7.3 Mean Absolute Deviation, or MAD, for the TMs considering a large part of the spectrum from [Ar] up to low-lying neutral excited states and the anion. The shaded region of half of chemical accuracy is not visible on this scale. We note that for Sc and Ti, our final ccECP is equivalent to our spectral ccECP.S

up to $n = 6, 7$, ie, 6Z,7Z and level of correlation up to CCSDT(Q) level (triples explicitly and quadruples perturbationally). These limited calculations suggest that it is difficult to ascertain accuracies better than roughly 0.02 eV for energy differences. In addition, for the heaviest atoms, we were not able to carry out the full sequence of basis set calculations up to 5Z, so this threshold is probably mildly higher. Clearly, ECPs with discrepancies below these inherent biases of ≈ 0.05 eV (that include also a number of our constructions presented below) are certainly interesting for methodology reasons. However, the ECPs within such bound should be considered as being of comparable quality.

We briefly discuss the results for every element in the following subsections. In section 7.3.1, we present an overview of the atomic spectra for all transition metals. In sections 7.3.2-7.3.11, for every atom we show the atomic spectrum discrepancies and transferability tests on monohydrides and monoxides. In these molecules, we test various core approximations near and out of equilibrium geometries, i.e., compressed bond lengths. In general, we show results for our spectral-only ECP (labeled ccECPS) as well as the ECP that compromises a small part of the spectrum for increase in the overall transferability (this is labeled as ccECP). For oxygen ECP, we use the ECP from the corresponding table; our oxygen ccECP can be found in our previous work[Ben17a]. Although for some of the oxide molecules we were unable to converge with 5Z basis sets due to technical difficulties. However, we note that in terms of the binding curves discrepancies are well converged at the VQZ level. Therefore, for each molecule we show only the largest basis set that we converged as opposed to the CBS extrapolation.

All ECP parameters are given in Table 7.4 and Table 7.5 while our hydrogen ECP is given in Table 7.2.

7.3.1 Atomic Spectra

We first provide an overview of the atomic spectra. The optimized spectrum for ccECPS included the neutral s^2d^n , s^1d^{n+1} , d^{n+2} , the first ionized s^1d^n and d^{n+1} , all ionizations down to the Ar core, and the [Ne]3s² state. As we will see in future sections, the inclusion of these highly ionized states can lead to non-negligible biases and a decrease in the overall transferability. Therefore, in cases where this full spectrum is insufficient, we add in constraints and work with a reduced spectrum as described in section 7.2.2.

Figure 7.2 shows the errors across the entire transition metal series for select atomic states ranging from some low lying excited states to a highly ionized state. In Figs. 7.2a and

7.2b, we show the neutral $s \leftrightarrow d$ transitions from the $s^2 d^n$ to $s^1 d^{n+1}$ and d^{n+2} , respectively. Both STU and BFD, result in discrepancies on the order of 0.1 eV throughout the entire transition metal series for these low lying atomic states. UC, eCEPP, and our ccECPS and ccECP all result in discrepancies within half of the chemical accuracy, indicated by the shaded region. Figure 7.2c shows the $s^2 d^n$ to $s^1 d^n$ ionization, which is well described with all core approximations. It is important to note that STU and BFD include many of these states directly into their optimization as well. However, the neglect of directly introducing correlation into the construction can result in large errors in the atomic spectra, despite being well optimized at the HF/DHF level as illustrated by both the eCEPP and our ccECPS.

As the ionization level is increased, all previous core approximations begin to show larger discrepancies with the AE reference. In order to illustrate this, we show the ionization down to the Ar core in Figure 7.2d. The STU and BFD ECPs we see significantly larger discrepancies, ranging from 1-10 eV throughout the series. The correlated eCEPPs show an improvement but still result in discrepancies of a few eV. The all-electron UC approximation results in a slightly increasing error as the atomic number increases up to roughly 1 eV for Zn. Our energy consistent ECPs (ccECPS), however, are able to maintain a uniform accuracy within half of the chemical accuracy for each atomic state and for all ionizations.

In Figure ??, we show the mean absolute deviation,

$$\text{MAD} = \frac{1}{N_s} \sum_{s=1}^{N_s} |\Delta_s^{\text{ECP}} - \Delta_s^{\text{AE}}| \quad (7.12)$$

across all of the states including the low-lying $s \leftrightarrow d$ transitions, anions, and ionizations down to the semicore [Ar]. Our ccECPS ECPs are able to maintain MADs of nearly 0.01 eV for all transition atoms. In the following sections focused on the individual atoms, we investigate energy consistency influences the transferability of the ECPs for each element. Our adjusted ccECP constructions, which slightly compromise the atomic spectrum when compared to the ccECPS (as illustrated in Fig. ??), still maintain higher accuracy for the MADs when compared against the other tested ECP constructions. In Table 7.1, we summarize the MADs for each atom and ECP. LMAD indicates we are only considering the low lying atomic spectrum, which includes the electron affinity, ionization potential, and the neutral and singly ionized $s \leftrightarrow d$ transitions. The MAD corresponds to the LMAD, as well as additional ionizations down to the Ar core. Details about the atomic excitation energies can be found in the supplemental material.

Table 7.1 A summary of the atomic state discrepancy data. For each atom, we provide mean absolute deviations (MADs). LMAD corresponds to the low-lying atomic spectrum, which includes the electron affinity, ionization potential, and the neutral and first ionized $s \leftrightarrow d$ transitions. The MAD includes the LMAD as well as ionizations down to the Ar core. A * indicates that the our final ccECP is equivalent to our spectral only ccECPs, as described in the atomic sections. N/A indicates that the particular ECP does not exist.

Atom	Quantity	UC	BFD	STU	eCEPP	ccECPs	ccECP
Sc	LMAD	0.009606	0.077487	0.093684	0.002683	0.006727	*
	MAD	0.013306	0.081214	0.099679	0.004097	0.008280	*
Ti	LMAD	0.014724	0.106347	0.09654	0.009031	0.008538	*
	MAD	0.030568	0.136353	0.123009	0.165998	0.012846	*
V	LMAD	0.016435	0.124708	0.086546	0.007886	0.006087	0.022334
	MAD	0.050886	0.189096	0.135905	0.207941	0.011069	0.019089
Cr	LMAD	0.011010	0.10847	0.089896	0.007706	0.006906	0.014199
	MAD	0.066494	0.272375	0.30147	0.065267	0.009445	0.041598
Mn	LMAD	0.005000	0.159928	0.02406	0.007142	0.007812	0.009088
	MAD	0.093381	0.283245	0.67035	0.025464	0.009995	0.024478
Fe	LMAD	0.007768	0.163985	0.049415	0.016317	0.007271	0.010904
	MAD	0.117865	0.341598	0.815454	0.156170	0.005497	0.061351
Co	LMAD	0.009465	0.137939	0.043948	N/A	0.004171	0.016915
	MAD	0.141522	0.370784	1.156205	N/A	0.008795	0.152958
Ni	LMAD	0.010379	0.143013	0.022406	N/A	0.002937	0.009916
	MAD	0.164481	0.363034	2.190893	N/A	0.009971	0.114590
Cu	LMAD	0.008540	0.026354	0.031670	0.003207	0.016600	0.007160
	MAD	0.201222	0.313908	0.919896	0.463213	0.015345	0.177998
Zn	LMAD	0.007157	0.112360	0.013524	N/A	0.008612	0.011117
	MAD	0.256321	0.532372	1.612408	N/A	0.014890	0.136370

7.3.2 Scandium

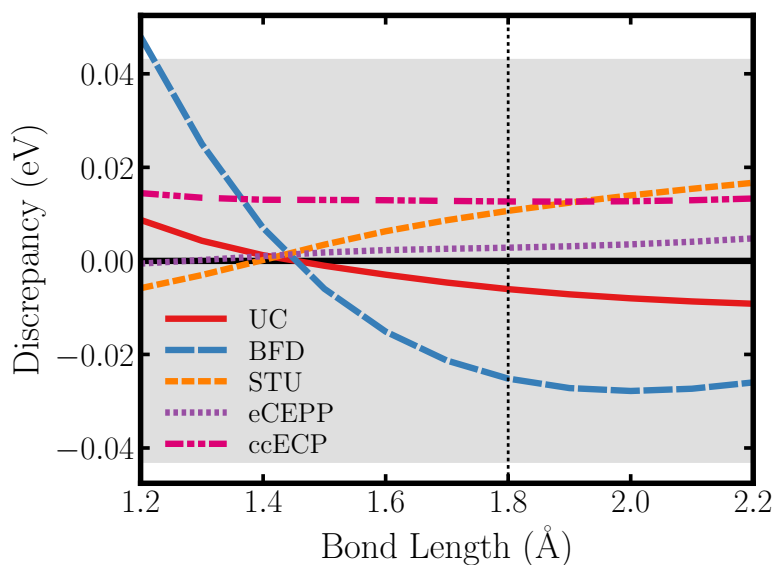
In the case of Sc, the atomic and molecular data is given in Table 7.1 and Fig. 7.4, respectively. Our spectral-only ccECPS results in a significant improvement over the atomic spectra from previously existing tables such as STU and BFD. In fact, when comparing to the all-electron uncorrelated core UC, we see an improvement for the low-lying spectrum (indicated by LMAD) as well as the full spectrum ionized down to the Ar semi-core. Compared to recently derived eCEPP [TN17], our spectral errors are only slightly larger, presumably due to our significantly smaller number of gaussian functions as well as shorter radial range of nonlocal channels. Since for Sc the higher order effects such as spin-orbit can reach 0.010 eV [Liu15], discrepancies below this level do not provide a genuine quality measure as we have noted also above.

For the molecular calculations, we plot discrepancies from the AE CCSD(T) binding curve for a range of bond lengths. The compressed bond regime is plotted near the dissociation threshold, i.e., where the binding energy approaches zero while the vertical line indicates the equilibrium bond length. For ScH, we see that all ECPs result in binding discrepancies that are well within chemical accuracy with regard to the AE results, indicated by the shaded region. Note that STU, BFD, and UC discrepancies vary as functions of the bond length. That might cause some shifts in the predicted vibrational frequencies, whereas a flat discrepancy should lead to better accuracy for vibrational properties. The ScO molecule probes the charge transfer and polar bond regime showing that both STU and BFD exhibit significantly larger deviations from the AE potential energy surface. Our ccECPS results in a relatively flat discrepancy throughout the entire binding region and a very small error for the binding energy. Due to the quality of the constructed operator, we decided that no further refinements to the ccECPS were needed. The parameters for this optimal ccECPS are given in Table 7.4.

7.3.3 Titanium

In the case of Ti, the atomic and molecular data is given in Table 7.1 and Fig. 7.5, respectively. For the atomic spectrum, our ccECPS outperforms all other core approximations, for both the low lying atomic spectra (LMAD) as well as all ionizations down to the Ar semi-core. We see that STU and BFD result in comparable MADs of roughly 0.1 eV. For the low-lying spectra, both the eCEPP and ccECPS have marginally smaller discrepancies than UC showing that

(a) ScH binding curve discrepancies.



(b) ScO binding curve discrepancies.

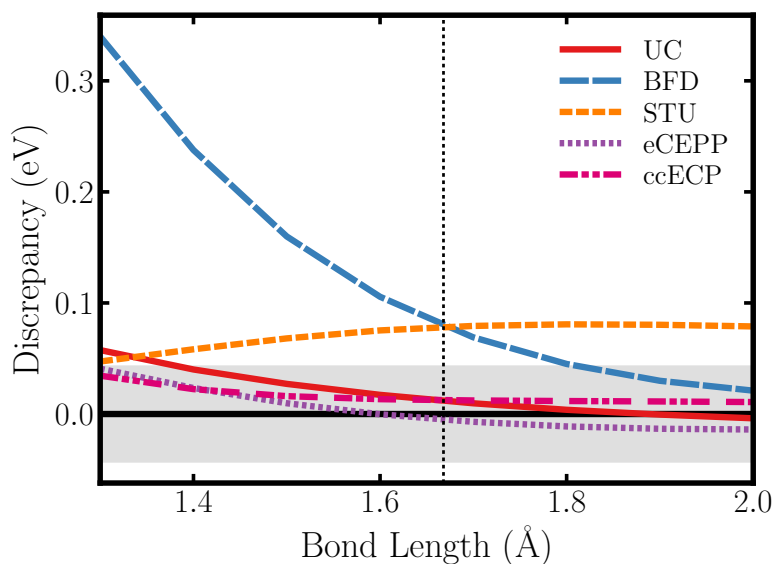


Figure 7.4 Binding energy discrepancies for (a) ScH and (b) ScO molecules. The binding curves are relative to the AE CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

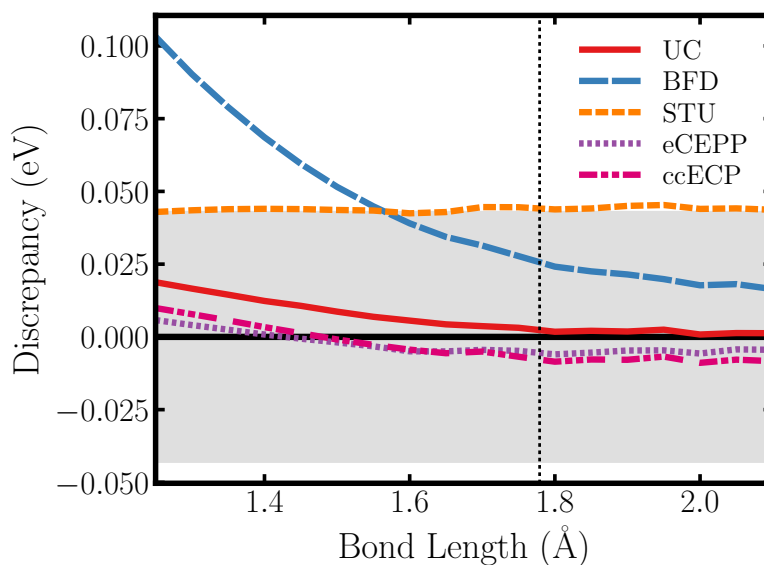
some of the core-valence correlations have been captured. Note that our ccECPS maintains uniform accuracy throughout the entire spectrum, resulting in a MAD of 0.013 eV, see the supplementary material for details about individual states.

When considering the hydride molecule, the BFD shows significant underbinding for the compressed geometries. On the other hand, STU maintains uniform accuracy throughout the entire binding region with mild 0.04 eV underbinding. UC, eCEPP, and our ccECPS are all very comparable for TiH, showing deviations within 0.02 eV for all bond lengths. For TiO with polar bond, we find that BFD is inadequate to describe the binding even at equilibrium and the error reaches up to 0.4 eV near in the short bond region. STU is also underbound near equilibrium, and has the opposite behavior to BFD in that the error decreases as the bond is compressed. While UC and eCEPP are well within chemical accuracy near equilibrium, each begins to underbind as the bond length is compressed. Near dissociation, both UC and eCEPP reach region outside the chemical accuracy. Our ccECPS, on the other hand, is well within the desired error margin throughout the entire bonding region. Additionally, the ccECPS has an extremely flat discrepancy near the equilibrium. Considering the accuracy of the constructed ccECPS, we did not pursue any refinements. The parameters for the Ti ccECPS are given in Table 7.4.

7.3.4 Vanadium

The atomic and molecular data for the V ECPs is given in Table 7.1 and Fig. 7.6, respectively. For the atomic spectra, our ccECPS has significantly smaller discrepancies compared to all other core approximations, including the all-electron UC approximation, for both the low-lying spectra and all ionizations. If we consider VO for our ccECPS, we see that we have quite favorable binding properties compared to other core approximations. Both BFD and STU exhibit relatively large errors throughout the entire bonding region, including near equilibrium. Both eCEPP and UC begin to underbind to nearly 0.1 eV near the dissociation threshold, whereas our ccECPS is well within chemical accuracy. However, when considering the hydride, we found that our spectral-only ECP overbinds by a small amount of ≈ 0.04 eV that reaches the borderline at dissociation. In fact, this is still acceptable considering the systematic biases present although less accurate than the eCEPP binding curve. Note that ccECPS ECPs fits a significant part of the atomic a spectrum that includes ionizations past the valence electrons and down into the $3s$ and $3p$ semi-core states. However, this nudges the ECP operator away from transferability optimum as can be observed on the

(a) TiH binding curve discrepancies.



(b) TiO binding curve discrepancies.

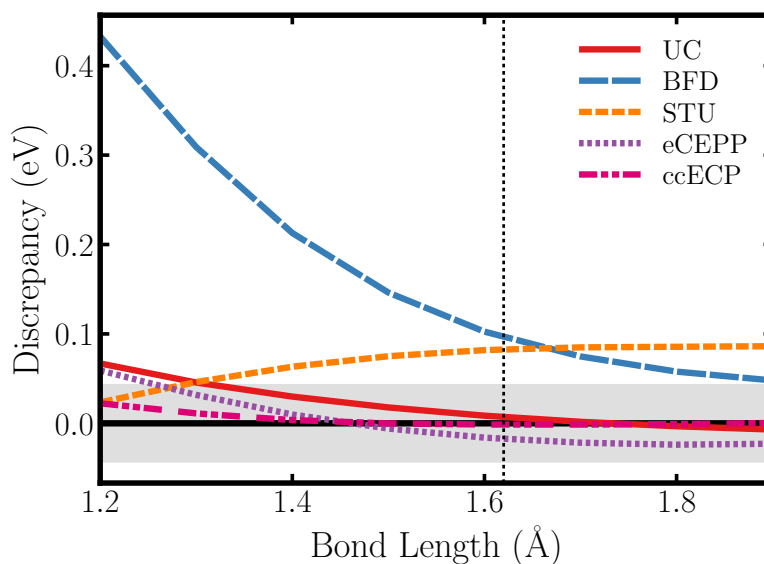


Figure 7.5 Binding energy discrepancies for (a) TiH and (b) TiO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

VH molecule. Therefore, we refined ccECPS by reducing the considered spectrum and modifying our objective function as described in part 7.2.2, resulting in refined ccECP. The charge-transfer physics is unchanged and it is almost identical to the spectral ccECPS. However, we see a significant improvement for the hydride molecule. In terms of the spectrum, we slightly compromise the low-lying spectrum (LMAD) while maintaining roughly the same overall MAD. The parameters for our ccECP of V are included in Table 7.4.

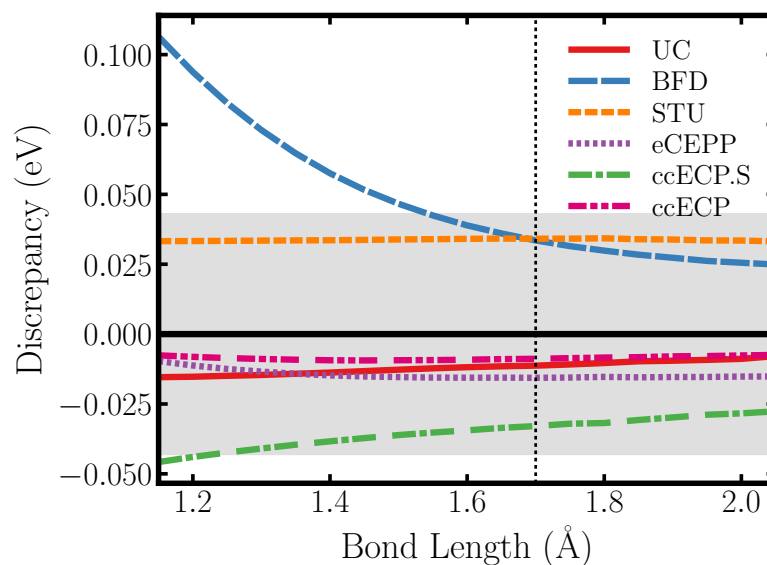
7.3.5 Chromium

The atomic and molecular data for Cr are given in Table 7.1 and Fig. 7.7, respectively. Our spectral ECP, ccECPS, has a significantly improved spectrum when compared to all other core approximations, including the all-electron UC approximation. The low-lying spectrum (LMAD) is slightly improved over eCEPP, while the MAD across the spectrum above [Ar] is significantly lower. The ccECPS performs reasonably well for CrH, being within chemical accuracy to AE for the entire binding region. However, when we consider the oxide CrO, we see that the ccECPS shows clear tendency to overbind, up to ≈ 0.08 eV near the dissociation threshold. Our refined ccECP, optimized using Eq. 7.11, shows well balanced atomic spectrum with mild increase in LMAD. In molecular calculations, we see a modest improvement for CrH and a very satisfactory reduction in overbinding when compared to ccECPS. The final ccECP is well within chemical accuracy throughout the entire binding region, including near the dissociation threshold. The parameters for our ccECP are given in Table 7.4.

7.3.6 Manganese

The atomic and molecular results for Mn are given in Table 7.1 and Figure 7.8, respectively. For the spectral ccECPS, we find similar uniform accuracy for both the low-lying atomic and full spectrum to [Ar]. It is interesting to see that despite this fidelity we find the overall transferability to be lacking. While ccECPS for MnH is within chemical accuracy for most of the binding curve we observe a slightly larger discrepancy near the dissociation threshold. This error is more severe for MnO with overbinding by roughly 0.04 eV near equilibrium that increases to 0.11 eV at dissociation. Therefore, despite having the best atomic spectrum, the ccECPS MnO molecule has the largest discrepancies from AE when compared to all other core approximations. We clearly observe that within the use parameterization, the

(a) VH binding curve discrepancies.



(b) VO binding curve discrepancies.

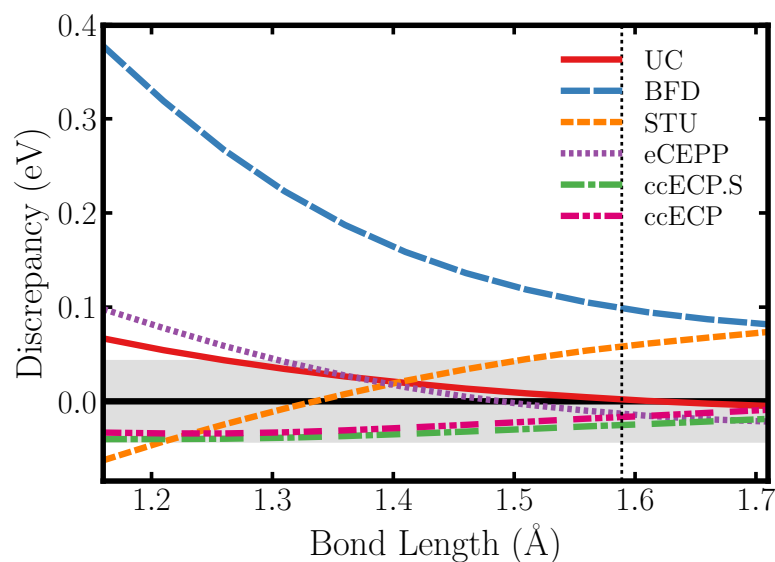
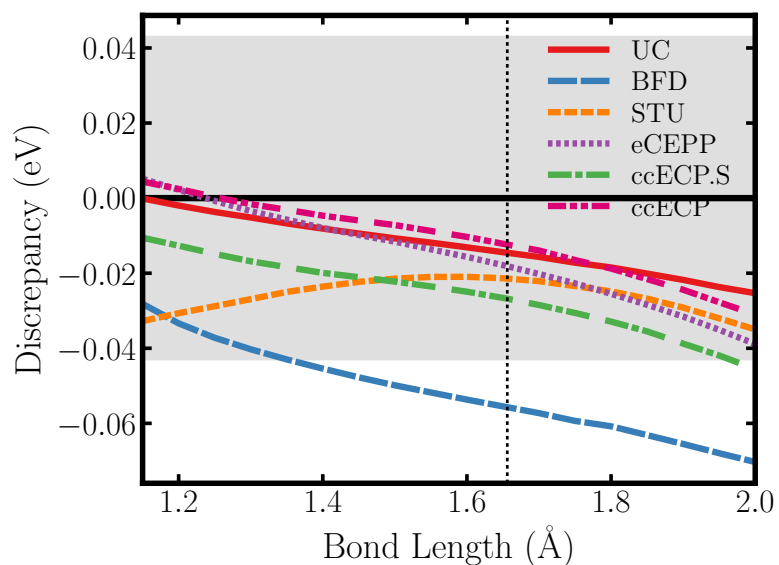


Figure 7.6 Binding energy discrepancies for (a) VH and (b) VO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

(a) CrH binding curve discrepancies.



(b) CrO binding curve discrepancies.

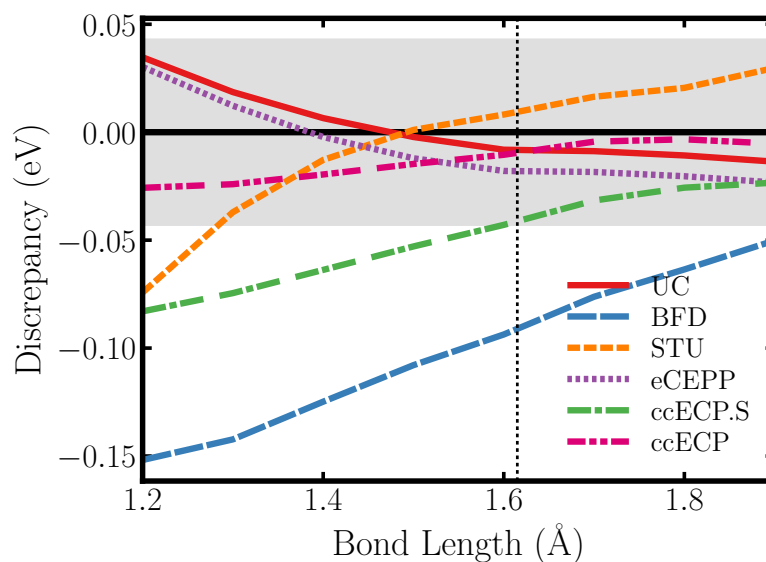


Figure 7.7 Binding energy discrepancies for (a) CrH and (b) CrO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

atomic accuracy and molecular accuracy pull in rather opposite directions.

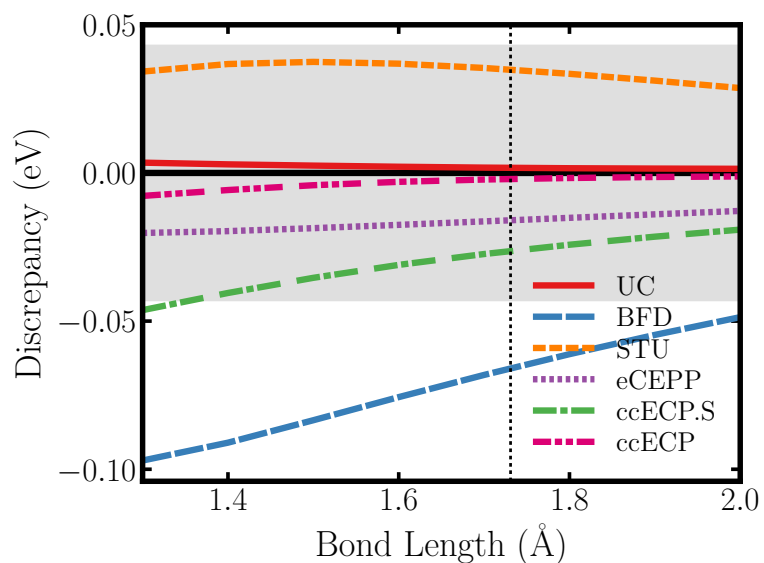
We improve upon our spectral only ccECPS and reoptimize via Eq. 7.11 to obtain our final ccECP. In terms of the atomic spectrum, we slightly increase the LMAD to ≈ 0.01 eV. Some error of 0.1 eV is introduced to the highly ionized [Ar] semi-core excitation. This increase results in a MAD on the entire spectrum of only 0.024 eV, which outperforms all other core approximations including the all-electron UC. For both the hydride and oxide we see a clear improvement. For MnH, the ccECP has below a 0.01 eV discrepancy for all bond lengths. The most drastic improvement comes for the oxide molecule, where the discrepancy is bounded by the chemical accuracy throughout the entire curve. The resulting ccECP has excellent atomic and molecular properties and the parameters are given in Table 7.4.

7.3.7 Iron

The atomic discrepancies for Fe atom are given in Table 7.1. Our spectral ccECPS produces an atomic spectrum with uniform accuracy throughout the entire valence spectrum, with a low-lying spectrum MAD of 7 meV, and 5 meV for the entire spectrum. In terms of the low-lying spectrum (LMAD), the ccECPS is comparable to the all-electron UC approximation. In addition, the ccECPS has a much lower MAD for the entire spectrum. While the ccECPS is well within chemical accuracy for FeH, when we consider the polar bond, illustrated by the FeO molecule, we find significant overbinding across the entire binding region (Figure 7.9). In fact, the ccECPS overbinds well outside chemical accuracy for most of the bond lengths. It should be noted that complications related to molecules with high-spin TM atoms are clearly visible for BFD and STU ECPs with significant discrepancies in the binding energies as well as in the shapes of the potential energy surfaces.

We achieve a remarkable balance between the atomic spectrum as well as the molecular properties with our refined ccECP. The ccECP results in an increase of only 3 meV on the LMAD, and we find larger errors only for the highest ionizations. It should be noted that the absolute error of 0.43 eV for the ionization to the [Ar] semi-core corresponds to relative error of only ≈ 0.08 % since the excitation is sizeable, 560 eV. Additionally, despite the compromise in the spectrum, the ccECP outperforms all other ECPs on the LMAD and all the core approximations on the entire MAD. The molecular properties are significantly improved for the ccECP. For FeH, there is near perfect agreement on the binding curve, both near equilibrium as well as near dissociation. For FeO, we see a significant reduction

(a) MnH binding curve discrepancies.



(b) MnO binding curve discrepancies.

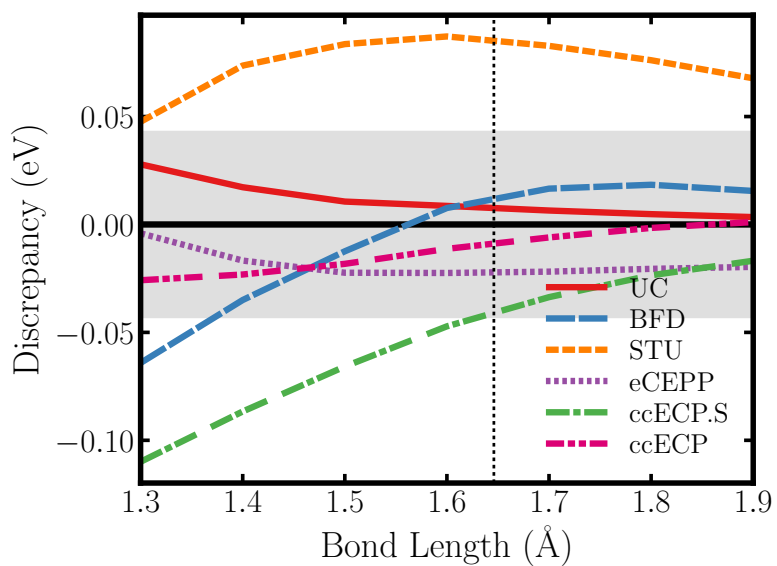


Figure 7.8 Binding energy discrepancies for (a) MnH and (b) MnO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

in the overbinding, putting the binding discrepancy within the chemical accuracy of the AE CCSD(T) binding curve. The parameters for this ccECP are given in Table 7.5.

7.3.8 Cobalt

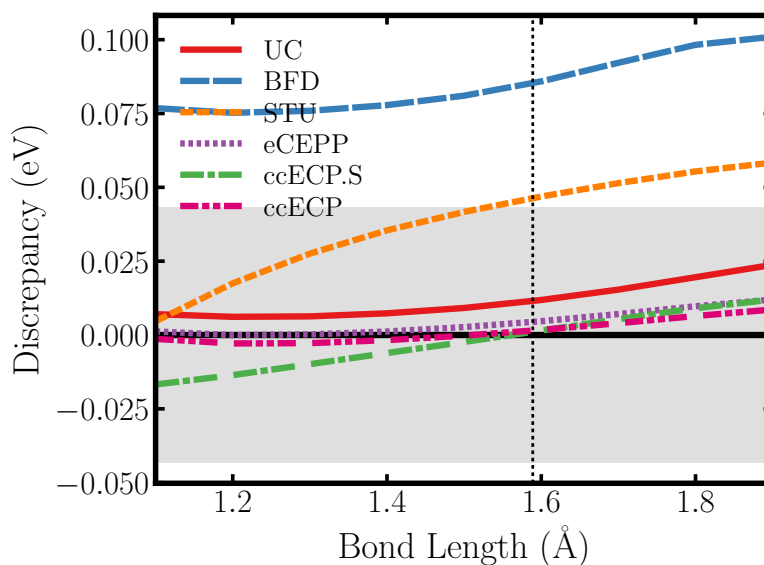
The atomic and molecular data for Co are given in Table 7.1 and Fig. 7.10, respectively. Our spectral construction ccECPS shows that significant improvements when compared with other core approximations. We are able to achieve for both the LMAD and MAD significantly lower discrepancies than UC, BFD, and STU. In terms of CoH, the ccECPS construction looks very satisfactory. CoH results in a discrepancy that is well within chemical accuracy for ccECPS across the entire binding region. However, for CoO we see that the spectral construction is less accurate and as in previous cases it results in significant overbinding. Note that other available ECPs have a similar behavior, resulting in overbinding over 0.1 eV near the dissociation threshold.

When we introduce additional constraints into our objective function, see section 7.2.2, we are able to obtain a dramatic improvement in the overall transferability. With regards to the spectrum, our ccECP is comparable to the UC approximation for the low-lying spectrum (LMAD) and even slightly better for the entire atomic spectrum (signaling a complicated landscape of the objective function with many minimas). For CoH, the binding energy discrepancy is comparable to our ccECPS, but the overall curve is flatter. The most profound improvement comes with the CoO polar bond, where we are able to obtain a nearly flat discrepancy throughout the entire binding region and very marginal 0.01 eV overbinding. Note the overall major boost in accuracy when compared to the other ECPs we studied. The parameters for the ccECP are given in Table 7.5.

7.3.9 Nickel

We show the atomic and molecular data for Ni in Table 7.1 and Fig. 7.11, respectively. Optimization for ccECPS results in an accurate spectrum with a 3 meV LMAD for the low-lying spectrum and 10 meV MAD for the entire spectrum. These discrepancies are significantly smaller than for the other core approximations of UC, BFD, and STU. For NiH, we see significant issues with BFD, whereas all other core-approximations agree with AE to within chemical accuracy. We find more sizeable errors for NiO for BFD, STU, and also for our spectral ccECPS.

(a) FeH binding curve discrepancies.



(b) FeO binding curve discrepancies.

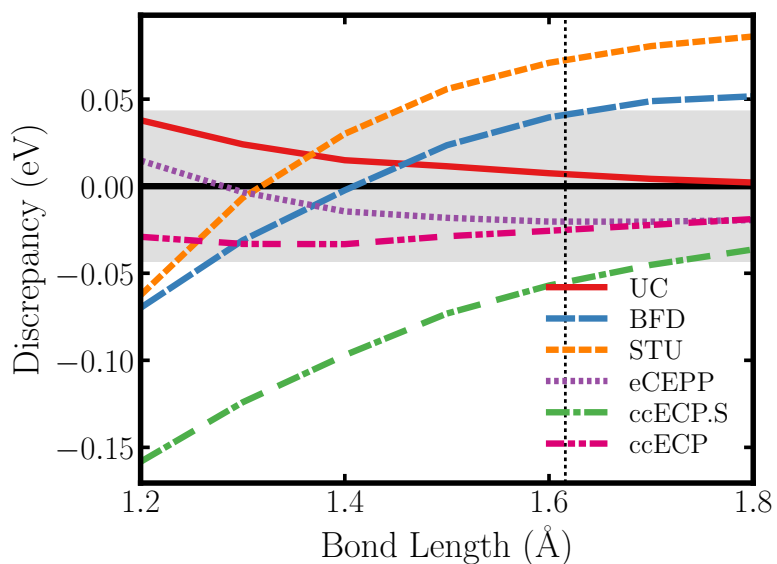
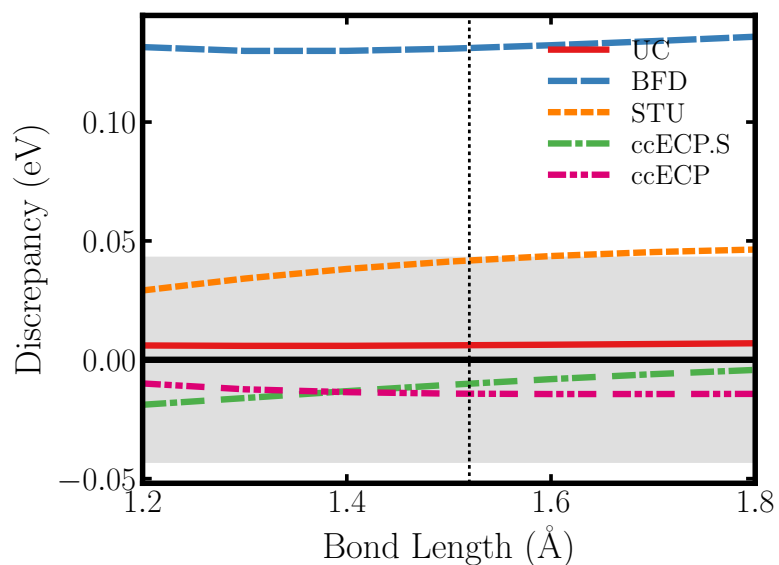


Figure 7.9 Binding energy discrepancies for (a) FeH and (b) FeO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

(a) CoH binding curve discrepancies.



(b) CoO binding curve discrepancies.

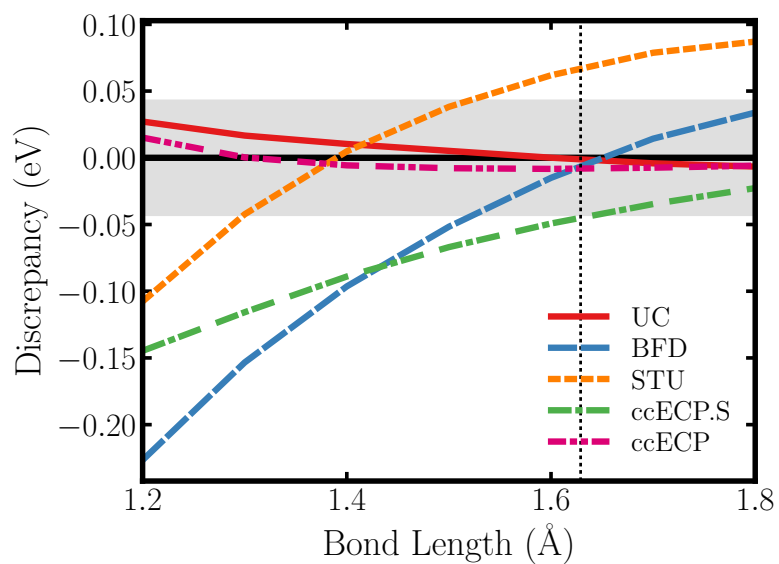


Figure 7.10 Binding energy discrepancies for (a) CoH and (b) CoO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

To obtain a more reasonable description of the molecular properties, we reoptimize using Eq. 7.11 that produces our ccECP. With refinement we compromise on both the LMAD and MAD that mildly increases relative to ccECPS, however, this increase is caused mainly by the very deeply lying states (> 500 eV) while the low-lying spectrum is still exceptionally well described, much better than for BFD, STU, and comparable to UC. With ccECP, we see a minor improvement for NiH, and a significant improvement for NiO, where we reduce the discrepancy to be well within chemical accuracy for all considered bond lengths and maintain a flat discrepancy throughout. The final parameters for our ccECP are given in Table 7.5.

7.3.10 Copper

We show the atomic properties for Cu in Table 7.1 and the molecular properties for CuH and CuO in Fig. 7.12. Due to the closed d-shell LMAD includes 4 excitations, whereas previous tables for Sc-Ni included 5. Focusing first on the spectral ccECPS, we are able to obtain an almost uniform agreement for the entire spectrum, with a LMAD of 0.018 eV and a MAD of 0.015 eV. While the LMAD is slightly higher than UC and eCEPP, the overall MAD is significantly improved since the high ionizations are in much better agreement to AE. As has been seen for the previous elements, a uniform accuracy on the atomic spectrum can decrease the overall transferability see the supplemental material for details about each state in the spectrum. While CuH is quite well described by the ccECPS, we again find significant overbinding near dissociation for our ccECPS. In fact, *all ECPs* including STU, BFD, and eCEPP significantly overbind for the CuO molecule, from roughly 0.1 eV for our ccECPS to over 0.25 eV for BFD.

We substantially improve the overall transferability with our refined ccECP. For both CuH and CuO, we have near perfect agreement with AE along the entire binding curve. For CuO, we are able to obtain the accuracy that stands out when compared to previous ECPs. Note that behavior of biases for this element suggests that keeping the accuracy for energies > 500 eV one would need more variational freedom and perhaps a more sophisticated construction. This is due to different energy and also length scales for semicore vs valence spaces.

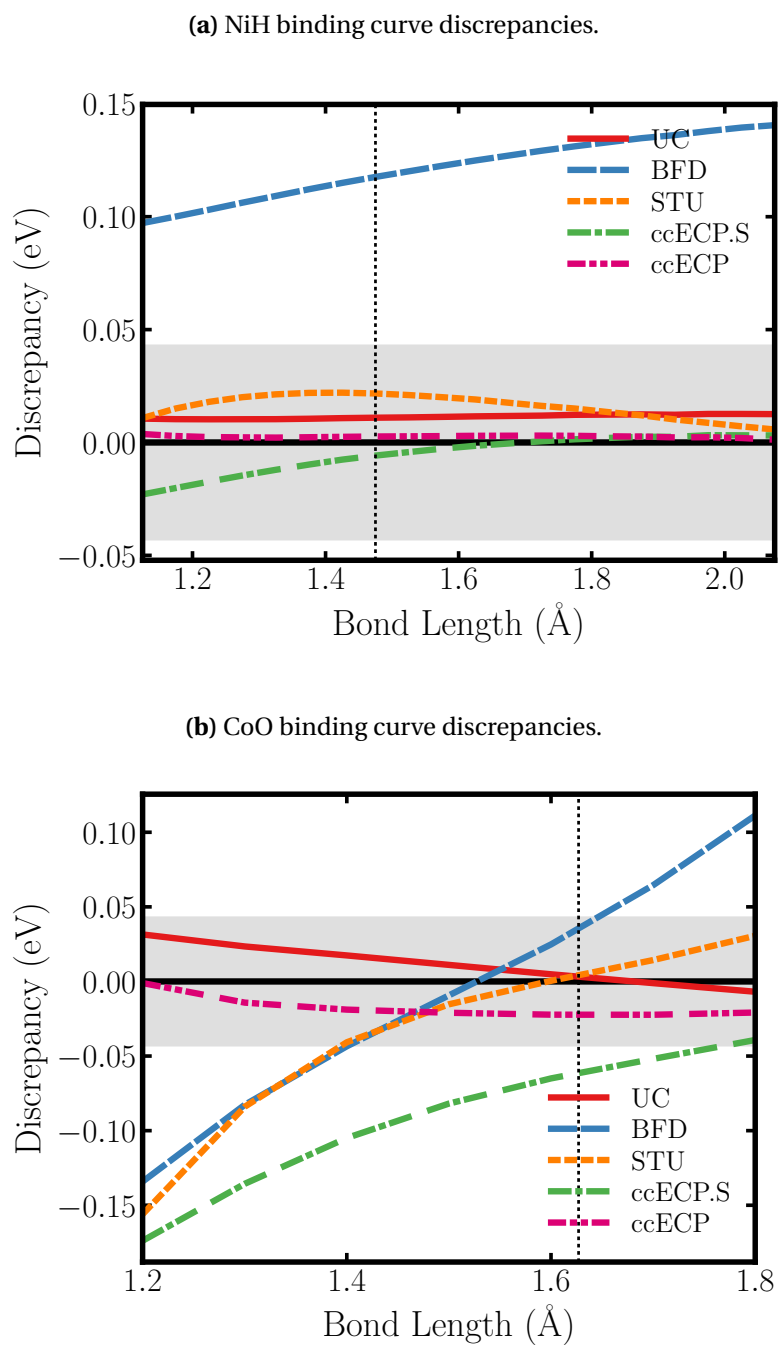
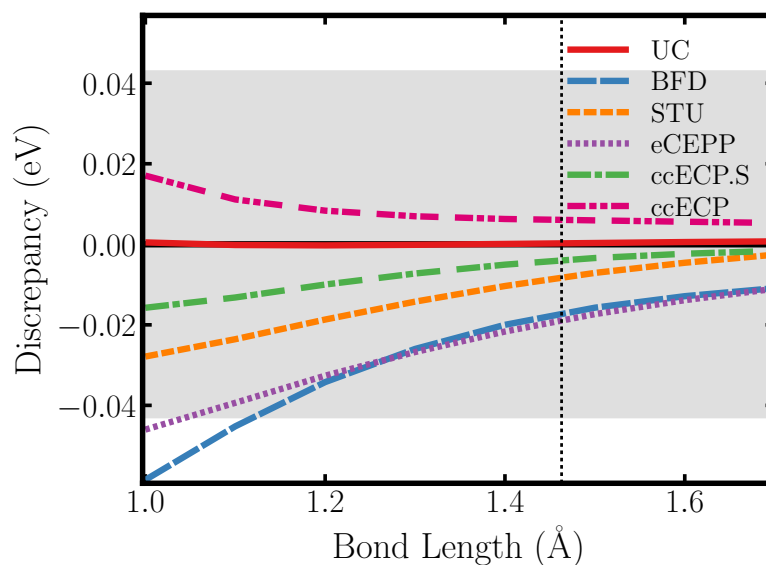


Figure 7.11 Binding energy discrepancies for (a) NiH and (b) NiO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

(a) CuH binding curve discrepancies.



(b) CuO binding curve discrepancies.

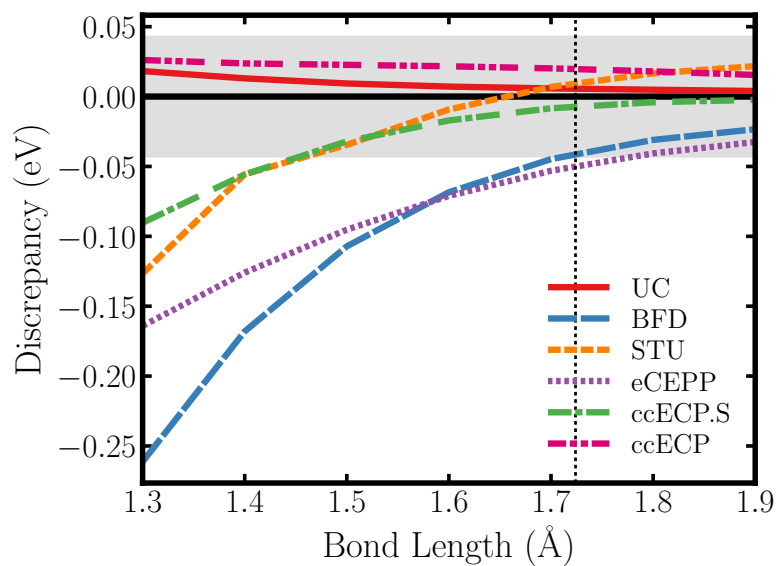


Figure 7.12 Binding energy discrepancies for (a) CuH and (b) CuO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

7.3.11 Zinc

The atomic and molecular data for Zn is given in Table 7.1 and Fig. 7.13, respectively. For our spectral ccECPs, we are able to achieve a uniform accuracy across the entire atomic spectrum, resulting in a MAD of 0.015 eV. This outperforms BFD, STU, and UC significantly, resulting in an ECP that on spectrum is very close to the original AE Hamiltonian. Additionally, the LMAD for Zn includes only the first two states, due to the fact that the anion and most of the $s \leftrightarrow d$ transitions do not exist. For the transferability tests, we find similar behavior for both ZnH and ZnO, where the ccECPs is within chemical accuracy near equilibrium but begins to overbind as the bond is compressed.

With the refinements from our ccECP, we are able to strike a reasonable balance between the atomic properties as well as the molecular transferability. In both the hydride and oxide, the ccECP has a relatively flat discrepancy resulting in excellent vibrational frequencies as well as energy differences well within chemical accuracy for the entire curve. In terms of the atomic spectrum, the lowest states are only 0.01 eV error and we achieve a MAD on the entire spectrum that is better than all other core approximations. The parameters for our final ccECP are given in Table 7.5.

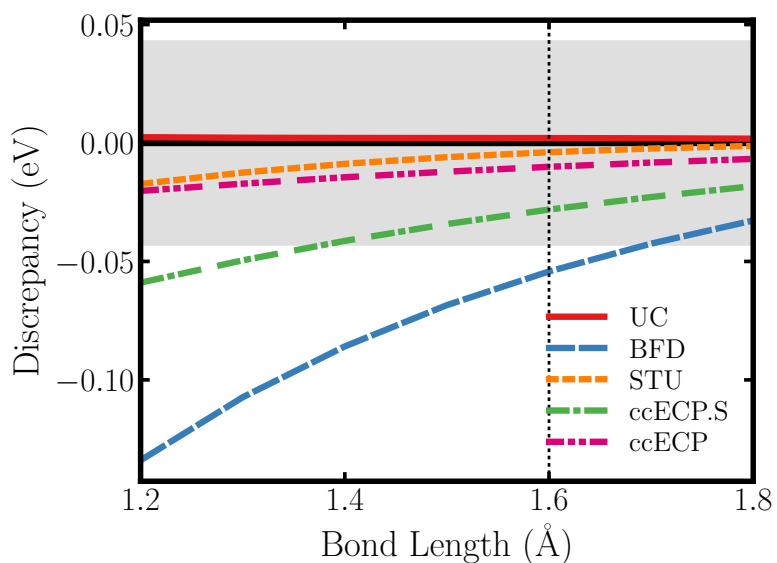
Table 7.2 Parameter values for hydrogen. The highest l value corresponds to the local channel. Each term takes the form $V_{\ell k}(r) = \beta_{\ell k} r^{n_{\ell k} - 2} \exp[-\alpha_{\ell k} r^2]$

Atom	Z_{eff}	ℓ	$n_{\ell k}$	$\alpha_{\ell k}$	$\beta_{\ell k}$
H	1	0	2	1.00000000	0.00000000
		1	1	21.24359508	1.00000000
		1	3	21.24359508	21.24359508
		1	2	21.77696655	-10.85192405

7.3.12 Average molecular discrepancies

In Tab. 7.3 we collect the results of molecular calculations for all elements and evaluate mean absolute deviations for the equilibrium molecular parameters. Note that even our ccECPs set achieves very respectable accuracy that is comparable or better than the available tables.

(a) ZnH binding curve discrepancies.



(b) ZnO binding curve discrepancies.

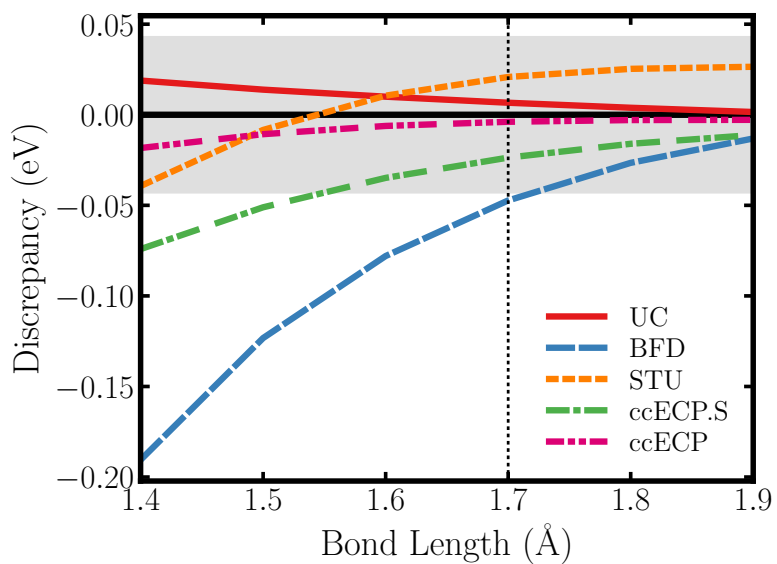


Figure 7.13 Binding energy discrepancies for (a) ZnH and (b) ZnO molecules. The binding curves are relative to the CCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

Clearly, the adjusted set ccECP is the most balanced overall, to the best of our knowledge shows overall the best consistency for all the calculated parameters.

Table 7.3 Mean absolute deviations of binding parameters for various core approximations with respect to AE data for transition metal hydride and oxide molecules. All parameters were obtained using Morse potential fit. The parameters shown are dissociation energy D_e , equilibrium bond length r_e , vibrational frequency ω_e and binding energy discrepancy at dissociation bond length D_{diss} .

	UC	BFD	STU	eCEPP	ccECPS	ccECP
D_e (eV)	0.0063(40)	0.0590(41)	0.0380(41)	0.0163(45)	0.0240(40)	0.0104(40)
r_e (Å)	0.0012(13)	0.0064(13)	0.0026(13)	0.0019(15)	0.0027(13)	0.0010(13)
ω_e (cm ⁻¹)	2.2(5.8)	10.4(5.9)	4.6(5.9)	3.9(6.9)	6.4(5.8)	2.9(5.8)
D_{diss} (eV)	0.021(41)	0.145(41)	0.036(41)	0.032(46)	0.054(40)	0.016(41)

Table 7.4 Parameter values for early transition metal ECPs. For all ECPs, the highest l value corresponds to the local channel. Each term takes the form $V_{\ell k}(r) = \beta_{\ell k} r^{n_{\ell k}-2} \exp[-\alpha_{\ell k} r^2]$

Atom	Z_{eff}	ℓ	$n_{\ell k}$	$\alpha_{\ell k}$	$\beta_{\ell k}$
Sc	11	0	2	11.49466541	153.96530175
		0	2	5.01031394	14.93675657
		1	2	11.45126730	97.21725690
		1	2	4.76798446	10.81704018
		2	1	16.02394388	11.00000000
		2	3	14.08647403	176.26338271
		2	2	11.93985121	-83.68149599
		2	2	3.69440111	0.432827647
Ti	12	0	2	12.70580614	173.94657236
		0	2	6.11178552	18.83768334
		1	2	12.64091930	111.45672882
		1	2	5.35437416	11.17702683
		2	1	18.41366202	12.00000000
		2	3	15.92292414	220.96394426
		2	2	13.65000623	-94.29025825
		2	2	5.09555211	0.09791142
V	13	0	2	15.12502151	195.56713891
		0	2	6.29898914	22.88642835
		1	2	15.93855113	126.42119501
		1	2	5.74006267	16.03597128
		2	1	20.32168914	13.00000000
		2	3	19.59698040	264.18195885
		2	2	17.33147348	-115.29293208
		2	2	5.12320658	-0.66288726
Cr	14	0	2	16.90078761	219.48146210
		0	2	7.33662151	28.07933177
		1	2	17.31974517	139.98396872
		1	2	6.92409758	19.54835786
		2	1	18.28091074	14.00000000
		2	3	17.09800655	255.93275041
		2	2	16.72267276	-132.01826317
		2	2	5.02865106	-0.77388761
Mn	15	0	2	18.92044966	244.66870493
		0	2	8.32764757	33.54162717
		1	2	20.17347020	162.35033686
		1	2	7.80047874	24.17956695
		2	1	21.91937433	15.00000000
		2	3	21.35527128	328.79061500
		2	2	21.27162654	-162.05172805
		2	2	7.93913962	-1.82694273

Table 7.5 Parameter values for late transition metal ECPs. For all ECPs, the highest l value corresponds to the local channel. Each term takes the form $V_{\ell k}(r) = \beta_{\ell k} r^{n_{\ell k}-2} \exp[-\alpha_{\ell k} r^2]$

Atom	Z_{eff}	ℓ	$n_{\ell k}$	$\alpha_{\ell k}$	$\beta_{\ell k}$
Fe	16	0	2	22.21062697	277.50032547
		0	2	9.51515801	46.20495585
		1	2	24.57000872	194.99875057
		1	2	8.86648777	31.67945133
		2	1	23.22091714	16.00000000
		2	3	23.54714680	371.53467418
		2	2	23.47256345	-181.22603445
		2	2	9.85238815	-2.37305236
Co	17	0	2	23.41427031	271.77708487
		0	2	10.76931694	54.26461122
		1	2	25.47446317	201.53430745
		1	2	10.68404901	38.99231927
		2	1	25.00124116	17.00000000
		2	3	22.83490097	425.02109972
		2	2	23.47468156	-195.48211283
		2	2	10.33794825	-2.81572866
Ni	18	0	2	26.43201939	321.24000243
		0	2	11.74896968	60.34700846
		1	2	29.49299982	236.53999900
		1	2	11.55698315	44.39698879
		2	1	28.26300010	18.00000000
		2	3	26.93602546	508.73400183
		2	2	27.08600753	-220.09999930
		2	2	12.21300013	-2.13493271
Cu	19	0	2	29.35562243	370.71371825
		0	2	12.77235920	66.27560813
		1	2	33.51694544	271.66281028
		1	2	12.52471485	49.76265057
		2	1	31.53811263	19.00000000
		2	3	31.06925531	599.22413998
		2	2	30.59035868	-244.68915484
		2	2	14.05141064	-1.29349526
Zn	20	0	2	35.02141357	431.70804303
		0	2	14.63498692	95.87640437
		1	2	42.22979235	313.57770564
		1	2	14.57429304	74.01270049
		2	1	35.80797616	20.00000000
		2	3	34.53646084	716.15952324
		2	2	28.62830178	-204.68393324
		2	2	7.96239683	0.76026614

7.4 Conclusions

In this paper, we present the correlation consistent ECPs for 3d elements. First, we constructed the ECPs using spectral optimizations for all the elements. This optimization was iterative and included states that ionize the given atom down to the $[\text{Ne}]3s^2$ ion, as well as $s \leftrightarrow d$ transfer energies for low-lying atomic and ionic states, including bounded anions. Quite unexpectedly, for most atoms, we were able to optimize the states within 0.01 eV or even smaller discrepancies for the full span of ionization energies, sometimes with accuracies better than the known spectroscopic data and also better than the inherent systematic biases in our methodology. As discussed above any discrepancies smaller than ≈ 0.05 eV are in fact comparable to estimated systematic errors in our calculations. These spectral-only results simply pointed out that such fits are indeed feasible using rather small number of free ECP parameters.

We further probed for accuracy and transferability of the constructed operators on transition metal-hydride and oxide molecules. The hydrides appear mostly less problematic and within the error bar window of chemical accuracy for all bond lengths with a few exceptions as analyzed above. The oxide molecules and their polar bonds revealed more complicated picture and most ECPs required refinement that has led to high fidelity and chemical accuracy along the binding curves. That set is labeled as the recommended ccECP. The spectral-only ccECP.S versions could still be useful for atomic calculations where the accuracy of highly ionized states would be of crucial interest. We note that in cases of Sc and Ti ccECP=ccECP.S since the spectral optimization provided desired accuracy without further refinement. It is quite remarkable that such accurate fits can be constructed using just the spectral information as an input.

For high spin elements, the spectral-only ECPs, while still respectable, were less accurate than desired. For geometries around the equilibrium and the stretched bond regions were described very well, however, for shorter bond length regions we observed in some cases overbinding of the order ≈ 0.1 eV. Although rather small in a relative sense we have opted for further improvement that took into account the decreasing importance of correlations of deeply ionized states vs. the key desired accuracy of the low-lying excitations most relevant for valence properties. Adding a shift towards the HF eigenvalues into the objective function enabled us to mildly retune the 3s and 3p channels and that proved to be sufficient to get the binding curves within the chemical accuracy threshold. This caused some increase in

spectral errors, however, only for very highly ionized states that are of the order of > 500 eV with relative errors being still very small ($\approx 0.08\%$ or smaller).

For the late transition elements namely, Co, Ni, Cu, and Zn we have observed similar behavior with an overbinding tendency for oxide molecules at short bond lengths. Similar refinement as for the high-spin elements have enabled to alleviate these deviations so that all the binding curves were within the chemical accuracy threshold. For these elements we also note additional complications with the largest basis sets for Co, Cu and Zn where we have encountered the feasibility limits of the used codes. Therefore we have restricted ourselves to the best results we could obtain with reasonable computational resources and we consider the achieved systematic comparisons as adequate for our present purposes.

Together with accuracy tests, we have also tried to estimate the exact atomic total energies for both the ground and excited states. The extrapolations from extensive basis sets aug-cc-pCV n Z with $n=2,3,4,5$ provide rather a consistent picture of the corresponding correlation energies, however, at present we do not claim better systematic accuracy than ≈ 10 mHa. In order to provide these energies with better uncertainties, we plan further investigation in a subsequent study that will be devoted to decrease the error margins to much lower values.

An observation on the legacy ECP constructions is that our results confirm that the Stuttgart-Koeln-Bonn ECPs [Dol87; DC12] established and derived by Stoll, Dolg and coworkers over the past three decades show systematic consistency and respectable accuracy for the 3d series. The discrepancies are mostly within the 0.1 - 0.2 eV margins for low-lying atomic excitations and also for most of the molecular binding curves. Since the STU table was constructed with Dirac-Hartree-Fock spectral fits, this confirms that the precise self-consistent energy differences data is the dominant factor in achieving consistent behavior. However, correlation effects become important at finer resolutions that we have targeted in this work.

Clearly, our present constructions raise the bar of accuracy higher and we believe that they provide a significant step forward for correlated calculations. Further data including the basis sets, including both cc-pV n Z and aug-cc-pV n Z for $n \in \{D, T, Q, 5\}$, can be found in web library <http://pseudopotentiallibrary.org> as well as in the Supplementary Information. Data for this work is also provided at the Materials Data Facility [Dat].

The presented results show that there is a room for sizeable improvements of both the construction methods and practical versions of ECPs. In agreement with previous papers

on first and second rows, we were aiming at offering a "minimal model" that is still more accurate than existing tables. For example, nonlocal s and p channels are described by two gaussians only (similarly to the STU table). We have demonstrated that such a combination of small variational space and accuracy is indeed feasible and the resulting constructions are ready for general use. There are several directions where this work could expand further. Heavier atoms such as Co, Cu and Zn could benefit from more variational parameters that would address differences between deeply lying 3s, 3p states vs large valence subshell. In addition, more testing and validation is needed in a variety of chemical systems such as larger molecules, solids and 2D materials. Indeed, we expect that such data would provide new insights and possibly point out the directions for further refinements and updates in future.

Projector quantum Monte Carlo with averaged vs. explicit spin-orbit effects: applications to tungsten molecular systems

Portions of this chapter also appeared in:

**Projector quantum Monte Carlo with averaged vs. explicit spin-orbit effects:
applications to tungsten molecular systems**

J. Phys. Chem. Sol. (2018)

Contributions:

In the work described in this chapter, the author performed the fixed-node DMC calculations of the AREP-containing WO and W₂ systems.

8.1 Introduction

Over the decades, there has been considerable interest in transition metal systems due to their intriguing physical properties and their potential applications. As a result of the versatility of d -bonding, a wide variety of physical phenomena have been observed in transition metal compounds such as superconductivity, ferroelectricity, a plethora of magnetic phases and more. Molecular nanostructures, surfaces and interfaces offer still more potential uses in both research and applications. For example, the versatility of the d -bonding finds utility in catalysts and bioenzymes [Cla98; MH84]. For 2D transition metal dichalcogenides, the optical and properties and high electron mobility make them ideal candidates for photonics, optoelectronics, spintronics, etc [Man17]. Apart from applications, transition metal compounds are of great theoretical interest due to their strong electronic correlations which are difficult for standard theoretical techniques to describe (e.g. Mott insulators). Between the family of electronic structure quantum Monte Carlo (QMC) methods, the diffusion Monte Carlo (DMC) has had significant success at accurately describing systems with strong correlations, including transition metal compounds [Ben17b; KM08; Foy14; MK10; ZW15; Luo16]. Despite its successes, DMC traditionally works with many-body Hamiltonians without spin terms. This implies that important physics of spin-orbit compounds, non-collinear spin phases, etc, has not been accessible to this unique methodology. While the strength of the spin-orbit may be rather modest in the 3rd row transition metals, as one begins to consider heavier elements in the 5th row and beyond, scalar relativistic and spin-orbit effects are no longer negligible and can be comparable to that of the correlation effects. Thus so far, heavy element systems with strong spin-orbit interactions have been left out from electronic structure QMC studies.

As discussed in chapter 4, we have generalized DMC algorithm to work with spinorbitals and spinors within a complex wave function framework under the fixed-phase approximation and we named the method fixed-phase spin-orbital diffusion Monte Carlo (FPSODMC) [Mel16a; Mel16b]. Using this construction, we showed that DMC is now capable of treating the spin-orbit terms in the Hamiltonian directly without relying on perturbation theory as is typically done in many mean-field techniques. We applied the methodology to some simple atoms and molecules, illustrating that DMC is able to directly include the spin-orbit corrections while still accurately describing the electronic correlations. In this paper, we apply the FNDMC and FPSODMC methods to molecular systems containing tungsten with

scalar relativity only (with averaged spin-orbit) and with explicit spin-orbit terms included, respectively. In particular, we study the tungsten oxide (WO) and tungsten dimer (W_2) molecules where limited experimental data exists.

This chapter is organized as follows. First, we briefly introduce the fixed-phase DMC method that must be employed for complex trial wave functions. After introducing the methodology used for complex wave functions, we introduce a complex representation for the electrons spin degree of freedom which allows for efficient sampling of this space. We also describe how the relativistic effects enter directly into the Hamiltonian through the use of non-local relativistic effective core potentials that are used in relativistic quantum chemistry. The non-local effective core potentials gives us the freedom to choose between including only scalar relativistic effects with averaged spin-orbit (AREP) or whether to include spin-orbit explicitly and directly (REP). We then present our results for the tungsten molecules WO and W_2 and compare our results against other theoretical investigations and experiment where applicable. We then conclude with a summary as well as an outlook on how FPSODMC can be improved for future studies of correlated materials with spin-orbit interactions included.

8.2 Fixed-Phase Diffusion Monte Carlo

Imaginary time projector methods, like DMC, work to project out the ground state in the infinite limit, namely $\lim_{\tau \rightarrow \infty} e^{-\tau H} \Psi \propto \Psi_0$. However, in order to avoid the notorious fermion sign problem, one must invoke an approximation utilizing a trial wave function. Traditionally, this approximation comes in the form of the fixed-node approximation, where the ground state wave function that is projected has the same zeros (nodes) as the trial wave function. While the unrestricted DMC projection is formally exact, the boundary conditions enforced by the fixed nodes introduce a bias. This manifests in higher total energy since the fixed-node approximation is variational. If the trial wave function happens to have the exact nodal structure, the energy obtained is the exact ground state energy. While fixed-node diffusion Monte Carlo (FNDMC) is the most common implementation [Fou01], it cannot be used for inherently complex wave functions as are needed, for example, for spin-orbit applications. For complex wave functions, a more general elimination of sign/complex amplitudes called the fixed-phase approximation [Ort93] was introduced some time ago. As outlined below, we use the fixed-phase framework to formulate DMC with variable

fermionic spins.

For simplicity, we sketch the fixed-phase method first considering just the spatial degrees of freedom while the spins will be added later. Let us denote our configuration space as $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$. The imaginary time Schrödinger equation is written as

$$-\frac{\partial \Psi(\mathbf{R}, \tau)}{\partial \tau} = \left[-\frac{1}{2} \nabla_{\mathbf{R}}^2 + V(\mathbf{R}) + W \right] \Psi(\mathbf{R}, \tau) \quad (8.1)$$

where $-1/2\nabla_{\mathbf{R}}^2$ is the kinetic energy, $V(\mathbf{R})$ is the local potential, and W is any non-local potential such as a pseudopotential (or effective core potential). Here, we assume a complex wave function $\Psi(\mathbf{R}, \tau) = \rho(\mathbf{R}, \tau) \exp[i\Phi(\mathbf{R}, \tau)]$, where $\rho \geq 0$ and Φ are the amplitude and phase, respectively. Substituting this into equation (8.1) yields two coupled equations for the real and imaginary parts of the Schrödinger equation, namely

$$-\frac{\partial \rho}{\partial \tau} = -\frac{1}{2} \nabla_{\mathbf{R}}^2 \rho + \frac{1}{2} |\nabla_{\mathbf{R}} \Phi|^2 \rho + V(\mathbf{R}) \rho + \text{Re}[\Psi^{-1} W \Psi] \rho \quad (8.2)$$

$$-\frac{\partial \Phi}{\partial \tau} = -\frac{1}{2} \nabla_{\mathbf{R}}^2 \Phi - \rho^{-1} \nabla_{\mathbf{R}} \rho \cdot \nabla_{\mathbf{R}} \Phi + \text{Im}[\Psi^{-1} W \Psi] \quad (8.3)$$

We first note that the expectation value of the Hamiltonian must be real since the Hamiltonian is Hermitian, i.e., the eigenvalue is determined from the real part given by equation (8.2). The two coupled equations cannot easily be solved, so we will approximate the phase by a time independent trial wave function $\Psi_T(\mathbf{R}) = \rho_T(\mathbf{R}) \exp[i\Phi_T(\mathbf{R})]$ and only consider solving the real part. An approximate phase can only raise the energy, so this approximation is variational as given elsewhere [Ort93; Mel16b]. An additional approximation must be invoked since we do not know the action of the nonlocal potential W on the wave function Ψ since we are trying to solve for Ψ . We therefore invoke the locality approximation [Mit91] and replace $\Psi^{-1} W \Psi$ by $\Psi_T^{-1} W \Psi_T$. It can be shown that the error from this approximation goes as $(\Psi_T - \Psi_0)^2$ where Ψ_0 is the ground state wave function. While the localization approximation is not variational, one can restore the upper bound property made by enforcing the so-called T-moves [Cas06; Mel16b].

As is traditionally done in DMC, we perform an importance sampling transformation [Rey82] by multiplying the real part by the trial amplitude ρ_T . Coupled with our various approximations, the importance sampled Schrödinger equation becomes

$$-\frac{\partial f}{\partial \tau} = -\frac{1}{2} \nabla_{\mathbf{R}}^2 f + \nabla_{\mathbf{R}} [f \nabla_{\mathbf{R}} \ln \rho_T] + E_L(\mathbf{R}) f \quad (8.4)$$

where the mixed distribution f is given as a product, $f(\mathbf{R}, \tau) = \rho_T(\mathbf{R})\rho(\mathbf{R}, \tau)$, and the local energy $E_L(\mathbf{R})$ is

$$E_L(\mathbf{R}) = -\frac{1}{2} \frac{\nabla_{\mathbf{R}}^2 \rho_T}{\rho_T} + \frac{1}{2} |\nabla_{\mathbf{R}} \Phi_T|^2 + V(\mathbf{R}) + \text{Re}[\Psi_T^{-1} W \Psi_T] \quad (8.5)$$

We solve equation (8.4) in integral form

$$f(\mathbf{R}', \tau + \delta\tau) = \int d\mathbf{R} G(\mathbf{R}', \mathbf{R}; \delta\tau) f(\mathbf{R}, \tau) \quad (8.6)$$

where the Green's function can be approximated for small time steps $\delta\tau$ as

$$G(\mathbf{R}', \mathbf{R}; \delta\tau) \approx (2\pi\delta\tau)^{-3N/2} \exp\left[\frac{-|\mathbf{R}' - \mathbf{R} - \delta\tau \nabla_{\mathbf{R}} \ln \rho_T|^2}{2\delta\tau}\right] \times \exp\left[-\frac{\delta\tau}{2} (E_L(\mathbf{R}') + E_L(\mathbf{R}) - 2E_T)\right] \quad (8.7)$$

where we have introduced an energy offset E_T to the Hamiltonian and N is the number of electrons. The error associated with this approximate Green's function goes as $O(\delta\tau^3)$ and can easily be extrapolated to zero. By repeated application of the Green's function, we can step forward in imaginary time to achieve the steady state solution which corresponds to the ground state wave function and energy.

8.3 Spin-Orbit Interactions and Dynamic Spins

In the previous discussion, we have restricted ourselves to static spins and only reference the spatial configuration space. In our previous works [Mel16a; Mel16b; MM16; Mel17], significant descriptions for the introduction of dynamic spins into the configuration space has been given, so we only briefly describe it here.

If we consider an arbitrary one-particle spinor

$$\psi(\mathbf{r}, s) = \alpha \phi^\uparrow(\mathbf{r}) \chi^\uparrow(s) + \beta \phi^\downarrow(\mathbf{r}) \chi^\downarrow(s) \quad (8.8)$$

where α, β are complex constants, $\phi^{\uparrow/\downarrow}$ are (complex) spatial orbitals and $\chi^{\uparrow/\downarrow}$ are the spin functions, where $\chi^{\uparrow/\downarrow}(s = \pm 1/2)$ is the standard S_z representation for the spin degrees of

freedom. DMC, however, traditionally works in a configuration space that is continuous whereas the spin variables are discrete. This begs the question as to whether the spin variables can be made continuous in order to sample them in a similar fashion to the spatial degrees of freedom. Indeed this can be achieved, through the introduction a complex representation for the spin states as well as a spin “kinetic energy.” The new spin representation is given as

$$\chi^\uparrow(s) = e^{is}, \quad \chi^\downarrow(s) = e^{-is}, \quad s \in [0, 2\pi) \quad (8.9)$$

which preserves the orthogonality of the $\chi^{\uparrow/\downarrow}$ states. It should be noted that using the discrete representation, the determinants in the trial wave function in Eq. 8.10 breaks up into a product of spin up and spin down determinants [Mel17], whereas using our complex spin representation we retain a single determinant.

Clearly, the results obtained depend on the quality of the trial phase, and thus, the trial wave function. We write the many-body trial wave function in the multi-reference Slater-Jastrow form

$$\Psi_T(\mathbf{R}, \mathbf{S}) = \exp[J(\mathbf{R})] \sum_{k=1}^M c_k \det_k[\psi_i(\mathbf{r}_j, s_j)] \quad (8.10)$$

where M is the number of determinants, $\psi_i(\mathbf{r}_j, s_j)$ is the i th single-particle spinor for electron j , and $J(\mathbf{R})$ is the Jastrow factor which explicitly includes inter-particle correlations to the wave function. If the trial phase happens to be the exact phase of a given symmetry, then the algorithm is exact and will reproduce the exact ground state energy within that symmetry channel.

In order to sample the continuous spins, we add to the Hamiltonian a term

$$T_{\mathbf{S}} = \sum_{i=1}^N T_{s_i}, \quad T_{s_i} = -\frac{1}{2\mu_s} \left[\frac{\partial^2}{\partial s_i^2} + 1 \right] \quad (8.11)$$

that generates the spin evolution while μ_s is a “spin mass”. Since $T_{s_i} \psi(\mathbf{r}_i, s_i) = 0$, it also annihilates any trial wave function generally written as a linear combination of spinor determinants. Note that the spin mass μ_s results in an effective time step for the spin degrees of freedom, namely $\delta\tau_s = \delta\tau/\mu_s$. With the addition of this spin kinetic energy, we are simulating a Hamiltonian of the form $H' = H + T_{\mathbf{S}}(\mu_s)$. As we discussed elsewhere [Mel16b; Mel17] the ratio of spin time step and the spatial time step can vary and provides different limits that imply faster (or slower) spin evolution with regard to the spatial degrees of freedom. Here we employ the limit of $H' \rightarrow H$ by taking $\mu_s \rightarrow \infty$, ie, $\tau_s \rightarrow 0$, which

corresponds to slowly evolving spins compared to the spatial degrees of freedom and enables to recover the original Hamiltonian without the contamination from the spin kinetic energy.

The spin kinetic energy acts to add dynamics for the spin degrees of freedom, but thus far no reference to the spin-orbit interaction has been made. For matter at the nuclei and electrons resolution a straightforward way to introduce scalar and spin-orbit effects is through the use of relativistic effective core potentials (REPs) [Lee77b; DC12]. The motivation for this is twofold. First, all-electron calculations within DMC scales poorly with the atomic number, namely $Z^{5.5-6.5}$ whereas they scale as N_{val}^3 for the valence electrons when an ECP is utilized [Cep86; Ham87]. Since the core electrons do not affect the valence properties directly it is favorable to remove the cores by an accurate effective model. This allows DMC to treat larger and more complicated systems with much more favorable scaling and resolution for valence energy differences. Second, if one considers relativity, one needs to solve the all-electron Hamiltonian within the four-component spinor formalism. For the deepest core electrons, all four components of the spinors can be non-zero. For valence electrons that are relevant in bonding, the 4-component spinors essentially decouple into major and minor 2-component spinors. The minor component spinors can be eliminated by using several approximations/transformations of the relativistic Hamiltonian (such as DKH or X2C), leaving only 2-component spinors for the valence electrons. If the core electrons are projected out and replaced with an effective operator, what is left are 2-component spinors for the valence electrons moving in an effective potential that represents the impact of the relativistic core on degrees of freedom. Given our two-component spinor representation above, we simply include these effective core operators into our Hamiltonian and drop the atomic cores and corresponding degrees of freedom.

For any Hamiltonian that does not depend on spin, both non-relativistic and scalar relativistic with averaged spin-orbit (AREP), the nonlocal ECP typically is written as a projector with spherical harmonics with a local radial dependence so that both spins feel the same projector. In cases where spin is included in the Hamiltonian such as spin-orbit, spin no longer commutes with the Hamiltonian and only the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$ is a good quantum number. Thus, the REPs utilize a more general form of the projectors,

$$W_i^{REP} = \sum_{\ell} \sum_{j=|\ell-1/2|}^{|\ell+1/2|} \sum_{m_j=-j}^j W_{\ell j}^{REP}(r_i) |\ell j m_j\rangle \langle \ell j m_j| \quad (8.12)$$

where $|\ell j m_j\rangle$ are the so-called spin spherical harmonics and the radial dependence is typically parametrized by

$$W_{\ell j}^{REP}(r_i) = \frac{1}{r_i^2} \sum_{\alpha} A_{\ell j \alpha} r_i^{n_{\ell j \alpha}} \exp[-B_{\ell j \alpha} r_i^2] \quad (8.13)$$

This form of the RECP captures both scalar and spin-orbit relativistic effects, and parameterizations have been generated by several research groups [Fig09; Ros90]. The contribution from both scalar and spin-orbit relativistic effects enter into the local energy in equation (8.5)

$$\begin{aligned} \text{Re}\left[\frac{W^{REP}\Psi_T}{\Psi_T}\right] = \text{Re}\left[\sum_{iI} \sum_{\ell j} W_{\ell j}^{REP}(r_{iI}) \int d\Omega'_{iI} \int ds' \right. \\ \left. \sum_m \langle \Omega_{iI} s_s | \ell j m \rangle \langle \ell j m | \Omega'_{iI} s'_i \rangle \frac{\Psi_T(\dots, \mathbf{r}'_i, s'_i, \dots)}{\Psi_T(\dots, \mathbf{r}_i, s_i, \dots)}\right] \end{aligned} \quad (8.14)$$

where $\mathbf{r} = (r, \Omega)$, $\mathbf{r}' = (r, \Omega')$ and Ω is the solid angle around a particular nucleus. While the form above is completely general for an arbitrary spin representation, utilization of the complex spin functions described above yields a simple complex functional form for the spin spherical harmonics [Mel16b].

The spin degrees of freedom then enter the Green's function in a similar way as spatial coordinates. Therefore they appear both in evolution represented by the diffusion and drifts as well as in the local energy. The evolution of the walker coordinates (space and spin) is governed by the Eq. (6).

8.4 Results

We illustrate the method on heavy transition metal molecules, namely tungsten oxide WO and the tungsten dimer W_2 . This is partially driven by the fact that the spinors (orbitals) and corresponding wave functions are much better developed in molecular codes that enable us to study the combined correlation and spin-orbit effects more systematically. For both cases, we study the molecules using scalar relativistic spin-averaged Hamiltonians in one-component conventional framework as well as with spin-orbit included in two-component framework outlined above. For scalar relativistic Hamiltonians we use FNDMC whereas for spin-orbit Hamiltonians we utilize FPSODMC [Mel16a; Mel16b] using a modified version of

QWALK [Wag09b]. The trial wave functions for the scalar relativistic Hamiltonians use either Hartree-Fock (HF) or Density Functional Theory (DFT) orbitals from GAMESS [Sch93b] to define the nodal surface. For the FPSODMC, the trial wave functions are generated by the DIRAC code [Dir], and are typically multi-reference in nature. For multi-reference wave functions, we use either complete open-shell configuration interaction (COSCI) or larger configuration interaction wave functions that use Dirac HF (DHF) single particle spinors. In closed-shell systems, the trial wave function is a single determinant and one can use either DHF spinors or relativistic DFT spinors. For the W atom, we use the REPs from the Stuttgart group [Fig09] whereas we use the O ECP of Burkatzki, Filippi, and Dolg [Bur07].

8.4.1 Tungsten Oxide

The tungsten atom and many tungsten containing compounds have a different electronic structure than their isovalent counterparts containing Cr and Mo. For example, the ground state configuration of W is $[\text{Xe}4f^{14}]6s^25d^4$, whereas both Cr and Mo have $ns^1(n-1)d^5$ ground state configurations [Mel16a]. The favored configuration and state results from subtle balance between electron correlation as well as the splitting due to spin-orbit. Interestingly, this trend manifests itself also in the WO molecule. CrO and MoO have $^5\Pi$ molecular ground states, and some of the original theoretical work on WO predicted a $^5\Pi$ ground state [NB85]. However, experimental determination as well as contracted multi-reference configuration interaction (CMRCI) calculations show a $^3\Sigma^-$ ground state for WO [Ram01]. Our results described below confirm this conclusion using both scalar relativistic and spin-orbit Hamiltonians, indicating that this effect is predominantly driven by correlations with smaller impact from the spin-orbit effects.

We first discuss the determination of the ground state without the spin-orbit interaction. For experiments the estimation of the equilibrium bond length was found to be $1.65807(6) \text{ \AA}$ [Ram01]. The CMRCI calculations predict a minimum at 1.67 \AA for the $^3\Sigma$ state, which is only in error by $\sim 0.01 \text{ \AA}$. In order to determine the ground state, we calculate the $^3\Sigma$ and $^5\Pi$ states at $r_e = 1.67 \text{ \AA}$. We also tested various nodal surfaces, namely PBE and hybrid PBE0 which includes (HF) exchange, which in many cases can have a significant impact on the nodal surface, especially for transition metals. The results are shown in Table 8.1. For the $^3\Sigma$ state, the PBE nodal surface results in the lowest energy overall by 2 mHa. Clearly, the $^3\Sigma$ state is correctly predicted to be the ground state, even in the absence of spin-orbit. Now that the ground state has been determined, we can determine the dissociation energy D_e at equilib-

Table 8.1 Total energies for WO for the $^3\Sigma$ and $^5\Pi$ states at $r_e = 1.67 \text{ \AA}$. These results only utilize a scalar relativistic Hamiltonian and neglects spin-orbit.

Ψ_T	SCF	VMC	FN-DMC
$^3\Sigma$ PBE	-83.501631	-83.2377(2)	-83.3149(3)
$^3\Sigma$ PBE0	-83.429649	-83.2397(1)	-83.3130(3)
$^5\Pi$ PBE	-83.484707	-83.2227(2)	-83.2870(4)
$^5\Pi$ PBE0	-83.398745	-83.2244(1)	-83.2862(4)

rium. We calculate the W atom and O atom within FN-DMC, and the dissociation energy can be determined by $D_e = E(\text{WO}) - E(\text{W}) - E(\text{O})$. Note that our dissociation energy does not include the zero-point motion. We find the dissociation energy to be $D_e^{\text{FN-DMC}} = 6.80(2) \text{ eV}$. To see how this compares to the experimental results, we first note that the experimental vibrational frequency is $\omega_e = 1065.6231(52) \text{ cm}^{-1}$ [Ram01] and the experimental binding energy (including zero point motion) is $D_0^{\text{expt}} = 6.85(44) \text{ eV}$ [DN70]. Correcting for the zero point motion, the experimental dissociation energy becomes $D_e^{\text{expt}} = 6.91(44) \text{ eV}$. Without the spin-orbit correction, the FN-DMC dissociation energy falls well within the experimental error bars. Regarding the excited state $^5\Pi$, previous CASSCF (complete active space self-consistent field) calculations predicted a minimum a $r_e = 2.11 \text{ \AA}$ and $D_e = 2.42 \text{ eV}$ whereas CISD found $r_e = 1.99 \text{ \AA}$ and $D_e = 3.25 \text{ eV}$ [NB85]. The more recent CMRCI finds an equilibrium bond length which is shifted toward the $^3\Sigma$ minimum at $r = 1.72 \text{ \AA}$. We calculate the $^5\Pi$ state at this geometry within FN-DMC using PBE0 nodes and find a total energy of $-83.2848(3) \text{ Ha}$. Comparing with the results in Table 8.1, it is clear that FN-DMC does not predict $r_e = 1.72 \text{ \AA}$ to be the minimum for the $^5\Pi$ state, and in our calculations it has a lower energy at $r_e = 1.67 \text{ \AA}$.

We now turn to the FPSODMC results with explicit inclusion of spin-orbit and two-component spinors. With the spin-orbit interaction included, the $^3\Sigma$ and $^5\Pi$ states are split. We take our trial wave functions from the lowest energy spin-orbit state coming from each of the term symbols calculated from COSCI, which we indicate by $^3\Sigma$ and $^5\Pi$ respectively. As above, we first investigate the difference between the ground states of the two manifolds at $r_e = 1.67 \text{ \AA}$. The results are shown in Figure 8.1. As mentioned above we extrapolate in the spin time step to find the desired expectation values. Note that the time step for the spatial degrees of freedom was conservatively small $\tau = 10^{-3} \text{ a.u.}$. It is clear that FPSODMC predicts the $^3\Sigma$ symmetry to be the ground state, in agreement with CMRCI results quoted above.

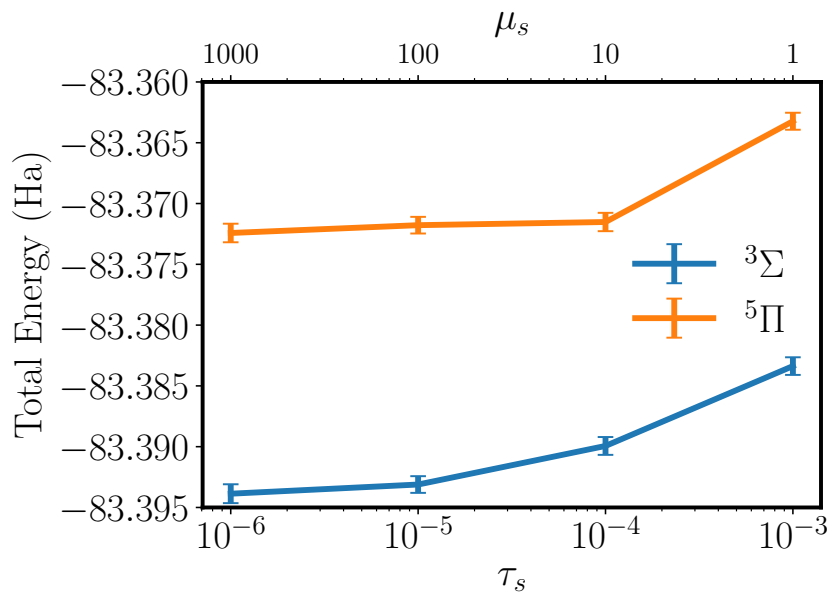


Figure 8.1 FPSODMC total energies of WO molecule as a function of spin time step τ_s /effective spin mass μ_s for the lowest states with ${}^3\Sigma$ and ${}^5\Pi$ symmetries. Note that the within the error bars the difference between the states remains very similar regardless of the time step.

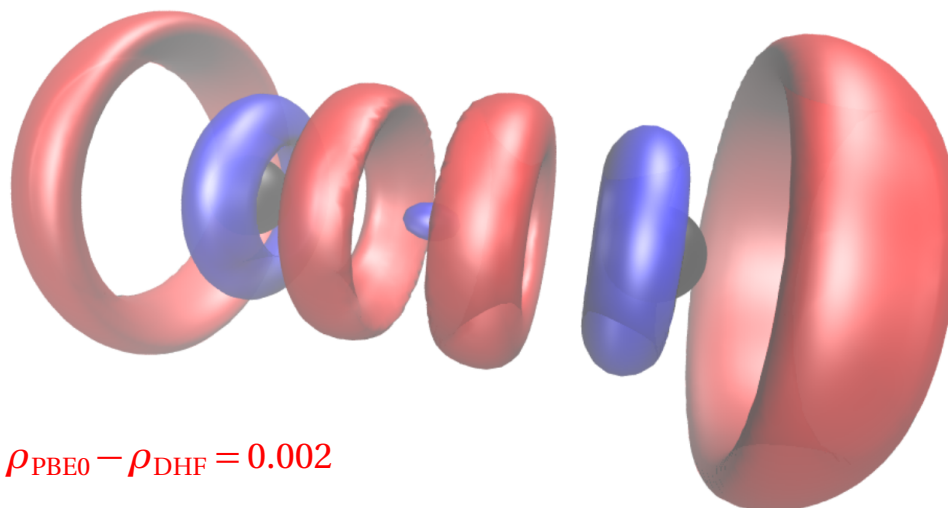
In order to predict the dissociation energy, we calculate the ground states of individual atoms and find a dissociation energy of $D_e^{FPSODMC} = 6.52(2)$ eV. We note that the predicted dissociation energy for the case without spin-orbit and with spin-orbit differ by roughly 0.3 eV. While the quality of the nodal surface or phase can shift the predicted value slightly (we use DFT nodal surfaces whereas we use COSCI phases due to difficulties in generating an open-shell DFT trial wave function within the relativistic framework), another factor contributes more strongly to this difference. While the spin-orbit interaction tends to lower the ground state energy when it splits a term symbol into various states, this effect is much less pronounced in the WO molecular system. The spin-orbit splitting leads to lower energy for an individual atom, ultimately resulting in a reduced estimation of the dissociation energy than what the scalar relativistic FNDMC will predict. Additionally, we note that the prediction from FPSODMC agrees with experiment to the experimental uncertainty. We also investigate the location of the ${}^5\Pi$ minimum. At $r_e = 1.72$ Å as predicted by CMRCI, FPSODMC finds a total energy of -83.3705(5) Ha. If we compare to the total energy shown in Figure 8.1 which saturates to -83.3724(7) Ha, we find that $r_e = 1.72$ Å is not the equilibrium

geometry for this excited state, in agreement with the scalar relativistic FNDMC calculations.

8.4.2 Tungsten Dimer

Unlike WO which differs from its isovalent counterparts, W_2 is expected to follow the trend of Mo_2 and Cr_2 and form a $^1\Sigma$ ground state configuration. To the best of our knowledge, experimental data does not exist for the binding energy of the W_2 . However a number of model estimates are in the literature predicting a wide range of binding energies in eV, namely 4.69(89) [MG79], 5.62(1.23) [Kra75], 4.68 [Mie80] and 5.00(69) [BW80]. From the scatter in the various estimates, an adopted prediction of the binding energy is given as $D_0^{est} = 5(1)$ eV [Mor86]. More recent B3LYP calculations predict a $^1\Sigma$ ground state over a $^3\Sigma$ calculation with an equilibrium bond length of 2.048 Å [Kra01]. Multi-reference perturbation theory estimates an equilibrium bond length of $r_e = 2.0561$ Å and $D_e = 4.5110$ eV [Ang07] in the $^1\Sigma$ state. Furthermore, a CASPT2 study found a dissociation energy of 5.37 eV and $r_e = 2.010$ Å [Bor10]. We aim to add DMC among the predictions of the dissociation energy for W_2 . A summary of the predictions are listed in Table 8.2. Although there is consistent discrepancy among various theories for the dissociation energies, the vibrational frequencies are in reasonable agreement and are quite small. Experimental detection of W_2 in an argon matrix found $\omega_e = 336.8$ cm^{-1} [Hu92], and theoretical estimates include $\omega_e = 401.2$ cm^{-1} [Kra01], $\omega_e = 326.69$ cm^{-1} [Ang07], and $\omega_e = 354$ cm^{-1} [Bor10]. Because of the very small vibrational frequencies, the binding well is relatively shallow and energetic changes for geometric differences on the order of 0.01 Å as is seen between the various other predictions, will be difficult to discern within statistical errors within DMC, so we adopt a geometry of the intermediate bond length of $r = 2.048$ Å to perform all our calculations.

We now turn to our DMC results. Utilizing a scalar relativistic Hamiltonian within FNDMC, we test both a PBE and PBE0 nodal surface in order to minimize the DMC energy. PBE0 results in a lower energy by 3 mHa, and produces a dissociation energy of 5.34(2) eV. Since W_2 is a closed shell molecule, when we include spin-orbit within FPSODMC we first try a single slater determinant built from DHF spinors. However, this only results in underbinding with dissociation energy of 2.51(2) eV that results from rather poor approximation of the phase and consequently a larger fixed-phase error. We attempt to improve the fixed-phase approximation by using the COSCI expansion where the anti-bonding spinors are included in the active space. While this lowers the total energy within FPSODMC, the dissociation energy increases only to 3.03(2) eV. As has been seen in other transition metal



$$\rho_{\text{PBE0}} - \rho_{\text{DHF}} = 0.002$$

$$\rho_{\text{DHF}} - \rho_{\text{PBE0}} = 0.002$$

Figure 8.2 The difference in the charge density between DHF and PBE0 trial wave functions for the outermost spinors $\pi^4\sigma_{z_2}^2\sigma_s^2\delta^4$. In red, we show where PBE0 has a greater charge density and in blue where DHF shows a greater charge density. The black spheres indicate the location of the Tungsten atoms.

Table 8.2 FNDMC (AREP) and FPSODMC (REP) dissociation energies of W_2 compared with other methods. In the fixed-node DMC calculations the trial nodes are from Slater determinants built from PBE0 single-particle orbitals.

Method	D_0/eV	$r_e/\text{\AA}$	Ref.
FNDMC/PBE0 [†]	5.34(2)	2.048*	This Work
PBE [†]	4.57	2.048*	This Work
PBE0 [†]	2.54	2.048*	This Work
B3LYP [†]		2.048	[Kra01]
SC-NEVPT3 [†]	4.5110	2.056	[Ang07]
CASPT2 [†]	5.37	2.010	[Bor10]
FPSODMC/DF [‡]	2.51(2)	2.048*	This Work
FPSODMC/COSCI [‡]	3.03(2)	2.048*	This Work
FPSODMC/PBE0 [‡]	4.17(2)	2.048*	This Work
Estimated	5(1)		[Mor86]

* Indicates the calculation was performed at this geometry.

[†] Indicates a spin-averaged calculation

[‡] Indicates an explicit spinor calculation

systems [Wag07], HF orbitals tend to localize charge too much close to the ions [WM07; Kol10]. This is due to the fact that the only way for HF to lower the energy is through exchange, which becomes larger when the charge density and particular states become more localized. The same bias is then built-in also in the excited virtual one-particle space. This results in a well-known "too ionic" bias of the HF or HF-based wave functions and it is expected that the same persists with spin-orbit included. We therefore consider further possibilities for improving the trial phases here. Since W_2 is a closed-shell molecule, we can utilize a single Slater determinant built from DFT relativistic spinors. In Figure 8.2, we illustrate the difference in the relativistic charge density between the PBE0 and DHF trial wave functions. As in the case of other FNDMC treatments of transition metals, HF localizes more on the atoms whereas the PBE0 pushes more charge density into the bonding region. For the PBE0 trial phase, we find a significantly lower energy resulting in a dissociation energy of 4.17(2) eV. To ensure the quality of the trial phase, we also try LDA, PBE, and B3LYP trial wave functions. All of these result in the same total energy as PBE0 within statistical error bars. Note that compared to the FNDMC without spin-orbit, there is a significant contribution to the dissociation energy coming from spin-orbit.

8.5 Conclusions

We have presented our recent development for QMC method with Hamiltonians that include particle spins. We applied these developments to two molecules with tungsten, WO and W_2 , and we compared the results obtained with traditional spin-orbit averaged FNDMC and with explicit treatment of spin-orbit in the FPSODMC methods. We illustrate that DMC is able to deal with systems containing heavy elements by using both scalar relativistic as well as spin-orbit interactions through the use of non-local ECPs and that the results enable to see the differences that result from proper treatment of spins as quantum variables.

For WO, we find that the ground state is in agreement with the experimental determination, and both the scalar relativistic and spin-orbit prediction for the dissociation energy are in agreement with the (large) experimental error. We also show that previous theoretical estimates for the excited state predict an equilibrium bond length appears larger by a few percent. For W_2 , there very limited amount of experimental data to compare against. Instead, we make a genuine prediction for the binding energy. Additionally we show that

for the heavy transition metals, trial phases that localize charge density too much on the atoms results in a larger fixed-phase bias. This agrees with previous investigations in transition metal systems where the use of the DFT/hybrid functionals generally improved the one-particle orbitals by easing too ionic character of Hartree-Fock based orbitals [WM07; Kol10].

DMC has been one of the most accurate methods for treating strong correlations in various materials, and with the addition of spin-orbit it can now treat heavy-element compounds to high accuracy. We note however that in order to reach higher accuracies, various improvements need to be addressed. In particular, more accurate relativistic ECPs that are designed to be used in a correlated framework with much higher accuracy and better benchmarking will be needed. Additionally, higher accuracy trial wave functions will be important. This will presumably involve combinations of configuration interaction expansions, improved open-shell DFT trial wave functions and their combinations. Nevertheless, the presented developments open new possibilities for accurate many-body calculations of systems with heavy atoms and significant spin orbit-effect by accurate QMC methods.

Within DMC, we investigated a number of heavy transition metal systems. In particular, we found for benzene systems with molybdenum and tungsten atoms in half- and full-sandwich geometries that nodal errors decrease significantly when moving from the $4d$ - to $5d$ -element systems which we attribute to the decrease in electronic density and a more favorable ordering of the s , p and d levels in the latter case – these results imply that accurate and predictive calculations are feasible for these heavy element systems with fixed-node DMC at the single-determinant level, which provides encouragement for future DMC studies of similar systems. We also elaborated on fixed-phase DMC with spin-orbit and compared its results with those from fixed-node DMC for a test bed of atoms and molecules – revealing the possibility for the inclusion of these effects in future studies of systems where spin-orbit is significant. We have also investigated the level of errors introduced from ECPs, a necessary approximation for heavy atoms, and found at the CCSD(T) level that previously tabulated potentials can have large errors relative to all-electron results in atomic spectra and molecular binding parameters. We therefore explored possibilities to generate more accurate potentials that account for many-body effects from the outset and found a strategies to generate ECPs that mimic the atomic and molecular properties of all-electron cores at the highly accurate CCSD(T) level of theory. We have used our many-body construction strategies to generate ECPs for 1st row elements, the entire set of 2nd row

elements, and the full $3d$ transition metal series.

Bibliography

- [Dir] DIRAC, a relativistic ab initio electronic structure program, Release DIRAC15 (2015), written by R. Bast, T. Saue, L. Visscher, and H. J. Aa. Jensen, with contributions from V. Bakken, K. G. Dyall, S. Dubillard, U. Ekstroem, E. Eliav, T. Enevoldsen, E. Fasshauer, T. Fleig, O. Fossgaard, A. S. P. Gomes, T. Helgaker, J. Henriksson, M. Ilias, Ch. R. Jacob, S. Knecht, S. Komorovsky, O. Kullie, J. K. Laerdahl, C. V. Larsen, Y. S. Lee, H. S. Nataraj, M. K. Nayak, P. Norman, G. Olejniczak, J. Olsen, Y. C. Park, J. K. Pedersen, M. Pernpointner, R. Di Remigio, K. Ruud, P. Salek, B. Schimmelpfennig, J. Sikkema, A. J. Thorvaldsen, J. Thyssen, J. van Stralen, S. Villaume, O. Visser, T. Winther, and S. Yamamoto (see <http://www.diracprogram.org>).
- [Nis] A. Kramida, Y. Ralchenko, J. Reader and NIST ASD Team (2013), *NIST Atomic Spectra Database (Version 5.1)*, NIST and references therein.
- [Eat] D.L. Carpenter, A.M. Covington, and J.S. Thompson, Unpublished data, (1999) cited in T. Andersen, H.K. Haugen, and H.J. Hotop, *Phys. Chem. Ref. Data*, **28**, 1511(1999).
- [AC94] Acioli, P. H. & Ceperley, D. M. "Generation of pseudopotentials from correlated wave functions". *J. Chem. Phys.* **100**.11 (1994), p. 8169.
- [Amb09] Ambrosetti, A. et al. "Quantum Monte Carlo study of the two-dimensional electron gas in presence of Rashba interaction". *Phys. Rev. B* **80**.12 (2009), p. 125306.
- [Amb12] Ambrosetti, A. et al. "Variational Monte Carlo for spin-orbit interacting systems". *Phys. Rev. B* **85**.4 (2012), p. 045115.
- [And99] Andersen, T. et al. "Binding energies in atomic negative ions: III". *Journal of Physical and Chemical Reference Data* **28** (1999), p. 1511.
- [And94] Andersson, K. et al. "The Cr₂ potential energy curve studied with multiconfigurational second-order perturbation theory". *Chem. Phys. Lett.* **230**.Dec (1994), p. 391.
- [Ang07] Angeli, C. et al. "Ground states of the Mo₂, W₂, and CrMo molecules: A second and third order multireference perturbation theory study". *The Journal of Chemical Physics* **127**.7 (2007), p. 074306.
- [AH82] Atha, P. M. & Hillier, I. H. "Correlation effects and the bonding in Mo₂ and Cr₂". *Molecular Physics* **45**.2 (1982), p. 285.

- [Bac82] Bachelet, G. B. et al. "Pseudopotentials that work: From H to Pu". *Phys. Rev. B* **26**.8–15 (1982), p. 4199.
- [Baj06] Bajdich, M. et al. "Pfaffian Pairing Wave Functions in Electronic-Structure Quantum Monte Carlo Simulations". *Phys. Rev. Lett.* **96**.13 (2006), p. 130201.
- [BM09] Bajdich, M. & Mitas, L. "Electronic structure quantum Monte Carlo". *Acta Physica Slovaca* **59** (2009), pp. 81–168.
- [Baj05] Bajdich, M. et al. "Approximate and exact nodes of fermionic wavefunctions: Coordinate transformations and topologies". *Phys. Rev. B* **72** (2005), p. 075131.
- [BP05] Balabanov, N. B. & Peterson, K. A. "Systematically convergent basis sets for transition metals. I. All-electron correlation consistent basis sets for the 3d elements Sc–Zn". *The Journal of Chemical Physics* **123**.6 (2005), p. 064107.
- [Bal02] Balasubramanian, K. "Spectroscopic properties of Mo_2^- and Mo_2^+ ". *Chem. Phys. Lett.* **365** (2002), p. 413.
- [BZ02] Balasubramanian, K. & Zhu, X. "Spectroscopic constants and potential energy curves of electronic states of Mo_2 ". *J. Chem. Phys.* **117**.10 (2002), p. 4861.
- [BT88] Bauschlicher, C. W. & Taylor, P. R. "Symmetry and equivalence restrictions in electronic structure calculations". *Theoretica chimica acta* **74**.1 (1988), pp. 63–73.
- [Bay84] Baykara, N. A. et al. "LCAO local-spin-density and $X\alpha$ calculations for Cr_2 and Mo_2 ". *Molecular Physics* **52**.4 (1984), p. 891.
- [Ben17a] Bennett, M. C. et al. "A new generation of effective core potentials for correlated calculations". *J. Chem. Phys.* **147** (2017), p. 224106.
- [Ben17b] Bennett, M. C. et al. "A quantum Monte Carlo study of mono(benzene) TM and bis(benzene) TM systems". *Chemical Physics Letters* **667** (2017), pp. 74–78.
- [Ben18] Bennett, M. C. et al. "New generation of effective core potentials from correlated calculations: 2nd row elements". *ArXiv e-prints* (2018).
- [BH01] Bilodeau, R. C. & Haugen, H. K. "Electron affinity of Bi using infrared laser photodetachment threshold spectroscopy". *Phys. Rev. A* **64**.2 (2001), p. 024501.
- [Boo13] Booth, G. et al. "Towards an exact description of electronic wavefunctions in real solids". *Nature* **493** (2013), p. 365.

- [Bor08] Borin, A. C. et al. "A theoretical study of the binding and electronic spectrum of the Mo₂ molecule". *Chem. Phys.* **343**.2-3 (2008), p. 210.
- [Bor10] Borin, A. C. et al. "Electronic structure and chemical bonding in W₂ molecule". *Chemical Physics Letters* **490**.1 (2010), pp. 24–28.
- [Bre29] Breit, G. "The Effect of Retardation on the Interaction of Two Electrons". *Phys. Rev.* **34**.4 (1929), p. 553.
- [BW80] Brewer, L. & Winn, J. S. "Models for calculation of dissociation energies of homonuclear diatomic molecules". *Faraday Symp. Chem. Soc.* **14**.0 (1980), pp. 126–135.
- [Bur07] Burkatzki, M. et al. "Energy-consistent pseudopotentials for quantum Monte Carlo calculations". *J. Chem. Phys.* **126**.23 (2007), p. 234105.
- [Bur08] Burkatzki, M. et al. "Energy-consistent small-core pseudopotentials for 3d-transition metals adapted to quantum Monte Carlo calculations". *The Journal of Chemical Physics* **129**.16 (2008), p. 164115.
- [Caf16] Caffarel, M. et al. "Communication: Toward an improved control of the fixed-node error in quantum Monte Carlo: The case of the water molecule". *J. Chem. Phys.* **144** (2016), p. 151103.
- [CK85] Carlson, J. & Kalos, M. H. "Variational Monte Carlo calculation of ¹⁶O". *Phys. Rev. C* **32**.6 (1985), p. 2105.
- [Cas06] Casula, M. "Beyond the locality approximation in the standard diffusion Monte Carlo method". *Physical Review B* **74**.16 (2006), p. 161102.
- [Cep78] Ceperley, D. M. "Ground state of the fermion one-component plasma: A Monte Carlo study in two and three dimensions". *Phys. Rev. B.* **18**.7 (1978), p. 3126.
- [Cep86] Ceperley, D. M. "The statistical error of green's function Monte Carlo". *Journal of Statistical Physics* **43**.5 (1986), pp. 815–826.
- [CA80] Ceperley, D. M. & Alder, B. J. "Ground State of the Electron Gas by a Stochastic Method". *Phys. Rev. Lett.* **45**.7 (1980), p. 566.
- [CM96] Ceperley, D. M. & Mitas, L. "Monte Carlo methods in quantum chemistry". *Advances in Chemical Physics, Vol. XCIII*. Ed. by Prigogine, I. & Rice, S. A. New York: John Wiley and Sons, 1996, pp. 1–38.

- [Cha77] Chang, T.-C. et al. "Accuracy and limitations of the pseudopotential method. II. First-row molecules". *Theoretica chimica acta* **44** (1977), p. 61.
- [Chr79] Christiansen, P. A. et al. "Improved ab initio effective core potentials for molecular calculations". *J. Chem. Phys.* **71** (1979), p. 4445.
- [Cla98] Claridge, J. B. et al. "New Catalysts for the Conversion of Methane to Synthesis Gas: Molybdenum and Tungsten Carbide". *Journal of Catalysis* **180.1** (1998), pp. 85–100.
- [Cla] *Clarkson University Relativistic Effective Potential Database*. <http://people.clarkson.edu/~pchristi/reps.html>.
- [Dac99] Dachsel, H. et al. "Multireference Configuration Interaction Calculations on Cr₂: Passing the One Billion Limit in MRCI/MRACPF Calculations". *J. Phys. Chem. A* **103** (1999), p. 152.
- [DN70] Darwent, B. deB. & National Standard Reference Data System (U.S.) *Bond dissociation energies in simple molecules [electronic resource] / B. deB. Darwent*. English. U.S. Dept. of Commerce, National Bureau of Standards [Washington, D.C.], 1970, 1 online resource (iv, 48 p.)
- [Dat] *Data for this paper, published in the Materials Data Facility*. <http://dx.doi.org/doi:10.18126/M2DK97> .
- [Des73] Desclaux, J. P. "Relativistic Dirac-Fock expectation values for atoms with $Z = 1$ to $Z = 120$ ". *At. Data Nucl. Data Tables* **12.4** (1973), pp. 311–406.
- [Des75] Desclaux, J. P. "A multiconfiguration relativistic DIRAC-FOCK program". *Comput. Phys. Commun.* **9.1** (1975), pp. 31–45.
- [DK75] Desclaux, J. P. & Kim, Y. K. "Relativistic effects in outer shells of heavy atoms". *J. Phys. B: At. Mol. Phys.* **8.8** (1975), p. 1177.
- [DF29] Dirac, P. A. M. & Fowler, R. H. "Quantum mechanics of many-electron systems". *Proc. Roy. Soc. A* **123.792** (1929), p. 714.
- [Dol05] Dolg, M. "Improved relativistic energy-consistent pseudopotentials for 3d-transition metals". *Theor. Chem. Acta* **114** (2005), p. 297.
- [Dol87] Dolg, M. et al. "Energy-adjusted ab initio pseudopotentials for the first row transition elements". *J. Chem. Phys.* **86.2** (1987), p. 866.

- [Dol96] Dolg, M. "Valence correlation energies from pseudopotential calculations". *Chem. Phys. Lett.* **250** (1996), p. 75.
- [DC12] Dolg, M. & Cao, X. "Relativistic Pseudopotentials: Their Development and Scope of Applications". *Chem. Rev.* **112.1** (2012), p. 403.
- [Dun02] Dunning Jr., T. H. et al. *Encyclopedia of Computational Chemistry*. Wiley, 2002. Chap. Basis Sets: Correlation Consistent Sets.
- [Efr78] Efremov, Y. M. et al. "On the electronic spectrum of the Mo₂ molecule observed after flash photolysis of Mo(CO)₆". *J. Mol. Spectrosc.* **73** (1978), p. 430.
- [Erm81] Ermler, W. C. et al. "Ab initio effective core potentials including relativistic effects. A procedure for the inclusion of spin-orbit coupling in molecular wavefunctions". *Chem. Phys. Lett.* **81.1** (1981), pp. 70–74.
- [Fan17] Fang, Z. et al. "Prediction of bond dissociation energies/heats of formation for diatomic transition metal compounds: CCSD (T) works". *Journal of chemical theory and computation* **13.3** (2017), pp. 1057–1066.
- [Fei81] Feigerle, C. S. et al. "Electron affinities of B, Al, Bi, and Pb". *J. Chem. Phys.* **74.2** (1981), p. 1513.
- [FP99] Feller, D. & Peterson, K. A. "Re-examination of atomization energies for the Gaussian-2 set of molecules". *The Journal of chemical physics* **110.17** (1999), pp. 8384–8396.
- [Fig09] Figgen, D. et al. "Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf–Pt". *The Journal of Chemical Physics* **130.16** (2009), p. 164108.
- [Fou01] Foulkes, W. M. C. et al. "Quantum Monte Carlo simulations of solids". *Rev. Mod. Phys.* **73** (2001), p. 33.
- [Foy14] Foyevtsova, K. et al. "Ab initio Quantum Monte Carlo Calculations of Spin Superexchange in Cuprates: The Benchmarking Case of Ca₂CuO₃". *Phys. Rev. X* **4.3** (2014), p. 031003.
- [Fri] Frisch, M. J. et al. *Gaussian 09 Revision E.01*. Gaussian Inc. Wallingford CT 2009.
- [Fro04] Fromager, E. et al. "An analysis of core effects on shape-consistent pseudopotentials". *J. Chem. Phys.* **121.18** (2004), p. 8687.

- [Fu90] Fu, Z. et al. "Spectroscopy and electronic structure of jet-cooled Al₂". *The Journal of Chemical Physics* **93**.12 (1990), pp. 8420–8441.
- [FP06] Furche, F. & Perdew, J. P. "The performance of semilocal and hybrid density functionals in 3 d transition-metal chemistry". *The Journal of chemical physics* **124**.4 (2006), p. 044103.
- [Gan06] Gandolfi, S. et al. "Auxiliary field diffusion Monte Carlo calculation of properties of oxygen isotopes". *Phys. Rev. C* **73**.4 (2006), p. 044304.
- [Gan07] Gandolfi, S. et al. "Auxiliary Field Diffusion Monte Carlo Calculation of Nuclei with $A \leq 40$ with Tensor Interactions". *Phys. Rev. Lett.* **99**.2 (2007), p. 022507.
- [GG82] Goodgame, M. M. & Goddard III, W. A. "Nature of Mo-Mo and Cr-Cr Multiple Bonds: A Challenge for the Local-Density Approximation". *Phys. Rev. Lett.* **48**.3 (1982), p. 135.
- [Gra70] Grant, I. P. "Relativistic calculation of atomic structures". *Adv. Phys.* **19**.82 (1970), pp. 747–811.
- [Haa95] Haaf, D. F. B. ten et al. "Proof for an upper bound in fixed-node Monte Carlo for lattice fermions". *Phys. Rev. B* **51**.19 (1995), p. 13039.
- [Ham13] Hamann, D. R. "Optimized norm-conserving Vanderbilt pseudopotentials". *Phys. Rev. B* **88**.8 (2013), p. 085117.
- [Ham87] Hammond, B. L. et al. "Valence quantum Monte Carlo with ab initio effective core potentials". *The Journal of Chemical Physics* **87**.2 (1987), pp. 1130–1136.
- [HW85] Hay, P. J. & Wadt, W. R. "Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg". *J. Chem. Phys.* **82**.1 (1985), p. 270.
- [HM12] Hongo, K. & Maezono, R. "A benchmark quantum Monte Carlo study of the ground state chromium dimer". *Int. J. Quantum Chem.* **112**.5 (2012), p. 1243.
- [Hor12] Horváthová, L. et al. "Spin Multiplicity and Symmetry Breaking in Vanadium-Benzene Complexes". *Phys. Rev. Lett.* **109**.5 (2012), p. 053001.
- [HL85] Hotop, H. & Lineberger, W. C. "Binding Energies in Atomic Negative Ions: II". *J. Phys. and Chem. Ref. Data* **14**.3 (1985), p. 731.

- [Hu92] Hu, Z. et al. "Optical and Raman spectroscopy of mass-selected tungsten dimers in argon matrices". *The Journal of Chemical Physics* **97**.11 (1992), pp. 8811–8812.
- [HH79] Huber, K. P. & Herzberg, G. *Molecular spectra and molecular structure: IV. Constants of diatomic molecules*. Van Nostrand Reinhold Co., 1979.
- [Stu] *Institute for Theoretical Chemistry*. <http://www.uni-stuttgart.de/theochem/>.
- [Iri07] Irikura, K. K. "Experimental vibrational zero-point energies: Diatomic molecules". *Journal of physical and chemical reference data* **36**.2 (2007), pp. 389–397.
- [Joh18] Johnson III, R. D. NIST Computational Chemistry Comparison and Benchmark Database NIST Standard Reference Database Number 101 Release 19, [Online]. Available: <http://cccbdb.nist.gov/> [2018, April 19]. National Institute of Standards and Technology, Gaithersburg, MD. 2018.
- [Kah76] Kahn, L. R. et al. "Ab initio effective core potentials: Reduction of all-electron molecular structure calculations to calculations involving only valence electrons". *J. Chem. Phys.* **65**.10 (1976), p. 3826.
- [KB82] Kleinman, L. & Bylander, D. M. "Efficacious Form for Model Pseudopotentials". *Phys. Rev. Lett.* **48**.20 (1982), p. 1425.
- [KS65] Kohn, W. & Sham, L. J. "Self-Consistent Equations Including Exchange and Correlation Effects". *Phys. Rev.* **140**.4A (1965), A1133–A1138.
- [KM11] Kolorenc, J. & Mitas, L. "Applications of quantum Monte Carlo methods in condensed systems". *Rep. Prog. Phys.* **74** (2011), p. 026502.
- [KM08] Kolorenč, J. & Mitas, L. "Quantum Monte Carlo calculations of structural properties of FeO under pressure". *Phys. Rev. Lett.* **101** (2008), p. 185502.
- [Kol10] Kolorenč, J. et al. "Wave functions for quantum Monte Carlo calculations in solids: Orbitals from density functional theory with hybrid exchange-correlation functionals". *Phys. Rev. B* **82**.11 (2010), p. 115108.
- [Kra18] Kramida, A. et al. NIST Atomic Spectra Database (ver. 5.5.6), [Online]. Available: <https://physics.nist.gov/asd> [2018, April 19]. National Institute of Standards and Technology, Gaithersburg, MD. 2018.
- [Kra75] Krasnov, K. "On transition metal molecules". *Teplofizika Vysokikh Temperatur* **13**.2 (1975), pp. 441–443.

- [Kra01] Kraus, D. et al. "On the dimers of the VIB group: a new NIR electronic state of Mo₂". *PhysChemComm* **4**.10 (2001), pp. 44–48.
- [KK17] Krogel, J. T. & Kent, P. R. C. "Magnitude of pseudopotential localization errors in fixed node diffusion quantum Monte Carlo". *J. Chem. Phys.* **146**.24 (2017), p. 244101.
- [Kro16] Krogel, J. T. et al. "Pseudopotentials for quantum Monte Carlo studies of transition metal oxides". *Phys. Rev. B* **93** (2016), p. 075143.
- [Kul14] Kulahlioglu, A. H. et al. "Density dependence of fixed-node errors in diffusion quantum Monte Carlo: Triplet pair correlations". *Chem. Phys. Lett.* **591** (2014), p. 170.
- [Lee77a] Lee, Y. S. et al. "Relativistic effects in outer shells of heavy atoms". *J. Chem. Phys.* **67**.12 (1977), p. 5861.
- [Lee77b] Lee, Y. S. et al. "Ab initio effective core potentials including relativistic effects. I. Formalism and applications to the Xe and Au atoms". *The Journal of Chemical Physics* **67**.12 (1977), pp. 5861–5876.
- [Lej14] Lejaeghere, K. et al. "Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals". *Critical Reviews in Solid State and Materials Sciences* **39**.1 (2014), pp. 1–24.
- [Lej16] Lejaeghere, K. et al. "Reproducibility in density functional theory calculations of solids". *Science* **351**.6280 (2016).
- [Liu15] Liu, Z. et al. "Application of the multireference equation of motion coupled cluster method, including spin-orbit coupling, to the atomic spectra of Cr, Mn, Fe and Co". *Molecular Physics* **113**.19-20 (2015), pp. 2999–3013.
- [Lou82] Louie, S. G. et al. "Nonlinear ionic pseudopotentials in spin-density-functional calculations". *Phys. Rev. B* **26**.4–16 (1982), p. 1738.
- [Lov05] Lovas, F. J. et al. NIST Diatomic Spectral Database 114, [Online]. Available: www.physics.nist.gov/PhysRefData/MolSpec/Diatomic/index.html [2005, November]. Institut für Quantenoptik, Universität Hannover, Hannover, Germany. 2005.
- [Lu71] Lu, C. C. et al. "Relativistic Hartree-Fock-Slater eigenvalues, radial expectation values, and potentials for atoms, $2 \leq Z \leq 126$ ". *At. Data Nucl. Data Tables* **3** (1971), pp. 1–31.

- [Lüc07] Lüchow, A. et al. “Direct optimization of nodal hypersurfaces in approximate wave functions”. *J. Chem. Phys.* **126** (2007), p. 144110.
- [Luo16] Luo, Y. et al. “Phase stability of TiO₂ polymorphs from diffusion Quantum Monte Carlo”. *New Journal of Physics* **18**.11 (2016), p. 113049.
- [Man17] Manzeli, S. et al. “2D Transition Metal Dichalcogenides”. *Nature Reviews Materials* **2**.17033 (2017).
- [MT98] Maron, L. & Teichteil, C. “On the accuracy of averaged relativistic shape-consistent pseudopotentials”. *Chem. Phys.* **237**.1–2 (1998), p. 105.
- [MH84] Masuda, T. & Higashimura, T. “Synthesis of high polymers from substituted acetylenes: exploitation of molybdenum- and tungsten-based catalysts”. *Accounts of Chemical Research* **17**.2 (1984), pp. 51–56.
- [Mel16a] Melton, C. et al. “Spin-orbit interactions in electronic structure quantum Monte Carlo methods”. *Phys. Rev. A* **93** (2016), p. 042502.
- [MM16] Melton, C. A. & Mitas, L. “Fixed-Node and Fixed-Phase Approximations and Their Relationship to Variable Spins in Quantum Monte Carlo”. *Recent Progress in Quantum Monte Carlo*. 2016. Chap. 1, pp. 1–13.
- [Mel16b] Melton, C. A. et al. “Quantum Monte Carlo with variable spins”. *The Journal of Chemical Physics* **144**.24 (2016), p. 244113.
- [Mel17] Melton, C. A. et al. “Quantum Monte Carlo with Variable Spins: Fixed-Phase and Fixed-Node Approximations”. *Phys. Rev. E* **96**.4 (2017), p. 043305.
- [Mel18] Melton, C. A. et al. *J. Phys. Chem. Sol.* (2018).
- [MS00] Metz, B. & Stoll, H. “Small-core multiconfiguration-Dirac-Hartree-Fock-adjusted pseudopotentials for post-*d* main group elements: Application to PbH and PbO”. *J. Chem. Phys.* **113**.7 (2000), p. 2563.
- [Met00] Metz, B. et al. “A small-core multiconfiguration Dirac–Hartree–Fock-adjusted pseudopotential for Tl – application to TlX (X = F, Cl, Br, I)”. *Theor. Chem. Acc.* **104**.1 (2000), pp. 22–28.
- [Mic82] Michalopoulos, D. L. et al. “The bond length of chromium dimer”. *J. Phys. Chem.* **86**.20 (1982), p. 3914.

- [MG79] Miedema, A. R. & Gingerich, K. A. "On the formation enthalpy of metallic dimers". *Journal of Physics B: Atomic and Molecular Physics* **12.13** (1979), p. 2081.
- [Mie80] Miedema, A. R. "Model predictions of the dissociation energies of homonuclear and heteronuclear diatomic molecules of two transition metals". *Faraday Symp. Chem. Soc.* **14.0** (1980), p. 136.
- [MK10] Mitas, L. & Kolorenc, J. "Quantum Monte Carlo Studies of Transition Metal Oxides". *Review in Mineralogy and Geochemistry* **71.1** (2010), pp. 137–145.
- [Mit91] Mitas, L. et al. "Nonlocal pseudopotentials and diffusion Monte Carlo". *The Journal of Chemical Physics* **95.5** (1991), p. 3467.
- [Mor01] Moravec, V. D. et al. "The electronic structure of ZnO and ZnF determined by anion photoelectron spectroscopy". *Chemical physics letters* **341.3-4** (2001), pp. 313–318.
- [Mor86] Morse, M. D. "Clusters of Transition-Metal Atoms". *Chemical Review* **86** (1986), pp. 1049–1109.
- [Mül09] Müller, T. "Large-Scale Parallel Uncontracted Multireference-Averaged Quadratic Coupled Cluster: The Ground State of the Chromium Dimer Revisited". *J. Phys. Chem. A* **113.45** (2009), p. 12729.
- [MM84] Muller, W. & Meyer, W. "Ground-state properties of alkali dimers and their cations (including the elements Li, Na, and K) from *ab initio* calculations with effective core polarization potentials". *J. Chem. Phys.* **80** (1984), p. 3311.
- [Mul84] Muller, W. et al. "Treatment of intershell correlation effects in *ab initio* calculations by use of core polarization potentials. Method and application to alkali and alkaline earth atoms". *J. Chem. Phys. Rev.* **80** (1984), p. 3297.
- [NI11] Nakada, K. & Ishii, A. "DFT Calculation for Adatom Adsorption on Graphene". *Graphene Simulation*. Ed. by Gong, J. 2011, pp. 3–15.
- [Naz16] Nazarov, R. et al. "Benchmarking the pseudopotential and fixed-node approximations in diffusion Monte Carlo calculations of molecules and solids". *Phys. Rev. B* **93.9** (2016), p. 094111.
- [NB85] Nelin, C. J. & Bauschlicher, C. W. "On the low-lying states of WO: A comparison with CrO and MoO". *Chemical Physics Letters* **118.2** (1985), pp. 221–225.

- [OD03] Oganov, A. R. & Dorogokupets, P. I. “All-electron and pseudopotential study of MgO: Equation of state, anharmonicity, and stability”. *Phys. Rev. B* **67** (2003), p. 224110.
- [Ort93] Ortiz, G. et al. “New stochastic method for systems with broken time-reversal symmetry: 2D fermions in a magnetic field”. *Phys. Rev. Lett.* **71**.17 (1993), p. 277.
- [Ovc01] Ovcharenko, I. et al. “Soft pseudopotentials for efficient quantum Monte Carlo calculations: From Be to Ne and Al to Ar”. *J. Chem. Phys.* **114**.18 (2001), p. 7790.
- [PC85] Pacios, L. F. & Christiansen, P. A. “Ab initio relativistic effective potentials with spin-orbit operators. I. Li through Ar”. *J. Chem. Phys.* **82**.6 (1985), p. 2664.
- [Pak88] Pak, K. et al. “Dissociation energy and excited-state properties of diatomic Sn₂”. *Faraday Discuss. Chem. Soc.* **86** (1988), pp. 153–162.
- [Ped04] Pederiva, F. et al. “Auxiliary field diffusion Monte Carlo calculation of ground state properties of neutron drops”. *Nucl. Phys. A* **742**.1–2 (2004), pp. 255–268.
- [Pet03] Peterson, K. A. “Systematically convergent basis sets with relativistic pseudopotentials. I. Correlation consistent basis sets for the post-*d* group 13–15 elements”. *J. Chem. Phys.* **119**.21 (2003), p. 11099.
- [Pet07] Peterson, K. A. et al. “Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4*d* elements Y-Pd”. *J. Chem. Phys.* **126**.12 (2007), p. 124101.
- [Pic89] Pickett, W. E. “Pseudopotential methods in condensed matter applications”. *Comp. Phys. Rep.* **9**.3 (1989), p. 115.
- [Web] *Pseudopotential Library*. URL: <http://www.pseudopotentiallibrary.org>.
- [Ram01] Ram, R. S. et al. “The X³Σ⁻ ground state of WO”. *Chemical Physics Letters* **343**.3 (2001), pp. 437–445.
- [RB96] Ram, R. & Bernath, P. “Fourier transform emission spectroscopy of the B 1Π–X 1Σ⁺, C 1Σ⁺–X 1Σ⁺, and G 1Π–X 1Σ⁺ systems of ScH and ScD”. *The Journal of chemical physics* **105**.7 (1996), pp. 2668–2674.
- [Ran15] Ranasinghe, D. S. et al. “Core-core and core-valence correlation energy atomic and molecular benchmarks for Li through Ar”. *J. Chem. Phys.* **143** (2015), p. 214110.

- [RM12] Rasch, K. M. & Mitas, L. “Impact of electron density on the fixed-node errors in Quantum Monte Carlo of atomic systems”. *Chem. Phys. Lett.* **528** (2012), p. 59.
- [Ras14] Rasch, K. M. et al. “Communication: Fixed-node errors in quantum Monte Carlo: Interplay of electron density and node nonlinearities”. *J. Chem. Phys.* **140.4** (2014), p. 041102.
- [RW04] Reiher, M. & Wolf, A. “Exact decoupling of the Dirac Hamiltonian. I. General theory”. *J. Chem. Phys.* **121** (2004), p. 2037.
- [RC82] Reynolds, P. J. & Ceperley, D. M. “Fixed-node quantum Monte Carlo for molecules”. *J. Chem. Phys.* **77.11** (1982), p. 5593.
- [Rey82] Reynolds, P. J. et al. “Fixed-node quantum Monte Carlo for molecules”. *The Journal of Chemical Physics* **77.11** (1982), pp. 5593–5603.
- [Ros79] Rosmus, P. “Molecular constants for the $1 \Sigma^+$ ground state of the ArH^+ ion”. *Theoretica chimica acta* **51.4** (1979), pp. 359–362.
- [Ros90] Ross, R. B. et al. “Ab initio relativistic effective potentials with spin-orbit operators. IV. Cs through Rn”. *J. Chem. Phys.* **93.9** (1990), p. 6654.
- [Sar03] Sarsa, A. et al. “Neutron matter at zero temperature with an auxiliary field diffusion Monte Carlo method”. *Phys. Rev. C* **68.2** (2003), p. 024308.
- [SG15] Schlipf, M. & Gygi, F. “Optimization algorithm for the generation of ONCV pseudopotentials”. *Computer Physics Communications* **196** (2015), pp. 36–44.
- [Sch01] *Quantum Monte Carlo: Recent Advances and Common Problems in Condensed Matter and Field Theory*. Pisa, Italy, 2001, pp. 143–170.
- [Sch93a] Schmidt, M. W. et al. “General Atomic and Molecular Electronic Structure System”. *J. Comput. Chem.* **44** (1993), p. 1347.
- [Sch93b] Schmidt, M. W. et al. “General atomic and molecular electronic structure system”. *Journal of Computational Chemistry* **14.11** (1993), pp. 1347–1363.
- [Shi] Shirley, E. L. private communication.
- [Shi91] Shirley, E. et al. “Core-valence partitioning and quasiparticle pseudopotentials”. *Phys. Rev. B* **44** (1991), p. 3396.
- [SM93] Shirley, E. L. & Martin, R. M. “Many-body core-valence partitioning”. *Phys. Rev. B* **47** (1993), p. 15413.

- [SM13] Shulenburger, L. & Mattsson, T. R. “Quantum Monte Carlo applied to solids”. *Phys. Rev. B* **88**.24 (2013), p. 245117.
- [SY06] Sohnlein, B. R. & Yang, D.-S. “Pulsed-field ionization electron spectroscopy of group 6 metal (Cr, Mo, and W) bis(benzene) sandwich complexes.” *J. Chem. Phys.* **124**.13 (2006), p. 134305.
- [Spe] Spellucci, P. *DONLP2 Nonlinear Optimization Code*. URL: <http://www.mathematik.tu-darmstadt.de/fbereiche/numerik/staff/spellucci/DONLP2/>.
- [Spe98a] Spellucci, P. “A new technique for inconsistent QP problems in the SQP method”. *Math. Meth. of Oper. Res.* **47**.3 (1998), pp. 355–400.
- [Spe98b] Spellucci, P. “An SQP method for general nonlinear programs using only equality constrained subproblems”. *Math. Prog.* **82**.3 (1998), p. 413.
- [Ste84] Stevens, W. J. et al. “Compact effective potentials and efficient shared-exponent basis sets for the first- and second-row atoms”. *J. Chem. Phys.* **81**.12 (1984), p. 6026.
- [Sto02] Stoll, H. et al. “Relativistic energy-consistent pseudopotentials—recent developments”. *J. Comput. Chem.* **23**.8 (2002), p. 767.
- [THD89] T. H. Dunning Jr., T. H. “Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen”. *J. Chem. Phys.* **90** (1989), p. 1007.
- [Tat09] Tatewaki, H. et al. “Electron affinity of lead: An ab initio four-component relativistic study”. *Chem. Phys. Lett.* **470**.4–6 (2009), pp. 158–161.
- [TU07] Toulouse, J. & Umrigar, C. J. “Optimization of quantum Monte Carlo wave functions by energy minimization”. *J. Chem. Phys.* **126** (2007), p. 084102.
- [TN05] Trail, J. R. & Needs, R. J. “Smooth relativistic Hartree–Fock pseudopotentials for H to Ba and Lu to Hg”. *The Journal of Chemical Physics* **122**.17 (2005), p. 174109.
- [TN13] Trail, J. R. & Needs, R. J. “Pseudopotentials for correlated electron systems”. *J. Chem. Phys.* **139**.1 (2013), p. 014101.
- [TN15] Trail, J. R. & Needs, R. J. “Correlated electron pseudopotentials for 3d-transition metals”. *J. Chem. Phys.* **142**.6 (2015), p. 064110.

- [TN17] Trail, J. R. & Needs, R. J. “Shape and energy consistent pseudopotentials for correlated electron systems”. *J. Chem. Phys.* **146**.20 (2017), p. 204107.
- [Umr07] Umrigar, C. J. et al. “Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions”. *Phys. Rev. Lett.* **98** (2007), p. 110201.
- [Wag09a] Wagner, L. K. et al. “QWalk: A quantum Monte Carlo program for electronic structure”. *J. Comput. Phys.* **228**.9 (2009), p. 3390.
- [Wag07] Wagner, L. K. “Transition metal oxides using quantum Monte Carlo”. *Journal of Physics: Condensed Matter* **19**.34 (2007), p. 343201.
- [WM07] Wagner, L. K. & Mitas, L. “Energetics and dipole moment of transition metal monoxides by quantum Monte Carlo”. *J. Chem. Phys.* **126** (2007), p. 034105.
- [Wag09b] Wagner, L. K. et al. “QWalk: A quantum Monte Carlo program for electronic structure”. *Journal of Computational Physics* **228**.9 (2009), pp. 3390–3404.
- [Wal] Walter, E. J. *Opium Pseudopotential Generation Project*. URL: <http://opium.sourceforge.net>.
- [Wer12] Werner, H.-J. et al. “Molpro: a general-purpose quantum chemistry program package”. *WIREs Comput Mol Sci* **2** (2012), pp. 242–253.
- [Xu15] Xu, X. et al. “Do Practical Standard Coupled Cluster Calculations Agree Better than Kohn–Sham Calculations with Currently Available Functionals When Compared to the Best Available Experimental Data for Dissociation Energies of Bonds to 3 d Transition Metals?” *Journal of chemical theory and computation* **11**.5 (2015), pp. 2036–2052.
- [Zha04] Zhang, W. et al. “The nonmetallicity of molybdenum clusters”. *J. Chem. Phys.* **121**.16 (2004), pp. 7717–7724.
- [ZW15] Zheng, H. & Wagner, L. K. “Computation of the Correlated Metal-Insulator Transition in Vanadium Dioxide from First Principles”. *Phys. Rev. Lett.* **114**.17 (2015), p. 176401.
- [Zol69] Zollweg, R. J. “Electron Affinities of the Heavy Elements”. *J. Chem. Phys.* **50**.10 (1969), p. 4251.

Appendices

Supplementary Material: A new generation of effective core potentials for correlated calculations

A.1 Basis Set Extrapolation

Throughout this work, we have used the uncontracted aug-cc-pCV5Z basis set for all atoms. In order to analyze the effect of extrapolating to the complete basis set (CBS) limit, we provide comparisons between this basis set and extrapolating to the CBS limit. We define ϵ_{AE} as the error between the uncontracted V5Z basis and the CBS extrapolated result from CCSD(T) for the AE calculation. ϵ_{BFD} is the error between the uncontracted V5Z basis and the CBS extrapolated result from CCSD(T) for the BFD ECP [Bur07]. While the error for a given excitation can be a few mHa for each individual calculation, the quantity that we are matching in the spectral ECP is the discrepancy between the AE and the ECP calculation. The ϵ_{gap} is the discrepancy $\epsilon_{BFD} - \epsilon_{AE}$, namely the basis set error between V5Z and CBS limit for the values that are contributing to our spectral objective functions. Since the basis error for the gap discrepancies is sub mHa due to very extensive uncontracted basis we use, we conclude that there is only a very minor bias coming from the finite basis set to our objective functions. Clearly for most properties the differences converge faster than the

total energies. Since we are matching energy differences, the impact of CBS extrapolation is marginal as can be shown in Tables I and II. An important thing to note is that the basis set error for a ϵ_{gap} is on the order of 1-2 mHa for most states presented in these tables. However, if we look at the discrepancy from the AE calculation within a given basis for an ECP (i.e., the errors provided in the text) we find that the discrepancies are much larger than the error coming from the incompleteness of the basis. Note also that in our experience the other criteria that we include in the construction, such as systematic agreement for binding curves for a range of geometries, immediately amplify the discrepancies by about an order of magnitude and therefore then dominate the ECP refinement process.

Table A.1 The uncontracted aug-cc-pCV5Z basis set errors (a.u.) corresponding to nitrogen’s CCSD(T) excitation energies calculated in the main text.

Excitation	ϵ_{AE}	ϵ_{BFD}	ϵ_{gap}
$2s^1(^2\text{S})$	-0.0027(14)	-0.0019(2)	0.00070(145)
$2s^2(^1\text{S})$	-0.0023(10)	-0.0018(2)	0.00050(102)
$2s^12p^1(^3\text{P})$	-0.0018(11)	-0.0015(2)	0.00032(108)
$2s^22p^1(^2\text{P})$	-0.0014(10)	-0.0012(1)	0.00024(108)
$2s^12p^2(^4\text{P})$	-0.0015(11)	-0.0014(2)	0.00009(113)
$2s^22p^2(^1\text{D})$	-0.0000(12)	0.0001(1)	0.00016(121)
$2s^22p^2(^3\text{P})$	-0.0005(12)	-0.0004(2)	0.00008(122)
$2s^12p^3(^5\text{S})$	-0.0010(13)	-0.0011(3)	-0.00005(132)
$2s^22p^3(^2\text{D})$	0.0004(12)	0.0005(2)	0.00006(123)
$2s^12p^33s^1(^6\text{S})$	0.0012(11)	0.0010(13)	-0.00016(169)

A.2 Correlation Consistent Basis Sets for the ECPs

In this section we present correlation consistent basis sets for each of our ECPs. We construct VnZ basis sets for $n = \{\text{D,T,Q,5}\}$ from existing basis sets for ECPs of the same form, namely from the Burkatzki, Filippi, and Dolg basis sets [Bur07]. Unlike many all-electron basis sets, the contracted basis functions within a given angular momentum channel remain unchanged for each n . Additionally, for each ECP there is only one contracted basis function

Table A.2 The uncontracted aug-cc-pCV5Z basis set errors (a.u.) corresponding to sulfur’s CCSD(T) excitation energies calculated in the main text.

Excitation	\mathcal{E}_{AE}	\mathcal{E}_{BFD}	$\mathcal{E}_{\text{gap discrepancy}}$
$2s^1(2S)$	-0.0043(21)	-0.0029(2)	0.00141(214)
$2s^2(1S)$	-0.0036(22)	-0.0027(3)	0.00086(217)
$2s^12p^1(3P)$	-0.0035(22)	-0.0027(2)	0.00074(217)
$2s^22p^1(2P)$	-0.0027(22)	-0.0024(2)	0.00037(221)
$2s^12p^2(4P)$	-0.0030(22)	-0.0027(3)	0.00029(219)
$2s^22p^2(1D)$	-0.0019(22)	-0.0014(2)	0.00045(224)
$2s^22p^2(3P)$	-0.0021(22)	-0.0019(2)	0.00017(223)
$2s^12p^3(5S)$	-0.0026(22)	-0.0025(3)	0.00011(220)
$2s^22p^3(2D)$	-0.0010(22)	-0.0009(2)	0.00016(223)
$2s^22p^3(4S)$	-0.0014(22)	-0.0014(2)	0.00003(225)
$2s^12p^33d^1(6D)$	-0.0019(21)	-0.0017(3)	0.00015(214)
$2s^22p^4(1D)$	0.0002(22)	0.0003(3)	0.00010(217)
$2s^22p^33d^1(5D)$	0.0031(31)	0.0031(22)	0.00006(379)
$2s^12p^33d^2(7G)$	0.0014(23)	0.0015(14)	0.00012(273)
$2s^22p^5(2P)$	0.0012(21)	0.0012(4)	-0.00002(210)

per angular momentum, otherwise they are all primitive functions. We have carried out a restricted reoptimization of the contracted basis function exponents for each angular momentum. The new contractions with updated exponents are based on Hartree-Fock for a given ground state for a particular atom. The primitive functions for $n = \{D, T, Q, 5\}$ are the same as in the original basis for BFD.

The optimized basis sets are presented in Tables A.3, A.4, A.5, A.6, and A.7. Since each basis set shares the same contraction, we label V0Z as the shared contracted function, and each column of the table refers to the additional functions needed to construct the basis. Note that within a given column, the first number refers to the exponent while the second is the contraction coefficient. In the VnZ columns, rows separated by a blank space indicates a separate primitive basis function.

Table A.3 Correlation Consistent basis sets for the Boron ECP

Type	V0Z	VDZ = V0Z +	VTZ = V0Z +	VQZ = V0Z +	V5Z = V0Z +
s	11.76050 -0.0036757 5.150520 0.0250517 2.578276 -0.1249228 1.290648 -0.0662874 0.646080 0.1007341 0.323418 0.3375492 0.161898 0.4308431 0.081044 0.2486558 0.040569 0.0317295	0.082513 1.000000	0.626026 1.000000 0.092094 1.000000	0.422217 1.000000 0.305133 1.000000 0.082968 1.000000	0.614105 1.000000 0.375720 1.000000 0.170896 1.000000 0.070664 1.000000
p	7.470701 0.0047397 3.735743 0.0376009 1.868068 0.0510600 0.934132 0.1456587 0.467115 0.2237933 0.233582 0.3199467 0.116803 0.2850185 0.058408 0.1448808 0.029207 0.0176962	0.086803 1.000000	0.235016 1.000000 0.082056 1.000000	0.447031 1.000000 0.196614 1.000000 0.066445 1.000000	0.566611 1.000000 0.436327 1.000000 0.143772 1.000000 0.057917 1.000000
d		0.349879 1.000000	0.699153 1.000000 0.207316 1.000000	1.142614 1.000000 0.410733 1.000000 0.149100 1.000000	1.022256 1.000000 0.808233 1.000000 0.380163 1.000000 0.134838 1.000000
f			0.478872 1.000000	0.870011 1.000000 0.315902 1.000000	1.002171 1.000000 0.799174 1.000000 0.272717 1.000000
g				0.710746 1.000000	0.824366 1.000000 0.486131 1.000000
h					0.632779 1.000000

Table A.4 Correlation Consistent basis sets for the Carbon ECP

Type	V0Z	VDZ = V0Z +	VTZ = V0Z +	VQZ = V0Z +	V5Z = V0Z +
s	13.073594 0.0051583 6.541187 0.0603424 4.573411 -0.1978471 1.637494 -0.0810340 0.819297 0.2321726 0.409924 0.2914643 0.231300 0.4336405 0.102619 0.2131940 0.051344 0.0049848	0.127852 1.000000	0.921552 1.000000 0.132800 1.000000	0.846879 1.000000 0.269659 1.000000 0.109576 1.000000	1.009914 1.000000 0.744448 1.000000 0.232034 1.000000 0.098302 1.000000
p	9.934169 0.0209076 3.886955 0.0572698 1.871016 0.1122682 0.935757 0.2130082 0.468003 0.2835815 0.239473 0.3011207 0.117063 0.2016934 0.058547 0.0453575 0.029281 0.0029775	0.149161 1.000000	0.376742 1.000000 0.126772 1.000000	0.313254 1.000000 0.804681 1.000000 0.105389 1.000000	1.006252 1.000000 0.576869 1.000000 0.216618 1.000000 0.084047 1.000000
d		0.561160 1.000000	1.141611 1.000000 0.329486 1.000000	2.013760 1.000000 0.684884 1.000000 0.240171 1.000000	1.504882 1.000000 1.001526 1.000000 0.606933 1.000000 0.206619 1.000000
f			0.773485 1.000000	1.324930 1.000000 0.457302 1.000000	1.501091 1.000000 1.099564 1.000000 0.400573 1.000000
g				1.034180 1.000000	1.401343 1.000000 0.797648 1.000000
h					1.001703 1.000000

Table A.5 Correlation Consistent basis sets for the Nitrogen ECP

Type	V0Z	VDZ = V0Z +	VTZ = V0Z +	VQZ = V0Z +	V5Z = V0Z +
s	42.693822 -0.0009357 19.963207 0.0063295 9.3345971 0.0105038 4.9278187 -0.1653735 2.040920 -0.0005352 0.967080 0.2452063 0.4761312 0.4582128 0.211443 0.3641224 0.098869 0.0620406	0.175123 1.000000	1.202183 1.000000 0.163243 1.000000	1.625001 1.000000 0.310826 1.000000 0.135764 1.000000	1.532221 1.000000 0.702011 1.000000 0.286632 1.000000 0.115320 1.000000
p	18.925871 0.0073505 9.225603 0.0292844 4.581431 0.0652168 2.300164 0.1405153 1.154825 0.2328188 0.582039 0.2989556 0.290535 0.2802507 0.145867 0.1527995 0.073234 0.0355475	0.223042 1.000000	0.517547 1.000000 0.170104 1.000000	1.020750 1.000000 0.413103 1.000000 0.140736 1.000000	1.272759 1.000000 0.978538 1.000000 0.322697 1.000000 0.120601 1.000000
d		0.832058 1.000000	1.712416 1.000000 0.483567 1.000000	3.028459 1.000000 1.009895 1.000000 0.346233 1.000000	2.798122 1.000000 1.542532 1.000000 0.891436 1.000000 0.305579 1.000000
f			1.093097 1.000000	2.024747 1.000000 0.691129 1.000000	2.443045 1.000000 1.592967 1.000000 0.587676 1.000000
g				1.357512 1.000000	2.842018 1.000000 1.038637 1.000000
h					2.272542 1.000000

Table A.6 Correlation consistent basis set for the Oxygen ECP.

Type	V0Z	VDZ = V0Z +	VTZ = V0Z +	VQZ = V0Z +	V5Z = V0Z +
s	54.775216 -0.0012444 25.616801 0.0107330 11.980245 0.0018889 6.992317 -0.1742537 2.620277 0.0017622 1.225429 0.3161846 0.577797 0.4512023 0.268022 0.3121534 0.125346 0.0511167	0.258551 1.000000	1.686633 1.000000 0.237997 1.000000	1.351771 1.000000 0.843157 1.000000 0.224380 1.000000	1.937532 1.000000 0.935157 1.000000 0.384526 1.000000 0.160664 1.000000
p	22.217266 0.0104866 10.74755 0.0366435 5.315785 0.0803674 2.660761 0.1627010 1.331816 0.2377791 0.678626 0.2811422 0.333673 0.2643189 0.167017 0.1466014 0.083598 0.0458145	0.267865 1.000000	0.600621 1.000000 0.184696 1.000000	1.106737 1.000000 0.452364 1.000000 0.148562 1.000000	1.589967 1.000000 1.178227 1.000000 0.372674 1.000000 0.130580 1.000000
d		1.232753 1.000000	2.404278 1.000000 0.669340 1.000000	4.008867 1.000000 1.344331 1.000000 0.455711 1.000000	4.292433 1.000000 2.823972 1.000000 1.174596 1.000000 0.401152 1.000000
f			1.423104 1.000000	2.763115 1.000000 0.876289 1.000000	3.223721 1.000000 2.006788 1.000000 0.708666 1.000000
g				1.759081 1.000000	3.584495 1.000000 1.207657 1.000000
h					2.615818 1.000000

Table A.7 Correlation Consistent basis sets for the Sulfur ECP

Type	V0Z	VDZ = V0Z +	VTZ = V0Z +	VQZ = V0Z +	V5Z = V0Z +
s	39.950656 0.0010418 18.776623 -0.0061684 8.824926 0.0245015 4.147674 0.0104170 2.066160 -0.3274411 0.916203 0.0194043 0.441791 0.5308075 0.202385 0.5085514 0.095120 0.1360073	0.098454 1.000000	0.792025 1.000000 0.146642 1.000000	0.651905 1.000000 0.315587 1.000000 0.123759 1.000000	1.338226 1.000000 0.626702 1.000000 0.282790 1.000000 0.113918 1.000000
p	16.477979 0.0001288 8.116307 -0.0006332 3.997726 -0.0043706 1.969099 -0.0435541 0.969889 0.1292081 0.477723 0.3293373 0.235305 0.3722786 0.115901 0.2560243 0.057087 0.0810372	0.128926 1.000000	0.247967 1.000000 0.088694 1.000000	0.301333 1.000000 0.202707 1.000000 0.078717 1.000000	0.924454 1.000000 0.522537 1.000000 0.210632 1.000000 0.079101 1.000000
d		0.514135 1.000000	0.950659 1.000000 0.292889 1.000000	1.588204 1.000000 0.560638 1.000000 0.215701 1.000000	2.334308 1.000000 0.955579 1.000000 0.462328 1.000000 0.186546 1.000000
f			0.573218 1.000000	0.961826 1.000000 0.356554 1.000000	1.389533 1.000000 0.661568 1.000000 0.274343 1.000000
g				0.694803 1.000000	1.166495 1.000000 0.486698 1.000000
h					0.839494 1.000000

Supplementary Material: A new generation of effective core potentials from correlated calculations: 2nd row elements

B.1 Atomic Data

In Tables I-XVI, we show total energy components from various states of each of our ECPs from CCSD(T) calculations. All atomic calculations utilized the uncontracted aug-cc-pCVnZ basis sets where $n = \{T, Q, 5\}$. Each energy component, Hartree-Fock (SCF) and correlation, is extrapolated to the complete basis set (CBS) limit with the parameterized functions

$$E^{\text{SCF}}(n) = E_{\text{CBS}}^{\text{SCF}} + A \exp(-Bn) \quad (\text{B.1})$$

and

$$E^{\text{Correlation}}(n) = E_{\text{CBS}}^{\text{Correlation}} + \frac{C}{(n+3/8)^3} + \frac{D}{(n+3/8)^5}, \quad (\text{B.2})$$

respectively. The variables A , B , C , D , $E_{\text{CBS}}^{\text{SCF}}$ and $E_{\text{CBS}}^{\text{Correlation}}$ are fitting parameters.

Table B.1 Total energy components of various states of the Na atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] 3s ¹	² S	-0.186167	-0.186202	-0.186202	-0.186203
[Ne] 3s ²	¹ S	-0.182241	-0.182308	-0.182308	-0.182313
Correlation		T	Q	5	CBS
[Ne] 3s ¹	² S				
[Ne] 3s ²	¹ S	-0.024246	-0.024271	-0.02428	-0.02429

Table B.2 Total energy components of various states of the Na atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] 3s ¹	² S	-47.355004	-47.356999	-47.357015	-47.357015
[Ne]	¹ S	-47.172935	-47.174928	-47.174944	-47.174944
[Ne] 3s ²	¹ S	-47.351159	-47.353181	-47.353197	-47.353197
Correlation		T	Q	5	CBS
[Ne] 3s ¹	² S	-0.278836	-0.304079	-0.313965	-0.325551
[Ne]	¹ S	-0.273772	-0.298106	-0.307423	-0.318131
[Ne] 3s ²	¹ S	-0.302615	-0.327913	-0.337849	-0.349521

Table B.3 Total energy components of various states of the Mg atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] 3s ²	¹ S	-0.788370	-0.788345	-0.788370	-0.788370
[Ne] 3s ¹	² S	-0.544766	-0.544744	-0.544763	-0.544763
Correlation		T	Q	5	CBS
[Ne] 3s ²	¹ S	-0.034907	-0.034994	-0.035031	-0.035077
[Ne] 3s ¹	² S				

Table B.4 Total energy components of various states of the Mg atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] 3s ²	¹ S	-62.927052	-62.927327	-62.927412	-62.927451
[Ne]	¹ S	-62.142427	-62.142705	-62.142790	-62.142828
[Ne] 3s ¹	² S	-62.684058	-62.684333	-62.684414	-62.684447
Correlation		T	Q	5	CBS
[Ne] 3s ²	¹ S	-0.315920	-0.342027	-0.352077	-0.363681
[Ne]	¹ S	-0.270494	-0.294814	-0.304048	-0.314584
[Ne] 3s ¹	² S	-0.279223	-0.304733	-0.314498	-0.325718

Table B.5 Total energy components of various states of the Al atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] 3s ² 3p ¹	² P	-1.876699	-1.876990	-1.876973	-1.876999
[Ne] 3s ¹	² S	-1.030941	-1.030946	-1.030933	-1.030949
[Ne] 3s ²	¹ S	-1.674980	-1.675006	-1.674996	-1.675010
[Ne] 3s ² 3p ²	³ P	-1.878405	-1.878616	-1.878601	-1.878623
Correlation		T	Q	5	CBS
[Ne] 3s ² 3p ¹	² P	-0.058680	-0.059582	-0.059783	-0.059869
[Ne] 3s ¹	² S				
[Ne] 3s ²	¹ S	-0.041715	-0.042057	-0.042137	-0.042175
[Ne] 3s ² 3p ²	³ P	-0.072470	-0.073840	-0.074155	-0.074305

Table B.6 Total energy components of various states of the Al atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] $3s^2 3p^1$	2P	-80.990584	-80.992910	-80.993046	-80.993054
[Ne]	1S	-79.112214	-79.114296	-79.114427	-79.114435
[Ne] $3s^1$	2S	-80.144308	-80.146349	-80.146470	-80.146478
[Ne] $3s^2$	1S	-80.789130	-80.791175	-80.791301	-80.791309
[Ne] $3s^2 3p^2$	3P	-80.992087	-80.994343	-80.994480	-80.994489
Correlation		T	Q	5	CBS
[Ne] $3s^2 3p^1$	2P	-0.305640	-0.355044	-0.373996	-0.395813
[Ne]	1S	-0.231923	-0.278793	-0.297044	-0.318327
[Ne] $3s^1$	2S	-0.243946	-0.291640	-0.310130	-0.331614
[Ne] $3s^2$	1S	-0.289386	-0.338035	-0.356793	-0.378485
[Ne] $3s^2 3p^2$	3P	-0.319051	-0.368907	-0.387970	-0.409853

Table B.7 Total energy components of various states of the Si atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] $3s^2 3p^2$	3P	-3.671951	-3.672403	-3.672394	-3.672423
[Ne] $3s^1$	2S	-1.639761	-1.639759	-1.639758	-1.639762
[Ne] $3s^2$	1S	-2.813851	-2.813866	-2.813855	-2.813870
[Ne] $3s^2 3p^1$	2P	-3.390158	-3.390719	-3.390718	-3.390749
[Ne] $3s^2 3p^3$	4S	-3.707923	-3.708178	-3.708158	-3.708187
Correlation		T	Q	5	CBS
[Ne] $3s^2 3p^2$	3P	-0.085857	-0.087646	-0.088188	-0.088666
[Ne] $3s^1$	2S				
[Ne] $3s^2$	1S	-0.047828	-0.048152	-0.048263	-0.048376
[Ne] $3s^2 3p^1$	2P	-0.068862	-0.069816	-0.070130	-0.070437
[Ne] $3s^2 3p^3$	4S	-0.100941	-0.103501	-0.104256	-0.104894

Table B.8 Total energy components of various states of the Si atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] $3s^23p^2$	3P	-101.622367	-101.625367	-101.625788	-101.625857
[Ne]	1S	-97.946272	-97.948793	-97.949212	-97.949296
[Ne] $3s^1$	2S	-99.589327	-99.591835	-99.592241	-99.592319
[Ne] $3s^2$	1S	-100.765362	-100.767880	-100.768268	-100.768339
[Ne] $3s^23p^1$	2P	-101.341340	-101.344430	-101.344845	-101.344910
[Ne] $3s^23p^3$	4S	-101.657754	-101.660560	-101.660972	-101.661043
Correlation		T	Q	5	CBS
[Ne] $3s^23p^2$	3P	-0.336020	-0.386298	-0.405726	-0.428232
[Ne]	1S	-0.232488	-0.278592	-0.296671	-0.317884
[Ne] $3s^1$	2S	-0.246107	-0.293177	-0.311571	-0.333091
[Ne] $3s^2$	1S	-0.298506	-0.346684	-0.365457	-0.387362
[Ne] $3s^23p^1$	2P	-0.319161	-0.368422	-0.387552	-0.409811
[Ne] $3s^23p^3$	4S	-0.351145	-0.402192	-0.421832	-0.444501

Table B.9 Total energy components of various states of the P atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] $3s^23p^3$	4S	-6.340560	-6.340906	-6.340875	-6.340917
[Ne] $3s^1$	2S	-2.362967	-2.362961	-2.362981	-2.363013
[Ne] $3s^2$	1S	-4.186007	-4.185996	-4.186002	-4.186015
[Ne] $3s^23p^1$	2P	-5.265263	-5.265481	-5.265501	-5.265510
[Ne] $3s^23p^2$	3P	-5.970169	-5.970641	-5.970646	-5.970670
[Ne] $3s^23p^4$	3P	-6.320061	-6.320225	-6.320176	-6.320236
Correlation		T	Q	5	CBS
[Ne] $3s^23p^3$	4S	-0.112732	-0.115948	-0.116922	-0.117781
[Ne] $3s^1$	2S				
[Ne] $3s^2$	1S	-0.053463	-0.053785	-0.053895	-0.054008
[Ne] $3s^23p^1$	2P	-0.077685	-0.078756	-0.079108	-0.079454
[Ne] $3s^23p^2$	3P	-0.097219	-0.099311	-0.099955	-0.100534
[Ne] $3s^23p^4$	3P	-0.155101	-0.161173	-0.162879	-0.164211

Table B.10 Total energy components of various states of the P atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] $3s^23p^3$	4S	-125.253282	-125.258161	-125.258394	-125.258405
[Ne]	1S	-118.899156	-118.903679	-118.903877	-118.903886
[Ne] $3s^1$	2S	-121.271178	-121.275661	-121.275885	-121.275897
[Ne] $3s^2$	1S	-123.100459	-123.104903	-123.105109	-123.105119
[Ne] $3s^23p^1$	2P	-124.180512	-124.185217	-124.185449	-124.185461
[Ne] $3s^23p^2$	3P	-124.884427	-124.889414	-124.889658	-124.889670
[Ne] $3s^23p^4$	3P	-125.233338	-125.238037	-125.238254	-125.238264
Correlation		T	Q	5	CBS
[Ne] $3s^23p^3$	4S	-0.365923	-0.418003	-0.438225	-0.461752
[Ne]	1S	-0.233066	-0.278627	-0.296756	-0.318285
[Ne] $3s^1$	2S	-0.247871	-0.294566	-0.313051	-0.334912
[Ne] $3s^2$	1S	-0.306293	-0.354270	-0.373190	-0.395491
[Ne] $3s^23p^1$	2P	-0.330315	-0.379711	-0.399104	-0.421880
[Ne] $3s^23p^2$	3P	-0.349999	-0.400799	-0.420623	-0.443787
[Ne] $3s^23p^4$	3P	-0.408274	-0.463170	-0.484121	-0.508132

Table B.11 Total energy components of various states of the S atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] $3s^23p^4$	3P	-9.917772	-9.918088	-9.918140	-9.918156
[Ne] $3s^1$	2S	-3.199330	-3.199320	-3.199373	-3.199413
[Ne] $3s^2$	1S	-5.787262	-5.787214	-5.787263	-5.787312
[Ne] $3s^23p^1$	2P	-7.490378	-7.490458	-7.490604	-7.490623
[Ne] $3s^23p^2$	3P	-8.745383	-8.745651	-8.745767	-8.745853
[Ne] $3s^23p^3$	4S	-9.587305	-9.587807	-9.587895	-9.587914
[Ne] $3s^23p^5$	2P	-9.951496	-9.951654	-9.951678	-9.951687
Correlation		T	Q	5	CBS
[Ne] $3s^23p^4$	3P	-0.165776	-0.173557	-0.175915	-0.177994
[Ne] $3s^1$	2S				
[Ne] $3s^2$	1S	-0.058888	-0.059247	-0.059376	-0.059518
[Ne] $3s^23p^1$	2P	-0.085847	-0.087076	-0.087513	-0.087980
[Ne] $3s^23p^2$	3P	-0.107509	-0.109937	-0.110748	-0.111559
[Ne] $3s^23p^3$	4S	-0.124018	-0.127843	-0.129087	-0.130294
[Ne] $3s^23p^5$	2P	-0.203222	-0.214387	-0.217619	-0.220276

Table B.12 Total energy components of various states of the S atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] $3s^23p^4$	3P	-151.909927	-151.916064	-151.916369	-151.916385
[Ne]	1S	-141.960834	-141.966569	-141.966790	-141.966799
[Ne] $3s^1$	2S	-145.177455	-145.183135	-145.183423	-145.183438
[Ne] $3s^2$	1S	-147.778004	-147.783595	-147.783878	-147.783893
[Ne] $3s^23p^1$	2P	-149.483717	-149.489469	-149.489861	-149.489890
[Ne] $3s^23p^2$	3P	-150.738321	-150.744306	-150.744672	-150.744696
[Ne] $3s^23p^3$	4S	-151.578352	-151.584628	-151.584970	-151.584989
[Ne] $3s^23p^5$	2P	-151.943481	-151.949466	-151.949744	-151.949757
Correlation		T	Q	5	CBS
[Ne] $3s^23p^4$	3P	-0.420401	-0.477685	-0.499872	-0.525628
[Ne]	1S	-0.233065	-0.278340	-0.296665	-0.318728
[Ne] $3s^1$	2S	-0.248757	-0.295307	-0.314039	-0.336489
[Ne] $3s^2$	1S	-0.312562	-0.360586	-0.379826	-0.402804
[Ne] $3s^23p^1$	2P	-0.339291	-0.389051	-0.408907	-0.432544
[Ne] $3s^23p^2$	3P	-0.361203	-0.412731	-0.433178	-0.457406
[Ne] $3s^23p^3$	4S	-0.378375	-0.431623	-0.452638	-0.477427
[Ne] $3s^23p^5$	2P	-0.458354	-0.518979	-0.542040	-0.568391

Table B.13 Total energy components of various states of the Cl atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] $3s^23p^5$	2P	-14.688897	-14.689151	-14.689385	-14.689386
[Ne] $3s^1$	2S	-4.153414	-4.153407	-4.153487	-4.153506
[Ne] $3s^2$	1S	-7.623920	-7.623836	-7.623936	-7.623950
[Ne] $3s^23p^1$	2P	-10.068949	-10.069158	-10.069416	-10.069416
[Ne] $3s^23p^2$	3P	-11.995324	-11.995572	-11.995854	-11.995856
[Ne] $3s^23p^3$	4S	-13.435190	-13.435587	-13.435816	-13.435820
[Ne] $3s^23p^4$	3P	-14.255149	-14.255690	-14.255918	-14.256085
[Ne] $3s^23p^6$	1S	-14.784864	-14.784943	-14.785142	-14.785146
Correlation		T	Q	5	CBS
[Ne] $3s^23p^5$	2P	-0.214761	-0.228169	-0.232225	-0.235794
[Ne] $3s^1$	2S				
[Ne] $3s^2$	1S	-0.064175	-0.064536	-0.064686	-0.064870
[Ne] $3s^23p^1$	2P	-0.093905	-0.095211	-0.095684	-0.096199
[Ne] $3s^23p^2$	3P	-0.117691	-0.120313	-0.121206	-0.122119
[Ne] $3s^23p^3$	4S	-0.135384	-0.139587	-0.140982	-0.142366
[Ne] $3s^23p^4$	3P	-0.179033	-0.187874	-0.190649	-0.193217
[Ne] $3s^23p^6$	1S	-0.247496	-0.264870	-0.269980	-0.274288

Table B.14 Total energy components of various states of the Cl atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] $3s^23p^5$	2P	-181.606009	-181.611999	-181.612377	-181.612402
[Ne]	1S	-166.876313	-166.881725	-166.882022	-166.882039
[Ne] $3s^1$	2S	-171.052574	-171.057971	-171.058355	-171.058385
[Ne] $3s^2$	1S	-174.540922	-174.546208	-174.546602	-174.546634
[Ne] $3s^23p^1$	2P	-176.989287	-176.994921	-176.995449	-176.995504
[Ne] $3s^23p^2$	3P	-178.914799	-178.920519	-178.921056	-178.921111
[Ne] $3s^23p^3$	4S	-180.351450	-180.357396	-180.357852	-180.357890
[Ne] $3s^23p^4$	3P	-181.172702	-181.178912	-181.179310	-181.179337
[Ne] $3s^23p^6$	1S	-181.700830	-181.706653	-181.706998	-181.707020
Correlation		T	Q	5	CBS
[Ne] $3s^23p^5$	2P	-0.472345	-0.535132	-0.559372	-0.587434
[Ne]	1S	-0.234937	-0.279240	-0.297555	-0.319976
[Ne] $3s^1$	2S	-0.251472	-0.297158	-0.315912	-0.338747
[Ne] $3s^2$	1S	-0.320306	-0.367597	-0.386907	-0.410321
[Ne] $3s^23p^1$	2P	-0.349504	-0.398752	-0.418757	-0.442915
[Ne] $3s^23p^2$	3P	-0.373428	-0.424710	-0.445416	-0.470303
[Ne] $3s^23p^3$	4S	-0.391882	-0.445213	-0.466621	-0.492226
[Ne] $3s^23p^4$	3P	-0.435816	-0.493986	-0.516894	-0.543868
[Ne] $3s^23p^6$	1S	-0.506041	-0.572752	-0.598045	-0.626863

Table B.15 Total energy components of various states of the Ar atom using our ccECP[Ne]

SCF		T	Q	5	CBS
[Ne] $3s^23p^6$	1S	-20.778934	-20.779555	-20.779573	-20.779601
[Ne] $3s^1$	2S	-5.207475	-5.207483	-5.207610	-5.207618
[Ne] $3s^2$	1S	-9.665000	-9.664896	-9.665081	-9.665103
[Ne] $3s^23p^1$	2P	-12.964117	-12.964630	-12.964915	-12.964920
[Ne] $3s^23p^2$	3P	-15.676104	-15.676707	-15.677049	-15.677055
[Ne] $3s^23p^3$	4S	-17.831518	-17.832105	-17.832395	-17.832680
[Ne] $3s^23p^4$	3P	-19.266038	-19.266814	-19.266949	-19.266979
[Ne] $3s^23p^5$	2P	-20.234894	-20.235814	-20.235893	-20.235923
Correlation		T	Q	5	CBS
[Ne] $3s^23p^6$	1S	-0.259079	-0.279406	-0.285173	-0.289756
[Ne] $3s^1$	2S				
[Ne] $3s^2$	1S	-0.069503	-0.069933	-0.070038	-0.070099
[Ne] $3s^23p^1$	2P	-0.101947	-0.103367	-0.103790	-0.104156
[Ne] $3s^23p^2$	3P	-0.127850	-0.130675	-0.131528	-0.132275
[Ne] $3s^23p^3$	4S	-0.146823	-0.151385	-0.152764	-0.153973
[Ne] $3s^23p^4$	3P	-0.192939	-0.202715	-0.205588	-0.208007
[Ne] $3s^23p^5$	2P	-0.229486	-0.244616	-0.248974	-0.252528

Table B.16 Total energy components of various states of the Ar atom using our ccECP[He]

SCF		T	Q	5	CBS
[Ne] $3s^23p^6$	1S	-214.882779	-214.891228	-214.891473	-214.891481
[Ne]	1S	-194.015736	-194.023303	-194.023467	-194.023471
[Ne] $3s^1$	2S	-199.262254	-199.269806	-199.270110	-199.270123
[Ne] $3s^2$	1S	-203.750624	-203.758010	-203.758366	-203.758384
[Ne] $3s^23p^1$	2P	-207.059126	-207.067196	-207.067647	-207.067674
[Ne] $3s^23p^2$	3P	-209.774739	-209.782966	-209.783460	-209.783491
[Ne] $3s^23p^3$	4S	-211.929923	-211.938199	-211.938633	-211.938657
[Ne] $3s^23p^4$	3P	-213.368960	-213.377490	-213.377782	-213.377792
[Ne] $3s^23p^5$	2P	-214.339213	-214.347933	-214.348210	-214.348219
Correlation		T	Q	5	CBS
[Ne] $3s^23p^6$	1S	-0.516297	-0.586767	-0.613131	-0.642806
[Ne]	1S	-0.234846	-0.279060	-0.297392	-0.319883
[Ne] $3s^1$	2S	-0.251903	-0.297609	-0.316402	-0.339315
[Ne] $3s^2$	1S	-0.325612	-0.373081	-0.392439	-0.415887
[Ne] $3s^23p^1$	2P	-0.356990	-0.406593	-0.426698	-0.450934
[Ne] $3s^23p^2$	3P	-0.382642	-0.434514	-0.455381	-0.480384
[Ne] $3s^23p^3$	4S	-0.402186	-0.456407	-0.478050	-0.503818
[Ne] $3s^23p^4$	3P	-0.448215	-0.507975	-0.531290	-0.558527
[Ne] $3s^23p^5$	2P	-0.485495	-0.550750	-0.575657	-0.604203

B.2 Molecular Data

Table B.17 All-electron (AE) UCCSD(T) Na₂ ground state (¹Σ_g) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	0.737(13) ^{† a}	0.730(6)	-0.012(9)	-0.016(9)	-0.011(9)	-0.010(9)	-0.013(9)	0.015(9)	-0.002(10)
ω_e (cm ⁻¹)	139.9 ^a	146(3)	-10(5)	-11(5)	-10(5)	-10(5)	-10(5)	-1(5)	-1(5)
r_e (Å)	3.079 ^b	3.069(18)	0.116(28)	0.103(28)	0.099(28)	0.094(28)	0.098(28)	-0.007(26)	0.005(29)
Potential Energy Surface									
U(2.20)		-0.0369	0.1300	0.0868	0.0887	0.0806	0.0843	-0.0460	0.00494
U(2.45)		-0.4296	0.0971	0.0773	0.0742	0.0686	0.0730	-0.0288	0.00487
U(2.70)		-0.6443	0.0628	0.0557	0.0500	0.0464	0.0510	-0.0201	0.00442
U(2.95)		-0.7324	0.0315	0.0322	0.0260	0.0241	0.0280	-0.0152	0.00528
U(3.20)		-0.7350	0.0056	0.0104	0.0047	0.0043	0.0070	-0.0133	-0.00323
U(3.45)		-0.6841	-0.0141	-0.0081	-0.0126	-0.0125	-0.0105	-0.0134	-0.00253

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [HH79]

Table B.18 All-electron (AE) UCCSD(T) NaO ground state (${}^2\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	2.67 ^{† a}	2.52(2)	0.05(3)	0.06(2)	0.02(3)	-0.00(2)	0.01(2)	-0.02(3)	-0.01(3)
ω_e (cm ⁻¹)	526 ^a	538(6)	-13(10)	-50(6)	-19(10)	-63(6)	-37(7)	-12(8)	0(8)
r_e (Å)	1.95 ^a	1.964(5)	0.044(9)	-0.012(5)	0.033(8)	-0.016(5)	0.007(6)	0.006(7)	0.001(6)
Potential Energy Surface									
U(1.475)		-0.0305	0.4054	-0.7805	0.1843	-0.8672	-0.3741	-0.0486	0.0252
U(1.725)		-2.1531	0.0936	-0.1597	0.0563	-0.1069	-0.0414	0.0124	0.0134
U(1.975)		-2.4941	-0.0390	-0.0732	-0.0169	-0.0127	-0.0181	0.0180	0.0090
U(2.225)		-2.2604	-0.0964	-0.0902	-0.0640	-0.0458	-0.0472	0.0047	0.0068

[†] Sum of experimental dissociation and zero-point energies

^a Reference [HH79]

Table B.19 All-electron (AE) UCCSD(T) MgO ground state (${}^1\Sigma^+$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	2.5(2) ^{† a}	2.616(8)	0.01(1)	0.09(1)	0.14(1)	-0.02(1)	0.04(1)	-0.03(1)	0.00(1)
ω_e (cm ⁻¹)	785.2 ^b	780(6)	-14(8)	-39(8)	-7(8)	-43(9)	-17(8)	-33(8)	-1(9)
r_e (Å)	1.749 ^c	1.738(2)	0.017(3)	0.005(3)	-0.020(3)	0.004(4)	-0.007(3)	0.007(3)	0.000(3)
Potential Energy Surface									
U(1.4)		-0.0876	0.2135	-0.2493	-0.6121	-0.2685	-0.3327	-0.1149	0.0013
U(1.5)		-1.5908	0.1230	-0.0830	-0.3654	-0.0738	-0.1695	-0.0054	0.0013
U(1.6)		-2.3414	0.0581	-0.0189	-0.2259	0.0012	-0.0844	0.0319	0.0016
U(1.7)		-2.6140	0.0113	-0.0037	-0.1484	0.0203	-0.0423	0.0368	0.0019
U(1.8)		-2.5929	-0.0230	-0.0132	-0.1096	0.0124	-0.0278	0.0239	0.0014

[†] Sum of experimental dissociation and zero-point energies

^a Reference [Joh18]

^b Reference [Iri07]

^c Reference [HH79]

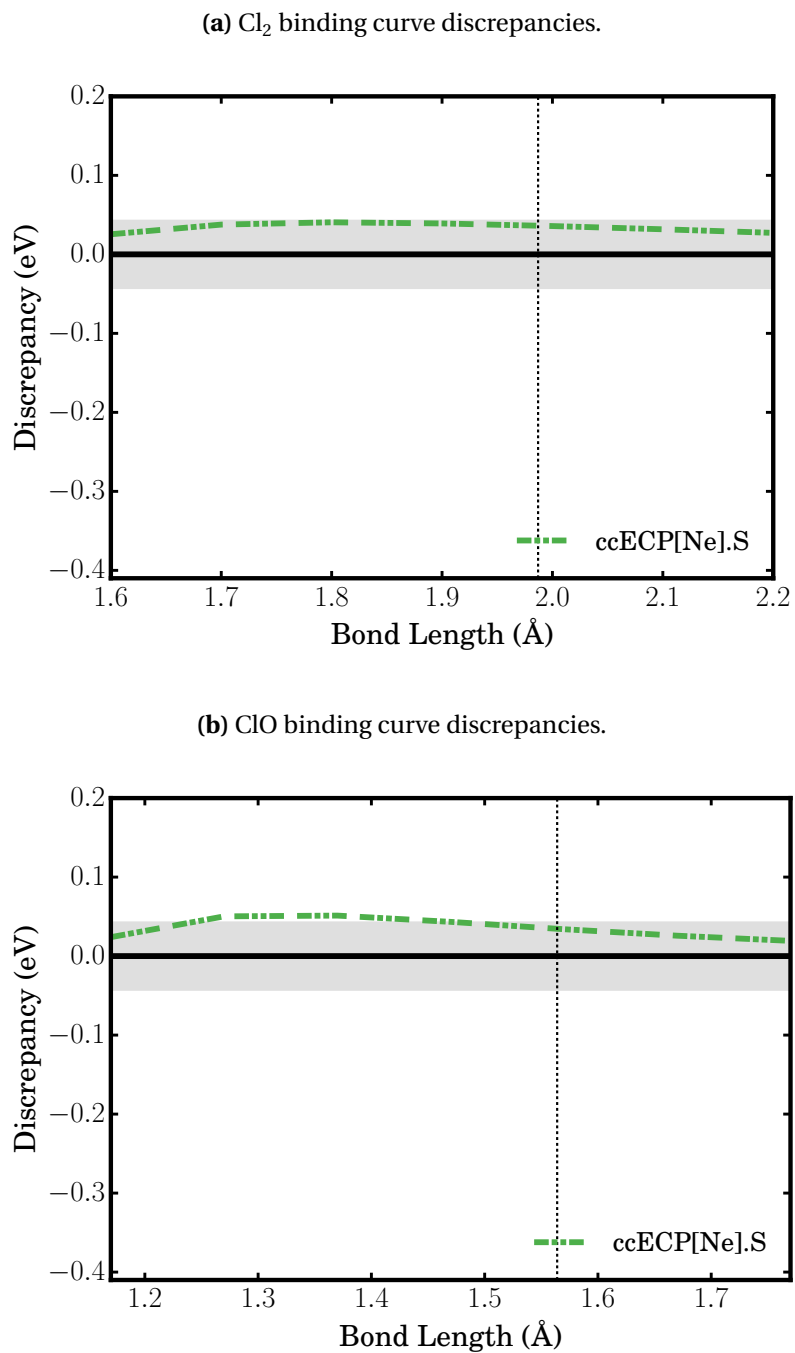


Figure B.1 Binding energy discrepancies for (a) Cl₂ and (b) ClO molecules in their ground states $^1\Sigma_g$ and $^2\Pi$, respectively. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

Table B.20 All-electron (AE) UCCSD(T) Al₂ ground state ($^3\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	1.34(13) ^{† a}	1.423(3)	0.002(5)	0.014(5)	0.035(5)	0.010(5)	0.048(5)	0.009(5)	0.003(5)
ω_e (cm ⁻¹)	285.8 ^b	285(1)	-1(2)	0(2)	2(2)	-1(2)	6(2)	-1(2)	0(2)
r_e (Å)	2.701 ^c	2.692(3)	0.014(4)	0.005(4)	-0.004(4)	0.004(4)	-0.021(4)	-0.005(4)	-0.000(4)
Potential Energy Surface									
U(2.1)		0.0717	0.0817	0.0101	-0.0473	-0.0009	-0.1253	0.0103	-0.0034
U(2.3)		-0.9124	0.0392	0.0002	-0.0409	-0.0007	-0.0900	0.0039	-0.0035
U(2.5)		-1.3265	0.0129	-0.0094	-0.0383	-0.0063	-0.0657	-0.0043	-0.0034
U(2.7)		-1.4265	-0.0023	-0.0154	-0.0359	-0.0113	-0.0477	-0.0102	-0.0033
U(2.9)		-1.3609	-0.0102	-0.0177	-0.0326	-0.0138	-0.0335	-0.0129	-0.0031

[†] Sum of experimental dissociation and zero-point energies

^a Reference [Joh18]

^b Reference [Iri07]

^c Reference [Fu90]

Table B.21 All-electron (AE) UCCSD(T) AlO ground state ($^2\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Potential Energy Surface									
D_e (eV)	5.26(9) ^{† a}	5.28(1)	-0.02(1)	0.01(1)	0.06(1)	-0.24(1)	0.07(1)	0.04(1)	-0.01(1)
ω_e (cm ⁻¹)	979.5 ^b	1028(5)	-4(7)	-36(6)	-26(6)	-90(6)	-12(6)	-36(6)	6(6)
r_e (Å)	1.618 ^c	1.611(2)	0.007(3)	0.004(2)	-0.002(2)	0.021(2)	-0.009(2)	0.004(2)	-0.003(2)
Potential Energy Surface									
U(1.22)		1.3650	0.2753	-0.4581	-0.6482	-0.3602	-0.7520	-0.4704	-0.0202
U(1.32)		-2.3179	0.1699	-0.1595	-0.3070	0.1097	-0.4055	-0.1652	-0.0153
U(1.42)		-4.2349	0.0980	-0.0432	-0.1495	0.2764	-0.2200	-0.0448	-0.0104
U(1.52)		-5.0781	0.0494	-0.0084	-0.0800	0.2913	-0.1202	-0.0064	-0.0064
U(1.62)		-5.2884	0.0179	-0.0056	-0.0508	0.2437	-0.0654	-0.0005	0.0209

[†] Sum of experimental dissociation and zero-point energies

^a Reference [Joh18]

^b Reference [Iri07]

^c Reference [HH79]

Table B.22 All-electron (AE) UCCSD(T) Si2 ground state ($^3\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	3.239 ^{† a}	3.258(4)	-0.012(6)	-0.004(6)	0.015(6)	0.007(6)	0.085(6)	0.006(6)	0.011(6)
ω_e (cm ⁻¹)	489.7 ^a	506(2)	-3(3)	-4(3)	-2(3)	-4(3)	9(3)	-4(3)	0(3)
r_e (Å)	2.246 ^b	2.242(2)	0.009(2)	0.004(3)	-0.004(3)	0.001(3)	-0.023(2)	0.003(2)	0.001(3)
Potential Energy Surface									
U(1.750)		0.1044	0.1444	-0.0001	-0.1319	-0.0554	-0.3778	-0.0134	-0.0011
U(1.875)		-1.7121	0.0935	0.0145	-0.0711	-0.0221	-0.2633	0.0072	-0.0052
U(2.000)		-2.7013	0.0563	0.0139	-0.0401	-0.0102	-0.1822	0.0112	-0.0060
U(2.125)		-3.1540	0.0298	0.0083	-0.0243	-0.0079	-0.1236	0.0087	-0.0076
U(2.250)		-3.2654	0.0113	0.0023	-0.0159	-0.0089	-0.0804	0.0047	-0.0142
U(2.375)		-3.1650	-0.0013	-0.0025	-0.0110	-0.0103	-0.0481	0.0010	-0.0138

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [HH79]

Table B.23 All-electron (AE) UCCSD(T) SiO ground state ($^1\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	8.313(87) ^{† a}	8.33(2)	-0.04(2)	-0.02(2)	0.02(2)	-0.16(2)	0.10(2)	0.01(2)	0.01(2)
ω_e (cm ⁻¹)	1259 ^a	1278(7)	-6(10)	-24(8)	-25(8)	-52(8)	-1(9)	-21(9)	1(10)
r_e (Å)	1.510 ^b	1.509(2)	0.005(3)	0.000(2)	-0.004(2)	0.005(2)	-0.013(2)	-0.001(2)	0.000(2)
Potential Energy Surface									
U(1.15)		-0.1504	0.2614	-0.3946	-0.7175	-0.3682	-0.9623	-0.3928	-0.0212
U(1.25)		-4.9302	0.1693	-0.1252	-0.3053	0.0205	-0.5351	-0.1373	-0.0186
U(1.35)		-7.2762	0.1053	-0.0272	-0.1258	0.1447	-0.2942	-0.0390	-0.0146
U(1.45)		-8.1833	0.0609	0.0012	-0.0498	0.1551	-0.1536	-0.0055	-0.0111
U(1.55)		-8.2557	0.0300	0.0037	-0.0194	0.1264	-0.0695	0.0022	-0.0086
U(1.65)		-7.8612	0.0084	-0.0019	-0.0084	0.0913	-0.0184	0.0005	-0.0070

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [Lov05]

Table B.24 All-electron (AE) UCCSD(T) P2 ground state ($^1\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	5.082(22) ^{† a}	4.96(2)	-0.03(2)	-0.06(2)	-0.03(2)	-0.03(2)	-0.13(2)	-0.05(2)	0.01(2)
ω_e (cm ⁻¹)	769.5 ^a	769(8)	-7(10)	-11(11)	-7(11)	-8(11)	-15(11)	-9(11)	0(11)
r_e (Å)	1.893 ^b	1.887(3)	0.008(4)	0.005(4)	-0.004(4)	0.001(4)	0.002(4)	0.003(4)	0.001(4)
Potential Energy Surface									
U(1.55)		-1.2900	0.1760	0.0708	-0.1612	-0.0236	0.0331	0.0354	-0.0046
U(1.65)		-3.4331	0.1197	0.0846	-0.0555	0.0191	0.1027	0.0571	-0.0093
U(1.75)		-4.5340	0.0767	0.0774	-0.0006	0.0330	0.1281	0.0580	-0.0122
U(1.85)		-4.9504	0.0442	0.0629	0.0263	0.0336	0.1311	0.0505	-0.0122
U(1.95)		-4.9286	0.0197	0.0475	0.0379	0.0288	0.1234	0.0406	-0.0130

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [HH79]

Table B.25 All-electron (AE) UCCSD(T) PO ground state ($^2\Pi$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	6.16(13) ^{† a}	6.14(1)	-0.03(2)	-0.03(2)	-0.01(2)	-0.10(2)	-0.09(2)	-0.03(2)	0.01(2)
ω_e (cm ⁻¹)	1233.3 ^b	1249(7)	-7(10)	-11(11)	-11(11)	-19(11)	-14(11)	-9(11)	0(10)
r_e (Å)	1.476 ^c	1.472(1)	0.004(2)	0.001(2)	-0.004(2)	0.003(2)	-0.002(2)	0.000(2)	0.000(2)
Potential Energy Surface									
U(1.2)		-1.9498	0.1467	-0.0306	-0.2437	0.0845	-0.1000	-0.0560	-0.0233
U(1.3)		-4.8025	0.0901	0.0229	-0.0790	0.1248	0.0246	0.0052	-0.0156
U(1.4)		-5.9524	0.0514	0.0319	-0.0108	0.1154	0.0732	0.0235	-0.0101
U(1.5)		-6.1369	0.0246	0.0270	0.0159	0.0944	0.0882	0.0258	-0.0068
U(1.6)		-5.7942	0.0059	0.0192	0.0248	0.0739	0.0881	0.0225	-0.0050

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [Iri07]

^c Reference [HH79]

B.3 Cl and Ar alternatives

In Table B.31, we share the alternative Ne-core ECP parameters for Cl and Ar which were generated with higher weights on the spectral component of the objective function. Addi-

Table B.26 All-electron (AE) UCCSD(T) S2 ground state ($^3\Sigma_g$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	4.408(3) ^{† a}	4.429(7)	-0.02(1)	0.00(1)	0.01(1)	0.01(1)	0.12(1)	-0.02(1)	0.02(1)
ω_e (cm ⁻¹)	699.5 ^a	717(3)	-4(5)	-3(5)	-3(5)	-3(5)	17(5)	-5(5)	1(5)
r_e (Å)	1.889 ^b	1.886(2)	0.006(2)	0.001(2)	-0.006(2)	0.001(2)	-0.025(2)	0.003(2)	-0.001(2)
Potential Energy Surface									
U(1.5)		0.3808	0.1600	-0.0256	-0.2756	-0.0329	-0.7207	0.0272	-0.0263
U(1.6)		-2.2587	0.1076	-0.0055	-0.1444	-0.0117	-0.4797	0.0354	-0.0258
U(1.7)		-3.6725	0.0692	-0.0013	-0.0705	-0.0081	-0.3102	0.0306	-0.0241
U(1.8)		-4.3034	0.0412	-0.0023	-0.0284	-0.0100	-0.1899	0.0223	-0.0221
U(1.9)		-4.4421	0.0211	-0.0041	-0.0043	-0.0127	-0.1042	0.0143	-0.0199
U(2.0)		-4.2789	0.0066	-0.0056	0.0094	-0.0147	-0.0430	0.0079	-0.0177

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [HH79]

Table B.27 All-electron (AE) UCCSD(T) SO ground state ($^3\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	5.421(13) ^{† a}	5.412(7)	-0.021(9)	-0.01(1)	0.02(1)	-0.06(1)	0.03(1)	-0.02(1)	0.01(1)
ω_e (cm ⁻¹)	1119 ^a	1165(3)	-5(4)	-6(5)	-7(5)	-13(5)	6(5)	-7(5)	0(5)
r_e (Å)	1.481 ^b	1.4772(8)	0.004(1)	0.001(1)	-0.005(1)	0.004(1)	-0.012(1)	0.001(1)	0.000(1)
Potential Energy Surface									
U(1.15)		0.7445	0.1464	-0.0520	-0.4014	0.1084	-0.5759	-0.0397	-0.0339
U(1.25)		-3.0835	0.0913	-0.0052	-0.1843	0.1147	-0.3000	0.0055	-0.0193
U(1.35)		-4.8278	0.0536	0.0053	-0.0830	0.0916	-0.1382	0.0164	-0.0103
U(1.45)		-5.3932	0.0277	0.0047	-0.0352	0.0667	-0.0429	0.0152	-0.0057
U(1.55)		-5.3060	0.0098	0.0014	-0.0128	0.0464	0.0128	0.0101	-0.0039

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [HH79]

tionally, the UCCSD(T) spectral errors and UCCSD(T) molecular errors from all-electron results are also provided. For Cl, the spectral errors are given in Table B.32 while the molec-

Table B.28 All-electron (AE) UCCSD(T) Cl₂ ground state (¹Σ_g) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPS. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	2.5143(4) ^{† a}	2.526(9)	-0.01(1)	-0.01(1)	0.02(1)	-0.01(1)	0.04(1)	-0.01(1)	0.01(1)
ω_e (cm ⁻¹)	559.6 ^a	533(5)	-3(7)	0(7)	1(7)	-2(7)	12(8)	-2(7)	1(7)
r_e (Å)	1.988 ^b	1.987(3)	0.005(5)	-0.001(5)	-0.009(5)	0.002(5)	-0.025(5)	0.002(5)	0.000(5)
Potential Energy Surface									
U(1.6)		0.5063	0.0825	-0.0122	-0.2120	0.0335	-0.4174	-0.0025	-0.0102
U(1.7)		-1.1542	0.0534	-0.0036	-0.1263	0.0279	-0.2635	0.0070	-0.0112
U(1.8)		-2.0502	0.0325	0.0011	-0.0736	0.0215	-0.1556	0.0110	-0.0117
U(1.9)		-2.4538	0.0174	0.0037	-0.0409	0.0162	-0.0796	0.0125	-0.0117
U(2.0)		-2.5435	0.0066	0.0053	-0.0208	0.0120	-0.0263	0.0127	-0.0114
U(2.1)		-2.4387	-0.0011	0.0059	-0.0085	0.0088	0.0102	0.0121	-0.0109

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [HH79]

Table B.29 All-electron (AE) UCCSD(T) ClO ground state (²Π) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPS. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	2.802(1) ^{† a}	2.753(5)	-0.009(7)	-0.012(7)	0.035(7)	-0.060(7)	-0.026(8)	-0.016(7)	-0.001(7)
ω_e (cm ⁻¹)	839.4 ^a	851(3)	-3(4)	-4(4)	0(4)	-13(4)	-4(4)	-6(4)	-2(4)
r_e (Å)	1.596 ^b	1.564(1)	0.004(1)	0.001(2)	-0.008(2)	0.007(2)	-0.007(2)	0.001(2)	0.001(2)
Potential Energy Surface									
U(1.17)		3.2728	0.1081	-0.1807	-0.5878	0.0054	-0.5093	-0.0260	-0.0119
U(1.27)		-0.1726	0.0674	-0.1125	-0.3375	0.0232	-0.2738	0.0076	-0.0002
U(1.37)		-1.8723	0.0400	-0.0694	-0.1919	0.0256	-0.1525	0.0172	0.0038
U(1.47)		-2.5904	0.0215	-0.0423	-0.1080	0.0244	-0.0766	0.0174	0.0034
U(1.57)		-2.7621	0.0089	-0.0259	-0.0592	0.0224	-0.0219	0.0146	0.0012
U(1.67)		-2.6326	0.0002	-0.0156	-0.0297	0.0205	0.0167	0.0114	-0.0010

[†] Sum of experimental dissociation and zero-point energies

^a Reference [FP99]

^b Reference [Lov05]

Table B.30 All-electron (AE) UCCSD(T) ArH+ ground state ($^1\Sigma$) binding parameters and potential energy surface along with the errors from uncorrelated core (UC) and ECPs. All energies in eV.

Qty.	Exp.	AE	Discrepancies from AE						
			UC	SBKJC	BFD	TN-DF	STU	ccECP[Ne]	ccECP[He]
Binding Parameters									
D_e (eV)	4.059 ^{† a}	4.072(4)	0.004(6)	-0.001(6)	0.026(6)	0.001(6)	-0.015(6)	0.000(6)	0.001(6)
ω_e (cm ⁻¹)	2723 ^a	2728(18)	-1(26)	2(25)	1(25)	0(26)	-8(25)	-1(25)	2(26)
r_e (Å)	1.286 ^a	1.282(1)	0.001(1)	0.000(1)	-0.004(1)	0.001(1)	-0.002(1)	0.000(1)	0.000(1)
Potential Energy Surface									
U(0.9)		0.4748	0.0307	0.0081	-0.1533	0.0213	-0.0770	-0.0033	0.0020
U(1.0)		-2.1495	0.0160	0.0067	-0.0913	0.0137	-0.0282	0.0014	0.0006
U(1.1)		-3.4435	0.0062	0.0041	-0.0562	0.0064	-0.0030	0.0014	0.0004
U(1.2)		-3.9723	-0.0003	0.0022	-0.0361	0.0012	0.0099	0.0004	0.0008
U(1.3)		-4.0653	-0.0045	0.0013	-0.0244	-0.0018	0.0161	-0.0006	0.0015
U(1.4)		-3.9164	-0.0072	0.0012	-0.0174	-0.0035	0.0186	-0.0013	0.0022

[†] Sum of experimental dissociation and zero-point energies

^a Reference [Ros79]

Table B.31 Parameter values for Cl and Ar ccECP[Ne].S potentials. The highest l value corresponds to the local channel.

Atom	Z_{eff}	l	$n_{l,k}$	$\alpha_{l,k}$	$\beta_{l,k}$
Cl	7	0	2	17.23708573	6.50888648
		0	2	4.31148447	46.76346700
		1	2	11.38275704	2.99464770
		1	2	3.83218762	28.01703410
		2	1	22.71655173	7.00000000
		2	3	78.57185685	159.01586213
		2	2	7.47352436	-15.65310650
		Ar	8	0	2
0	2			5.17516183	60.31303710
1	2			17.09350095	3.47705723
1	2			4.64108131	36.47236220
2	1			8.98034652	8.00000000
2	3			26.23635698	71.84277216
2	2			7.10917092	-22.13000430

ular errors are given in Figure B.1 and Tables B.33 and B.34. And for Ar, the spectral errors are given in Table B.35 while the molecular errors are given in Figure B.2 and Table B.36. In

Table B.32 All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Cl along with the errors from ccECP[Ne].S. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V in main text for further description.

Qty.	AE	Discrepancy from AE ccECP[Ne].S
IP(I)	12.9388	-0.0427
IP(II)	23.7309	-0.0830
IP(III)	39.6980	0.0233
IP(IV)	53.1811	-0.0321
IP(V)	67.6102	-0.1362
IP(VI)	96.8961	-0.1292
IP(VII)	114.2079	-0.4249
EA	3.6210	-0.0184
AMAD		0.1112
LMAD		0.0480
MARE		0.0025

Table B.33 All-electron (AE) UCCSD(T) Cl₂ ground state ($^1\Sigma_g$) binding parameters and potential energy surface along with the errors from ccECP[Ne].S. All energies in eV.

Qty.	AE	Discrepancy from AE ccECP[Ne].S
Binding Parameters		
D_e (eV)	2.526(9)	-0.04(1)
ω_e (cm ⁻¹)	533(5)	-5(7)
r_e (Å)	1.987(3)	0.002(5)
Potential Energy Surface		
U(1.6)	0.5063	0.0254
U(1.7)	-1.1542	0.0377
U(1.8)	-2.0502	0.0406
U(1.9)	-2.4538	0.0391
U(2.0)	-2.5435	0.0358
U(2.1)	-2.4387	0.0317

the figures and tables, we have labeled these alternative ECPs as ccECP[Ne].S.

Table B.34 All-electron (AE) UCCSD(T) ClO ground state ($^2\Pi$) binding parameters and potential energy surface along with the errors from ccECP[Ne].S. All energies in eV.

Qty.	AE	Discrepancy from AE ccECP[Ne].S
Binding Parameters		
D_e (eV)	2.753(5)	-0.036(7)
ω_e (cm^{-1})	850(3)	-10(4)
r_e (Å)	1.564(1)	0.003(2)
Potential Energy Surface		
U(1.17)	3.2728	0.0240
U(1.27)	-0.1726	0.0503
U(1.37)	-1.8723	0.0513
U(1.47)	-2.5904	0.0433
U(1.57)	-2.7621	0.0340
U(1.67)	-2.6326	0.0258

Table B.35 All-electron (AE) UCCSD(T) ionization potentials and electron affinity of Ar along with the errors from ccECP[Ne].S. The uncontracted aug-cc-pCV5Z basis was used for all calculations. All values in eV. See Tab. V in main text for further description.

Qty.	AE	Discrepancy from AE ccECP[Ne].S
IP(I)	15.7829	0.0202
IP(II)	27.6005	-0.0126
IP(III)	40.6121	-0.0862
IP(IV)	59.2804	0.0239
IP(V)	74.7437	-0.1283
IP(VI)	91.1085	-0.3698
IP(VII)	124.3356	-0.1401
IP(VIII)	143.4706	-0.5716
AMAD		0.1112
LMAD		0.0480
MARE		0.0025

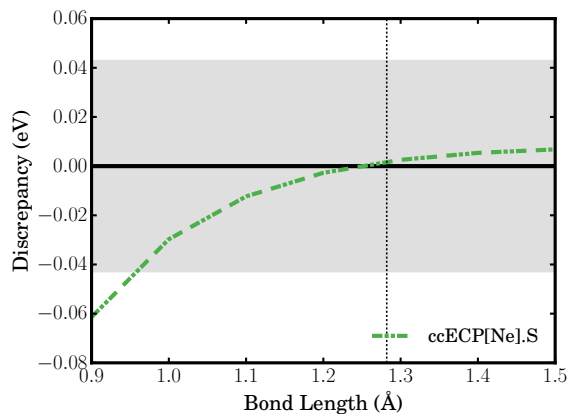


Figure B.2 Binding energy discrepancies for the ArH+ molecule in its ground state $^1\Sigma$. The binding curves are relative to the AE UCCSD(T) binding curve. The shaded region indicates a discrepancy of chemical accuracy in either direction.

Table B.36 All-electron (AE) UCCSD(T) ArH+ ground state ($^1\Sigma$) binding parameters and potential energy surface along with the errors from ccECP[Ne].S. All energies in eV.

Qty.	AE	Discrepancy from AE ccECP[Ne].S
Binding Parameters		
D_e (eV)	4.072(4)	-0.002(6)
ω_e (cm^{-1})	2728(18)	-2(25)
r_e (Å)	1.282(1)	-0.002(1)
Potential Energy Surface		
U(0.9)	0.4748	-0.0614
U(1.0)	-2.1495	-0.0297
U(1.1)	-3.4435	-0.0123
U(1.2)	-3.9723	-0.0027
U(1.3)	-4.0653	0.0026
U(1.4)	-3.9164	0.0054

Supplementary Material: New generation of effective core potentials from correlated calculations: 3d transition metal series

C.1 Basis Sets

For each of the correlation consistent ECPs, (ccECP), we have constructed correlation consistent basis sets. To optimize the basis sets, we use the procedure of Balabanov *et al.* [BP05] to construct both cc-pVnZ and aug-cc-pVnZ basis sets for $n \in \{D, T, Q, 5\}$ quality. We first optimize a set of primitive exponents for the s , p , d functions using 13 primitives for the s and p , and 11 primitives for the d . The optimal exponents are found by minimizing the HF energy for the $[\text{Ar}]3d^n 4s^2$ state, which is the ground state for Sc-V, Mn-Ni, and Zn.

Once we have determined the optimal set of primitive exponents, we generate the minimal basis set by first minimizing the state-averaged HF energy for the $[\text{Ar}]3d^n 4s^2$ state. This generates the first 2 s contractions. For the first two p contractions, we minimize the state-averaged HF energy of the $[\text{Ar}]3d^{n-1} 4s^2 4p^1$ state. To generate the d contraction, we consider a set of neutral states, namely the $[\text{Ar}]3d^n 4s^2$, $[\text{Ar}]3d^{n+1} 4s^1$, and $[\text{Ar}]3d^{n+2}$ states. We minimize the state-averaged HF energy of each of these states and obtain the contraction coefficients from the natural orbitals of the averaged density matrix from the

individual state-averaged density matrices. This procedure provides two s , two p , and one d contraction. To generate the cc-pVnZ basis, we must add an additional $n - 1$ functions to the s , p , and d levels as well as the polarization functions for the f , g , h , and i functions (e.g. DZ has 1 f , TZ has 2 f 1 g , QZ has 3 f 2 g 1 h , and 5Z has 4 f 3 g 2 h 1 i). The additional s , p and d contracted functions are obtained from the averaged CISD density matrix from the $[\text{Ar}]3d^n 4s^2$, $[\text{Ar}]3d^{n+1} 4s^1$, and $[\text{Ar}]3d^{n+2}$ states. We ignore the first two natural orbitals for the s and p and first natural orbital for the d from the averaged density matrix, and then take as the additional contracted functions the coefficients of the next highest occupation numbers (DZ takes the next highest, TZ takes the next two highest, QZ takes the next three highest, and 5Z takes the next four highest occupation number of the natural orbitals). The polarization functions are obtained from minimizing the average CISD energy of the $[\text{Ar}]3d^n 4s^2$, $[\text{Ar}]3d^{n+1} 4s^1$, and $[\text{Ar}]3d^{n+2}$ states. The polarization functions are even-tempered, i.e., we only optimize the smallest exponent and the ratio between the additional terms. Additionally, we also add an uncontracted function for the s , p , and d basis sets using the smallest exponent that was in the original contraction, thus finalizing our cc-pVnZ basis sets.

Lastly, we generate the augmented basis set as follows. For the s , p , and d functions, we take the smallest exponent and add an additional uncontracted function, with the exponent being (roughly) a factor of two smaller than the smallest exponent. For the f , g , h , and i functions, we use our optimized ratios from the even-tempered cc-pVnZ polarizations function to add an additional smaller exponent using the same ratio. For the terms that only have one exponent (f for DZ, g for TZ, h for QZ, and i for 5Z), we minimize the average CISD energy of the $[\text{Ar}]3d^n 4s^2$, $[\text{Ar}]3d^{n+1} 4s^1$, and $[\text{Ar}]3d^{n+2}$ states by adding an additional primitive exponent to be optimized. The final cc-pVnZ and aug-cc-pVnZ basis sets will be included on our website, <http://pseudopotentiallibrary.org>.

C.2 Atomic Spectra

In the following subsections, we show the atomic excitation information and binding parameters for all molecule used in this work.

C.2.1 Sc

Table C.2, shows atomic excitation discrepancies for Sc atom. ScH and ScO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.3 and Table C.4 respectively.

Table C.1 Total energy components of the $[\text{Ar}] 3d^1 4s^2 \ ^2D$ ground state of the Sc atom using our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-46.119887	-46.120165	-46.120191	-46.120194	-46.120116	-0.000078
Correlation	-0.409235	-0.422091	-0.428269	-0.435320		

Table C.2 Sc AE gaps and relative errors for various core approximations. All gaps are relative to the $[\text{Ar}] 3d^1 4s^2 \ ^2D$ state. All values in eV.

Gaps	AE	UC	BFD	STU	eCEPP	ccECPS
$[\text{Ar}] 3d^2 4s^2 \ ^3F$	0.509479	0.003048	0.525507	0.064818	-0.000680	0.008299
$[\text{Ar}] 3d^2 4s^1 \ ^4F$	1.419374	0.011864	0.671360	0.084764	-0.002449	0.006857
$[\text{Ar}] 3d^3 \ ^4F$	4.283618	0.011646	0.728095	0.152030	-0.006531	0.008463
$[\text{Ar}] 3d^1 4s^1 \ ^3D$	6.542273	0.002095	0.604447	0.028273	-0.003293	-0.009606
$[\text{Ar}] 3d^2 \ ^3F$	7.144951	0.019375	0.705238	0.138533	-0.000463	-0.000408
$[\text{Ar}] 3d^1 \ ^2D$	19.344148	0.004626	0.560909	0.105471	0.001551	-0.021660
$[\text{Ar}] \ ^1S$	44.069078	-0.040491	0.426239	0.123866	-0.013715	0.002667
$[\text{Ne}] 3s^2 \ ^1S$	795.976959	-0.762055	-0.281420	0.424253	-2.238355	0.004354

C.2.2 Ti

In Table C.5, we give the total energy components for our ccECP of the Ti ground state. Table C.6, shows atomic excitation discrepancies for Ti atom. TiH and TiO AE molecular parameters and corresponding discrepancies for various core approximations are shown

Table C.3 ScH AE molecular binding parameters and discrepancies for various core approximations. All parameters were obtained using Morse potential fit. The parameters shown are dissociation energy D_e , equilibrium bond length r_e , vibrational frequency ω_e and dissociation energy discrepancy at dissociation bond length D_{diss} .

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	2.06 ^a	1.775 ^b	1547 ^b	
AE	2.30(1)	1.755(5)	1548(16)	
UC	0.01(2)	0.002(7)	-1(23)	0.01(13)
BFD	0.02(2)	0.004(7)	4(23)	0.04(13)
STU	-0.01(2)	-0.003(7)	3(23)	-0.00(13)
eCEPP	-0.00(2)	-0.001(7)	1(23)	-0.00(13)
ccECPS	-0.01(2)	-0.000(7)	-1(23)	0.01(13)

^a Reference [FP06]

^b Reference [RB96]

Table C.4 ScO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	7.07 ^a	1.668 ^a	965 ^a	
AE	6.95(1)	1.659(1)	975.6(7.1)	
UC	-0.01(2)	0.002(2)	-2(10)	0.06(22)
BFD	-0.08(2)	0.008(2)	-4(10)	0.42(24)
STU	-0.08(2)	-0.001(2)	-1(10)	0.04(22)
eCEPP	0.00(2)	0.001(2)	-1(10)	0.05(22)
ccECPS	-0.01(2)	0.000(2)	0(10)	0.04(22)

^a Reference [FP06]

in Table C.7 and Table C.8 respectively.

Table C.5 Total energy components of the $[\text{Ar}] 3d^2 4s^2 {}^3F$ ground state of the Ti atom using our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-57.606631	-57.607057	-57.607104	-57.607110	-57.607088	-0.000022
Correlation	-0.452225	-0.467988	-0.474562	-0.482664		

Table C.6 Ti AE gaps and relative errors for various core approximations. All in eV.

Gaps		AE	UC	BFD	STU	eCEPP	ccECP
$[\text{Ar}] 3d^3 4s^2$	4F	0.020661	0.008558	0.084231	0.073673	0.005225	0.007623
$[\text{Ar}] 3d^3 4s^1$	5F	0.796464	0.015718	0.110225	0.082066	0.006530	0.006560
$[\text{Ar}] 3d^4$	5D	3.616908	0.019344	0.185767	0.159201	0.009038	0.012720
$[\text{Ar}] 3d^2 4s^1$	4F	6.804955	0.003393	0.009862	0.026440	-0.003403	-0.009247
$[\text{Ar}] 3d^3$	4F	6.915550	0.026609	0.141649	0.141318	0.020958	0.006538
$[\text{Ar}] 3d^2$	3F	20.387116	0.009622	0.015878	0.112828	0.008137	-0.017891
$[\text{Ar}] 3d^1$	2D	47.867620	-0.037232	-0.151758	0.130371	-0.040542	-0.029817
$[\text{Ar}]$	1S	91.091323	-0.124071	-0.391450	0.258178	-1.234149	0.012368
$[\text{Ne}] 3s^2$	1S	1,028.587483	-0.954480	-0.171680	0.493782	-1.703671	0.014604

C.2.3 V

In Table C.9, we give the total energy components for our ccECP of the V ground state. Table C.10, shows atomic excitation discrepancies for V atom. VH and VO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.11 and Table C.12 respectively.

Table C.7 TiH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	2.168 ^a	1.779 ^a	1405 ^a	
AE	2.105(4)	1.762(2)	1516.8(8.7)	
UC	-0.003(5)	0.001(3)	-0(12)	0.019(61)
BFD	-0.027(5)	0.007(3)	-4(12)	0.100(63)
STU	-0.045(5)	-0.000(3)	-3(13)	0.046(61)
eCEPP	0.005(5)	0.000(3)	2(12)	0.006(60)
ccECP	0.007(5)	0.001(3)	1(12)	0.010(60)

^a Reference [FP06]

Table C.8 TiO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	6.98 ^a	1.620 ^a	1009 ^a	
AE	6.85(1)	1.615(2)	999.1(6.8)	
UC	-0.01(2)	0.002(3)	-2.0(9.6)	0.06(22)
BFD	-0.10(2)	0.007(3)	-4.1(9.9)	0.41(24)
STU	-0.08(2)	-0.001(3)	-1.3(9.9)	0.03(23)
eCEPP	0.02(2)	0.001(3)	0.3(9.8)	0.06(23)
ccECP	0.00(2)	0.000(3)	1.0(9.7)	0.02(22)

^a Reference [FP06]

Table C.9 Total energy components of the [Ar] 3d³4s² 4F ground state of the V atom using our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-70.895598	-70.896086	-70.896115	-70.896117	-70.896038	-0.000079
Correlation	-0.503389	-0.522224	-0.529968	-0.539407		

Table C.10 V AE gaps and relative errors for various core approximations. All data in eV.

Gaps	AE	UC	BFD	STU	eCEPP	ccECPS	ccECP
[Ar] $3d^4 4s^2$ 5D	-0.476250	0.010698	0.104690	0.069543	-0.002350	0.005158	0.018813
[Ar] $3d^4 4s^1$ 6D	0.225411	0.016419	0.122574	0.072073	-0.012547	0.000018	0.016268
[Ar] $3d^5$ 6S	2.477803	0.022536	0.223108	0.153314	-0.012134	0.009320	0.040209
[Ar] $3d^3 4s^1$ 5F	7.040745	0.003875	0.011638	0.014400	-0.012391	-0.012788	-0.005745
[Ar] $3d^4$ 5D	6.702435	0.028646	0.161530	0.123399	0.000007	0.003149	0.030633
[Ar] $3d^3$ 4F	21.353719	0.012672	0.037204	0.086384	0.004329	-0.009391	0.011301
[Ar] $3d^2$ 3F	50.673774	-0.030859	-0.104118	0.088290	-0.015417	-0.022190	-0.008337
[Ar] $3d^1$ 2D	97.385632	-0.106347	-0.297488	0.181266	-0.065904	-0.024749	-0.014226
[Ar] 1S	162.607898	-0.225926	-0.639514	0.434473	-1.746391	0.012854	0.026267
[Ne] $3s^2$ 1S	1,304.413862	-1.135737	0.940843	1.273876	-3.906868	-0.005499	0.026110

Table C.11 VH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm $^{-1}$)	D_{diss} (eV)
Exp.	2.229 ^a			
	2.229(69) ^b	1.730 ^b		
	2.28(19), 2.433(65) ^c	1.73 ^c		
AE	2.622(2)	1.6799(9)	1652.7(3.7)	
UC	0.011(3)	-0.001(1)	2.9(5.2)	-0.016(29)
BFD	-0.035(3)	0.005(1)	-3.7(5.4)	0.101(31)
STU	-0.035(3)	-0.000(1)	-2.7(5.4)	0.034(30)
eCEPP	0.016(3)	0.000(1)	2.3(5.1)	-0.011(28)
ccECPS	0.033(3)	-0.002(1)	5.3(5.2)	-0.045(29)
ccECP	0.009(3)	-0.000(1)	2.1(5.2)	-0.008(29)

^a Reference [FP06]

^b Reference [Xu15]

^c Reference [Fan17]

Table C.12 VO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	6.505 ^a	1.589 ^a	1011 ^a	
	6.548(87) ^b	1.589 ^b		
	6.548(87), 6.474 ^c	1.589 ^c		
AE	6.686(9)	1.585(2)	1005.1(5.1)	
UC	-0.00(1)	0.002(2)	-1.4(7.2)	0.05(18)
BFD	-0.10(1)	0.005(3)	-3.1(7.5)	0.32(19)
STU	-0.06(1)	-0.003(2)	2.2(7.3)	-0.04(18)
eCEPP	0.01(1)	0.002(2)	-1.0(7.3)	0.08(18)
ccECPS	0.03(1)	-0.001(2)	2.9(7.3)	-0.04(18)
ccECP	0.02(1)	-0.001(2)	3.2(7.3)	-0.03(18)

^a Reference [FP06]

^b Reference [Xu15]

^c Reference [Fan17]

C.2.4 Cr

In Table C.13, we give the total energy components for our ccECP of the Cr ground state. Table C.14, shows atomic excitation discrepancies for Cr atom. CrH and CrO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.15 and Table C.16 respectively.

Table C.13 Total energy components of the [Ar] $3d^5 4s^1 \ ^7S$ ground state of the Cr atom for our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-86.047811	-86.048454	-86.048542	-86.048556	-86.048401	-0.000155
Correlation	-0.542224	-0.566030	-0.575994	-0.586753		

Table C.14 Cr AE gaps and relative errors for various ECPs. All values in eV

Gaps	AE	UC	BFD	STU	eCEPP	ccECPS	ccECP
[Ar] $3d^5 4s^2$ 6S	-1.639160	0.008708	0.100464	0.075539	0.002912	0.002231	-0.007320
[Ar] $3d^5 4s^1$ 7S	-1.026768	0.012626	0.109308	0.070858	-0.002395	-0.001850	-0.015864
[Ar] $3d^6$ 5D	3.444963	0.006449	0.189582	0.145908	-0.006204	-0.002068	-0.024681
[Ar] $3d^4 4s^1$ 6D	7.256491	0.003483	0.003102	0.016517	-0.016762	-0.011701	-0.017497
[Ar] $3d^5$ 6S	5.735591	0.023783	0.139894	0.140656	0.010259	0.016681	-0.005632
[Ar] $3d^4$ 5D	22.265020	0.013198	0.032490	0.127077	-0.010531	0.006177	-0.003946
[Ar] $3d^3$ 4F	53.342070	-0.022558	-0.082097	0.181337	-0.015293	-0.003129	0.008626
[Ar] $3d^2$ 3F	102.464577	-0.084655	-0.217174	0.344850	0.022232	-0.012436	0.032518
[Ar] $3d^1$ 2D	171.920939	-0.176820	-0.446866	0.669754	0.143731	-0.017361	0.082695
[Ar] 1S	262.444038	-0.312659	-1.402775	1.242200	0.422348	0.020817	0.217202
[Ne] $3s^2$ 1S	1,627.174829	-1.275153	2.973117	0.528908	0.506350	0.036490	-2.483884

Table C.15 CrH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm $^{-1}$)	D_{diss} (eV)
Exp.	2.029 ^a	1.656 ^a	1581 ^a	
	2.029(69) ^b	1.656 ^b		
	2.11(14), 2.238(65) ^c	1.656 ^c		
AE	2.120(6)	1.614(2)	1740(13)	
UC	0.015(8)	0.003(3)	-8(17)	-0.004(73)
BFD	0.057(7)	0.005(3)	-7(16)	-0.038(70)
STU	0.024(7)	0.002(3)	-9(16)	-0.037(69)
eCEPP	0.019(7)	0.005(3)	-14(16)	-0.002(70)
ccECPS	0.028(7)	0.004(3)	-12(16)	-0.018(69)
ccECP	0.013(7)	0.005(3)	-13(16)	-0.002(70)

^a Reference [FP06]^b Reference [Xu15]^c Reference [Fan17]

Table C.16 CrO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	4.467 ^a	1.615 ^a	898 ^a	
	4.540(52) ^b	1.621 ^b		
	4.818(87) ^c	1.615 ^c		
AE	4.573(4)	1.6139(6)	899.2(1.9)	
UC	0.008(5)	0.0010(9)	-0.2(2.9)	0.024(52)
BFD	0.090(5)	-0.0044(9)	9.8(2.9)	-0.146(51)
STU	-0.012(6)	-0.0024(9)	2.0(3.0)	-0.043(54)
eCEPP	0.018(5)	0.0011(9)	0.2(2.9)	0.018(52)
ccECPS	0.041(5)	-0.0025(9)	4.8(2.9)	-0.079(51)
ccECP	0.010(5)	-0.0009(9)	1.5(2.9)	-0.027(51)

^a Reference [FP06]

^b Reference [Xu15]

^c Reference [Fan17]

C.2.5 Mn

In Table C.17, we give the total energy components for our ccECP of the Mn ground state. Table C.18, shows atomic excitation discrepancies for Mn atom. MnH and MnO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.19 and Table C.20 respectively.

Table C.17 Total energy components of the [Ar] $3d^5 4s^2 \ ^6S$ ground state of the Mn atom for our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-103.243494	-103.244240	-103.244299	-103.244305	-103.244161	-0.000144
Correlation	-0.591277	-0.617442	-0.628553	-0.642432		

Table C.18 Mn AE gaps and relative error for various core approximations. All in eV.

Gaps	AE	UC	BFD	STU	eCEPP	ccECPS	ccECP
[Ar] $3d^6 4s^2$ 5D	1.299360	0.001050	0.146607	0.028564	0.004187	-0.003555	-0.006144
[Ar] $3d^6 4s^1$ 6D	2.151429	0.006575	0.169347	0.010605	0.002082	-0.006991	-0.013559
[Ar] $3d^7$ 4F	5.772304	0.007078	0.285947	0.059634	0.018798	0.004177	-0.003478
[Ar] $3d^5 4s^1$ 7S	7.414670	0.003239	-0.003639	-0.004177	-0.006771	-0.008937	-0.008985
[Ar] $3d^6$ 5D	9.245516	0.007056	0.194099	0.017318	0.003874	-0.015398	-0.013276
[Ar] $3d^5$ 6S	23.059159	0.014572	0.022398	0.081569	0.014290	0.030141	0.050613
[Ar] $3d^4$ 5D	56.822155	-0.015390	-0.072555	0.256077	-0.026524	0.016815	0.033293
[Ar] $3d^3$ 4F	108.315370	-0.074830	-0.158482	0.600224	-0.059576	0.010784	0.017702
[Ar] $3d^2$ 3F	180.815120	-0.162619	-0.291000	1.159615	-0.070647	0.003077	0.004779
[Ar] $3d^1$ 2D	276.392004	-0.282966	-0.586852	1.989336	-0.043384	-0.008268	0.013606
[Ar] 1S	395.519402	-0.451819	-1.184772	3.166735	0.029970	0.001807	0.103825
[Ne] $3s^2$ 1S	2001.865958	-1.490682	3.394075	0.790442	-2.370058	0.004111	-6.624472

Table C.19 MnH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm $^{-1}$)	D_{diss} (eV)
Exp.	1.349 ^a	1.731 ^a	1548 ^a	
AE	1.644(5)	1.721(3)	1521(18)	
UC	-0.002(7)	0.000(5)	-0(26)	0.004(90)
BFD	0.067(7)	-0.008(4)	27(26)	-0.100(87)
STU	-0.035(7)	0.002(5)	-13(26)	0.034(91)
eCEPP	0.016(7)	-0.001(4)	5(26)	-0.021(89)
ccECPS	0.027(7)	-0.004(4)	10(26)	-0.048(88)
ccECP	0.002(7)	-0.001(5)	0(26)	-0.009(90)

^a Reference [FP06]

Table C.20 MnO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	
Exp.	3.881 ^a	1.646 ^a	840 ^a	
AE	3.773(9)	1.640(2)	841.1(4.8)	
UC	-0.01(1)	0.001(2)	-0.2(6.6)	0.03(11)
BFD	-0.01(1)	-0.003(2)	1.0(6.1)	-0.06(10)
STU	-0.09(1)	0.001(2)	-8.2(6.7)	0.05(11)
eCEPP	0.02(1)	-0.000(2)	2.6(6.6)	-0.01(11)
ccECPS	0.04(1)	-0.004(2)	6.8(6.6)	-0.11(11)
ccECP	0.01(1)	-0.002(2)	3.3(6.7)	-0.03(11)

^a Reference [FP06]

C.2.6 Fe

In Table C.21, we give the total energy components for our ccECP of the Fe ground state. Table C.22, shows atomic excitation discrepancies for Fe atom. FeH and FeO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.23 and Table C.24 respectively.

Table C.21 Total energy components of the [Ar] $3d^6 4s^2 \ ^5D$ ground state of the Fe atom.

	T	Q	5	Extrap.	Num.	Diffs
SCF	-122.641201	-122.642105	-122.642174	-122.642183	-122.641993	-0.000190
Correlation	-0.677200	-0.708431	-0.719770	-0.732173		

C.2.7 Co

In Table C.25, we give the total energy components for our ccECP of the Co ground state. Table C.26, shows atomic excitation discrepancies for Co atom. CoH and CoO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.27 and Table C.28 respectively.

Table C.22 Fe AE gaps and relative errors for various ECPs. All data in eV

Gaps	AE	UC	BFD	STU	eCEPP	ccECPS	ccECP
[Ar]4s ² 3d ⁷ ⁴ F	-0.057987	0.001660	0.156504	0.044160	0.011820	0.007696	-0.005231
[Ar] 4s ¹ 3d ⁷ ⁵ F	0.888551	0.006370	0.171349	0.033548	0.008323	-0.002355	-0.014157
[Ar] 3d ⁸ ³ F	4.157110	0.010139	0.291230	0.092898	0.030171	-0.003989	-0.017050
[Ar] 4s ¹ 3d ⁶ ⁶ D	7.885826	0.004183	0.001180	0.005114	-0.004736	-0.020115	-0.015930
[Ar] 3d ⁷ ⁴ F	8.162522	0.016486	0.199663	0.071353	0.026536	0.002202	-0.002154
[Ar]3d ⁶ ⁵ D	24.090858	0.014936	0.033942	0.099740	0.010690	0.008389	0.013777
[Ar]3d ⁵ ⁶ S	54.661521	-0.005615	-0.168899	0.206248	-0.017071	0.004946	0.021878
[Ar]3d ⁴ ⁵ D	109.577638	-0.051901	-0.192542	0.455601	-0.007441	-0.005419	-0.008571
[Ar]3d ³ ⁴ F	185.022592	-0.129771	-0.222149	0.890969	0.067005	-0.000924	-0.005565
[Ar]3d ² ³ F	284.232562	-0.235873	-0.354853	1.557802	0.247372	0.001835	0.040188
[Ar]3d ¹ ² D	409.151353	-0.374283	-0.745345	2.506921	0.518200	-0.005224	0.160413
[Ar] ¹ S	560.024014	-0.563164	-1.561520	3.821096	0.924669	0.002865	0.431303
[Ne]3s ² ¹ S	2,426.492957	-1.657243	3.573802	0.956335	-0.680726	-0.017867	-5.966712

Table C.23 FeH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	1.70 ^a	1.589 ^a	1827 ^a	
	1.600(35) ^b	1.630 ^c		
	1.712(13),1.943(65) ^d	1.63 ^d		
AE	1.89(1)	1.538(4)	1833(28)	
UC	-0.01(2)	-0.003(7)	14(42)	0.01(15)
BFD	-0.09(2)	-0.006(7)	17(44)	0.09(16)
STU	-0.05(2)	-0.006(7)	11(43)	0.02(15)
eCEPP	-0.00(2)	-0.002(6)	9(41)	0.00(15)
ccECPS	0.00(2)	-0.004(6)	13(41)	-0.01(14)
ccECP	-0.00(2)	-0.002(6)	9(41)	-0.00(14)

^a Reference [FP06]^b Reference [Xu15]^c Reference [Fan17]

Table C.24 FeO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	4.224 ^a	1.616 ^a	880 ^a	
AE	4.12(1)	1.611(3)	878.7(8.2)	
UC	-0.01(2)	0.001(4)	-1(11)	0.03(19)
BFD	-0.04(2)	-0.003(4)	1(11)	-0.04(19)
STU	-0.07(2)	-0.003(4)	-0(12)	-0.02(20)
eCEPP	0.02(2)	0.000(4)	2(12)	-0.00(20)
ccECPS	0.06(2)	-0.004(4)	6(11)	-0.13(19)
ccECP	0.03(2)	-0.001(4)	3(12)	-0.03(19)

^a Reference [FP06]

Table C.25 Total energy components of the [Ar] $3d^7 4s^2 \ ^4F$ ground state of the Co atom using our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-144.325544	-144.327075	-144.327246	-144.327268	-144.327074	-0.000194
Correlation	-0.746000	-0.782672	-0.799273	-0.820951		

Table C.26 Co AE gaps and relative errors in eV

Gaps	AE	UC	BFD	STU	ccECPS	ccECP
[Ar] $3d^8 4s^2 \ ^3F$	-0.648824	0.003846	0.138580	0.038570	0.003500	-0.011140
[Ar] $3d^8 4s^1 \ ^4F$	0.404998	0.008382	0.146827	0.029158	0.000131	-0.018138
[Ar] $3d^9 \ ^2D$	3.290224	0.013495	0.245172	0.082758	0.001712	-0.027850
[Ar] $3d^7 4s^1 \ ^5F$	8.285103	0.004751	-0.004194	0.008731	-0.007956	-0.014990
[Ar] $3d^8 \ ^3F$	7.852169	0.016852	0.154921	0.060521	0.007556	-0.012459
[Ar] $3d^7 \ ^4F$	24.959596	0.014340	0.012356	0.096283	0.010968	0.005804
[Ar] $3d^6 \ ^5D$	58.474931	-0.000315	-0.137538	0.215407	0.000966	0.010511
[Ar] $3d^5 \ ^6S$	109.967312	-0.035818	-0.288764	0.449440	-0.020260	0.016728
[Ar] $3d^4 \ ^5D$	189.676472	-0.096964	-0.229092	0.878645	-0.018478	0.002135
[Ar] $3d^3 \ ^4F$	292.523396	-0.190846	-0.217323	1.539629	0.001712	0.062171
[Ar] $3d^2 \ ^3F$	421.774512	-0.312948	-0.383028	2.475055	0.017756	0.218327
[Ar] $3d^1 \ ^2D$	579.286933	-0.467425	-0.905317	3.738709	0.010326	0.520393
[Ar] $\ ^1S$	765.100530	-0.673798	-1.957082	5.417765	-0.013019	1.067807
[Ne] $3s^2 \ ^1S$	2910.339236	-1.819632	5.795586	1.122129	-0.018151	-8.537068

Table C.27 CoH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	2.099 ^a	1.52 ^a	1925 ^a	
	1.973(52) ^b	1.530 ^b		
	2.19(13), 2.368(65) ^c	1.53 ^c		
AE	2.123(4)	1.515(2)	1813(12)	
UC	-0.006(5)	-0.000(3)	1(17)	0.008(67)
BFD	-0.133(6)	-0.002(3)	1(21)	0.155(80)
STU	-0.042(6)	-0.003(3)	2(18)	0.029(69)
ccECPS	0.010(5)	-0.002(2)	7(17)	-0.021(66)
ccECP	0.014(5)	0.000(2)	2(17)	-0.008(65)

^a Reference [FP06]

^b Reference [Xu15]

^c Reference [Fan17]

Table C.28 CoO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	3.998 ^a	1.616 ^a	880 ^a	
AE	3.79(2)	1.607(5)	891(16)	
UC	0.00(3)	0.001(7)	-2(22)	0.02(35)
BFD	0.01(3)	-0.009(7)	13(23)	-0.16(35)
STU	-0.06(3)	-0.005(7)	3(23)	-0.05(35)
ccECPS	0.05(3)	-0.004(7)	8(23)	-0.12(35)
ccECP	0.01(3)	-0.000(7)	1(22)	0.00(35)

^a Reference [FP06]

Table C.29 Total energy components for the $[\text{Ar}]3d^84s^2\ ^3F$ state of the Ni atom for our ccECP.

SCF	T	Q	5	Extrap.	Numerical	Diffs
SCF	-168.409192	-168.411003	-168.411106	-168.411112	-168.410808	-0.000304
Correlation	-0.822084	-0.864373	-0.883557	-0.908639		

C.2.8 Ni

In Table C.29, we show the total energy components for the Ni ground state using our ccECP. Table C.30, shows atomic excitation discrepancies for Ni atom. NiH and NiO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.31 and Table C.32 respectively.

Table C.30 Ni AE gaps and relative errors for various ECPs. All values in eV

Gaps		AE	UC	BFD	STU	ccECPS	ccECP
$[\text{Ar}] 3d^94s^2$	2D	-1.209029	0.005065	0.146336	0.013536	-0.003339	0.000304
$[\text{Ar}] 3d^94s^1$	3D	-0.039999	0.010094	0.154368	-0.007817	-0.003776	-0.003715
$[\text{Ar}] 3d^{10}$	1S	1.665329	0.014848	0.258005	0.022213	-0.004163	-0.006370
$[\text{Ar}] 3d^84s^1$	4F	8.678287	0.004880	-0.004271	-0.026728	0.001234	-0.011767
$[\text{Ar}] 3d^9$	2D	7.580538	0.017010	0.152086	-0.041737	0.002174	-0.002632
$[\text{Ar}] 3d^8$	3F	25.813278	0.013459	-0.005649	-0.039091	0.014570	-0.007257
$[\text{Ar}] 3d^7$	4F	61.040014	-0.000243	-0.144218	0.100049	0.016374	-0.042470
$[\text{Ar}] 3d^6$	5D	116.199196	-0.026699	-0.226084	0.435981	0.004456	-0.105248
$[\text{Ar}] 3d^5$	6S	192.239024	-0.075327	-0.281178	1.016989	-0.018535	-0.175304
$[\text{Ar}] 3d^4$	5D	300.110087	-0.148578	-0.080333	2.014774	-0.012141	-0.209976
$[\text{Ar}] 3d^3$	4F	433.615540	-0.256215	0.003373	3.416095	0.010293	-0.172026
$[\text{Ar}] 3d^2$	3F	596.091441	-0.391620	-0.212840	5.276018	0.023892	-0.054807
$[\text{Ar}] 3d^1$	2D	789.334381	-0.559135	-0.961215	7.649629	0.010737	0.163535
$[\text{Ar}]$	1S	1013.195350	-0.779566	-2.452519	10.611841	-0.013917	0.540565
$[\text{Ne}] 3s^2$	1S	3455.730422	-1.969547	5.538940	20.075182	-0.001036	-13.848153

Table C.31 NiH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	2.6582 ^a	1.475 ^a	1927 ^a	
AE	3.249(2)	1.4502(7)	2039.3(5.7)	
UC	-0.014(3)	-0.0004(9)	2.6(7.7)	0.024(39)
BFD	-0.122(2)	-0.0045(8)	11.3(6.7)	0.130(34)
STU	-0.025(3)	0.001(1)	-10.9(8.1)	-0.001(40)
ccECPS	0.004(3)	-0.002(1)	6.9(7.9)	-0.022(40)
ccECP	0.031(3)	-0.001(1)	2.9(8.2)	-0.056(41)

^a Reference [FP06]

Table C.32 NiO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	3.92 ^a	1.627 ^a	838 ^a	
AE	4.23(1)	1.619(2)	863.7(7.3)	
UC	-0.00(1)	0.002(3)	-3(10)	0.02(18)
BFD	-0.04(2)	-0.012(4)	20(14)	-0.08(23)
STU	-0.01(2)	-0.004(4)	4(12)	-0.09(20)
ccECPS	0.06(2)	-0.004(4)	6(11)	-0.14(18)
ccECP	0.02(2)	0.000(4)	2(11)	0.01(19)

^a Reference [FP06]

C.2.9 Cu

In Table C.33, we show the total energy components for the Cu atom ground state using our ccECP. Table C.34, shows atomic excitation discrepancies for Cu atom. CuH and CuO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.35 and Table C.36 respectively.

Table C.33 Total energy components of the $[\text{Ar}] 3d^{10}4s^1 \ ^2S$ ground state of the Cu atom using our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-195.335042	-195.337118	-195.337366	-195.337400	-195.336990	-0.000410
Correlation	-0.957822	-1.008029	-1.030905	-1.060903		

Table C.34 Cu AE and relative errors for various core approximations. All values in eV

Gaps	AE	UC	BFD	STU	eCEPP	ccECPS	ccECP
$[\text{Ar}] 3d^{10}4s^2 \ ^1S$	-2.760729	0.004101	0.053157	0.039397	-0.000870	-0.016530	-0.006277
$[\text{Ar}] 3d^{10}4s^1 \ ^2S$	-1.513939	0.007503	0.047758	0.042300	0.004682	-0.020428	-0.007892
$[\text{Ar}] 3d^9 4s^1 \ ^3D$	9.062324	0.005137	-0.006833	0.004934	0.003888	-0.003649	-0.002273
$[\text{Ar}] 3d^{10} \ ^1S$	6.223451	0.013241	0.014972	0.047085	0.003103	-0.032589	-0.008363
$[\text{Ar}] 3d^9 \ ^2D$	26.646257	0.012720	-0.009051	0.024636	0.003493	-0.009803	0.010994
$[\text{Ar}] 3d^8 \ ^3F$	63.558190	-0.000543	-0.027268	0.020881	0.003248	0.010925	0.017731
$[\text{Ar}] 3d^7 \ ^4F$	121.008477	-0.025039	-0.028883	0.055165	0.013036	0.018442	-0.001066
$[\text{Ar}] 3d^6 \ ^5D$	201.489726	-0.062319	-0.024452	0.156973	0.047778	0.010412	-0.036229
$[\text{Ar}] 3d^5 \ ^6S$	305.535974	-0.122740	-0.058675	0.364107	0.118660	-0.008278	-0.054449
$[\text{Ar}] 3d^4 \ ^5D$	444.867210	-0.207748	0.024219	0.733521	0.331561	-0.002246	-0.067000
$[\text{Ar}] 3d^3 \ ^4F$	612.257887	-0.328280	-0.034713	1.317769	0.665239	0.022817	0.031193
$[\text{Ar}] 3d^2 \ ^3F$	811.134546	-0.476918	-0.377656	2.138293	1.124162	0.032874	0.255770
$[\text{Ar}] 3d^1 \ ^2D$	1043.247944	-0.657795	-1.166121	3.235735	1.709602	0.012528	0.655110
$[\text{Ar}] \ ^1S$	1308.271906	-0.893022	-2.520960	4.697743	2.455655	-0.013314	1.337623
$[\text{Ne}] 3s^2 \ ^1S$	4066.688500	-2.130124	8.869749	5.120704	15.171119	-0.025267	-13.809237

Table C.35 CuH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	2.749 ^a	1.463 ^a	1941 ^a	
	2.715(61) ^b	1.463 ^a		
	2.736(208) ^c	1.463 ^a		
AE	2.83(3)	1.47(1)	1937(52)	
UC	-0.00(4)	-0.00(2)	1(73)	0.00(40)
BFD	0.02(4)	-0.00(2)	4(74)	-0.05(40)
STU	0.01(4)	-0.00(2)	6(73)	-0.03(40)
eCEPP	0.02(4)	-0.00(2)	8(73)	-0.04(40)
ccECPS	0.00(4)	-0.00(2)	2(73)	-0.02(40)
ccECP	-0.01(4)	0.00(2)	1(73)	0.02(40)

^a Reference [FP06]

^b Reference [Xu15]

^c Reference [Fan17]

Table C.36 CuO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	2.888 ^a	1.724 ^a	640 ^a	
AE	2.769(7)	1.731(3)	651.5(5.8)	
UC	-0.01(1)	0.001(4)	-0.3(8.3)	0.01(13)
BFD	0.04(1)	-0.008(4)	1.6(8.6)	-0.20(12)
STU	-0.01(1)	-0.006(4)	4.1(9.1)	-0.07(13)
eCEPP	0.05(1)	-0.007(4)	6.0(8.6)	-0.14(13)
ccECPS	0.01(1)	-0.003(4)	-0.4(8.5)	-0.07(13)
ccECP	-0.02(1)	0.001(4)	-2.1(8.4)	0.02(13)

^a Reference [FP06]

Table C.37 Total energy components for the $[\text{Ar}] 3d^{10}4s^2 \ ^1S$ ground state of the Zn atom using our ccECP.

	T	Q	5	Extrap.	Numerical	Diffs
SCF	-225.274095	-225.274983	-225.275051	-225.275057	-225.274600	-0.000457
Correlation	-0.977277	-1.030767	-1.055331	-1.087707		

C.2.10 Zn

In Table C.37, we show the total energy components for the Zn atom ground state using our ccECP. Table C.38, shows atomic excitation discrepancies for Zn atom. ZnH and ZnO AE molecular parameters and corresponding discrepancies for various core approximations are shown in Table C.39 and Table C.40 respectively.

Table C.38 Zn AE gaps and relative errors for various ECPs. All values in eV

Gaps	AE	UC	BFD	STU	ccECPS	ccECP
$[\text{Ar}] 3d^{10}4s^1 \ ^2S$	9.411709	0.004952	0.018765	0.004463	0.007538	0.000215
$[\text{Ar}] 3d^94s^2 \ ^2D$	17.323005	-0.009362	-0.205955	-0.022585	0.009686	0.022019
$[\text{Ar}] 3d^{10} \ ^1S$	27.392261	0.011411	0.036333	0.006989	0.002052	-0.009078
$[\text{Ar}] 3d^9 \ ^2D$	67.262367	0.003722	-0.050548	0.003788	0.010425	0.013184
$[\text{Ar}] 3d^8 \ ^3F$	126.975193	-0.018903	-0.090639	0.051139	0.013474	0.028731
$[\text{Ar}] 3d^7 \ ^4F$	210.344966	-0.052708	-0.055643	0.173830	0.006980	0.031967
$[\text{Ar}] 3d^6 \ ^5D$	319.614848	-0.099359	0.033786	0.406930	-0.008892	0.026004
$[\text{Ar}] 3d^5 \ ^6S$	455.004148	-0.170049	0.098292	0.795761	-0.026917	0.025026
$[\text{Ar}] 3d^4 \ ^5D$	629.026722	-0.264676	0.351989	1.415417	-0.015099	0.072364
$[\text{Ar}] 3d^3 \ ^4F$	833.485035	-0.396401	0.369407	2.316968	0.018092	0.171157
$[\text{Ar}] 3d^2 \ ^3F$	1,071.906266	-0.556213	-0.076598	3.529685	0.034067	0.296417
$[\text{Ar}] 3d^1 \ ^2D$	1,346.006047	-0.748315	-2.414393	5.101447	0.007904	0.438073
$[\text{Ar}] \ ^1S$	1,655.288042	-0.996097	-3.118490	7.132298	-0.032440	0.638571
$[\text{Ne}] 3s^2 \ ^1S$	4,748.163279	-2.271911	7.708246	5.787312	-0.022816	-8.705975

Table C.39 ZnH AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.	0.951 ^a	1.595 ^a	1608 ^a	
	0.937(22) ^b	1.590 ^b		
	0.945(22) ^c	1.595 ^a		
AE	0.928(5)	1.580(3)	1614(35)	
UC	-0.002(7)	0.000(4)	-1(49)	0.002(86)
BFD	0.060(7)	-0.014(4)	56(50)	-0.113(81)
STU	0.006(7)	-0.002(4)	6(49)	-0.016(85)
ccECPS	0.031(7)	-0.006(4)	27(49)	-0.053(84)
ccECP	0.012(7)	-0.002(4)	9(49)	-0.020(85)

^a Reference [HH79]

^b Reference [Xu15]

^c Reference [Fan17]

Table C.40 ZnO AE molecular binding parameters and discrepancies for various core approximations. Labeling as in Table C.3

	D_e (eV)	r_e (Å)	ω_e (cm ⁻¹)	D_{diss} (eV)
Exp.			720 ^a	
	1.644(39) ^b	1.800 ^b		
	1.657(43) ^c	1.8 ^b		
AE	1.46(2)	1.697(9)	669(22)	
UC	-0.01(3)	0.00(1)	-2(31)	0.02(27)
BFD	0.05(3)	-0.01(1)	16(31)	-0.16(25)
STU	-0.02(3)	-0.00(1)	-0(31)	-0.03(26)
ccECPS	0.02(3)	-0.00(1)	6(31)	-0.06(26)
ccECP	0.00(3)	-0.00(1)	0(31)	-0.01(26)

^a Reference [Mor01]

^b Reference [Xu15]

^c Reference [Fan17]