Generating WFX Files: A Guide for Programmers

Thomas A. Manz

email: tmanz@nmsu.edu

June 14, 2020

CONTENTS:

1. The purpose of WFX files

The WFX file has a standardized format that facilitates the exchange of data between quantum chemistry programs and atomic population analysis programs. A standardized format means there is not a different set of rules for reading WFX files generated by different quantum chemistry programs — all WFX files are read according to exactly the same set of rules irrespective of the program that generated them.

WFX files are text files containing tags separating blocks of information. For example, the system's net charge (-1.0 in this example) is indicated by the following lines

```
<Net Charge>
-1.0
</Net Charge>
```

"The opening and closing tags of a data section MUST appear alone on separate lines before and after the corresponding data. Thus, unlike XML, single-line statements like <Net Charge>1.0</Net Charge> are NOT allowed."[1]
"There is no required order for the data sections." [1]

WFX files contain the system's geometry, its net charge, the spin-up and spin-down natural orbitals (i.e., the eigenvectors of the spin-up and spin-down first-order density matrices), the natural orbital occupancies, information that allows the number of electrons on each atom replaced by a relativistic effective core potential (RECP) to be computed, etc. For density functional theory (DFT) and Hartree-Fock calculations, the orbital energies are also listed.

The WFX file does NOT store the basis set that was used in the quantum chemistry calculation used to generate the natural orbitals. Rather, the natural orbitals are expressed as coefficients of unnormalized Cartesian Gaussian primitives.

What is the advantage of WFX files compared to using cube files? Cube files can store the total electron density, the valence electron density, the core electron density, or the spin electron density for a set of uniformly spaced grid points. There are four reasons why WFX files are more convenient than cube files to use as input for atomic population analysis. First, cube files usually have much larger file sizes than WFX files. Second, it is often more cumbersome to generate cube files than to print WFX files. Third, a WFX file is designed to include all required information to perform atomic population analysis, while several cube files (e.g., valence density cube file and spin density cube file) may be required along with additional information specified in yet another file. Fourth, **WFX files provide access to the natural orbitals which allows bond order components and s-p-d-f-g populations to be computed; these cannot be computed from cube files containing valence and/or spin densities. (Bond orders but not bond order components can be computed from the electron and spin density distributions.) For density functional theory (DFT) or Hartree-Fock calculations, the WFX files also allow projected density of states plots to be prepared.**

What is the advantage of WFX files compared to using molden files? Unfortunately, molden files written by different quantum chemistry programs may not follow a standardized format. This means there may be different rules for reading a molden file generated by one quantum chemistry program compared to the rules for reading a molden file generated by another quantum chemistry program. Because molden files do not follow a standardized format, Chargemol is not set up to read them. For similar reasons, Chargemol is also not set up to read fchk files.

2. Summary of WFX tags used by Chargemol

Please note these tags are comprised without leading or trailing spaces. For example, <keywords> is used instead of < keywords >. Capitalization is not important. For example, <Keywords> and <keywords> are both fine.

(a) keywords – GTO stands for Gaussian type orbitals and is currently the only kind read by Chargemol.

```
<Keywords>
GTO
</Keywords>
```

(b) net charge – The net charge of the unit cell. This is typically a decimalized integer (e.g., -2.0, -1.0, 0.0, 1.0, 2.0, etc.) but some programs write it as a true integer (e.g., -2, -1, 0, 1, 2).

```
<Net Charge>
0.0
</Net Charge>
```

(c) number of nuclei – The number of atomic nuclei including Ghost atoms, but excluding dummy atoms.

```
<Number of Nuclei>
2
</Number of Nuclei>
```

(d) number of primitives – The number of Cartesian Gaussian primitives.

```
<Number of Primitives>
50
</Number of Primitives>
```

(e) number of occupied orbitals – The number of fully or partially occupied molecular orbitals. Traditionally, empty (completely unoccupied) molecular orbitals are not included in the WFX file. For periodic materials, this is the total number of reported orbitals, which includes the total combinations of bands and kpoints. For example, if orbitals are reported for 8 kpoints and 7 bands, then the number of occupied orbitals = 8 × 7 = 56.

```
<Number of Occupied Orbitals>
6
</Number of Occupied Orbitals>
```

This keyword can also be equivalently indicated by

```
<Number of Occupied Molecular Orbitals>
6
</Number of Occupied Molecular Orbitals>
```

(f) number of translation vectors – The number of periodic dimensions (0, 1, 2, or 3). For nonperiodic materials, this can either be specified as 0 or this section can be omitted.

```
<Number of Translation Vectors>
2
</Number of Translation Vectors>
```

(g) translation vectors – The lattice vectors for periodic systems; omit this section for non-periodic materials. One lattice vector is specified per line as X, Y, Z coordinates in bohrs. The number of lattice vectors must match the number of translation vectors specified in (f) above.

```
<Translation Vectors>
  4.75034738E+00 -6.93889390E-18  0.00000000E+00
  2.37517655E+00  4.11420313E+00  0.00000000E+00
</Translation Vectors>
```

(h) atomic numbers – The atomic number for each atom; one per line.

```
<Atomic Numbers>
 5
 7
</Atomic Numbers>
```

(i) nuclear charges – The effective nuclear charge for each atom, which **equals the atomic number minus core electrons replaced by a relativistic effective core potential (RECP).** For example, a Zr atom with 18 core electrons replaced by a RECP has nuclear charge = 40 - 18 = 22. **The nuclear charges are zero for ghost atoms.**

```
<Nuclear Charges>
  5.000000000000e+00
  7.000000000000e+00
</Nuclear Charges>
```

(j) nuclear cartesian coordinates – Regular atoms and ghost atoms are included in this list. X, Y, Z coordinates in bohrs for one atom per line.

```
<Nuclear Cartesian Coordinates>
 -1.187587226830e+00 -6.856894978828e-01  0.000000000000e+00
  1.187587226830e+00  6.856894978828e-01  0.000000000000e+00
</Nuclear Cartesian Coordinates>
```

(k) primitive centers – The number of entries in this section equals the number of primitives in section (d) above. The number here indicated on which center (i.e., atom or ghost atom) the Cartesian Gaussian primitive resides. In the below example, the first 25 Cartesian Gaussian primitives reside on center # 1 and the second 25 Cartesian Gaussian primitives reside on center # 2. Note that the locations of these centers are specified in section (j) above.

```
<Primitive Centers>
                 1                  1                  1                  1                  1
                 1                  1                  1                  1                  1
                 1                  1                  1                  1                  1
                 1                  1                  1                  1                  1
                 1                  1                  1                  1                  1
```

```
                2           2           2           2           2
                2           2           2           2           2
                2           2           2           2           2
                2           2           2           2           2
                2           2           2           2           2
</Primitive Centers>
```

(l) primitive exponents – Exponents α of the Cartesian Gaussian primitives in units of bohr $^{-2}$.

```
<Primitive Exponents>
 2.788000000000e+03  4.190000000000e+02  9.647000000000e+01  2.807000000000e+01  9.376000000000e+00
 3.406000000000e+00  1.306000000000e+00  3.406000000000e+00  3.245000000000e-01  1.022000000000e-01
 1.134000000000e+01  2.436000000000e+00  6.836000000000e-01  2.134000000000e-01  1.134000000000e+01
 2.436000000000e+00  6.836000000000e-01  2.134000000000e-01  1.134000000000e+01  2.436000000000e+00
 6.836000000000e-01  2.134000000000e-01  7.011000000000e-02  7.011000000000e-02  7.011000000000e-02
 5.909000000000e+03  8.875000000000e+02  2.047000000000e+02  5.984000000000e+01  2.000000000000e+01
 7.193000000000e+00  2.686000000000e+00  7.193000000000e+00  7.000000000000e-01  2.133000000000e-01
 2.679000000000e+01  5.956000000000e+00  1.707000000000e+00  5.314000000000e-01  2.679000000000e+01
 5.956000000000e+00  1.707000000000e+00  5.314000000000e-01  2.679000000000e+01  5.956000000000e+00
 1.707000000000e+00  5.314000000000e-01  1.654000000000e-01  1.654000000000e-01  1.654000000000e-01
</Primitive Exponents>
```

(m) primitive types – Each unnormalized Cartesian Gaussian primitive has the form

$(X-X_{center})^a(Y-Y_{center})^b(Z-Z_{center})^c\exp(-\alpha(\text{distance}^2))$ where distance is the distance from (X,Y,Z) to $(X_{center}, Y_{center}, Z_{center})$. The primitive type is defined by the powers a, b, c. As explained in Part 5 below, each (a, b, c) combination is assigned an integer code. For example, (a = 0, b = 0, c = 0) is assigned the code 1. The number of entries in this section must equal the number of primitives in section (d) above.

```
<Primitive Types>
                1           1           1           1           1
                1           1           1           1           1
                2           2           2           2           3
                3           3           3           4           4
                4           4           2           3           4
                1           1           1           1           1
                1           1           1           1           1
                2           2           2           2           3
                3           3           3           4           4
                4           4           2           3           4
</Primitive Types>
```

(n) orbital occupation numbers – The occupancy for each orbital. The number of lines in this section must equal the number of orbitals. For an idempotent first-order density matrix, the occupancy will be either 0.0, 1.0, or 2.0. (However, molecular orbitals with an occupancy of 0.0 are traditionally not included in the WFX file.) For a non-idempotent first-order density matrix the occupancy can be non-integer (e.g., 1.9723476, 0.02387497234, etc.). These should preferably be non-negative values.

```
<Orbital Occupation Numbers>
 2.000000000000e+00
 2.000000000000e+00
 2.000000000000e+00
```

```
  2.000000000000e+00
  2.000000000000e+00
  2.000000000000e+00
</Orbital Occupation Numbers>
```

This keyword can also be equivalently indicated by < Molecular Orbital Occupation Numbers>, </Molecular Orbital Occupation Numbers>.

(o) molecular orbital spin types – The allowed values are “Alpha” (for a spin-up natural orbital), “Beta” (for a spin-down natural orbital), or “Alpha and Beta” (for a spin-unpolarized natural orbital).

```
<Orbital Spin Types>
 Alpha and Beta
 Alpha and Beta
 Alpha and Beta
 Alpha and Beta
 Alpha and Beta
 Alpha and Beta
</Orbital Spin Types>
```

This keyword can also be equivalently indicated by <Molecular Orbital Spin Types>, </Molecular Orbital Spin Types>.

(p) orbital primitive coefficients – Cartesian Gaussian primitive coefficients for the normalized orbitals. The keyword “<Orbital Primitive Coefficients>, </Orbital Primitive Coefficients>” can also be equivalently indicated by “<Molecular Orbital Primitive Coefficients>, </Molecular Orbital Primitive Coefficients>”. The keyword “<Orbital Number>, </Orbital Number>” can also be equivalently indicated by “<MO Number>, </ MO Number>”. The number of orbitals presented in this section must equal the “<Number of Occupied Orbitals>” specified in section e above.

```
<Orbital Primitive Coefficients>
<Orbital Number>
 1
</Orbital Number>
  1.102988263573e-04  2.032920206924e-04  3.273590383081e-04  4.356868904548e-04
  4.339887162675e-04  2.426275865638e-04  3.819664053834e-05 -6.062683112437e-04
  6.160370363144e-04  5.372140446504e-04 -2.982819748005e-09 -2.675895843322e-09
 -1.897408270643e-09 -7.488165938694e-10  1.946393375059e-07  1.746114878471e-07
  1.238124727528e-07  4.886288078307e-08 -1.373233139373e-15 -1.231931246271e-15
 -8.735304632174e-16 -3.447408321277e-16  4.826826278356e-08 -3.401113658002e-06
 -2.026947122266e-16  5.704301497920e-01  1.052300930223e+00  1.699182087393e+00
  2.302482349995e+00  2.399636946984e+00  1.394941912040e+00  2.172800474759e-01
 -6.448604759851e-03  7.412466710769e-03 -2.582152332310e-03  3.638559467697e-09
  3.543883326144e-09  2.489775670296e-09  9.453474240244e-10 -3.973394729781e-07
 -3.870006098863e-07 -2.718895105198e-07 -1.032342196350e-07 -2.426669950728e-14
 -2.363527448923e-14 -1.660509840996e-14 -6.304819825660e-15 -1.526410020413e-09
  1.648051005897e-07  1.777920275393e-15
<Orbital Number>
 2
</Orbital Number>
  3.511385448134e-01  6.471842600288e-01  1.042154120214e+00  1.387018028759e+00
  1.381611856884e+00  7.724098296533e-01  1.215997777886e-01 -7.643989255480e-03
  7.767155893247e-03  4.429327870471e-05  3.089376488087e-09  2.771488189475e-09
  1.965190321523e-09  7.755669381419e-10 -2.140093193063e-07 -1.919883520776e-07
 -1.361339560388e-07 -5.372558286382e-08 -7.654426683612e-14 -6.866807342075e-14
 -4.869074809577e-14 -1.921591715718e-14 -1.380330647512e-07  1.245751812959e-05
 -7.779994212658e-16 -9.582997202106e-05 -1.767823260004e-04 -2.854557599260e-04
 -3.868077787603e-04 -4.031293604906e-04 -2.343446335200e-04 -3.650217450452e-05
 -1.709649784959e-03  1.965188221332e-03 -3.259081539872e-03  4.485982552702e-09
  4.369256270521e-09  3.069646192745e-09  1.165519510692e-09 -4.346452527851e-07
 -4.233356848522e-07 -2.974169270014e-07 -1.129267705345e-07 -1.291643049430e-15
 -1.258034204702e-15 -8.838391862848e-16 -3.355864980023e-16 -5.859669594882e-09
  5.344506893548e-07  1.569897259970e-14
<Orbital Number>
 3
</Orbital Number>
 -5.435581756154e-02 -1.001833324380e-01 -1.613241840466e-01 -2.147086955830e-01
 -2.138718267843e-01 -1.195680975602e-01 -1.882349697757e-02 -7.898910459551e-02
  8.026184610627e-02 -6.240863367486e-05 -1.255894366519e-07 -1.126666308706e-07
 -7.988898288884e-08 -3.152837319204e-08  1.023087616986e-05  9.178147299977e-06
  6.507985966503e-06  2.568391821510e-06 -1.960275073436e-12 -1.758568188467e-12
 -1.246955046332e-12 -4.921137137164e-13  9.484234931103e-07  5.711249709090e-06
 -9.668388059895e-14 -1.146810340204e-01 -2.115578196954e-01 -3.416087996789e-01
 -4.628981424059e-01 -4.824304017836e-01 -2.804434178828e-01 -4.368264988381e-02
 -1.976500965241e-01  2.271925192232e-01  8.536387141201e-02  2.509865395739e-07
  2.444558129609e-07  1.717438413061e-07  6.520972949802e-08 -3.438233594225e-05
 -3.348769977278e-05 -2.352697661725e-05 -8.933000272088e-06  2.851467798672e-12
  2.777271960639e-12  1.951188433994e-12  7.408502628843e-13 -9.403791243440e-09
  4.655177834556e-06  7.436915463021e-13
<Orbital Number>
 4
</Orbital Number>
  4.782279931856e-13  8.814231147976e-13  1.419346525357e-12  1.889028869668e-12
  1.881666013140e-12  1.051972243458e-12  1.656110346262e-13  3.273356499560e-12
 -3.326099681267e-12 -4.029512431410e-12 -1.440310217079e-11 -1.292106278147e-11
 -9.161990160508e-12 -3.615800758964e-12 -1.428858660996e-11 -1.281833055521e-11
 -9.089145406016e-12 -3.587052406917e-12  2.292260986724e-01  2.056393669207e-01
  1.458135362552e-01  5.754565174402e-02 -3.349996010673e-10  8.650251589000e-11
  2.541068341569e-03 -1.183178129029e-12 -2.182667669740e-12 -3.524419394336e-12
 -4.775776245315e-12 -4.977292932914e-12 -2.893368736195e-12 -4.506791938356e-13
  3.814449017199e-12 -4.384588203630e-12  1.187118788766e-11  5.600765444485e-11
  5.455032258940e-11  3.832464375673e-11  1.455155325209e-11  2.232197620558e-11
  2.174115332836e-11  1.527437266396e-11  5.799554162142e-12  7.761396563940e-01
  7.559443267243e-01  5.310930466846e-01  2.016516787397e-01 -3.568300161137e-13
  2.191033500532e-12  1.407904506559e-02
<Orbital Number>
 5
</Orbital Number>
  2.962583932332e-06  5.460345263544e-06  8.792737502513e-06  1.170238182778e-05
  1.165676952411e-05  6.516883390635e-06  1.025947031942e-06  4.540554093341e-06
 -4.613715470546e-06  1.272791270819e-05 -2.439268008695e-02 -2.188274075086e-02
 -1.551648334469e-02 -6.123616296384e-03 -2.943529604520e-01 -2.640648546966e-01
 -1.872415327891e-01 -7.389530707931e-02  2.693196997738e-11  2.416074473261e-11
 1.713175682640e-11  6.761087738588e-12  5.041761310350e-02  6.080577396727e-01
 1.257920879816e-12  2.605488958847e-06  4.806475352117e-06  7.761160887775e-06
 1.051678692481e-05  1.096054893467e-05  6.371517619428e-06  9.924453763228e-07
 3.913776066128e-06 -4.498761497092e-06 -1.777378838882e-05  9.808539972402e-02
 9.553319540499e-02  6.711739742397e-02  2.548392592913e-02  1.183779045205e+00
 1.152976846300e+00  8.100305332161e-01  3.075619571244e-01 -1.214382595278e-11
-1.182784085068e-11 -8.309717807271e-12 -3.155131772389e-12  4.839631142426e-03
 5.841973192997e-02 -1.328536253699e-11
<Orbital Number>
 6
</Orbital Number>
-2.696452441569e-07 -4.969837700465e-07 -8.002878246919e-07 -1.065114669236e-06
-1.060963178159e-06 -5.931466088885e-07 -9.337853179481e-08 -3.912753849560e-07
 3.975799560373e-07 -1.102408419955e-06 -2.943933271010e-01 -2.641010677290e-01
-1.872672105102e-01 -7.390544085176e-02  2.438925116911e-02  2.187966465919e-02
 1.551430216795e-02  6.122755489901e-03  1.370621242593e-11  1.229588106447e-11
 8.718690036011e-12  3.440851332199e-12  6.084714227662e-01 -5.038072025627e-02
-4.885228469544e-13 -2.185849352760e-07 -4.032345253972e-07 -6.511149642607e-07
-8.822955021423e-07 -9.195244797874e-07 -5.345322081385e-07 -8.326022937535e-08
-3.765275763735e-07  4.328065107869e-07  1.494499005212e-06  1.183789171913e+00
 1.152986709509e+00  8.100374626703e-01  3.075645881814e-01 -9.808492367419e-02
-9.553273174209e-02 -6.711707167492e-02 -2.548380224488e-02 -5.316370663127e-11
-5.178037494213e-11 -3.637860106134e-11 -1.381265678400e-11  5.840865862847e-02
-4.840454513515e-03 -5.594129630292e-12
</Orbital Primitive Coefficients>
```

In cases where the orbital primitive coefficients are complex-valued rather than real-valued, this should be indicated using the <Type of Orbital Primitive Coefficients> tag described in section t below. In such case, the complex-valued orbital primitive coefficients can be written using the following number format, where the first number in parentheses is the real component and the second number in parentheses is the imaginary component of each complex coefficient:

```
(-3.81038447719695283E-006,1.59125299984949267E-005) (1.19515037610326272E-008,2.21217663248210112E-008) (5.49037981401733665E-009,2.95365788457435277E-008) (-4.92450498616101652E-009,4.96072793785747650E-009)
 (-1.91310429506973068E-006,-3.98265150397573701E-006) (3.14721851085448147E-011,2.57148944883954202E-010) (1.84015389104044636E-007,7.61947568334268565E-007) (7.16543599784785718E-006,1.25616440864185361E-006)
 (1.77613328166988703E-008,-1.53883578169081806E-008) (-4.04038302470644678E-003,-8.94470585299387573E-009) (-4.39278837862110388E-009,1.44335033361548073E-009) (4.86679189642241208E-005,2.02768118530349843E-005)
 (-6.06141449186623632E-002,4.75645548046981850E-002) (1.24047969404551939E-006,-1.50388474854887191E-006)  (-6.08354386068579289E-003,2.80506105026706804E-003)  (7.28887969404789939E-004,1.50312349244887191E+013)
```

example using unformatted Fortran read statement:

The array should be declared using one of the following according to whether the orbital primitive coefficients are real-valued or complex-valued:

```
REAL(kind=dp), ALLOCATABLE, DIMENSION(:,:) :: orbital_coefficients
```

```
COMPLEX(kind=dp), ALLOCATABLE, DIMENSION(:,:) :: orbital_coefficients
```

(Note: kind=dp was previously defined as the kind parameter associated with double precision.)

The array was first allocated using

```
ALLOCATE orbital_coefficients(norbitals,nprimitives)
```

Then read using

! Note that line_of_text is a line read from WFX file that has been converted into lowercase characters

```
IF (line_of_text == '<orbital primitive coefficients>') THEN
   DO i=1,norbitals
```

```
      READ(input_FID,'(a)')text ! Reads the line <Orbital Number> or <MO Number>
      READ(input_FID,*)text  ! Reads the orbital number
      READ(input_FID,*)text  ! Reads the line </Orbital Number> or </MO Number>
      READ(input_FID,*)(orbital_coefficients(i,j),j=1,nprimitives)
   END DO
END IF
```

(q) orbital energies – This section is meaningful only for canonical natural orbitals (e.g., produced by a DFT or Hartree-Fock calculation). For natural orbitals produced by a multi-electronic correlated wavefunction (e.g., CCSD, CAS-SCF, SAC-CI, CISD, MP2, etc.) or for localized natural orbitals produced by a DFT or Hartree-Fock calculation, set each orbital energy equal to zero or skip this section entirely.

```
<Orbital Energies>
 -1.399356731433e+01
 -6.560443027570e+00
 -8.706021450601e-01
 -4.110596113746e-01
 -2.744047464135e-01
 -2.744034373012e-01
</Orbital Energies>
```

This keyword can also be equivalently indicated by <Molecular Orbital Energies>,
</Molecular Orbital Energies>.

(r) number of kpoints – This section is only included when the system has 1, 2, or 3 periodic boundary conditions. This is the number of sampled k-points located in the **first Brillouin zone**. ***This should not be confused with the irreducible Brillouin zone***. The irreducible Brillouin zone reduces the number of sampled k-points by the symmetry operations of the crystal's point group. ***If the quantum chemistry calculation was performed using k-points sampled from the irreducible Brillouin zone, then these should be used to regenerate the full set of equivalent k-points located in the first Brillouin zone for purposes of writing the WFX file***. In other words, the WFX file contains the full non-reduced set of k-points.

```
<Number of Kpoints>
8
</Number of Kpoints>
```

(s) kpoint weights – This section is only included when the system has 1, 2, or 3 periodic boundary conditions. As described in Part 8 below, these weights describe the relative contribution of the first-order density matrix at each k-point to the overall first-order density matrix. **These weights must be non-negative and sum to 1.0. Also, a weight of 0.00 makes no sense, because that k-point could simply be omitted from the WFX file.** The number of lines in this section must equal the <Number of Kpoints> listed in section r above.

```
<Kpoint Weights>
0.0625
0.1375
0.2
0.1
```

```
0.0625
0.1375
0.2
0.1
</Kpoint Weights>
```

(s) kpoint index for orbitals – This section is only included when the system has 1, 2, or 3 periodic boundary conditions. This section lists the kpoint index for each orbital. The number of lines in this section equals the <Number of Occupied Orbitals> in section e above, which equals the number of kpoints times the number of bands. On each line, the corresponding index of the kpoint associated with that orbital is listed. The index indicates the order in which the kpoint is listed in sections s and t. For example, the kpoint having index value = 3 is the third kpoint listed in sections s and t. In the example below, the fourth orbital and the fifteenth orbital are associated with the fifth kpoint.

```
<Kpoint Index for Orbitals>
1
2
3
5
4
8
7
6
2
1
3
7
8
6
5
4
... etc.
</Kpoint Index for Orbitals>
```

(t) kpoint fractional coordinates – This section is only included when the system has 1, 2, or 3 periodic boundary conditions. The number of lines in this section must match the <Number of Kpoints> in section r above. These fractional coordinates indicate the position of each k-point as a fraction of the corresponding reciprocal lattice vectors. For each k-point, the number of fractional coordinate components must equal the <Number of Translation Vectors> listed in section f above. For example, when Number of Translation Vectors = 2, then 0.25 -0.37 denotes a k-point whose value equals 0.25 times the first reciprocal lattice vector plus -0.37 times the second reciprocal lattice vector. These k-points should belong to the first Brillouin zone.

Example for 1-dimensional periodic boundary conditions

```
<Kpoint Fractional Coordinates>
0.1
0.0
0.25
0.37
-0.27
... etc.
</Kpoint Fractional Coordinates>
```

Example for 2-dimensional periodic boundary conditions

```
<Kpoint Fractional Coordinates>
0.1 0.0
0.0 0.4
0.25 0.25
0.37 -0.27
-0.27 0.0
... etc.
</Kpoint Fractional Coordinates>
```

Example for 3-dimensional periodic boundary conditions

```
<Kpoint Fractional Coordinates>
0.1 0.0 -0.3
0.0 0.4 -0.2
0.25 0.25 -0.25
0.37 -0.27 0.0
-0.27 0.0 0.31
... etc.
</Kpoint Fractional Coordinates>
```

(t) type of orbital primitive coefficients – Whether the <Orbital Primitive Coefficients> listed in section p above are real-valued or complex-valued. If this tag is missing, they are assumed to be real-valued.

```
<Type of Orbital Primitive Coefficients>
real
<Type of Orbital Primitive Coefficients>

<Type of Orbital Primitive Coefficients>
complex
</Type of Orbital Primitive Coefficients>
```

(u) orbital type – This optional tag is mainly for informational purposes. It allows any program reading the WFX file to read and print the orbital type information for the user’s benefit. For example, a particular method (e.g., Pipek-Mezey, Intrinsic Bond Orbial, Foster-Boys, etc.) by which localized natural orbitals were produced. The tag could also indicate whether canonical, natural, or canonical natural orbitals were used. For correlated wavefunction methods (e.g., CCSD, CISD, SAC-CI, CAS-SCF, MP2, etc.) the natural orbitals are preferred, because they yield the correct first-order electron and spin density matrices while the canonical orbitals do not. For DFT and Hartree-Fock methods, either canonical natural or localized natural orbitals may be used, because both of these yield the correct first-order electron and spin density matrices. (Note that for DFT and Hartree-Fock calculations, the canonical orbitals are also natural orbitals, so these are indicated as “canonical natural” orbitals. For DFT and Hartree-Fock calculations in which each spin-orbital is either fully occupied or fully unoccupied, the localized orbitals are also natural orbitals, so these are indicated as “localized natural” orbitals.)

```
<Orbital Type>
natural
</Orbital Type>

<Orbital Type>
localized natural: Pipek-Mezey
</Orbital Type>

<Orbital Type>
localized natural: Foster-Boys
</Orbital Type>

<Orbital Type>
localized natural: Intrinsic Bond Orbitals
</Orbital Type>

<Orbital Type>
canonical
</Orbital Type>

<Orbital Type>
canonical natural
</Orbital Type>
```

3. Additional WFX tags (currently not used by Chargemol)

The following tags are currently not read by the Chargemol program. Their presence or absence does not effect the Chargemol calculation in any way. These tags may be used by some other programs that read WFX files. Please refer to http://aim.tkgristmill.com/wfxformat.html for more information about these tags.

```
<Title>
 BN_plane
</Title>

<Number of Perturbations>
 0
</Number of Perturbations>

<Number of Electrons>
 12
</Number of Electrons>
```

```
<Number of Alpha Electrons>
 6
</Number of Alpha Electrons>

<Number of Beta Electrons>
 6
</Number of Beta Electrons>

<Electronic Spin Multiplicity>
 1
</Electronic Spin Multiplicity>

<Number of Core Electrons>
 0
</Number of Core Electrons>

<Nuclear Names>
 B1
 N2
</Nuclear Names>

<Energy = T + Vne + Vee + Vnn>
 -7.969671098519e+01
</Energy = T + Vne + Vee + Vnn>

<Virial Ratio (-V/T)>
 1.986181658648e+00
</Virial Ratio (-V/T)>

<Model>
Restricted HF
</Model>

<Nuclear Cartesian Energy Gradients>
O1          6.09070231000000E-016 -5.55187875000000E-016 -2.29270172000000E-004
H2         -2.46849911000000E-016 -1.18355659000000E-004 1.14635086000000E-004
H3         -3.62220320000000E-016 1.18355659000000E-004 1.14635086000000E-004
</Nuclear Cartesian Energy Gradients>

<Nuclear Virial of Energy-Gradient-Based Forces on Nuclei, W>
6.14477996016389E-004
</Nuclear Virial of Energy-Gradient-Based Forces on Nuclei, W>

<Full Virial Ratio, -(V - W)/T>
2.00600662884992E+000
</Full Virial Ratio, -(V - W)/T>
```

Note: Chargemol does not require EDF functions for the relativistic effective core potential (RECP) electrons to be printed in the WFX file, because a complete set of core reference densities are internally stored in the Chargemol program. Chargemol uses these to re-insert the core electrons that were replaced by RECPs during the quantum chemistry calculation. In such manner, Chargemol always performs an effective all-electron partitioning irrespective of whether the quantum chemistry calculation of the wavefunction was performed using pseudopotentials (RECPs), frozen-core electrons, or relaxed all-electron calculations.

```
<Additional Electron Density Function (EDF)>
<Number of EDF Primitives>
49
</Number of EDF Primitives>
<EDF Primitive Centers>
4 4 4 4 4
4 4 4 4 4
4 4 4 4 4
4 4 4 4 4
4 4 4 4 4
4 5 5 5 5
5 5 5 5 5
5 5 5 5 5
5 5 5 5 5
5 5 5 5
</EDF Primitive Centers>
<EDF Primitive Types>
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1
</EDF Primitive Types>
<EDF Primitive Exponents>
4.459566346600e+007 2.277438917020e+007 1.163056588380e+007 5.939569301590e+006 3.033255977470e+006
1.549041918310e+006 7.910743051340e+005 4.039907176480e+005 2.063124777110e+005 1.053609318230e+005
5.380637214860e+004 2.747817083330e+004 1.403272218880e+004 7.166317336820e+003 3.659739249530e+003
1.868978269460e+003 9.544613792240e+002 4.874302389260e+002 2.489238883750e+002 1.271219905030e+002
6.491944415170e+001 3.315346316010e+001 1.693101556660e+001 4.415615257680e+000 2.254993703160e+000
1.151594127780e+000 5.939569301590e+006 3.033255977470e+006 1.549041918310e+006 7.910743051340e+005
4.039907176480e+005 2.063124777110e+005 1.053609318230e+005 5.380637214860e+004 2.747817083330e+004
1.403272218880e+004 7.166317336820e+003 3.659739249530e+003 1.868978269460e+003 9.544613792240e+002
4.874302389260e+002 2.489238883750e+002 1.271219905030e+002 6.491944415170e+001 3.315346316010e+001
1.693101556660e+001 8.646435720220e+000 4.415615257680e+000 2.254993703160e+000
</EDF Primitive Exponents>
<EDF Primitive Coefficients>
7.487900912270e+002 1.995198887630e+004 -1.051775310550e+004 1.511945457420e+004 -1.530581814380e+003
7.473773106890e+003 2.557337101310e+003 4.879670599210e+003 3.593392191520e+003 7.566316297390e+003
2.045060403160e+003 8.346936656130e+003 4.464742448850e+003 4.802771166360e+003 6.825921849576e+003
4.521493682700e+003 3.134312247727e+003 1.372563088957e+003 1.722377326180e+002 2.859393067010e+000
3.262155912880e+002 1.533389423812e+002 -2.112187740940e-001 2.679767784061e+001 1.855573804510e-002
4.098181567201e-001 1.871504693790e+002 4.702293155340e+002 -1.570032436790e+002 3.338771301530e+002
1.064310876220e+002 1.671641111610e+002 1.928741244180e+002 1.664547652900e+002 2.221025750370e+002
```

```
6.990568432720e+002 2.866893784760e+002 3.345152960450e+002 4.155348021290e+002 5.999669639920e+002
4.836882984020e+002 3.760977922740e+002 1.786090102550e+002 3.350395287220e+001 1.205528578710e+000
7.956941591424e+000 2.584199626783e+001 1.898716467978e+000 3.283109046630e-001
</EDF Primitive Coefficients>
</Additional Electron Density Function (EDF)>
```

4. Correlated wavefunction versus density functional theory (DFT) calculations

Orbitals stored within a WFX file should be the natural orbitals, which are the eigenvectors of the first-order density matrix. For spin-polarized quantum chemistry calculations, the spin-up natural orbitals and the spin-down natural orbitals should be separately computed as the eigenvectors of the spin-up first-order density matrix and the spin-down first-order density matrix, respectively. For spin-unpolarized quantum chemistry calculations, the spin-up and spin-down natural orbitals are equal.

The orbital occupancies are the corresponding eigenvalues of the first-order density matrix. For an N-representable wavefunction, the spin-up natural orbital occupancies are between 0 and 1, and the spin-down natural orbital occupancies are between 0 and 1.[2] Hence, in a spin-unpolarized calculation, the combined spin-up plus spin-down natural orbital occupancy would be between 0 and 2 for an N-representable wavefunction.

The electron density can be regenerated from the natural orbitals by

$$\rho(\vec{r}) = \sum_i f_i \phi_i^*(\vec{r}) \phi_i(\vec{r}) \tag{1}$$

where $\rho(\vec{r})$ is the electron density, $\phi_i(\vec{r})$ is a natural orbital, $\phi_i^*(\vec{r})$ is the complex conjugate, and f_i is the corresponding natural orbital occupation. The spin density can be regenerated from the natural orbitals by

$$\rho^{\mathrm{spin}}(\vec{r}) = \sum_{i \in \mathrm{alpha}} f_i^\alpha \phi_i^*(\vec{r}) \phi_i(\vec{r}) - \sum_{j \in \mathrm{beta}} f_j^\beta \phi_j^*(\vec{r}) \phi_j(\vec{r}) \tag{2}$$

where $\rho(\vec{r})$ is the electron density, $\phi_i(\vec{r})$ is a natural orbital, $\phi_i^*(\vec{r})$ is the complex conjugate, f_i^α is the spin-up occupation of the natural orbital, and f_j^β is the spin-down occupation of the natural orbital.

It is important to understand the distinctions between natural orbitals, canonical orbitals, and localized orbitals. Spin-up natural orbitals and spin-down natural orbitals are defined as the eigenvectors of the spin-up first-order density matrix and spin-down first-order density matrix, respectively. Canonical orbitals are the orbitals that (approximately) solve the multi-electronic Schrodinger equation for a particular exchange-correlation theory; each canonical orbital has an associated orbital energy ε_i. Localized orbitals are a unitary transformation of the natural orbitals that maximizes some measure of orbital localization. Localized orbitals are also natural orbitals if and only if the first-order density matrix is idempotent. (A first-order density matrix is idempotent if and only if each natural orbital is either fully occupied or unoccupied.)

For correlated wavefunction calculations (e.g., MP2, CCSD, CISD, CAS-SCF, SAC-CI, etc.) of multi-electronic systems, the natural orbitals are not the same as canonical orbitals which are not the same as localized orbitals. **It is improper to write canonical instead of natural orbitals to the WFX file for correlated wavefunction calculations, because the canonical**

orbitals are not related to the correlated electron density via eqn (1). For most DFT methods, the canonical and natural orbitals are equal.

The <molecular orbital energies> have a clear meaning only for canonical orbitals. Hence <molecular orbital energies> should be written to the WFX file only for those calculations in which the natural orbitals and the canonical orbitals coincide: (a) Hartree-Fock calculations and (b) DFT calculations. If it is desired to also write a <molecular orbital energies> section for correlated wavefunction calculations (e.g., MP2, CCSD, CISD, CAS-SCF, SAC-CI, etc.) of multi-electronic systems, then in this case each natural orbital energy should be printed as 0.0.

When the first-order density matrix is idempotent (e.g., DFT or Hartree-Fock calculations), three important consequences result:

(1) Any unitary linear combination of the natural orbitals is also a set of natural orbitals.

(2) The canonical orbitals are usually natural orbitals.

(3) Localized orbitals are also natural orbitals.

(Please note that all three of the above are not true when the first-order density matrix is not idempotent.) Since the WFX file contains natural orbitals, it follows from (2) and (3) above that either canonical natural orbitals or localized natural orbitals can be written to the WFX file when the first-order density matrix is idempotent. When writing localized natural orbitals to the WFX file (which can only be done if the first-order density matrix is idempotent), then the <molecular orbital energies> section should be omitted from the WFX file or else each orbital energy should be printed as 0.0.

Note on the electron density for double-hybrid functionals: Double-hybrid functionals combine some aspects of correlated wavefunction theory with DFT. Many double-hybrid functionals incorporate a MP2-like energy term based on perturbation theory. There are two potential ways to treat the electron density for double-hybrid functionals: (a) use a Kohn-Sham like ansatz that neglects the MP2-like contribution to the electron density or (b) include MP2-like perturbative corrections to the electron density. For (a), the canonical and natural orbitals would be equal and the electron density matrix would be idempotent (similar to DFT). For (b), the canonical and natural orbitals would be unequal and the electron density matrix would not be idempotent (similar to correlated wavefunction theories). Therefore, depending on how the electron density distribution is computed, double-hybrid functionals can potential behave either like a DFT functional or else like a correlated wavefunction theory.

5. Description of Cartesian primitives

Each unnormalized Cartesian Gaussian primitive has the form $(X-X_{center})^{a}(Y-Y_{center})^{b}(Z-Z_{center})^{c}\exp(-\alpha(\text{distance}^{2}))$ where distance is the distance from (X,Y,Z) to $(X_{center}, Y_{center}, Z_{center})$. The primitive type is defined by the powers a, b, c. The following codes taken from http://aim.tkgristmill.com/wfxformat.html define Cartesian Gaussian primitive types:

```
# 1 S
# 2 PX
# 3 PY
# 4 PZ
# 5 DXX
# 6 DYY
# 7 DZZ
# 8 DXY
# 9 DXZ
# 10 DYZ
# 11 FXXX
# 12 FYYY
# 13 FZZZ
# 14 FXXY
# 15 FXXZ
# 16 FYYZ
# 17 FXYY
# 18 FXZZ
# 19 FYZZ
# 20 FXYZ
# 21 GXXXX
# 22 GYYYY
# 23 GZZZZ
# 24 GXXXY
# 25 GXXXZ
# 26 GXYYY
# 27 GYYYZ
# 28 GXZZZ
# 29 GYZZZ
# 30 GXXYY
# 31 GXXZZ
# 32 GYYZZ
# 33 GXXYZ
# 34 GXYYZ
# 35 GXYZZ
#
# For H and higher-shells, the order is given by:
#
# Do IX = 0, L
# Do IY = 0, (L-IX)
# IZ = L - IX - IY
#
# Thus, for H shells (L = 5), this leads to
# 36 HZZZZZ (005)
# 37 HYZZZZ (014)
# 38 HYYZZZ (023)
```

# 39 HYYYZZ (032)
# 40 HYYYYZ (041)
# 41 HYYYYY (050)
# 42 HXZZZZ (104)
# 43 HXYZZZ (113)
# 44 HXYYZZ (122)
# 45 HXYYYZ (131)
# 46 HXYYYY (140)
# 47 HXXZZZ (203)
# 48 HXXYZZ (212)
# 49 HXXYYZ (221)
# 50 HXXYYY (230)
# 51 HXXXZZ (302)
# 52 HXXXYZ (311)
# 53 HXXXYY (320)
# 54 HXXXXZ (401)
# 55 HXXXXY (410)
# 56 HXXXXX (500)

For example, type 27 is GYYYZ which corresponds to (a = 0, b = 3, c = 1).

6. Converting basis set coefficients to Cartesian primitive coefficients

The WFX file does not store basis set information; rather it stores molecular orbital coefficients of Cartesian Gaussian primitives. The key aspect for programmers is to first expand each basis set function as a linear combination of unnormalized Cartesian Gaussian primitives:

$$b_j\left(\vec{r}-\vec{R}_{center\_k}\right)=\sum_q \Lambda_{jq} p_q\left(\vec{r}-\vec{R}_{center\_k}\right) \qquad (3)$$

where $b_j\left(\vec{r}-\vec{R}_{center\_k}\right)$ is a basis set function and $\left\{\Lambda_{jq}\right\}$ are the associated coefficients for expanding $b_j\left(\vec{r}-\vec{R}_{center\_k}\right)$ in terms of $\left\{p_q\left(\vec{r}-\vec{R}_{center\_k}\right)\right\}$. The basis set center position $\vec{R}_{center\_k}$ represents an atom or ghost atom. The summation over q in eqn (3) only needs to consider those Cartesian Gaussian primitives that have the same center position $\vec{R}_{center\_k}$ as the basis set function b_j. $\Lambda_{jq}=0$ for any Cartesian Gaussian primitive p_q that does not have the same center position as b_j. In other words, each basis set function is expanded in terms of Cartesian Gaussian primitives that reside on the same center as the basis set function. **The eqn (3) expansion occurs only at the time the WFX file is written, which is after the quantum chemistry calculation on the material has already been completed; therefore, this expansion has no effect whatsoever on the quantum chemistry calculation itself.**

Each Cartesian Gaussian primitive has the form

$$p_q\left(\vec{r}-\vec{R}_{center\_k}\right)=\left(X-X_{center}\right)^a\left(Y-Y_{center}\right)^b\left(Z-Z_{center}\right)^c \exp(-\alpha\left\|\vec{r}-\vec{R}_{center\_k}\right\|^2) \qquad (4)$$

An individual Cartesian Gaussian primitive is completely specified by (a, b, c, α), where a, b, and c are non-negative integers and α is a positive real number.

$b_j\left(\vec{r}-\vec{R}_{center\_k}\right)$ can be either a contracted or uncontracted basis set function. Because a contracted basis set function contains multiple exponents, there will be the corresponding number of distinct α values associated with its linear expansion into Cartesian Gaussian primitives (eqn (3)). $b_j\left(\vec{r}-\vec{R}_{center\_k}\right)$ can be either a normalized or unnormalized basis set function. Regardless of whether normalized or unnormalized basis functions are used, the linear expansion in eqn (3) is always into unnormalized Cartesian Gaussian primitives. $b_j\left(\vec{r}-\vec{R}_{center\_k}\right)$ can be either a spherical-style (e.g., 5d, 7f, 9g) or Cartesian-style basis function (e.g., 6d, 10f,) or indeed any combination of spherical-style and Cartesian-style basis functions (e.g., spherical for d and f functions but Cartesian for g functions). Regardless of whether spherical-style or Cartesian-style basis functions or any combination thereof is used, the linear expansion in eqn (3) is always into unnormalized Cartesian Gaussian primitives. In other words, you can use any style Gaussian basis set functions that can be dreamed up when performing the quantum chemistry calculation on your material; regardless, these basis set functions are always linearly expanded in terms of unnormalized Cartesian Gaussian primitives for the purpose of writing the WFX file (eqn (3)).

Table 1 expands normalized spherical Gaussian functions in terms of unnormalized Cartesian Gaussian primitives. This table may be helpful when the basis set functions are based on normalized spherical Gaussian functions.

In terms of the basis set functions, each natural orbital can be expanded as

$$\phi_i\left(\vec{r}\right)=\sum_{k=1}^{ncenters}\sum_{j\in center\_k}C_{ij}b_j\left(\vec{r}-\vec{R}_{center\_k}\right) \tag{5}$$

where $j \in center\_k$ means a basis function centered around $\vec{R}_{center\_k}$. Substituting eqn (3) into (5) gives

$$\phi_i\left(\vec{r}\right)=\sum_{k=1}^{ncenters}\sum_{q\in center\_k}\Omega_{iq}p_q\left(\vec{r}-\vec{R}_{center\_k}\right) \tag{6}$$

where

$$\Omega_{iq}=\sum_j C_{ij}\Lambda_{jq} \tag{7}$$

In other words, the matrix Ω equals the matrix C times the matrix Λ. $\left\{\Omega_{iq}\right\}$ are written to the WFX file in the <Molecular Orbital Primitive Coefficients> section.

After deriving the coefficients $\left\{\Lambda_{jq}\right\}$ for eqn (3), how does one check to see if these were correctly computed? This can be checked by randomly choosing some position $\vec{r}$, followed by computing natural orbital values at $\vec{r}$ in two different ways. If the natural orbital values at $\vec{r}$ computed using the basis set functions and their coefficients (eqn (5)) match those computed using the Cartesian Gaussian primitives and their coefficients (eqn (6)), then everything is in order. This should be checked at a large number of randomly selected positions $\vec{r}$ during initial programming and testing of the wfx file writing routines. If a mismatch occurs, this should be investigated and the underlying cause corrected. If a mismatch occurs, the best strategy is to start with the simplest tests possible using systems containing only s basis functions, then to advance to systems containing both s and p basis functions, then to systems containing s and p and d functions, then to systems containing s and p and d and f functions, etc. In this way, the precise source of the problem can be identified. It might also be useful to start with single atom systems and advance to multi-atom systems to pinpoint the problem source.

The expansion in eqn (3) is the same as was required when writing WFN files. In fact, the WFX file format is meant to succeed and replace the prior WFN file format.[1] By using tags to mark data sections, the WFX file format achieves more flexibility compared to the WFN file format. This makes WFX files easier to read, write, and more extensible compared to WFN files. **Because of the similarity of data stored in WFN and WFX files, it should be trivial to modify a quantum chemistry program that writes WFN files so that it writes WFX files instead.**

Table 1: Real-valued spdfg spherical harmonics. These spherical harmonics are normalized.

$Y_{0,0} = s = \frac{1}{2}\sqrt{\frac{1}{\pi}}$	$Y_{3,1} = f_{xz^2} = \frac{1}{4}\sqrt{\frac{21}{2\pi}} \cdot \frac{x\left(4z^2 - x^2 - y^2\right)}{r^3}$
$Y_{1,-1} = p_y = \sqrt{\frac{3}{4\pi}} \cdot \frac{y}{r}$	$Y_{3,2} = f_{z\left(x^2-y^2\right)} = \frac{1}{4}\sqrt{\frac{105}{\pi}} \cdot \frac{\left(x^2 - y^2\right)z}{r^3}$
$Y_{1,0} = p_z = \sqrt{\frac{3}{4\pi}} \cdot \frac{z}{r}$	$Y_{3,3} = f_{x\left(x^2-3y^2\right)} = \frac{1}{4}\sqrt{\frac{35}{2\pi}} \cdot \frac{\left(x^2 - 3y^2\right)x}{r^3}$
$Y_{1,1} = p_x = \sqrt{\frac{3}{4\pi}} \cdot \frac{x}{r}$	$Y_{4,-4} = g_{xy\left(x^2-y^2\right)} = \frac{3}{4}\sqrt{\frac{35}{\pi}} \cdot \frac{xy\left(x^2 - y^2\right)}{r^4}$
$Y_{2,-2} = d_{xy} = \frac{1}{2}\sqrt{\frac{15}{\pi}} \cdot \frac{xy}{r^2}$	$Y_{4,-3} = g_{zy^3} = \frac{3}{4}\sqrt{\frac{35}{2\pi}} \cdot \frac{\left(3x^2 - y^2\right)yz}{r^4}$
$Y_{2,-1} = d_{yz} = \frac{1}{2}\sqrt{\frac{15}{\pi}} \cdot \frac{yz}{r^2}$	$Y_{4,-2} = g_{z^2xy} = \frac{3}{4}\sqrt{\frac{5}{\pi}} \cdot \frac{xy\left(7z^2 - r^2\right)}{r^4}$
$Y_{2,0} = d_{z^2} = \frac{1}{4}\sqrt{\frac{5}{\pi}} \cdot \frac{-x^2 - y^2 + 2z^2}{r^2}$	$Y_{4,-1} = g_{z^3y} = \frac{3}{4}\sqrt{\frac{5}{2\pi}} \cdot \frac{yz\left(7z^2 - 3r^2\right)}{r^4}$
$Y_{2,1} = d_{xz} = \frac{1}{2}\sqrt{\frac{15}{\pi}} \cdot \frac{xz}{r^2}$	$Y_{4,0} = g_{z^4} = \frac{3}{16}\sqrt{\frac{1}{\pi}} \cdot \frac{35z^4 - 30z^2r^2 + 3r^4}{r^4}$
$Y_{2,2} = d_{x^2-y^2} = \frac{1}{4}\sqrt{\frac{15}{\pi}} \cdot \frac{x^2 - y^2}{r^2}$	$Y_{4,1} = g_{z^3x} = \frac{3}{4}\sqrt{\frac{5}{2\pi}} \cdot \frac{xz\left(7z^2 - 3r^2\right)}{r^4}$
$Y_{3,-3} = f_{y\left(3x^2-y^2\right)} = \frac{1}{4}\sqrt{\frac{35}{2\pi}} \cdot \frac{\left(3x^2 - y^2\right)y}{r^3}$	$Y_{4,2} = g_{z^2\left(x^2-y^2\right)} = \frac{3}{8}\sqrt{\frac{5}{\pi}} \cdot \frac{\left(x^2 - y^2\right)\left(7z^2 - r^2\right)}{r^4}$
$Y_{3,-2} = f_{xyz} = \frac{1}{2}\sqrt{\frac{105}{\pi}} \cdot \frac{xyz}{r^3}$	$Y_{4,3} = g_{zx^3} = \frac{3}{4}\sqrt{\frac{35}{2\pi}} \cdot \frac{\left(x^2 - 3y^2\right)xz}{r^4}$

$Y_{3,-1} = f_{yz^2} = \frac{1}{4}\sqrt{\frac{21}{2\pi}}\cdot\frac{y\left(4z^2-x^2-y^2\right)}{r^3}$	$Y_{4,4} = g_{x^4+y^4} = \frac{3}{16}\sqrt{\frac{35}{\pi}}\cdot\frac{x^2\left(x^2-3y^2\right)-y^2\left(3x^2-y^2\right)}{r^4}$
$Y_{3,0} = f_{z^3} = \frac{1}{4}\sqrt{\frac{7}{\pi}}\cdot\frac{z\left(2z^2-3x^2-3y^2\right)}{r^3}$	

7. Ghost atoms

Ghost atoms have zero nuclear charge and serve as the center for basis functions. Ghost atoms are often used in calculations that evaluate basis set superposition errors (BSSE). The distinction between a “Ghost atom” and a “Dummy atom” is that a “Dummy atom” is used only to simplify the material’s Z-matrix and does not carry any basis functions, while a “Ghost atom” carries basis functions. Dummy atoms are completely ignored in WFX files, because they have no effect.

Ghost atoms are included in the WFX files as follows: (a) ghost atoms must be counted in the <number of nuclei> section, (b) ghost atoms are specified with atomic number 0 and nuclear charge 0.0 (c) coordinates of ghost nuclei must appear in the appropriate position of the <nuclear cartesian coordinates> section, and (d) if the optional <nuclear names> section (not read by Chargemol) is included, then ghost atom names begin with Bq (e.g., Bq7 denotes the seventh ghost atom).

8. Constructing the density matrix and natural orbitals with periodic boundary conditions

For periodic materials, the natural orbitals are Bloch waves defined as

$$\varphi_i\left(\vec{r},\vec{k}\right) = e^{\sqrt{-1}\left(\vec{k}\bullet\vec{r}\right)}u_i\left(\vec{r},\vec{k}\right) \qquad (8)$$

where $u_i\left(\vec{r},\vec{k}\right)$ is a spatial function with the same periodicity as the crystal. In collinear magnetism, there will be one set of Bloch waves for the spin-up electrons and a different set of Bloch waves for the spin-down electrons. Under 1D, 2D, or 3D periodic boundary conditions, the electron density is given by

$$\rho\left(\vec{r}\right) = \sum_{i\in\text{alpha}}\int_{\text{FBZ}} f_i^{\alpha}\left(\vec{k}\right)\varphi_i^{*}\left(\vec{r},\vec{k}\right)\varphi_i\left(\vec{r},\vec{k}\right)d^{D}\vec{k} + \sum_{j\in\text{beta}}\int_{\text{FBZ}} f_j^{\beta}\left(\vec{k}\right)\varphi_j^{*}\left(\vec{r},\vec{k}\right)\varphi_j\left(\vec{r},\vec{k}\right)d^{D}\vec{k} \qquad (9)$$

where FBZ is the first Brillouin zone, $f_i^{\alpha}\left(\vec{k}\right)$ is the spin-up occupation of the alpha natural orbital $\left(i,\vec{k}\right)$, and D = 1, 2, or 3 is the number of periodic boundary conditions. The spin density is given by

$$\rho^{\text{spin}}\left(\vec{r}\right) = \sum_{i\in\text{alpha}}\int_{\text{FBZ}} f_i^{\alpha}\left(\vec{k}\right)\varphi_i^{*}\left(\vec{r},\vec{k}\right)\varphi_i\left(\vec{r},\vec{k}\right)d^{D}\vec{k} - \sum_{j\in\text{beta}}\int_{\text{FBZ}} f_j^{\beta}\left(\vec{k}\right)\varphi_j^{*}\left(\vec{r},\vec{k}\right)\varphi_j\left(\vec{r},\vec{k}\right)d^{D}\vec{k} \qquad (10)$$

When periodic boundary conditions are employed, it is common to use a smearing function that smooths the occupation numbers $\left\{f_i\left(\vec{k}\right)\right\}$ to prevent discontinuous behavior between two orbitals having infinitesimal energy difference. The Bloch functions are orthonormal

$$\left\langle \varphi_i\left(\vec{r},\vec{k}\right) \middle| \varphi_j\left(\vec{r},\vec{k}'\right)\right\rangle = \oint \varphi_i^*\left(\vec{r},\vec{k}\right)\varphi_j\left(\vec{r},\vec{k}'\right)d^3\vec{r} = \delta_{i,j}\delta\left(\left|\vec{k}-\vec{k}'\right|\right) \tag{11}$$

For DFT and Hartree-Fock calculations, the spin-up canonical orbitals are also the spin-up natural orbitals that diagonalize the spin-up first-order density matrix. For DFT and Hartree-Fock calculations, the spin-down canonical orbitals are also the spin-down natural orbitals that diagonalize the spin-down first-order density matrix. For correlated wavefunction calculations (e.g., CCSD, CISD, CAS-SCF, MP2, etc.), the natural orbitals and the canonical orbitals are two different sets of Block waves. The Bloch waves corresponding to the natural orbitals should be the ones written to the WFX file, because these are the Block waves that give the electron and spin density distributions according to eqn (9) – (10).

Integration over the first Brillouin zone is usually approximated by numeric quadrature using a k-point mesh. Specifically, continuous integration over $\left\{\vec{k}\right\}$ in the first Brillouin zone can be approximated as a weighted sum over discrete k-points:

$$\rho\left(\vec{r}\right) = \sum_i \sum_{\tau\in\text{FBZ}} w_\tau\left(f_{i,\tau}^{\alpha}u_{i,\tau}^{\alpha\,*}\left(\vec{r}\right)u_{i,\tau}^{\alpha}\left(\vec{r}\right) + f_{i,\tau}^{\beta}u_{i,\tau}^{\beta\,*}\left(\vec{r}\right)u_{i,\tau}^{\beta}\left(\vec{r}\right)\right) \tag{12}$$

$$\rho^{\text{spin}}\left(\vec{r}\right) = \sum_i \sum_{\tau\in\text{FBZ}} w_\tau\left(f_{i,\tau}^{\alpha}u_{i,\tau}^{\alpha\,*}\left(\vec{r}\right)u_{i,\tau}^{\alpha}\left(\vec{r}\right) - f_{i,\tau}^{\beta}u_{i,\tau}^{\beta\,*}\left(\vec{r}\right)u_{i,\tau}^{\beta}\left(\vec{r}\right)\right) \tag{13}$$

where τ is the k-point index, $f_{i,\tau}$ is the fractional occupation of spin-orbital $\left(i,\tau\right)$. w_τ is the weight factor associated with k-point τ. These sum to 1:

$$\sum_{\tau\in\text{FBZ}} w_\tau = 1 \tag{14}$$

Note that

$$\phi_{i,\tau}^*\left(\vec{r}\right)\phi_{i,\tau}\left(\vec{r}\right) = u_{i,\tau}^*\left(\vec{r}\right)u_{i,\tau}\left(\vec{r}\right) \tag{15}$$

because the phase factor $e^{\sqrt{-1}\left(\vec{k}\bullet\vec{r}\right)}$ times its complex conjugate is unity.

As described in Part 5 above, the WFX file stores the list of k-points, their weights $\left\{w_\tau\right\}$, their fractional occupations $\left\{f_{i,\tau}\right\}$, and the orbitals $\left\{u_{i,\tau}\left(\vec{r}\right)\right\}$. Note that the WFX-stored orbitals are properly $\left\{u_{i,\tau}\left(\vec{r}\right)\right\}$ not $\left\{\phi_{i,\tau}\left(\vec{r}\right)\right\}$, because $\left\{u_{i,\tau}\left(\vec{r}\right)\right\}$ has the same periodicity as the reference unit cell while $\left\{\phi_{i,\tau}\left(\vec{r}\right)\right\}$ does not. In the WFX file, the $\left\{u_{i,\tau}\left(\vec{r}\right)\right\}$ are stored in terms of their Cartesian Gaussian primitive coefficients. In this representation, each Cartesian Gaussian primitive includes all of its periodic images; that is, the same orbital coefficient applies to a Cartesian Gaussian primitive and all of its periodic images. Note that an individual spin-orbital stored in the WFX file refers to a specific $\left(i,\tau\right)$ combination. For example, if there were 10 bands and 8 k-points, then there would be 10 × 8 = 80 orbitals stored in the WFX file, with the 80 orbital occupations, orbital energies, k-point value for each orbital, and Cartesian Gaussian primitives specified as described in the above Sections.

9. Example wfx files

As summarized in Table 2, please find attached example wfx files contained in example_wfx_files.zip. Some version of each wfx file has been successfully used in tests of a Chargemol development version.

Table 2: Summary of wfx files contained in example_wfx_files.zip. For orbital type, NAT=natural, CN = canonical natural, LN = localized natural.

filename	PBC[a]	Ghost atoms?	exchange-correlation method	Idempotent density matrix?	orbital type	orbital energies?[b]
CsO_CCSD _def2QZVPPD.wfx	0	N	CCSD	N	NAT	N
C4H6_M06L_augccpvtz _save_PM.wfx	0	N	DFT	Y	LN	N
O2_CCSD _def2QZVPPD.wfx	0	N	CCSD	N	NAT	N
Br3anion_CCSD _def2QZVPPD.wfx	0	N	CCSD	N	NAT	N
H2S_with_ghost.wfx	0	Y	DFT	Y	CN	Y
cystine_PBE_augccpvtz.wfx	0	N	DFT	N	CN	Y
distal-otrimeS-b3lyp-6311++gss_Multiwfn _Pipek_Mezey.wfx	0	N	DFT	Y	LN	N
polyfluoroprene_PBE _6311ppGss_KPOINTS.wfx	1	N	DFT	Y	CN	Y
BN_plane_lanl2dz_KPOINTs.wfx	2	N	DFT	Y	CN	Y
GaAs_solid_lanl2dz_KPOINTs.wfx	3	N	DFT	Y	CN	Y

[a] Number of periodic boundary conditions. [b] Whether orbital energies are written as non-zero values.

References

[1] Todd A. Keith, http://aim.tkgristmill.com/wfxformat.html

[2] A. J. Coleman, “Structure of Fermion Density Matrices,” Rev. Mod. Phys. 35 (1963) 668–687. DOI: 10.1103/RevModPhys.35.668