QM/MM computations with GAMESS and TINKER
Written by Cheol Ho Choi - Nov 16, 1999
Most recently edited - June 14, 2003

GAMESS can perform either Shoemaker et al's SIMOMM (Surface Integrated
Molecular Orbital/Molecular Mechanics) or Morokuma et al's IMOMM (Integrated
Molecular Orbital/Molecular Mechanics) scheme. Although SIMOMM was
developed especially for the study of surface chemical systems, it is a
general method so that one may find it useful in other systems too.

The existence of a $LINK group triggers these QM/MM computations.

SIMOMM and IMOMM are models in which some atoms embedded in a larger
set of atoms are treated by quantum mechanics (any such method available
in GAMESS), while the remaining atoms are treated by molecular mechanics.
Capping atoms (usually hydrogens) connected by -single- bonds are
included in the QM treatment, to satisfy any dangling valences created
by the boundary between the QM atoms and the total set of atoms.
IMOMM places these capping atoms at fixed angles and distances, along
the cut bonds, whereas SIMOMM allows the positions of these capping
atoms to be optimized. Accordingly, most applications we have performed
to date use the SIMOMM model.

The original IMOMM model is described in
    F.Maseras, K.Morokuma J.Comput.Chem. 16, 1170(1995).

The following paper describes the SIMOMM method:
    (a) J.R.Shoemaker, L.W.Burggraf, M.S.Gordon
        J.Phys.Chem.A 103, 3245-51(1999).
An initial application was to the controversial dimers found on the
reconstructed, bare silicon 100 surface:
    (b) J.R.Shoemaker, L.W.Burggraf, M.S.Gordon
        J.Chem.Phys. 112, 2994-3005(2000)
Since then, this surface and the structures and reactivity of a variety
of species adsorbed on it surfaces have been studied:
    (c) C.H.Choi, M.S.Gordon J.Am.Chem.Soc. 121, 11311-7(1999)
    (d) Y.Jung, C.H.Choi, M.S.Gordon J.Chem.Phys.B 105,4039-44(2001)
    (e) C.H.Choi, D.-J.Liu, J.W.Evans, M.S.Gordon

J.Am.Chem.Soc. 124, 8730(2002)
       (f) Y.Jung, Y.Akinaga, K.D.Jordan, M.S.Gordon
             Theoret.Chem.Acc. 109, 268-73(2003)
       (g) C.H.Choi, M.S.Gordon J.Am.Chem.Soc. 124, 6162(2002)
and consequently there is a review article,
       (h) "Chemistry on Silicon Surfaces" C.H.Choi, M.S.Gordon
             in "The chemistry of organic silicon compounds",
             ed. Z.Rappaport, Y.Apeloig, p. 821 (2001)
Note that SIMOMM is not necessarily just a surface model, as shown by the
QM treatment of a small part of a larger organic molecule examples below.

Program TINKER has been integrated into GAMESS for the MM part of the
computations. This molecular mechanics program is the product of
Jay Ponder's research group, and more information about TINKER can be
found on their web page,
      http://dasher.wustl.edu/tinker
Two journal articles relating to TINKER are
 a) J. W. Ponder and F. M Richards, "An Efficient Newton-like
    Method for Molecular Mechanics Energy Minimization of
    Large Molecules", Journal of Computational Chemistry,
    8, 1016-1024 (1987)
 b) C. E. Kundrot, J. W. Ponder and F. M. Richards, "Algorithms for
    Calculating Excluded Volume and Its Derivatives as a Function
    of Molecular Conformation and Their Use in Energy Minimization",
    Journal of Computational Chemistry, 12, 402-409 (1991)

The TINKER program is normally obtained directly from Washington
University through the web page mentioned above, under an agreement
not to make copies of it (much like GAMESS itself). Since the
TINNKER program had to be modified somewhat to interface with GAMESS,
Jay Ponder graciously gave permission to distribute this program in
its modified form:
          You should read the licensing requirements on the
          TINKER web page, and consider this source code to
          be bound by this same agreement !!!
The manual for TINKER is not include here, but may be downloaded
from the TINKER web site.

TINKER is offered as an optional extension to GAMESS. This is primarily
due to its use of numerous include files to dimension memory by static
storage allocations (changeable by adjusting sizes.i and recompiling).
This differs greatly from the usual GAMESS dynamic memory, so TINKER is
not included in GAMESS normally. The use of static storage, with 1,000
atoms and a Hessian store of 500,000, leads to roughly a doubling of the
memory image of GAMESS, from 23 MB to 44 MB, of course not counting the
user input for the MEMORY= amount used by the QM portion of the run.

The package consists of twelve *.f FORTRAN source files, and numerous
include files (*.i). The files named Libt*.f are lightly modified
TINKER code, while the three T*.f files were written by Jim Shoemaker
and Cheol Choi to perform the QM/MM interface.

The include file sizes.i can be used to adjust dimensions. Note that
the parameter MAXATM in this file is used in GAMESS' file INPUTB.SRC,
and the values given must match exactly.

To add TINKER to GAMESS, you must recompile and relink. The steps
for installing GAMESS are fully described in the file
    ~/gamess/misc/readme.unix
so here it is only necessary to say what small changes are needed in
this procedure. Obtain the TINKER add-on by downloading from
     http://php.scl.ameslab.gov/GAMESS/tinker/tinker.tar.Z
and unpack this file by
    chdir ~/gamess
    zcat /full/path/to/file/tinker.tar.Z | tar -xvf -
After this, the files for TINKER will be present below the GAMESS
directory tree, namely in the directories
    ~/gamess/tinker
    ~/gamess/tinker/params
The procedure for installing GAMESS is modified as follows:
    1. edit INPUTB.SRC in the standard GAMESS source directory so that
       the parameter MAXATM is set to 1000, in order to match sizes.i
    2. edit "compall" to set the variable TINKER to true.
    3. edit "lked" to set the variable TINKER to true.
Now repeat the usual directions in ~/gamess/misc/readme.unix, and TINKER
will be included into GAMESS.

TINKER itself does not run in parallel. The quantum mechanical portion
of the run handled by GAMESS will speed up in parallel, while the much
shorter molecular mechanics computation handled by TINKER is simply
repeated on every node.

TINKER + GAMESS is not tested as thoroughly as GAMESS alone, for example
on as many kinds of computers. We have installed it on RedHat Linux,
IBM AIX in both 32 and 64 bit modes, and on AXP computers. There may
be troubles with anything else! Be sure to be cautious about the
significance of any results you obtain with GAMESS+TINKER.

Since we are using TINKER for the MM calculations, the initial input
format has to be TINKER format. If your starting format is not TINKER,
program BABEL is strongly recommended to convert any file into TINKER
format. BABEL can be downloaded from
 ftp://ccl.osc.edu/pub/chemistry/software/UNIX/babel/

For example, one is presumed to use a GUI molecular editor (such as Chem3D)
to draw a molecule and save it in a SYBYL format (or whatever format BABEL
can understand). After that one uses BABEL to convert files in SYBYL format
into TINKER format.

It is easiest to start this way, with only a TINKER input file defining
the total system. Correspondingly, in the CONTRL group we need to set
COORD=TINKER. The coordinate information is in a new data group, $TINXYZ,
where one inputs coordinates in TINKER format. Note that the first
example is this way, with no $DATA at all. A $DATA group containing
the quantum mechanical atoms will be in the xxx.dat file of any such
run, so that later runs (see the 2nd example) may well have the usual
$DATA group. Note that the order of atoms must match the original $TINXYZ
so that it is easiest to do an initial run with COORD=TINKER, let the
code generate $DATA in the precise order, and use that $DATA later (for
example to change the basis set to a manual specification). The capping
atoms will always be at the end of this $DATA group, and by default they
will be hydrogen atoms!

Since it is very convenient to use DLC (delocalized coordinates), it is

useful to set NZVAR in CONTRL group and add DLC=.True. and AUTO=.True.
to the $ZMAT group.

Suppose we want to treat the double bond of cyclopentene quantum
mechanically, and the rest with molecular mechanics on the basis of
IMOMM scheme. The input file may be

```
!  An example file to illustrate the general input file to perform
!  Morokuma's MOMM scheme.
!
!  The results of the geometry optimization are:
!  NSERCH=0:
!  FINAL QM E= -77.0655534370, RMS grad= 0.0127255, MM energy= 5.3199258027
!  combined QM+MM energy is -77.0570755959
!  NSERCH=7:
!  FINAL QM E= -77.0675805805, RMS grad= 0.0000101, MM energy= 5.4640573966
!  combined QM+MM energy is -77.0588730511
!
!  The energy of the QM + MM system is lowered 1.273 - 0.144 = 1.129 kcal/mol
!
!  Note that the existence of a LINK group triggers MOMM calculation.
!  COORD=TINKER, if TINNKER coords are in $TINXYZ, means no $DATA given.
!  NZVAR used to trigger use of delocalized coordinates.
!
 $CONTRL SCFTYP=RHF RUNTYP=optimize COORD=TINKER NZVAR=33 $END
 $BASIS  GBASIS=sto NGAUSS=3  $END
 $ZMAT   DLC=.T. AUTO=.T. $END
!  $DATA group is not provided, because input format is only TINKER.
 $TINXYZ
    13 Cyclopentene, Maseras and Morokuma J. Comp. Chem. 1995, 16, 1170
     1  C     -0.809200  -1.126700  -0.107500     2    2    3   11
     2  C     -1.400300   0.245100  -0.087400     1    1    4    5    6
     3  H     -1.420900  -2.040000  -0.149000     5    1
     4  C     -0.180000   1.163900  -0.012100     1    2    7    8    9
     5  H     -2.049300   0.379500   0.806900     5    2
     6  H     -1.974700   0.435900  -1.021500     5    2
     7  H     -0.214800   1.764100   0.924700     5    4
     8  H     -0.164900   1.842400  -0.894100     5    4
```

```
      9  C         1.068700     0.280400    -0.018000      1     4    10    11   12
     10  H         1.688800     0.488300    -0.918500      5     9
     11  C         0.519000    -1.107700    -0.070000      2     1     9    13
     12  H         1.662100     0.432500     0.911300      5     9
     13  H         1.157900    -2.003100    -0.076500      5    11
 $END
!   Because IMOMM=.T., Morokuma's MOMM scheme is used.
!   IQMATM contains the atoms to be calculated quantum mechanically.
 $LINK IMOMM=.T. IQMATM(1)=1,11,3,13 $END
!   Due to non-GAMESS input format, each option in TINKEY group
!   should be on a separate line.
 $TINKEY
parameters /u1/mike/gamess/tinker/params/mm2
 $END
```

The presence of the $LINK group in the input file is what triggers
the QM/MM calculation. In this group, you must choose between IMOMM
and SIMOMM, by entering one of these as .TRUE. There is no default
selection, you must make one.

For either procedure, IQMATM specifies the atoms which are going to be
treated by quantum mechanics. Any atoms not given in IQMATM are only
in the molecular mechanics region. In the above example, the two C's
of the double bond, and their bonded hydrogen atoms are numbers 1, 11,
3, and 13. The $TINXYZ group shows that each C is originally hooked
to one other C in the ring. To be precise, the information in $TINXYZ
shows that the C atoms were attached to two other atoms, numbers 2 and
9. Therefore the cut between QM and MM region separates bonds 1 with 2,
and bonds 11 with 9.   The IMOMM method will cut these bonds off, and
add an H atom to cap off the dangling valence. The H atom will be
placed at the same angle as the original CC bond being cut, and at a
distance that is determined by the sum of the covalent radii (C+H in
this example). The capping H will be frozen at that location, while
the rest of the double bond is geometry optimized.

Prior to June 2003, it was necessary to define the linkage explicitly,
but now that information will be deduced from the data in $TINXYZ and
the $LINK's IQMATM input. It is no longer possible to define the

distance at which the capping hydrogens are fixed. This means that
older input files, containing LINKGE, BLQM, and/or BLMM in the $LINK
group, no longer work.

SIMOMM removes the constraint about the angular and distance location
of the capping atoms, and as a result may provide a more realistic
geometry optimization within the quantum mechanical zone.

The capping atom is necessarily monovalent, and the any automatically
generated $DATA will cap with hydrogen atoms. The capping atom may
be changed to some other monovalent atom, e.g. fluorine, by editing
this $DATA. Be careful not to change the order of atoms within this
group, during such a transmutation. The monovalence implies the bonds
that are cut by the boundary between QM and MM must be single bonds,
and the type of bond cut must be reasonably well modeled by a M-H or
M-Cl (etc.) bond in place of the original M-M'. In other words, the
electronegativity match must be reasonable, as is the case of replacing
the severed Si-Si bonds by capping Si-H in the references quoted above.
Transition metal surfaces are likely to be poor candidates for SIMOMM
modeling, as that sort of surface is probably a poor match for H or
halogens in electronegativity, and the capping process necessarily
interferes with the long range delocalization in metals. SIMOMM, in
other words, is better suited to insulator or semiconductor type systems.

The group $TINKEY replaces the TINKER key file. Users are encouraged to
read the TINKER manual for the details. You must specify the complete
path to the molecular mechanics parameter file in this group, but
not including the .prm extension of these files. In the above example,
we used the MM2 parameter set. Currently, TINKER has 14 different
parameter sets. In addition, various other TINKER options can be
specified in the $TINKEY group. A group $TOPTMZ can be used, when
TINKER Newton optimization needs extra options such as MODE, METHOD
and GRDMIN. Please read the TINKER manual for detailed options.

The above input yields optimized structures of QM part and MM part. The
QM part has extra hydrogens for the dangling bonds. Sometimes it is
easier to use a previously optimized QM coordinates, which are easily
provided in $DATA (compared to regenerating a $TINXYZ!). Or, we may

simply want to add a user defined basis set, which we can only do with
a $DATA. The second example shows a case where the cyclopentene
example was already optimized with the SIMOMM scheme and we want to
input the optimized QM part separately:

```
!  An example file to illustrate the general input file to perform
!  Shoemaker et al.'s SIMOMM scheme.
!
!  Note that the existence of a $LINK group triggers SIMOMM calculation.
!
!  COORD=UNIQUE, if you want to input QM atoms separately from the
!  MM atoms, although the same atoms must still be in $TINXYZ.  The
!  coordinates for the QM atoms will be those of the $DATA group,
!  not what $TINXYZ says.  Note that $DATA must have all the capping
!  hydrogens in it, at the end.
!
!  This particular optimization happens to input a previously
!  optimized geometry, so it runs only the initial NSERCH=0 point:
!  FINAL E=-77.0739142459, 7 ITERS, RMS GRAD= .0000089,
!  MM Energy=8.2680467843, combined QM+MM energy= -77.0607382752
!
 $CONTRL SCFTYP=RHF RUNTYP=optimize MULT=1 EXETYP=RUN
  COORD=UNIQUE NZVAR=33 $END
 $BASIS GBASIS=sto NGAUSS=3  $END
 $ZMAT DLC=.T. AUTO=.T.  $END
!  The optimized QM part. Note that it MUST include the hydrogens
!  that replace dangling bonds.
 $DATA
Ab initio atoms in QM region
C1
C            6.0   -.7980989479  -1.1214243765   -.1070524532
C            6.0    .5077499641  -1.1027520491   -.0702239088
H            1.0  -1.3723767399  -2.0367069593   -.1475026326
H            1.0   1.1093101048  -2.0012199099   -.0775923300
H            1.0  -1.3810058083   -.2097263990   -.0989003921
H            1.0   1.0629880755   -.1747785792   -.0300254581
 $END
 $TINXYZ
```

```
     13 Cyclopentene, Maseras and Morokuma J. Comp. Chem. 1995, 16, 1170
    1  C      -0.809200    -1.126700    -0.107500     2     2     3    11
    2  C      -1.400300     0.245100    -0.087400     1     1     4     5     6
    3  H      -1.420900    -2.040000    -0.149000     5     1
    4  C      -0.180000     1.163900    -0.012100     1     2     7     8     9
    5  H      -2.049300     0.379500     0.806900     5     2
    6  H      -1.974700     0.435900    -1.021500     5     2
    7  H      -0.214800     1.764100     0.924700     5     4
    8  H      -0.164900     1.842400    -0.894100     5     4
    9  C       1.068700     0.280400    -0.018000     1     4    10    11    12
   10  H       1.688800     0.488300    -0.918500     5     9
   11  C       0.519000    -1.107700    -0.070000     2     1     9    13
   12  H       1.662100     0.432500     0.911300     5     9
   13  H       1.157900    -2.003100    -0.076500     5    11
 $END
!   Because SIMOMM=.T., Shoemaker's SIMOMM scheme is used.
!   IQMATM contains the atoms to be calculated quantum mechanically.
 $LINK SIMOMM=.T. IQMATM(1)=1,11,3,13 $END
!   Due to non-GAMESS input format, each option in TINKEY group
!   must have its own line.
 $TINKEY
parameters /u1/mike/gamess/tinker/params/mm2
 $END
```

By now, the above example should be self-explanatory.

Remember the total energy is QM energy + MM energy. QM energy is in
atomic unit (hartree) and negative, while MM energy is in kcal/mol
and positive. Note that the string "FINAL" gives the QM energy,
at a point before the MM energy is known, so it only contains the
QM energy. The MM region is later fully relaxed during computation
of the gradient at each QM geometry step, giving a fully optimized
MM energy for this QM geometry. This energy is printed as the
"MM ENERGY", and at this point the program can sum "QM+MM ENERGY".
It is this total system energy that is optimized by the program,
and this should go down. Note that the MM energy can and often
does go up during the system geometry minimization.

ADVICE:

If one wants to freeze some of the geometric parameters in QM region, note that numbering MUST be based on the sequence in QM region not the one in MM region.

Frequently, GAMESS complains by saying there are no parameter values for certain MM parameters. It is simply because the parameter set you are using does not have such parameters. If these parameters are for the QM region atoms, one can fool the program by providing arbitrary parameter values in the parameter file.