U.S. DEPARTMENT OF THE INTERIOR
U.S. GEOLOGICAL SURVEY

USER'S GUIDE TO PHREEQC (VERSION 2)

(Equations on which the program is based)

By **David L. Parkhurst** *and* **C.A.J. Appelo**[1]

Water-Resources Investigations Report 99-4259

[1]Hydrochemical Consultant
Valeriusstraat 11
1071 MB Amsterdam, NL
appt@xs4all.nl
http://www.xs4all.nl/~appt/index.html

Denver, Colorado
1999

EQUATIONS FOR SPECIATION AND FORWARD MODELING

In this section of the report, the algebraic equations used to define thermodynamic activities of aqueous species, ion-exchange species, surface-complexation species, gas-phase components, solid solutions, and pure phases are presented. First, thermodynamic activities and mass-action equations are described for aqueous, exchange, and surface species. Then, a set of functions, denoted f, are defined that must be solved simultaneously to determine equilibrium for a given set of conditions. Many of these functions are derived from mole-balance equations for each element or element valence state, exchange site, and surface site or from mass-action equations

for pure phases and solid solutions. Additional functions are derived for alkalinity, activity of water, aqueous charge balance, gas-phase equilibria, ionic strength, and surface-complexation equilibria. Each function is reduced to contain a minimum number of variables, such that the number of functions equals the number of variables. The program uses a modified Newton-Raphson method to solve the simultaneous nonlinear equations. This method uses the residuals of the functions and an array of partial derivatives of each function with respect to the set of master unknowns or master unknowns. For clarity, the set of variables used in partial differentiation are referred to as "master unknowns". The total derivatives of each function, f, will be presented without derivation. In the following equations, lack of a subscript or the subscript "(aq)" will refer to entities in the aqueous phase, "(e)" refers to exchangers, "(g)" refers to gases, "(s)" refers to surfaces, "(ss)" refers to solid solutions, and "(p)" refers to phases.

Activities and Mass-Action Equations

In this section the activities of aqueous, exchange, and surface species are defined and the mass-action relations for each species are presented. Equations are derived from the mass-action expression for the moles of each species in the chemical system in terms of the master unknowns. These equations are then differentiated with respect to the master unknowns. Later, these equations for the moles of a species and the partial derivatives will be substituted into the constituent mole-balance, charge-balance, and phase-equilibria functions.

Aqueous Species

PHREEQC allows speciation or equilibration with respect to a single aqueous phase. However, multiple aqueous phases may be defined in the course of a run and an aqueous phase may be defined as a mixture of one or more aqueous phases (see **MIX** keyword in "Description of Data Input"). The dissolved species in the aqueous phase are assumed to be in thermodynamic equilibrium, with one exception; in initial solution calculations, disequilibrium among valence states of redox elements is allowed. The unknowns for each aqueous species i are the activity, a_i, activity coefficient, γ_i, molality, m_i, and moles in solution, n_i.

PHREEQC rewrites all chemical equations in terms of master species. There is one master aqueous species associated with each element (for example, Ca^{+2} for calcium) or element valence state (for example, Fe^{+3} for ferric iron) plus the activity of the hydrogen ion, the activity of the aqueous electron, and the activity of water. Some programs, for example MINTEQA2 (Allison and others, 1990) and MINEQL$^+$ (Schecher and McAvoy, 1991) use the term "component" for these species, but that terminology is not used here because of confusion with the definition of component for the Gibbs' phase rule. For PHREEQC, the identity of each aqueous master species is defined with **SOLUTION_MASTER_SPECIES** data block (see "Description of Data Input"). The numerical method reduces the number of unknowns to be a minimum number of master unknowns, and iteratively refines the values of these master unknowns until a solution to the set of algebraic equations is found. The master unknowns for aqueous solutions are the natural log of the activities of master species, the natural log of the activity of water, a_{H_2O}, the ionic strength, μ, and the mass of solvent water in an aqueous solution, W_{aq}.

The following relationships apply to all aqueous species (except aqueous electrons and water itself): $a_i = \gamma_i m_i$ and $n_i = m_i W_{aq}$. Equilibrium among aqueous species in an ion-association model requires that all mass-action equations for aqueous species are satisfied. For example, the association reaction for the aqueous

species $CaSO_4^0$ is $Ca^{2+} + SO_4^{2-} = CaSO_4^0$. The log K for this reaction at 25°C is 2.3, which results in the mass-action equation:

$$10^{2.3} = \frac{a_{CaSO_4^0}}{a_{Ca^{2+}}a_{SO_4^{2-}}}.$$
(1)

In general, mass-action equations can be written as

$$K_i = a_i \prod_m^{M_{aq}} a_m^{-c_{m,i}},$$
(2)

where K_i is a temperature-dependent equilibrium constant, $c_{m,i}$ is the stoichiometric coefficient of master species m in species i and M_{aq} is the total number of aqueous master species. The values of $c_{m,i}$ may be positive or negative. For PHREEQC, terms on the right-hand side of an association reaction are assigned negative coefficients and terms on the left-hand side are assigned positive coefficients. The same formalism applies to master species, where the mass-action equation is simply $1 = \dfrac{a_m}{a_m}$.

The total moles of an aqueous species i can be derived from the mass-action expression:

$$n_i = m_i W_{aq} = K_i W_{aq} \frac{\displaystyle\prod_m^{M_{aq}} a_m^{c_{m,i}}}{\gamma_i}.$$
(3)

The Newton-Raphson method uses the total derivative of moles with respect to the master unknowns. The total derivative is

$$dn_i = n_i \left[d\ln(W_{aq}) + \sum_m^{M_{aq}} c_{m,i} d\ln(a_m) - \frac{\partial}{\partial \mu} \ln(\gamma_i) d\mu \right].$$
(4)

Activity coefficients of aqueous species are defined with the Davies equation:

$$\log\gamma_i = -Az_i^2 \left(\frac{\sqrt{\mu}}{1+\sqrt{\mu}} - 0.3\mu \right),$$
(5)

or the extended or WATEQ Debye-Hückel equation:

$$\log\gamma_i = -\frac{Az_i^2\sqrt{\mu}}{1+Ba_i^o\sqrt{\mu}} + b_i\mu,$$
(6)

where z_i is the ionic charge of aqueous species i, and A and B are constants dependent only on temperature. Equation 6 is the extended Debye-Hückel equation, if b_i is zero, or the WATEQ Debye-Hückel equation (see Truesdell and Jones, 1974), if b_i is not equal to zero. In the extended Debye-Hückel equation, a_i^o is the ion-size parameter, whereas in the WATEQ Debye-Hückel equation a_i^o and b_i are ion-specific parameters fitted from mean-salt activity-coefficient data. Unless otherwise specified in the database file or the input data set, the Davies equation is used for charged species. For uncharged species, the first term of the activity coefficient equation is zero, and the WATEQ Debye-Hückel equation reduces to the Setchenow equation ($\ln\gamma_i = b_i\mu$) (see Langmuir, 1997 for discussion). Unless otherwise specified, b_i is assumed to be 0.1 for all uncharged species.

The partial derivatives of these activity coefficient equations with respect to ionic strength are

$$\frac{\partial}{\partial \mu}\ln\gamma_i = -\ln(10)\left[Az_i^2\left(\frac{1}{2\sqrt{\mu}(\sqrt{\mu}+1)^2} - 0.3\right)\right],\tag{7}$$

for the Davies equation and

$$\frac{\partial}{\partial \mu}\ln\gamma_i = -\ln(10)\left(\frac{Az_i^2}{2\sqrt{\mu}(Ba_i^o\sqrt{\mu}+1)^2} + b_i\right),\tag{8}$$

for the extended or WATEQ Debye-Hückel equation.

For data input to PHREEQC, the chemical equation for the mole-balance and mass-action expressions, the log K and its temperature dependence, and the activity coefficient parameters for each aqueous species are defined through the **SOLUTION_SPECIES** data block. Master species for elements and element valence states are defined with the **SOLUTION_MASTER_SPECIES** data block. Composition of a solution is defined with the **SOLUTION** or **SOLUTION_SPREAD** data block (see "Description of Data Input").

Exchange Species

Ion-exchange equilibria are included in the model through heterogeneous mass-action equations and mole-balance equations for exchange sites. PHREEQC allows multiple exchangers, termed an "exchange assemblage", to exist in equilibrium with the aqueous phase. The approach uses mass-action expressions based on half-reactions between aqueous species and a fictive unoccupied exchange site (Appelo and Postma, 1993) for each exchanger. This unoccupied exchange site is the master species for the exchanger and the log of its activity is an additional master unknown. Its identity is defined with **EXCHANGE_MASTER_SPECIES** data block (see "Description of Data Input"). However, the master species is not included in the mole-balance equation for the exchanger, forcing its physical concentration to be zero. Its activity is also physically meaningless, but is such that all of the exchange sites are filled by other exchange species.

The unknowns for exchange calculations are the activity, a_{i_e}, which is defined to be the equivalent fraction in PHREEQC times an activity coefficient, γ_{i_e}, and the moles, n_{i_e}, of each exchange species, i_e, of exchanger e. The equivalent fraction is the moles of sites occupied by an exchange species divided by the total number of exchange sites. The activity of an exchange species is $a_{i_e} = \gamma_{i_e}\dfrac{b_{e,i_e}n_{i_e}}{T_e}$, where b_{e,i_e} is the number of equivalents of exchanger, e, occupied by the exchange species i_e, and T_e is the total number of exchange sites for the exchanger, in equivalents. Note that T_e is the total number of equivalents of the exchanger in the system, which is not necessarily equal to the number of equivalents per kilogram of water (eq/kgw) because the mass of water in the system may be more or less than 1 kg. By default, the activity coefficient for an exchange species is 1.0, but optionally, a Davies, extended Debye-Hückel, or WATEQ Debye-Hückel activity coefficient can be used, which is based on the aqueous ionic strength and the number of equivalents of exchange sites occupied by the exchange species.

Equilibrium among aqueous and exchange species requires that all mass-action equations for the exchange species are satisfied. The association reaction for the exchange species CaX_2 is $Ca^{2+} + 2X^- = CaX_2$, where X^- is the exchange master species for the default database. The use of equivalent fractions for activities and this form for the chemical reaction is known as the Gaines-Thomas convention (Gaines and Thomas, 1953) and is the

convention used in the databases *phreeqc.dat* and *wateq4f.dat*, which are distributed with PHREEQC. [It is also possible to use the Gapon convention in PHREEQC, which also uses equivalent fraction, but writes the exchange reaction as $0.5Ca^{2+} + X^- = Ca_{0.5}X$. See Appelo and Postma (1993) for more discussion.] The log K for calcium exchange in the default database file is 0.8, which results in the following mass-action equation:

$$10^{0.8} = \frac{a_{CaX_2}}{a_{Ca^{2+}}a_{X^-}^2}.$$ (9)

In general, mass-action equations can be written as

$$K_{i_e} = a_{i_e} \prod_m^M a_m^{-c_{m,i_e}},$$ (10)

where *m* varies over all master species, including exchange master species, c_{m,i_e} is the stoichiometric coefficient of master species, *m*, in the association half-reaction for exchange species i_e, and K_{i_e} is a half-reaction selectivity constant. The values of c_{m,i_e} may be positive or negative. For PHREEQC, terms on the right-hand side of an association reaction are assigned negative coefficients and terms on the left-hand side are assigned positive coefficients.

For an exchange species, the equation for the total moles of species i_e is

$$n_{i_e} = K_{i_e} \frac{\prod_m^M a_m^{c_{m,i_e}}}{\gamma_{i_e}\left(\dfrac{b_{e,i_e}}{T_e}\right)}.$$ (11)

The natural log of the activity of the master species of the exchanger is a master unknown in the numerical method. The total derivative of the moles of species i_e with respect to the master unknowns is

$$dn_{i_e} = n_{i_e}\left(\sum_m^M c_{m,i_e}d\ln(a_m) - \frac{\partial}{\partial\mu}\ln(\gamma_{i_e})d\mu\right).$$ (12)

For data input to PHREEQC, the chemical equation for the mole-balance and mass-action expressions, the log K and its temperature dependence, and, optionally, the activity-coefficient expression for each exchange species are defined through the **EXCHANGE_SPECIES** data block. Exchange master species are defined with the **EXCHANGE_MASTER_SPECIES** data block. The number of exchange sites and exchanger composition are defined with the **EXCHANGE** data block (see "Description of Data Input").

Surface Species

Surface-complexation processes are included in the model through heterogeneous mass-action equations, mole-balance equations for surface sites, and charge-potential relations for each surface. PHREEQC allows multiple surfaces and surface-site types, termed a "surface assemblage", to exist in equilibrium with the aqueous phase. Two formulations of the mass-action equations for surface species are available in PHREEQC: (1) one that includes electrostatic potential terms and (2) another that excludes all electrostatic potential terms. If the Dzombak and Morel (1990) model, which includes electrostatic potential terms, is used, additional equations and mass-action terms become operational because of surface charge and surface electrostatic potential.

The two principle differences between the formulation of exchange reactions and surface reactions are that exchange reactions are formulated as half-reactions, which causes the master species not to appear in any mole-balance equations, and the exchange species are expected to be neutral. Surface reactions are not half-reactions, so the master species is a physically real species and appears in mole-balance equations, and surface species may be anionic, cationic, or neutral.

The basic theory for surface-complexation reactions including electrostatic potentials is presented in Dzombak and Morel (1990). The theory assumes that the number of active sites, T_s (eq), the specific area, A_s (m^2/g), and the mass, S_s (g), of the surface are known. The two additional master unknowns are (1) the quantity,

$$\ln a_{\Psi_s} = \ln\left(e^{\frac{F\Psi_s}{2RT}}\right) = \frac{F\Psi_s}{2RT}, \text{ where } F \text{ is the Faraday constant (96493.5 J V}^{-1} \text{ eq}^{-1}), \Psi_s \text{ is the potential at sur-}$$

faces (volts), R is the gas constant (8.3147 J mol^{-1} K^{-1}), and T is temperature (Kelvin) and (2) the natural log of the activity of the master surface species. Note that the quantity $\ln a_{\Psi_s}$ is defined with a 2 in the denominator of the term on the right-hand side. This is a different master unknown than that used in Dzombak and Morel (1990), but produces the same results as their model because all equations are written to be consistent with this master unknown.

The activity of a surface species is assumed to be equal to the mole fraction of a given surface-site type that is occupied. In other words, a surface species is in the standard state (has activity of 1) when it completely covers a given kind of surface site. This convention differs from Dzombak and Morel (1990) who assumed that activity of a surface species (conceptually in the solid phase) is numerically equal to molarity (concentration in solution). If only monodentate complexes are considered (as is done by Dzombak and Morel, 1990), terms cancel in the mass-action equation and identical numerical results are obtained irrespective of the convention for standard state. However, a notable difference in surface site concentration exists when the molarity convention is used for multidentate complexes (bidentate, tridentate, and others, cf. Appelo and Postma, 1999). If a vessel contains a solution in equilibrium with a surface containing multidentate species, and more of exactly the same solution is added, the composition of solution and surface would change with the molarity convention. The molarity convention is clearly not correct in this case.

"*Hfo*" (*H*ydrous *f*erric *o*xide) is used in the default database files with "_w", which indicates a low affinity or weak site and "_s", which indicates a high affinity or strong site. "*Hfo_wOH*" is used to represent a neutral surface species at a weak site and the association reaction for the formation of a negatively charged weak site (it is an association reaction in the sense that the defined species is on the right hand side of the equation) can be written as

$$Hfo\_wOH \rightarrow Hfo\_wO^- + H^+. \tag{13}$$

The mass-action expression, which includes the electrostatic potential term, is

$$K^{int}_{Hfo\_wO^-} = \frac{a_{Hfo\_wO^-} a_{H^+}}{a_{Hfo\_wOH}} e^{-\frac{F\Psi_s}{RT}}, \tag{14}$$

where $K^{int}_{Hfo\_wO^-}$ is the intrinsic equilibrium constant for the reaction, and $e^{-\frac{F\Psi_s}{RT}}$ is a factor that accounts for the work involved in moving a charged species (H$^+$) away from a charged surface. In general, the mass-action equation for surface species i_{s_k} is

$$K_{i_{(s_k)}}^{int} = \left(a_{i_{(s_k)}} \prod_m^M a_m^{-c_{m, i_{(s_k)}}}\right) e^{\frac{F\Psi_s}{RT}\Delta z_{i_{(s_k)}}} \tag{15}$$

where $K_{i_{(s_k)}}^{int}$, is the intrinsic equilibrium constant; $i_{(s_k)}$ is the ith surface species for surface-site type k (weak or strong in Dzombak and Morell, 1990) in surface s; m varies over all master species, M, including surface master species; c_{m, i_s} is the stoichiometric coefficient of master species, m, in the association reaction for surface species $i_{(s_k)}$, and $\Delta z_{i_{(s_k)}}$ is the net change in surface charge due to the formation of the surface species. The values of $c_{m, i_{(s_k)}}$ may be positive or negative. For PHREEQC, terms on the right-hand side of an association reaction are assigned negative coefficients and terms on the left-hand side are assigned positive coefficients.

For a surface species, the equation for the total moles of species $i_{(s_k)}$ is

$$\begin{aligned} n_{i_{(s_k)}} &= a_{i_{(s_k)}}\frac{T_{s_k}}{b_{i_{(s_k)}}} = K_{i_{(s_k)}} T_{s_k} e^{\left(-\frac{F\Psi_s}{RT}\Delta z_{i_{(s_k)}}\right)}\prod_m^M a_m^{c_{m, i_{(s_k)}}} \\ &= K_{i_{(s_k)}}\frac{T_{s_k}}{b_{i_{(s_k)}}}a_{\Psi_s}^{-2\Delta z_{i_{(s_k)}}}\prod_m^M a_m^{c_{m, i_{(s_k)}}} \end{aligned} \tag{16}$$

where T_{s_k} is the total number of a type of surface site, and $b_{i_{(s_k)}}$ is the number of surface sites bounded to the species. The total derivative of the moles of species $i_{(s_k)}$ with respect to the master unknowns is

$$dn_{i_{(s_k)}} = n_{i_{(s_k)}}\left[\sum_m^M c_{m, i_{(s_k)}}d\ln a_m - 2\Delta z_{i_{(s_k)}}d\ln a_{\Psi_s}\right]. \tag{17}$$

The second formulation of mass-action equations for surface species excludes the electrostatic potential term in the mass-action expression (**-no_edl** identifier in the **SURFACE** data block). The equation for the moles of a surface species is the same as equation 16, except the factor involving a_{Ψ_s} does not appear. Likewise, the total derivative of the moles is the same as equation 17, except the final term is absent.

For data input to PHREEQC, the chemical equation for the mole-balance and mass-action expressions and the log K and its temperature dependence of surface species are defined through the **SURFACE_SPECIES** data block. Surface master species or types of surface sites are defined with the **SURFACE_MASTER_SPECIES** data block. The identity of the surfaces and the number of equivalents of each site type, the composition of the surface, the specific surface area, and the mass of the surface are defined with the **SURFACE** data block (see "Description of Data Input").

Gas-Phase Components

Equilibrium between a multicomponent gas phase and the aqueous phase is modeled with heterogeneous mass-action equations and an equation for total pressure (fixed-pressure gas phase only). Only one gas phase can exist in equilibrium with the aqueous phase, but the gas phase may contain multiple components. All gas components are assumed to behave ideally and the gas phase is assumed to be an ideal mixture of gas components.

If a gas phase is specified to have a fixed volume, then the pressure in the gas volume will vary with reaction extent, but each gas component will always be present in the gas phase. For a fixed-volume gas phase, no additional

master unknowns are needed, and the moles of a component in the gas phase can be calculated from the activities of the aqueous master species.

If a gas phase is specified to have a fixed pressure, the gas phase is a fixed-pressure bubble that will vary in volume with reaction extent. If the sum of the partial pressures of the component gases is less than the specified total pressure, the fixed-pressure gas phase will not exist and none of the gas components will be present in the gas phase. For a fixed-pressure gas phase, one additional master unknown is included in the equations, which is the total moles of gas components in the gas phase, N_{gas}.

By the assumption of ideality, the fugacity (activity) of a gas component is equal to its partial pressure. PHREEQC uses dissolution equations, in the sense that the gas component is assumed to be on the left-hand side of the chemical reaction. For carbon dioxide, the dissolution reaction may be written as

$$CO_{2(g)} = CO_{2(aq)}. \tag{18}$$

The Henry's law constant relates the partial pressure of the gas component (numerically equal to fugacity for ideal gases) to the activity of aqueous species. For carbon dioxide, the Henry's law constant is $10^{-1.468}$ [following the ideal gas assumption, units are atmospheres (atm)], and the following mass-action equation applies at equilibrium:

$$P_{CO_2} = 10^{1.468} a_{CO_{2(aq)}}, \tag{19}$$

where P_{CO_2} is the partial pressure (atm) calculated using activities in the aqueous phase. In general, the partial pressure of a gas component may be written in terms of aqueous phase activities as

$$P_g = \frac{1}{K_g} \prod_m^{M_{aq}} a_m^{c_{m,g}}, \tag{20}$$

where P_g is the partial pressure of gas component g, calculated using activities in the aqueous phase; K_g is the Henry's law constant for the gas component; and $c_{m,g}$ is the stoichiometric coefficient of aqueous master species, m, in the dissolution equation. The values of $c_{m,g}$ may be positive or negative. For PHREEQC, terms on the left-hand side of a dissolution reaction are assigned negative coefficients and terms on the right-hand side are assigned positive coefficients.

For a fixed-volume gas phase, the total volume of the gas phase is specified to be V_{total}, but the pressure of the gas phase is variable. At equilibrium, the number of moles of a gas component in the gas n_g is calculated as

$$n_g = \frac{V_{total} P_g}{RT} = \frac{V_{total}}{RTK_g} \sum_m^{M_{aq}} a_m^{c_{m,g}}. \tag{21}$$

The total derivative of the moles of a gas component in the gas phase is

$$dn_g = \frac{V_{total}}{RT} \sum_m^{M_{aq}} n_g c_{m,g} d\ln a_m. \tag{22}$$

For a fixed-pressure gas phase, the total pressure is specified as P_{total}, but the volume of the gas phase is variable. At equilibrium, the number of moles of a gas component in the gas phase is equal to the fraction of the total pressure for the gas times the total moles of gas in the gas phase:

$$n_g = N_{gas}\frac{P_g}{P_{total}} = \frac{N_{gas}}{P_{total}K_g}\prod_m^{M_{aq}}a_m^{c_{m,g}}. \tag{23}$$

The total derivative of the moles of a gas component in the gas phase is

$$dn_g = \frac{P_g}{P_{total}}dN_{gas} + \sum_m^{M_{aq}}n_g c_{m,g}d\ln a_m. \tag{24}$$

For data input to PHREEQC, the mass-action equations, Henry's law constant, and temperature dependence of the constant are defined with the **PHASES** data block. The type of gas phase (fixed-volume or fixed-pressure), the components to include in gas-phase calculations, and initial gas-phase composition are defined with the **GAS_PHASE** data block (see "Description of Data Input").

Equations for the Newton-Raphson Method

A series of functions, denoted by f, are used to describe heterogeneous equilibrium. These equations are derived primarily by substituting the equations for the moles of species (derived from mass-action equations in the previous section) into mole- and charge-balance equations. When equilibrium is satisfied, all of the functions relevant to a specific equilibrium calculation are equal to zero. The zeros of the functions are found by the Newton-Raphson method, by which each function is differentiated with respect to each master unknown to form the Jacobian matrix. A set of linear equations is formed from the Jacobian matrix that can be solved to approximate a solution to the nonlinear equations. By iteratively solving successive sets of linear equations, a solution to the nonlinear equations can be found. Each of the f functions that is used in the numerical method is presented in this section along with the total derivative with respect to the master unknowns that is used to form the Jacobian matrix.

Activity of Water

The activity of water is calculated from an approximation that is based on Raoult's law (Garrels and Christ, 1965, p. 65-66):

$$a_{H_2O} = 1 - 0.017\sum_i^{N_{aq}}\frac{n_i}{W_{aq}}. \tag{25}$$

The function f_{H_2O} is defined as

$$f_{H_2O} = W_{aq}(a_{H_2O} - 1) + 0.017\sum_i^{N_{aq}}n_i, \tag{26}$$

and the total derivative of this function is

$$df_{H_2O} = W_{aq}a_{H_2O}d\ln(a_{H_2O}) + (a_{H_2O} - 1)W_{aq}d\ln(W_{aq}) + 0.017\sum_i^{N_{aq}}dn_i. \tag{27}$$

The master unknown is the natural log of the activity of water $\ln a_{H_2O}$.

Ionic Strength

The ionic strength of the aqueous solution is a master unknown and is defined as

$$\mu = \frac{1}{2}\sum_i^{N_{aq}} z_i^2 \frac{n_i}{W_{aq}}. \tag{28}$$

The function f_μ is defined as

$$f_\mu = W_{aq}\mu - \frac{1}{2}\sum_i^{N_{aq}} z_i^2 n_i, \tag{29}$$

and the total derivative of this function is

$$df_\mu = \mu W_{aq} d\ln(W_{aq}) + W_{aq} d\mu - \frac{1}{2}\sum_i^{N_{aq}} z_i^2 dn_i. \tag{30}$$

Equilibrium with a Fixed-Volume Multicomponent Gas Phase

For a fixed-volume gas phase, the moles of each gas component can be calculated from the activities of the aqueous master species, and the numerical model treats the gas phase components in the same way that it treats aqueous species. The terms for the moles of each gas components, n_g, appear in the mole-balance equations for elements and the terms dn_g appear in the Jacobian matrix for the mole-balance equations. No additional equation labeled f is required to calculate equilibrium with the fixed-volume gas phase.

For data input to PHREEQC, the mass-action equations, Henry's law constant, and temperature dependence of the constant are defined with the **PHASES** data block. The type of gas phase (fixed-volume or fixed-pressure), the components to include in gas-phase calculations, and initial gas-phase composition are defined with the **GAS_PHASE** data block (see "Description of Data Input").

Equilibrium with a Fixed-Pressure Multicomponent Gas Phase

For a fixed-volume gas phase, the number of moles of each gas component is calculated from the activities of the aqueous master species and the total moles of gas components in the gas phase, N_g. The terms for the moles of each gas components, n_g, appear in the mole-balance equations for elements and the terms dn_g appear in the Jacobian matrix for the mole-balance equations. Equilibrium between a fixed-pressure multicomponent gas phase and the aqueous phase requires one new equation--the sum of the partial pressures of the component gases is equal to the total pressure, P_{total}. The function $f_{P_{total}}$ is defined as

$$f_{P_{total}} = P_{total} - \sum_g^{N_g} P_g, \tag{31}$$

where N_g is the total number of gas components in the gas phase.

The total derivative of $f_{P_{total}}$ with respect to the master unknowns, with the convention that positive dN_{gas} are increases in solution concentration, is

$$df_{P_{total}} = -\sum_{g}^{N_g} \sum_{m}^{M_{aq}} c_{m,g} P_g d\ln a_m. \tag{32}$$

For data input to PHREEQC, the mass-action equations, Henry's law constant, and temperature dependence of the constant are defined with the **PHASES** data block. The type of gas phase (fixed-volume or fixed-pressure), the components to include in gas-phase calculations, and initial gas-phase composition are defined with the **GAS_PHASE** data block (see "Description of Data Input").

Equilibrium with Pure Phases

Equilibrium between the aqueous phase and pure phases, including gases with fixed partial pressures, is included in the model through heterogeneous mass-action equations. PHREEQC allows multiple pure phases, termed a pure-phase assemblage, to exist in equilibrium with the aqueous phase, subject to the limitations of the Gibbs' Phase Rule. The activity of a pure phase is assumed to be identically 1.0. The additional master unknown for each pure phase is the moles of the pure phase that is present in the system, n_p, where p refers to the pth phase. Terms representing the changes in the moles of each pure phase occur in the mole-balance equations for elements. PHREEQC also allows a calculation where equilibrium with a pure phase is produced by adding or removing a specified reactant (*alternative formula* and *alternative phase* in **EQUILIBRIUM_PHASES** data block); the mole transfer of the reactant that is necessary to produce equilibrium with the pure phase is calculated. In this type of calculation, the terms in the mole-balance equations are derived from the stoichiometry of the reactant rather than the stoichiometry of the pure phase, and the unknown is the number of moles of reactant that enter or leave solution.

The new function corresponding to each of the new unknowns is a mass-action expression for each pure phase. PHREEQC uses dissolution reactions, in the sense that the pure phase is on the left-hand side of the chemical equation. For calcite, the dissolution reaction may be written as

$$CaCO_3 = Ca^{2+} + CO_3^{2-}, \tag{33}$$

and, using log K of $10^{-8.48}$ and activity of the pure solid of 1.0, the resulting mass-action expression is

$$K_{calcite} = 10^{-8.48} = a_{Ca^{2+}} a_{CO_3^{2-}}. \tag{34}$$

In general, pure-phase equilibria can be represented with the following equation:

$$K_p = \prod_{m}^{M_{aq}} a_m^{c_{m,p}}, \tag{35}$$

where $c_{m,p}$ is the stoichiometric coefficient of master species m in the dissolution reaction. The values of $c_{m,p}$ may be positive or negative. For PHREEQC, terms on the left-hand side of a dissolution reaction are assigned negative coefficients and terms on the right-hand side are assigned positive coefficients. The saturation index for the mineral, SI_p, is defined to be

$$SI_p = \log \prod_{m}^{M_{aq}} a_m^{c_{m,p}}. \tag{36}$$

The function used for phase equilibrium in the numerical method is

$$f_p = (\ln K_p + [\ln(10)]SI_{p,\,target}) - \sum_m^{M_{aq}} c_{m,\,p}\ln(a_m)\,, \tag{37}$$

where $SI_{p,\,target}$ is the target saturation index for the phase, and $\ln(10)$ converts base-10 log to natural log. The target saturation index is specified by the user; a positive, zero, or negative value specifies supersaturation, equilibrium, or undersaturation for the mineral with respect to the solution. For fixed-partial-pressure gas component, $SI_{p,\,target}$ is equivalent to the log of the partial pressure of the gas component. The total derivative with respect to the master unknowns is

$$df_p = -\sum_m^{M_{aq}} c_{m,\,p}d\ln a_m\,. \tag{38}$$

For data input to PHREEQC, the mass-action equations, equilibrium constant, and temperature dependence of the constant for a pure phase are defined with the **PHASES** data block. Initial composition of a pure-phase assemblage and target saturation indices are defined with the **EQUILIBRIUM_PHASES** data block.

Equilibrium with Solid Solutions

Modeling of ideal, multicomponent or nonideal, binary solid solutions is based on the work of Glynn (Glynn and Reardon, 1990; Glynn and others, 1990; Glynn, 1991; Glynn and Parkhurst, 1992). Equilibrium between the aqueous phase and solid solutions is included in the model through heterogeneous mass-action equations. PHREEQC allows multiple solid solutions, termed a solid-solution assemblage, to exist in equilibrium with the aqueous phase, subject to the limitations of the Gibbs' Phase Rule. Modeling of nonideal solid solutions is limited to two-component (binary) solid solutions; ideal solid solutions may have two or more components. The additional master unknowns for solid solutions are the moles of each component in each solid solution $n_{p_{ss}}$, where *ss* refers to solid solution *ss*. Terms representing the changes in the moles of each component occur in the Jacobian matrix of the mole-balance equations for elements.

Unlike pure phases, the activity of a component in a solid solution is not identically 1.0. The activity of a component is defined to be $a_{p_{ss}} = \lambda_{p_{ss}} x_{p_{ss}}$, where $x_{p_{ss}}$ is the mole fraction of component *p* in the solid solution *ss*, and $\lambda_{p_{ss}}$ is the activity coefficient. The mole fraction of a component in a solid solution is defined as

$$x_{p_{ss}} = \frac{n_{p_{ss}}}{\sum\limits_{p_{ss}=1}^{N_{ss}} n_{p_{ss}}}\,, \text{ where } N_{ss} \text{ is the number of components in solid solution } ss. \text{ For ideal solid solutions, the}$$

activity coefficient is 1.0; for nonideal, binary solid solutions, the activity coefficients for the components are defined with the Guggenheim expressions:

$$\lambda_1 = \exp((a_0 - a_1(4x_1 - 1))x_2^2) \text{ and} \tag{39}$$

$$\lambda_2 = \exp((a_0 + a_1(4x_2 - 1))x_1^2)\,, \tag{40}$$

where λ_1 and λ_2 are the activity coefficients of components 1 and 2, and a_0 and a_1 are nondimensional Guggenheim parameters. The nondimensional parameters are calculated from dimensional parameters for the

excess free energy g_0 and g_1 (kJ/mol) by the equations: $a_0 = \dfrac{g_0}{RT}$ and $a_1 = \dfrac{g_1}{RT}$. The parameters a_0 and a_1 for the excess free energy may be defined directly or by a variety of means including the mole fractions of component 2 delimiting the miscibility gap, the mole fractions of component 2 delimiting the spinodal gap, the mole fraction of component 2 at the critical point and the critical temperature, Thompson and Waldbaum parameters, Margules parameters, mole fraction of component 2 and the log of the total solubility product of an alyotropic point, solid-phase activity coefficients for trace concentrations of component 1 and component 2, or two distribution coefficients for component 2 (Glynn, 1991).

The new function corresponding to each of the new unknowns is a mass-action expression for each component in each solid solution. PHREEQC uses dissolution reactions, in the sense that the solid-solution component is on the left-hand side of the chemical equation. For aragonite in an aragonite-strontianite solid solution, the dissolution reaction may be written as

$$CaCO_3 = Ca^{2+} + CO_3^{2-}, \tag{41}$$

and, using log K of $10^{-8.34}$ and activity coefficient for the solid, the resulting mass-action expression is

$$K_{Arag} = 10^{-8.34} = \frac{a_{Ca^{2+}} a_{CO_3^{2-}}}{a_{Arag}} = \frac{a_{Ca^{2+}} a_{CO_3^{2-}}}{\lambda_{Arag}\left(\dfrac{n_{Arag}}{n_{Arag} + n_{Stront}}\right)}. \tag{42}$$

In general, solid-solution phase equilibria can be represented with the following equation for each component:

$$K_{p_{ss}} = \frac{\displaystyle\prod_{m}^{M_{aq}} a_m^{c_{m,\,p_{ss}}}}{a_{p_{ss}}}, \tag{43}$$

where $K_{P_{ss}}$ is the equilibrium constant of component p in pure form, and $c_{m,\,p_{ss}}$ is the stoichiometric coefficient of master species m in the dissolution reaction for component p in solid solution ss. The values of $c_{m,\,p_{ss}}$ may be positive or negative. For PHREEQC, terms on the left-hand side of a phase dissolution reaction are assigned negative coefficients and terms on the right-hand side are assigned positive coefficients. The solubility quotient for a component of the solid solution is defined to be

$$Q_{p_{ss}} = \frac{\displaystyle\sum_{m}^{M_{aq}} a_m^{c_{m,\,p_{ss}}}}{K_{p_{ss}} a_{p_{ss}}}, \tag{44}$$

where $Q_{p_{ss}}$ is equal to 1 and $\ln Q_{p_{ss}}$ is equal to 0 at equilibrium. The functions used in the numerical method for each component of a nonideal, binary solid solution are

$$f_1 = \sum_{m}^{M_{aq}} c_{m,\,1} \ln a_m - \ln K_1 - \ln \frac{n_1}{n_1 + n_2} - \ln \lambda_1 \text{ and} \tag{45}$$

$$f_2 = \sum_{m}^{M_{aq}} c_{m,\,2} \ln a_m - \ln K_2 - \ln \frac{n_2}{n_1 + n_2} - \ln \lambda_2. \tag{46}$$

The total derivative with respect to the master unknowns is

$$df_1 = \sum_m^{M_{aq}} c_{m,1} d\ln a_m + \left(-\frac{x_2}{n_1} + \frac{2a_0 x_2^2 - 6a_1 x_2^2 + 12a_1 x_2^3}{n_1 + n_2} \right) dn_1 +$$

$$\frac{-2a_0 x_2 + 2a_0 x_2^2 + 6a_1 x_2 - 18a_1 x_2^2 + 12a_1 x_2^3 + 1}{n_1 + n_2} dn_2 \qquad (47)$$

and

$$df_2 = \sum_m^{M_{aq}} c_{m,2} d\ln a_m + \frac{-2a_0 x_1 + 2a_0 x_1^2 - 6a_1 x_1 + 18a_1 x_1^2 - 12a_1 x_1^3 + 1}{n_1 + n_2} dn_1 +$$

$$\left(-\frac{x_1}{n_2} + \frac{2a_0 x_1^2 + 6a_1 x_1^2 - 12a_1 x_1^3}{n_1 + n_2} \right) dn_2 \qquad . \qquad (48)$$

The function used in the numerical method for each component of an ideal solid solution is

$$f_{p_{ss}} = \ln Q_{p_{ss}} = \ln \left(\frac{\sum_m^{M_{aq}} a_m^{c_{m,p_{ss}}}}{K_{p_{ss}}} \right) - \ln \left(\frac{n_{p_{ss}}}{N_{total}} \right), \qquad (49)$$

where $N_{total} = \sum_{j_{ss}}^{N_{ss}} n_{j_{ss}}$ and j_{ss} ranges over all the components in solid solution ss. The total derivative with respect to the master unknowns is

$$df_{p_{ss}} = \sum_m^{M_{aq}} c_{m,p_{ss}} d\ln a_m - \frac{1}{n_{p_{ss}}} \left(\frac{N_{total} - n_{p_{ss}}}{N_{total}} \right) dn_{p_{ss}} + \sum_{j_{ss}}^{N_{ss}, j_{ss} \neq p_{ss}} \frac{1}{N_{total}} dn_{j_{ss}}. \qquad (50)$$

For data input to PHREEQC, the mass-action equations, equilibrium constant, and temperature dependence of the constant for each pure phase are defined with the **PHASES** data block. Initial composition of a solid-solution assemblage and Guggenheim parameters for nonideal solid solutions are defined with the **SOLID_SOLUTIONS** data block (see "Description of Data Input").

Mole Balance for Surface Sites

Mole balance for a surface site is a special case of the general mole-balance equation. The surface assemblage is a set of one or more surfaces, each of which may have one or more site types. The total number of moles of a surface site type is specified by input to be one of the following: (1) fixed, (2) proportional to the moles of a pure phase, or (3) proportional to the moles of a kinetic reactant. The sum of the moles of surface sites occupied by the surface species of a site type must equal the total moles of that surface site type. The following function is derived from the mole-balance relation for a surface site type s_k of surface s:

$$f_{s_k} = T_{s_k} - \sum_{i_{(s_k)}}^{N_{s_k}} b_{s_k, i_{(s_k)}} n_{i_{(s_k)}}, \tag{51}$$

where the value of the function f_{s_k} is zero when mole balance is achieved, T_{s_k} is the moles of the surface site type, N_{s_k} is the number of surface species for the site type, and $b_{s_k, i_{(s_k)}}$ is the number of surface sites occupied by the surface species $i_{(s_k)}$. The total derivative of f_{s_k} is

$$df_{s_k} = \Delta T_{s_k} - \sum_{i_{(s_k)}}^{N_{s_k}} b_{s_k, i_{(s_k)}} dn_{i_{(s_k)}}. \tag{52}$$

If the total number of sites is proportional to the moles of a pure phase, then $\Delta T_{s_k} = -c_{s_k, p} dn_p$, where $c_{s_k, p}$ is the moles of surface sites per mole of phase p. If the phase dissolves, then dn_p is positive and the number of surface sites decreases. If the total number of sites is proportional to the moles of a kinetic reactant, $\Delta T_{s_k} = 0$ in the total derivative equation. The change in the number of sites is included as part of the reaction that is integrated with the rate equations and no term is included in the Jacobian matrix. As the kinetic reaction increases or decreases the moles of reactant, the number of surface sites is adjusted proportionately. If the number of surface sites is fixed, $\Delta T_{s_k} = 0$.

For data input to PHREEQC, the number of moles of each type of surface site is defined with the **SURFACE** data block and may be a fixed quantity or it may be related to the moles of a pure phase or a kinetic reactant. Surface site types are defined with the **SURFACE_MASTER_SPECIES** data block and surface species are defined with the **SURFACE_SPECIES** data block (see "Description of Data Input").

Mole Balance for Exchange Sites

Mole balance for an exchange site is a special case of the general mole-balance equation. The total number of moles of an exchange site is specified by input to be one of the following: (1) fixed, (2) proportional to the moles of a pure phase, or (3) proportional to the moles of a kinetic reactant. The sum of the moles of sites occupied by exchange species must equal the total moles of the exchange site. The following function is derived from the mole-balance relation for an exchange site:

$$f_e = T_e - \sum_{i_e}^{N_e} b_{e, i_e} n_{i_e}, \tag{53}$$

where the value of the function f_e is zero when mole balance is achieved, T_e is the total moles of exchange sites for exchanger e, and b_{e, i_e} is the number of exchange sites occupied by the exchange species. The total derivative of f_e is

$$df_e = \Delta T_e - \sum_{i_e}^{N_e} b_{e, i_e} dn_{i_e}. \tag{54}$$

If the total number of sites is proportional to the moles of a pure phase, then $\Delta T_e = -c_{e, p} dn_p$, where $c_{e, p}$ is the moles of exchange sites per mole of phase p. If the phase dissolves, then dn_p is positive and the number of exchange sites decreases. If the total number of sites is proportional to the moles of a kinetic reactant, $\Delta T_e = 0$ in

the total derivative equation. The change in the number of sites is included as part of the reaction that is integrated with the rate equations and no term is included in the Jacobian matrix. As the kinetic reaction increases or decreases the moles of the reactant, the number of exchange sites is adjusted proportionately. If the number of exchange sites is fixed, $\Delta T_e = 0$.

For data input to PHREEQC, the moles of exchange sites are defined in the **EXCHANGE** data block and may be a fixed quantity or it may be related to the moles of a pure phase or a kinetic reactant. Exchanger sites are defined with the **EXCHANGE_MASTER_SPECIES** data block and exchange species are defined with the **EXCHANGE_SPECIES** data block (see "Description of Data Input").

Mole Balance for Alkalinity

The mole-balance equation for alkalinity is used only in speciation calculations and in inverse modeling. Mole balance for alkalinity is a special case of the general mole-balance equation where the coefficients are defined by the alkalinity contribution of each aqueous species. Alkalinity is defined as an element in PHREEQC and a master species is associated with this element (see **SOLUTION_MASTER_SPECIES** keyword in "Description of Data Input"). In the default databases for PHREEQC, the master species for alkalinity is CO_3^{2-}. The master unknown for alkalinity is $\ln a_{Alk}$, or for the default databases, $\ln a_{CO_3^{2-}}$.

The total number of equivalents of alkalinity is specified by input to the model. The sum of the alkalinity contribution of each aqueous species must equal the total number of equivalents of alkalinity. The following function is derived from the alkalinity-balance equation:

$$f_{Alk} = T_{Alk} - \sum_{i}^{N_{aq}} b_{Alk,i} n_i, \tag{55}$$

where the value of the function f_{Alk} is zero when mole balance is achieved, T_{Alk} is the number of equivalents of alkalinity in solution, and $b_{Alk,i}$ is the alkalinity contribution of the aqueous species i (eq/mol). The total derivative of f_{Alk} is

$$df_{Alk} = -\sum_{i}^{N_{aq}} b_{Alk,i} dn_i. \tag{56}$$

The value of T_{Alk} must be positive, provided a carbonate species is the master species for alkalinity. Conceptually, a measured alkalinity differs from the alkalinity calculated by PHREEQC. In the default database files for PHREEQC the values of $b_{Alk,i}$ have been chosen such that the reference state ($b_{Alk,i} = 0$) for each element or element valence state is the predominant species at a pH of 4.5. It is assumed that all of the element or element valence state is converted to this predominant species in an alkalinity titration. However, significant concentrations of aqueous species that are not in the reference state (that is species that have nonzero alkalinity contributions) may exist at the endpoint of a titration, and the extent to which this occurs causes the alkalinity calculated by PHREEQC to be a different quantity than the measured alkalinity. Hydroxide complexes of iron and aluminum are the most common examples of species that may not be converted to the defined reference state. Thus, the alkalinity of a solution as calculated by PHREEQC, though it will be numerically equal to the measured alkalinity, is an approximation because of the assumption that a titration totally converts elements and element valence states to

their reference state. In most solutions, where the alkalinity is derived predominantly from carbonate species, the approximation is valid.

For data input to PHREEQC, the alkalinity of each species is calculated from the association reaction for the species, which is defined in the **SOLUTION_SPECIES** data block, and the alkalinity contributions of the master species, which are defined with the **SOLUTION_MASTER_SPECIES** data block. Total alkalinity is part of the solution composition defined with the **SOLUTION** or **SOLUTION_SPREAD** data block (see "Description of Data Input").

Mole Balance for Elements

The total moles of an element in the system are the sum of the moles initially present in the pure-phase and solid-solution assemblages, aqueous phase, exchange assemblage, surface assemblage, gas phase, and diffuse layers of the surfaces. The following function is the general mole-balance equation:

$$
f_m = \left(T_m - \sum_p^{N_p} b_{m,p} n_p - \sum_{ss}^{SS} \sum_{p_{ss}}^{N_{ss}} b_{m,p_{ss}} n_{p_{ss}} \right) - \sum_i^{N_{aq}} b_{m,i} n_i - \sum_e^{E} \sum_{i_e}^{N_e} b_{m,i_e} n_{i_e} -
$$
$$
\sum_s^{S} \sum_k^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} b_{m,i_{(s_k)}} n_{i_{(s_k)}} - \sum_g^{N_g} b_{m,g} n_g - \sum_s^{S} \sum_i^{N_{aq}} b_{m,i} n_{i,s}
$$

(57)

where the value of the function f_m is zero when mole-balance is achieved, T_m is the total moles of the element in the system, N_p is the number of phases in the pure-phase assemblage, SS is the number of solid solutions in the solid-solution assemblage, N_{ss} is the number of components in solid solution ss, N_{aq} is the number of aqueous species, E is the number of exchangers in the exchange assemblage, N_e is the number of exchange species for exchange site e, S is the number of surfaces in the surface assemblage, K_s is the number of surface types for surface s, N_{s_k} is the number of surface species for surface type s_k, and N_g is the number of gas-phase components. The moles of each entity in the system are represented by n_p for phases in the pure-phase assemblage, $n_{p_{ss}}$ for components in a solid solution, n_i for aqueous species, n_{i_e} for the exchange species of exchange site e, $n_{i_{(s_k)}}$ for surface species for surface site type s_k, n_g for the gas components, and $n_{i,s}$ for aqueous species in the diffuse layer of surface s. The moles of element m per mole of each entity are represented by b_m, with an additional subscript to define the relevant entity; b_m is usually, but not always, equal to c_m (the coefficient of the master species for m in the mass-action equation).

To avoid solving for small differences between large numbers, the quantity in parenthesis in equation 57 is not explicitly included in the solution algorithm and the value of T_m is never actually calculated. Instead the quantity

$$
{}^*T_m = T_m - \sum_p^{N_p} b_{m,p} n_p - \sum_{ss}^{SS} \sum_{p_{ss}}^{N_{ss}} b_{m,p_{ss}} n_{p_{ss}}
$$
is used in the function f_m. Initially, ${}^*T_m$ is calculated from the total

moles of m in the aqueous phase, the exchange assemblage, the surface assemblage, the gas phase, and the surface diffuse layers:

$$\, ^{*}T_m = \sum_i^{N_{aq}} b_{m,i} n_i + \sum_e^{E} \sum_{i_e}^{N_e} b_{m,i_e} n_{i_e} + \sum_s^{S} \sum_k^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} b_{m,i_{(s_k)}} n_{i_{(s_k)}} + \sum_g^{N_g} b_{m,g} n_g + \sum_s^{S} \sum_i^{N_{aq}} b_{m,i} n_{i,s}. \tag{58}$$

During the iterative solution to the equations, $\, ^{*}T_m$ is updated by the mole transfers of the pure phases and components of the solid solutions:

$$\, ^{*}T_{k+1}^m = \, ^{*}T_k^m + \sum_p^{N_p} b_{m,p} dn_p + \sum_{ss}^{SS} \sum_{p_{ss}}^{N_{ss}} b_{m,p_{ss}} dn_{p_{ss}}, \tag{59}$$

where k refers to the iteration number. It is possible for $\, ^{*}T_m$ to be negative in intermediate iterations, but it must be positive when equilibrium is attained.

The total derivative of the function f_m is

$$df_m = -\sum_p^{N_p} b_{m,p} dn_p - \sum_{ss}^{SS} \sum_{p_{ss}}^{N_{ss}} b_{m,p_{ss}} dn_{p_{ss}} - \sum_i^{N_{aq}} b_{m,i} dn_i - \sum_e^{E} \sum_{i_e}^{N_e} b_{m,i_e} dn_{i_e} -$$
$$- \sum_s^{S} \sum_k^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} b_{m,i_{(s_k)}} dn_{i_{(s_k)}} - \sum_g^{N_g} b_{m,g} dn_g - \sum_s^{S} \sum_i^{N_{aq}} b_{m,i} dn_{i,s} \tag{60}$$

For data input to PHREEQC, total moles of elements are initially defined for an aqueous phase with the **SOLUTION** or **SOLUTION_SPREAD** data block, for an exchange assemblage with the **EXCHANGE** data block, for a surface assemblage with the **SURFACE** data block, for the gas phase with a **GAS_PHASE** data block. The moles of each phase in a pure-phase assemblage are defined with the **EQUILIBRIUM_PHASES** data block. The moles of each component in each solid solution in a solid-solution assemblage are defined with the **SOLID_SOLUTIONS** data block. Total moles of elements may also be modified by batch-reaction and transport calculations (see "Description of Data Input").

Aqueous Charge Balance

The charge-balance equation sums the equivalents of aqueous cations and anions and, in some cases, the charge imbalances developed on surfaces and exchangers. When specified, a charge-balance equation is used in initial solution calculations to adjust the pH or the activity of a master species (and consequently the total concentration of an element or element valence state) to produce electroneutrality in the solution. The charge-balance equation is necessary to calculate pH in batch reactions and transport simulations.

In real solutions, the sum of the equivalents of anions and cations must be zero. However, analytical errors and unanalyzed constituents in chemical analyses generally cause electrical imbalances to be calculated for solutions. If a charge imbalance is calculated for an initial solution, the pH is adjusted in subsequent batch reactions or transport simulations to maintain the same charge imbalance. If mixing is performed, the charge imbalance for the batch-reaction step is the sum of the charge imbalances of each solution weighted by its mixing factor. If a surface is used in a simulation and the explicit diffuse-layer calculation is not specified, then the formation of charged surface species will result in a surface charge imbalance. Similarly, if exchange species are not electrically

neutral (all exchange species in the default databases are electrically neutral), the exchanger will accumulate a charge. The charge imbalances of surfaces and exchangers are included in the general charge-balance equation.

The charge imbalance for a solution is calculated in each initial solution calculation, in each batch-reaction step, and for each cell during each time step of transport simulations with the equation:

$$T_{z, q} = \sum_{i}^{N_{aq}} z_i n_i,$$ (61)

where q identifies the aqueous phase, $T_{z, q}$ is the charge imbalance for aqueous phase q, and z_i is the charge on aqueous species i. If charged surfaces or exchangers are not present, the charge imbalance for a solution at the end of a batch-reaction or transport simulation will be the same as at the beginning of the simulation.

The charge imbalance on a surface is calculated in the initial surface-composition calculation, in each batch-reaction step, and for each cell during each time step of transport simulations with the equation:

$$T_{z, s} = \sum_{k}^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} z_{i_{(s_k)}} n_{i_{(s_k)}} + \sum_{i}^{N_{aq}} z_i n_{i, s},$$ (62)

where $T_{z, s}$ is the charge imbalance for the surface, $z_{i_{(s_k)}}$ is the charge on the surface species i of surface type s_k of surface s, and the final term in the equation represents the charge accumulated in the diffuse layer. The final term is used only if the diffuse-layer composition is explicitly included in the calculation (**-diffuse_layer** in the **SURFACE** data block). When the diffuse-layer composition is calculated explicitly, it is required that all solutions be charge balanced, and $T_{z, s}$ will always be equal to zero.

Normally, exchange species have no net charge, but for generality, this is not required. However, the activity of exchange species (the equivalent fraction) is not well defined if the sum of the charged species is not equal to the total number of equivalents of exchange sites (exchange capacity). If charged exchange species exist, then the charge imbalance on an exchanger is calculated in the initial exchange-composition calculation, in each batch-reaction step, and for each cell during each time step of transport simulations with the equation:

$$T_{z, e} = \sum_{i_e}^{N_e} z_{i_e} n_{i_e},$$ (63)

where $T_{z, e}$ is the charge imbalance for the exchanger, and z_{i_e} is the charge on the exchange species i of exchanger e.

The charge imbalance for the system is defined at the beginning of each batch-reaction step and for each cell at the beginning of each time step in transport simulations to be:

$$T_z = \sum_{q}^{Q} \alpha_q T_{z, q} + \sum_{s}^{S} T_{z, s} + \sum_{e}^{E} T_{z, e},$$ (64)

where T_z is the charge imbalance for the system, Q is the number of aqueous phases that are mixed in the batch-reaction step or in the cell for a transport step, α_q is the mixing fraction for aqueous phase q, S is the number of surfaces, and E is the number of exchangers.

The charge-balance function is

$$f_z = T_z - \sum_i^{N_{aq}} z_i n_i - \left(\sum_s^S \sum_k^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} z_{i_{(s_k)}} n_{i_{(s_k)}} + \sum_s^S \sum_i^{N_{aq}} z_i n_{i, s} \right) - \sum_e^E \sum_{i_e}^{N_e} z_{i_e} n_{i_e}, \tag{65}$$

where f_z is zero when charge balance has been achieved. If the diffuse-layer composition is explicitly calculated, a separate charge-balance equation is included for each surface and the sum of the terms in the parentheses will be zero when surface charge balance is achieved. If the diffuse-layer composition is not calculated, the second term inside the parentheses is zero. The total derivative of f_z is

$$df_z = -\sum_i^{N_{aq}} z_i dn_i - \sum_s^S \sum_k^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} z_{i_{(s_k)}} dn_{i_{(s_k)}} - \sum_e^E \sum_{i_e}^{N_e} z_{i_e} dn_{i_e}, \tag{66}$$

where the triple summation for surfaces is present only if the diffuse-layer composition is not explicitly calculated.

For data input to PHREEQC, charge imbalance is defined by data input for **SOLUTION** or **SOLUTION_SPREAD**, **EXCHANGE**, and **SURFACE** data blocks combined with speciation, initial exchange-composition, and initial surface-composition calculations. The charge on a species is defined in the balanced chemical reaction that defines the species in **SOLUTION_SPECIES**, **EXCHANGE_SPECIES**, or **SURFACE_SPECIES** data blocks (see "Description of Data Input").

Surface Charge-Potential Equation with No Explicit Calculation of the Diffuse-Layer Composition

By default, PHREEQC uses the approach described by Dzombak and Morel (1990) to relate the charge density on the surface, σ_s, with the potential at the surface, Ψ_s. The surface-charge density is the amount of charge per area of surface material, which can be calculated from the distribution of surface species:

$$\sigma_s = \frac{F}{A_{surf}} \sum_k^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} z_{i_{(s_k)}} n_{i_{(s_k)}}, \tag{67}$$

where σ_s is the charge density for surface s in coulombs per square meter (C/m^2), F is the Faraday constant in coulombs per mole (96,493.5 C/mol), A_{surf} is the surface area of the material (m^2). The surface area is calculated by one of the following formulas: (1) $A_{surf} = A_s S_s$, where A_s is the specific area of the surface material (m^2/g), and S_s is the mass of surface material (g), or (2) $A_{surf} = A_r n_r$, where A_r is the surface area per mole of a pure phase or kinetic reactant (m^2/mol), and n_r is the moles of the pure phase or reactant. At 25°C, the surface-charge density is related to the electrical potential at the surface by:

$$\sigma_s = (8000 \varepsilon \varepsilon_0 RT)^{\frac{1}{2}} \mu^{\frac{1}{2}} \sinh\left(\frac{\nu F \psi_s}{2RT}\right), \tag{68}$$

where ε is the dielectric constant of water (78.5, dimensionless), ε_0 is the permittivity of free space (8.854×10^{-12} $CV^{-1}m^{-1}$ or C^2/m-J), ν is the ionic charge of a symmetric electrolyte, R is the gas constant (8.314 J mol^{-1} K^{-1}), T is temperature (K), μ is the ionic strength, and F is the Faraday constant (J V^{-1} eq^{-1} or C/mol), Ψ_s is the potential at the surface in volts. At 25°C, $(8000 \varepsilon \varepsilon_0 RT)^{\frac{1}{2}} = 0.1174$. The charge of the electrolyte ions is assumed to be 1.

The charge-potential function is

$$f_{\Psi_s} = (8000\varepsilon\varepsilon_0 RT)^{\frac{1}{2}}\mu^{\frac{1}{2}}\sinh\left(\frac{F\Psi_s}{2RT}\right) - \frac{F}{A_{surf}}\sum_{k}^{K_s}\sum_{i_{(s_k)}}^{N_{s_k}} z_{i_{(s_k)}} n_{i_{(s_k)}},$$ (69)

and the total derivative of this function is

$$df_{\Psi_s} = \frac{(8000\varepsilon\varepsilon_0 RT)^{\frac{1}{2}}}{2}\mu^{-\frac{1}{2}}\sinh\left(\frac{F\Psi_s}{2RT}\right)d\mu + (8000\varepsilon\varepsilon_0 RT)^{\frac{1}{2}}\mu^{\frac{1}{2}}\cosh\left(\frac{F\Psi_s}{2RT}\right)d\ln a_{\Psi_s} -$$

$$\frac{F}{A_{surf}}\sum_{k}^{K_s}\sum_{i_{(s_k)}}^{N_{s_k}} z_{i_{(s_k)}} dn_{i_{(s_k)}}$$. (70)

For data input to PHREEQC, calculation without an explicit diffuse layer is the default. Specific surface area (A_s or A_r) and mass of surface (S_s) are defined in the **SURFACE** data block. The moles of surface sites are defined (1) in the **SURFACE** data block if the number of sites is fixed, (2) by a proportionality factor in the **SURFACE** data block and the moles of a phase in **EQUILIBRIUM_PHASES** data block, or (3) by a proportionality factor in the **SURFACE** data block and the moles of a kinetic reactant in **KINETICS** data block. The charge on a surface species is specified in the balanced chemical reaction that defines the species in the **SURFACE_SPECIES** data block (see "Description of Data Input").

Surface Charge-Balance Equation with Explicit Calculation of the Diffuse-Layer Composition

As an alternative to the previous model for the surface charge-potential relation, PHREEQC optionally will use the approach developed by Borkovec and Westall (1983). Their development solves the Poisson-Boltzmann equation to determine surface excesses of ions in the diffuse layer at the oxide-electrolyte interface. Throughout the derivation that follows, it is assumed that a volume of one liter (L) contains 1 kg of water.

The surface excess is:

$$\Gamma_{i,s} = \int_{x_{d,s}}^{\infty} (c_{i,s}(x) - c_i^o)dx,$$ (71)

where $\Gamma_{i,s}$ is the surface excess in mol m^{-2} of aqueous species i on surface s, $x_{d,s}$ is the location of the outer Helmholtz plane, $c_{i,s}(x)$ is concentration as a function of distance from the surface in mol m^{-3}, and c_i^o is the concentration in the bulk solution. The surface excess is related to concentration in the reference state of 1.0 kg of water by

$$m_{i,s} = A_{surf}\Gamma_{i,s},$$ (72)

where $m_{i,s}$ is the surface excess of aqueous species i in moles per kilogram water (mol/kgw). This surface-excess concentration can be related to the concentration in the bulk solution by

$$m_{i, s} = g_{i, s} m_i, \tag{73}$$

where $g_{i, s}$ is a function of the potential at the surface and the concentrations and charges of all ions in the bulk solution:

$$g_{i, s} = A_{surf} \text{sign}(X_{d, s} - 1)\alpha \int_1^{X_{d, s}} \frac{(X^{z_i} - 1)}{\left[X^2 \sum_l^{N_{aq}} m_l (X^{z_l} - 1) \right]^{1/2}} dX, \tag{74}$$

where $X = e^{-\frac{F\Psi_s}{RT}}$, $X_{d, s}$ is the value of X at the outer Helmholtz plane, A_{surf} is the surface area (m^2), sign($X_{d, s} - 1$) is +1 or -1 depending on the sign of the term in parentheses, i is the aqueous species for which the surface excess is being calculated, z_i is the charge on aqueous species i, l ranges over all aqueous species, m_l is the molality and z_l is the charge of aqueous species l, and $\alpha = (\varepsilon\varepsilon_0 RT/2)^{1/2}$. The value of α at 25°C is 0.02931 (L/mol)$^{1/2}$ C m^{-2}. The relation between the unknown (X) used by Borkovec and Westall (1983) and the master unknown used by PHREEQC is $a_{\Psi_s} = X^{-2}$.

The development of Borkovec and Westall (1983) calculates only the total excess concentration in the diffuse layer of each aqueous species. A problem arises in batch-reaction and transport modeling when a solution is removed from the surface, for example, in an advection simulation when the water in one cell advects into the next cell. In this case, the total moles that remain with the surface need to be known. In PHREEQC, an arbitrary assumption is made that the diffuse layer is a specified thickness and that all of the surface excess resides in the diffuse layer. The total moles of an aqueous species in the diffuse layer are then the sum of the contributions from the surface excess plus the bulk solution in the diffuse layer:

$$n_{i, s} = n_{i, s, excess} + n_{i, s, aq} = W_{bulk} g_{i, s} \frac{n_i}{W_{aq}} + W_s \frac{n_i}{W_{aq}} \cong g_{i, s} n_i + W_s \frac{n_i}{W_{aq}}, \tag{75}$$

where $n_{i, s, aq}$ refers to the moles of aqueous species i that are present in the diffuse layer due to the contribution from the bulk solution, $n_{i, s, excess}$ refers to the surface excess, W_{aq} is the mass of water in the system excluding the diffuse layer, W_s is the mass of water in the diffuse layer of surface s. It is assumed that the amount of water in the aqueous phase is much greater than in the diffuse layers, such that $W_{bulk} \cong W_{aq}$, (In version 1,

$W_{bulk} = W_{aq} + \sum_s^S W_s$). The mass of water in the diffuse layer is calculated from the thickness of the diffuse

layer and the surface area, assuming 1 L contains 1 kg water:

$$W_s = t_s A_{surf}, \tag{76}$$

where t_s is the thickness of the diffuse layer in meters. If the moles of surface sites are related to the moles of a pure phase or kinetic reactant, then $A_{surf} = A_r n_r$, otherwise A_{surf} is constant and calculated from the specific area and the mass of the surface that are specified on input. According to electrostatic theory, the thickness of the diffuse layer should be greater at low ionic strength and smaller at high ionic strength. The default value used in PHREEQC for the thickness of the diffuse layer is 1x10^{-8} m, which is approximately the thickness calculated by Debye theory for an ionic strength of 0.001 molal. For ionic strength 0.00001, the Debye length of the diffuse layer is calculated to be 1x10^{-7} m. The assumption that the amount of water in the diffuse layer is small will be

invalid if the surface area is sufficiently large; for a thickness of 1×10^{-7} m, a surface area of 1000 m^2 results in a diffuse-layer volume of 0.1 L, which is a significant portion of 1 L of bulk solution.

The total derivative of the moles of an aqueous species in the diffuse layer is

$$dn_{i,s} = \left(g_{i,s} + \frac{W_s}{W_{aq}}\right)dn_i + n_i\frac{\partial g_{i,s}}{\partial X}\left(-2e^{\left(\frac{F\Psi_s}{2RT}\right)^{-2}}\right)d\ln a_{\Psi_s} - n_i\frac{W_s}{W_{aq}}d\ln W_{aq} + n_i\frac{t_sA_r}{W_{aq}}dn_r, \tag{77}$$

where the second term is the partial derivative with respect to the master unknown for the potential at the surface, $\ln a_{\Psi_s}$, and the last term is present only if the number of surface sites is related to the moles of a pure phase or kinetic reactant. The partial derivative, $\frac{\partial g_{i,s}}{\partial X}$, is equal to the integrand from equation 74 evaluated at $X_{d,s}$:

$$\left.\frac{\partial g_{i,s}}{\partial X}\right|_{X_{d,s}} = A_{surf}\text{sign}(X_{d,s}-1)\alpha\frac{(X_{d,s}^{z_i}-1)}{\left[X_{d,s}^2\sum\limits_{l}^{N_{aq}}m_l(X_{d,s}^{z_l}-1)\right]^{1/2}}, \tag{78}$$

and the partial derivative of the function $g_{i,s}$ with respect to the master unknown is

$$\frac{\partial g_{i,s}}{\partial\ln a_{\Psi_s}} = \frac{\partial g_{i,s}}{\partial X}\left(-2e^{\left(\frac{F\Psi_s}{2RT}\right)^{-2}}\right) = -A_{surf}\text{sign}(X_{d,s}-1)\alpha\left(2e^{\left(\frac{F\Psi_s}{2RT}\right)^{-2}}\right)\frac{(X_{d,s}^{z_i}-1)}{\left[X_{d,s}^2\sum\limits_{l}^{N_{aq}}m_l(X_{d,s}^{z_l}-1)\right]^{1/2}}. \tag{79}$$

In the numerical method, it is computationally expensive to calculate the functions $g_{i,s}$, so the same approach as Borkovec and Westall (1983) is used in PHREEQC to reduce the number of function evaluations. A new level of iterations is added when the diffuse layer is explicitly included in the calculations. The functions and their partial derivatives are explicitly evaluated once at the beginning of each of these diffuse-layer iterations. During the model iterations, which occur within the diffuse-layer iterations, the values of the functions are updated using the following equation:

$$g_{i,s}^{k+1} = g_{i,s}^{k} + \frac{\partial g_{i,s}}{\partial\ln a_{\Psi_s}}d\ln a_{\Psi_s}, \tag{80}$$

where k is the model iteration number and $g_{i,s}^{0}$ is the value that is evaluated explicitly at the beginning of the diffuse-layer iteration. The model iterations end when the Newton-Raphson method has converged on a solution; however, convergence is based on the values of the functions $g_{i,s}$ that are estimates. Thus, diffuse-layer iterations continue until the values of the functions are the same on successive diffuse-layer iterations within a specified tolerance.

When explicitly calculating the composition of the diffuse layer, the function involving the *sinh* of the potential unknown (equation 69) is replaced with a charge-balance function that includes the surface charge and the diffuse-layer charge:

$$f_{z,s} = \sum\limits_{k}^{K_s}\sum\limits_{i_{(s_k)}}^{N_{s_k}}z_{i_{(s_k)}}n_{i_{(s_k)}} + \sum\limits_{i}^{N_{aq}}z_in_{i,s}, \tag{81}$$

where the function $f_{z,s}$ is zero when charge balance is achieved. The total derivative of $f_{z,s}$ is

$$df_{z,s} = \sum_{k}^{K_s} \sum_{i_{(s_k)}}^{N_{s_k}} z_{i_{(s_k)}} dn_{i_{(s_k)}} + \sum_{i}^{N_{aq}} z_i dn_{i,s}. \tag{82}$$

For data input to PHREEQC, explicit calculation of the diffuse layer is invoked using the **-diffuse_layer** identifier in the **SURFACE** data block. Specific surface area (A_s or A_r) and mass of surface (S_s) are defined in the **SURFACE** data block. The moles of surface sites are defined (1) in **SURFACE** if the number of sites is fixed, or (2) by a proportionality factor in the **SURFACE** data block and the moles of a phase in **EQUILIBRIUM_PHASES** data block, or (3) by a proportionality factor in the **SURFACE** data block and the moles of a kinetic reactant in **KINETICS** data block. The charge on a surface species is specified in the balanced chemical reaction that defines the species in the **SURFACE_SPECIES** data block (see "Description of Data Input").

Non-Electrostatic Surface Complexation

Davis and Kent (1990) describe a non-electrostatic surface-complexation model. In this model, the electrostatic term is ignored in the mass-action expressions for surface complexes. In addition, no surface charge-balance or surface charge-potential relation is used; only the mole-balance equation is included for each surface site type.

For data input to PHREEQC, the non-electrostatic model for a surface is invoked by using the **-no_edl** identifier in the **SURFACE** data block (see "Description of Data Input").

NUMERICAL METHOD FOR SPECIATION AND FORWARD MODELING

The formulation of any chemical equilibrium problem solved by PHREEQC is derived from the set of functions denoted f in the previous sections. These include f_{Alk}, f_e, f_g, f_H, f_{H_2O}, $f_{m'}$, f_O, $f_{P_{total}}$, f_p, $f_{p_{ss}}$, f_{s_k}, f_z, $f_{z,s}$, f_μ, and f_{Ψ_s}, where f_H and f_O are the simply the mole-balance functions for hydrogen and oxygen and m' refers to all aqueous master species except H^+, e^-, H_2O and the alkalinity master species. The corresponding set of master unknowns is $\ln a_{Alk}$, $\ln a_e$, n_g, $\ln a_{e^-}$, $\ln a_{H_2O}$, $\ln a_{m'}$, $\ln W_{aq}$, N_{gas}, n_p (or possibly $\ln a_{m'}$ in speciation calculations), n_{ss}, $\ln a_{s_k}$, $\ln a_{H^+}$ (or possibly $\ln a_{m'}$ in speciation calculations), $\ln a_{\Psi_s}$ (explicit diffuse-layer calculation), μ, and $\ln a_{\Psi_s}$ (implicit diffuse-layer calculation). When the residuals of all the functions that are included for a given calculation are equal to zero, a solution to the set of nonlinear equations has been found, and the equilibrium values for the chemical system have been determined. (Note that some equations that are initially included in a given calculation may be dropped if a pure phase or gas phase does not exist at equilibrium.) The solution technique assigns initial values to the master unknowns and then uses a modification of the Newton-Raphson method iteratively to revise the values of the master unknowns until a solution to the equations has been found within specified tolerances.

For a set of equations, $f_i = 0$, in unknowns x_j the Newton-Raphson method involves iteratively revising an initial set of values for the unknowns. Let $r_i = f_i$ be the residuals of the equations for the current values of the unknowns. The following set of equations is formulated:

$$r_i = -\sum_j^J \frac{\partial f_i}{\partial x_j} dx_j,$$ (83)

where J is the total number of master unknowns for the calculation. The set of equations is linear and can be solved simultaneously for the unknowns, dx_j. New values of the unknowns are calculated, $^{k+1}x_j = {}^k x_j + dx_j$, where k refers to the iteration number, after which, new values of the residuals are calculated. The process is repeated until the values of the residuals are less than a specified tolerance.

Two problems arise when using the Newton-Raphson method for chemical equilibria. The first is that the initial values of the unknowns must be sufficiently close to the equilibrium values, or the method does not converge, and the second is that a singular matrix may arise if the chemical reactions for a set of phases are not linearly independent. PHREEQC uses an optimization technique developed by Barrodale and Roberts (1980) to avoid the occurrence of singular matrices. The optimization technique also allows inequality constraints to be added to the problem, which are useful for constraining the total amounts of phases and solid solutions that can react.

The selection of initial estimates for the master unknowns is described for each type of modeling in the following sections. Regardless of the strategy for assigning the initial estimates, the estimates for the activities of the master species for elements or element valence states are revised, if necessary, before the Newton-Raphson iterations to produce approximate mole balance. The procedure for aqueous master species is as follows. After the initial estimates have been made, the distribution of species is calculated for each element (except hydrogen and oxygen) and, in initial solution calculations only, for the individual valence states which were defined. Subsequently, the ratio of the calculated moles to the input moles is calculated. If the ratio for a master species m' is greater than 1.5 or less than 10^{-5}, the following equation is used to revise the value of the master unknown:

$$\ln a_{m'}^{k+1} = \ln a_{m'}^k + w \ln \left(\frac{\sum\limits_i^{N_{aq}} b_{m',i} n_i}{T_{m'}} \right),$$ (84)

where w is 1.0 if the ratio is greater than 1.5 and 0.3 if the ratio is less than 10^{-5}, and $T_{m'}$ is the total concentration of an element or element valence state. Analogous equations are used for exchange and surface master species. After revisions to the initial estimates, the distribution of species is calculated. The iterations continue until the ratios are within the specified ranges, at which point the modified Newton-Raphson technique is used. If the successive revisions fail to find activities such that the ratios are within the specified bounds, then a second set of iterations tries to reduce the ratios below 1.5 with no lower limit to these ratios. Whether or not the second set of iterations succeeds, the Newton-Raphson technique is then used.

The optimization technique of Barrodale and Roberts (1980) is a modification of the simplex linear programming algorithm that minimizes the sum of absolute values of residuals (L1 optimization) on a set of linear equations subject to equality and inequality constraints. The general problem can be posed with the following matrix equations:

$$AX = B$$
$$CX = D$$ (85)
$$EX \leq F \ .$$

The first matrix equation is minimized in the sense that $\sum_{i}^{I_B} \left| b_i - \sum_{j}^{J} a_{i,j} x_j \right|$ is a minimum, where I_B is the number

of equations to be optimized, subject to the equality constraints of the second matrix equation and the inequality constraints of the third matrix equation.

The approach of PHREEQC is to include some of the Newton-Raphson equations in the optimization equations ($AX = B$), rather than include all of the Newton-Raphson equations as equalities ($CX = D$). Equations that are included in the A matrix may not be solved for exact equality at a given iteration, but will be optimized in the sense given above. Thus, at a given iteration, an approximate mathematical solution to the set of Newton-Raphson equations can be found even if no exact equality solution exists, for example when forcing equality for all equations would result in an unsolvable singular matrix. The equations for alkalinity, total moles of gas in the gas phase, pure phases, and solid-solution components are included in the A matrix. All mole-balance, charge-balance, and surface-potential equations are included in the B matrix. Inequalities that limit the dissolution of pure phases, solid-solution components, and gas components to the amounts present in the system are included in the C matrix.

In an attempt to avoid some numerical problems related to small numbers in the B matrix, a row of the matrix that represents a mole-balance equation is scaled if all coefficients (a column of A and B) of the corresponding unknown (change in the log activity of the element master species) are less than 1e-10. In this case, the equation is scaled by 1e-10 divided by the absolute value of the largest coefficient. Alternatively, when specified, (**-diagonal_scale** in **KNOBS**), a mole-balance equation is scaled by 1e-10 divided by the coefficient of the corresponding unknown if the coefficient of the unknown in the mole-balance equation is less than 1e-10.

The scaled matrix is solved by the optimizing solver, and the solution that is returned is a vector of changes to the values of the master unknowns. The values of the changes are checked to ensure that the changes to the unknowns are less than criteria that limit the maximum allowable size of changes. These criteria are specified by default in the program or by input in the **KNOBS** data block. If any of the changes are too large, then all the changes to the unknowns, except the mole transfers of pure phases and solid-solution components, are decreased proportionately to satisfy all of the criteria. Pure-phase and solid-solution mole transfers are not altered except to produce nonnegative values for the total moles of the pure phases and solid-solution components. After suitable changes to the unknowns have been calculated, the master unknowns are updated; new molalities and activities of all the aqueous, exchange, and surface species are calculated, and residuals for all of the functions are calculated. The residuals are tested for convergence (convergence criteria are defined internally in the program, but can be switched to an alternate set with the **-convergence_tolerance** in **KNOBS** or **-high_precision** option in **SELECTED_OUTPUT** data blocks), and a new iteration is begun if convergence has not been attained.

Aqueous Speciation Calculations

Aqueous speciation calculations use a chemical composition for a solution as input and calculate the distribution of aqueous species and saturation indices for phases. Aqueous speciation calculations include the equations $f_{m'}$, f_{H_2O}, and f_{μ}, which are equations for mole balance for elements or element valence states, activity of water, and ionic strength. Mole-balance equations for hydrogen and oxygen are not included, because the total masses of hydrogen and oxygen generally are not known. Instead, the mass of water is assumed to be 1.0 kg or is specified (**-water** in the **SOLUTION** or **SOLUTION_SPREAD** data block) and the total masses of

hydrogen and oxygen are calculated in the speciation calculation from the mass of water and the concentrations of all hydrogen and oxygen containing aqueous species.

If pH, pe, or the master unknown for an element or element valence state is specified to be adjusted to obtain charge balance for the solution, f_z is included to calculate the value of the master unknown ($\ln a_{H^+}$, $\ln a_{e^-}$, or $\ln a_{m'}$) that produces charge balance. In this case, the calculated pH, pe, or total concentration of m' will differ from the input value. If f_z is included for the master unknown $\ln a_{m'}$, the equation $f_{m'}$ is excluded.

If pH, pe, or the master unknown for an element or element valence state is specified to be adjusted to obtain a specified saturation index for a pure phase, f_p is included to calculate the value of the master unknown ($\ln a_{H^+}$, $\ln a_{e^-}$, or $\ln a_{m'}$) that produces the target saturation index. In this case, the calculated pH, pe, or total concentration of m' will differ from the input value. If f_p is included for the master unknown $\ln a_{m'}$, the equation $f_{m'}$ is excluded.

If total alkalinity is specified in the input, the mole-balance equation for alkalinity, f_{Alk}, is included to calculate $\ln a_{Alk}$ and the total molality of the element associated with alkalinity (carbon in the default database). If the problem definition contains a mole-balance equation for both carbon [or carbon(+4)] and alkalinity, then the two master unknowns associated with these equations are $\ln a_{Alk} = \ln a_{CO_3^{-2}}$ (for the default database files) and $\ln a_{H^+}$. In this case, the pH will be calculated in the speciation calculation and will not be equal to the input pH.

For speciation calculations, if the alkalinity mole-balance equation is included in the problem formulation, it is included as the only optimization equation for the solver. All other equations are included as equality constraints. No inequality constraints are included for speciation calculations.

Partial redox disequilibrium is allowed in initial solution calculations, and redox options in the **SOLUTION** or **SOLUTION_SPREAD** data block affect the aqueous speciation and saturation index calculations. By default, whenever a value of the activity of the electron is needed to calculate the molality or activity of an aqueous species, the input pe is used. If a default redox couple is given (**-redox**) or a redox couple is specified for an element (or combination of element valence states) (see **SOLUTION** keyword in "Description of Data Input"), then the mass-action expression for each aqueous species of the redox element is rewritten to remove the activity of the electron from the expression and replace it with the activities of the redox couple. For example, if iron (Fe) is to be distributed using the sulfate-sulfide redox couple [S(+6)/S(-2)], then the original chemical reaction for Fe^{+3}:

$$Fe^{+2} = Fe^{+3} + e^- \tag{86}$$

would be rewritten using the association reaction for sulfide,

$$SO_4^{-2} + 9H^+ + 8e^- = HS^- + 4H_2O, \tag{87}$$

to produce the following chemical reaction that does not include electrons:

$$Fe^{+2} + \frac{1}{8}SO_4^{-2} + \frac{9}{8}H^+ = Fe^{+3} + \frac{1}{8}HS^- + \frac{1}{2}H_2O. \tag{88}$$

The mass-action expression for this final reaction would be used as the mass-action expression for the species Fe^{+3}, and the differential for the change in the moles of Fe^{+3}, $dn_{Fe^{+3}}$, would also be based on this mass-action expression. However, the original mass-action expression (based on equation 86) is used to determine the mole-balance equations in which the term $dn_{Fe^{+3}}$ appears, that is, the species Fe^{+3} would appear in the mole-balance equation for iron, but not in the mole-balance equations for S(+6) or S(-2). The effect of these manipulations is that ferrous iron, ferric iron, sulfate, and sulfide are in redox equilibrium. Another set of redox elements (for

example oxygen and nitrogen) may also be defined to be in equilibrium among themselves, but not necessarily in redox equilibrium with iron and sulfur.

By default, if a saturation-index calculation requires a value for pe (or activity of the electron), then the input pe is used. If a default redox couple has been defined (**-redox**), then the dissolution reaction for the phase is rewritten as above to eliminate the activity of the electron and replace it with the activities of the redox couple.

The set of master unknowns may change for redox elements during a calculation. The process, which is termed "basis switching", occurs if the activity of the master species which is the master unknown for a mole-balance equation becomes ten orders of magnitude smaller than the activity of another master species included in the same mole-balance equation. In this case, all of the mass-action expressions involving the current master unknown (including aqueous, exchange, gas, and surface species, and pure phases) are rewritten in terms of the new master species that has the larger activity. An example of this process is, if nitrogen is present in a system that becomes reducing, the master unknown for nitrogen would switch from nitrate, which would be present in negligible amounts under reducing conditions, to ammonium, which would be the dominant species. Basis switching does not affect the ultimate equilibrium distribution of species, but it does speed calculations and avoid numerical problems in dealing with small concentrations.

Initial values for the master unknowns are estimated and then revised according to the strategy described in the previous section. For initial solution calculations, the input values for pH and pe are used as initial estimates. The mass of water is 1.0 kg unless otherwise specified, and the activity of water is estimated to be 1.0. Ionic strength is estimated assuming the master species are the only species present and their concentrations are equal to the input concentrations (converted to units of molality). The activity of the master species of elements (except hydrogen and oxygen) and element valence states are set equal to the input concentration (converted to molality). If the charge-balance equation or a phase-equilibrium equation is used in place of the mole-balance equation for an element or element valence state, then the initial activity of the master species is set equal to one thousandth of the input concentration (converted to molality).

For data input to PHREEQC all options for a speciation calculation--use of an alkalinity equation, charge-balance equation, phase-equilibrium equation, and redox couples--are defined in a **SOLUTION** or **SOLUTION_SPREAD** data block (see "Description of Data Input").

Calculation of the Initial Composition of an Exchanger

An initial exchange-composition calculation is needed if the composition of an exchanger is not defined explicitly, but rather, is indicated to be in equilibrium with a specified solution composition. In this case, the composition of the exchanger is not known, only that it is in equilibrium with a solution. The equations for an initial exchange-composition calculation are f_e, $f_{m'}$, f_{H_2O}, and f_μ, which are equations for mole balance for each exchanger, mole balance for each element or element valence state, activity of water, and ionic strength.

For initial exchange-composition calculations, the values of $T_{m'}$ include only the aqueous concentrations and the mole-balance equations $f_{m'}$ do not contain terms for the contribution of the exchangers to the total element concentrations. All quantities related to the aqueous phase are the same as for the solution without the exchanger present. Essentially, only the values of the master unknowns of the exchange assemblage, $\ln a_e$, are adjusted to achieve mole balance for the exchanger. Once mole balance is achieved, the composition of each exchanger is known.

All equations for initial exchange-composition calculations are included as equality constraints in the solver. No equations are optimized and no inequality constraints are included.

An initial exchange-composition calculation is performed only if the exchanger is defined to be in equilibrium with a specified solution. The distribution of species for this solution has already been calculated, either by an initial solution calculation or by a batch-reaction or transport calculation. Thus, the values of all master unknowns related to the aqueous phase are known and are used as initial estimates for the exchange calculation. The initial estimate of the master unknown for each exchanger is set equal to the moles of exchange sites for that exchanger.

For data input to PHREEQC, definition of the initial exchange-composition calculation is made with the **EXCHANGE** data block (see "Description of Data Input").

Calculation of the Initial Composition of a Surface

An initial surface-composition calculation is needed if the composition of a surface is not defined explicitly, but is indicated to be in equilibrium with a specified solution composition. In this case, the composition of the surface is not known, only that it is in equilibrium with a solution. The equations for the initial surface-composition calculation are f_{s_k}, f_{Ψ_s} or $f_{z,s}$, $f_{m'}$, f_{H_2O}, and f_μ, which are equations for mole-balance for each type of surface site in the surface assemblage, the charge-potential relation or charge-balance for each surface (both of these equations are excluded in the non-electrostatic model), mole balance for each element or element valence state, activity of water, and ionic strength.

For initial surface-composition calculations, the values of $T_{m'}$ include only the aqueous concentrations and the corresponding mole-balance equations $f_{m'}$ do not contain terms for the contribution of the surfaces to the total element concentrations. All quantities related to the aqueous phase are the same as for the solution without the surface assemblage present.

For the explicit calculation of the diffuse layer, a charge-balance equation is used for each surface, $f_{z,s}$; the values of the master unknowns for each surface type of the surface assemblage, $\ln a_{s_k}$ and the potential unknowns $\ln a_{\Psi_s}$, are adjusted to achieve mole balance and charge balance for each surface. If the diffuse-layer composition is not explicitly included in the calculation, then the charge-potential equation f_{Ψ_s} is used in place of the surface charge-balance equation. If the non-electrostatic model is used for the surface assemblage, then neither the surface charge-balance nor the charge-potential equation is included in the set of equations to be solved.

All equations for initial surface-composition calculations are included as equality constraints in the solver. No equations are optimized and no inequality constraints are included.

An initial surface-composition calculation is performed only if the initial surface is defined to be in equilibrium with a specified solution. The distribution of species for this solution has already been calculated, either by an initial solution calculation or by a batch-reaction or transport calculation. Thus, the values of all master unknowns related to the aqueous phase are known and are used as starting estimates for the surface calculation. The initial estimate of the activity of the master species for each surface is set equal to one tenth of the moles of surface sites for that surface. For explicit and implicit diffuse-layer calculations, the initial estimate of the potential unknown $\ln a_{\Psi_s}$ is zero for each surface, which implies that the surface charge is zero.

For data input to PHREEQC, definition of the initial surface-composition calculation is made with the **SURFACE** data block (see "Description of Data Input").

Calculation of the Initial Composition of Fixed-Volume Gas Phase

An initial gas-phase-composition calculation is needed if the composition of a gas phase is not defined explicitly, but rather, the composition of a fixed-volume gas phase is defined to be that which is in equilibrium with a specified solution composition. The equations for the initial gas-phase-composition calculation are the same as an initial solution calculation and are $f_{m'}$, f_{H_2O}, and f_μ, which are equations for mole balance for each element or element valence state, activity of water, and ionic strength.

For initial gas-phase-composition calculations, the values of $T_{m'}$ include only the aqueous concentrations and the corresponding mole-balance equations $f_{m'}$ do not contain terms for the contribution of the gas components to the total element concentrations. The values calculated for all quantities related to the aqueous phase are the same as for the solution without the gas phase present. Once the distribution of species in the aqueous phase is determined, the partial pressures of all components in the gas phase can be calculated. The partial pressures and the specified fixed volume are used with the ideal gas law to calculate the moles of each component in the gas phase.

All equations for initial gas-phase-composition calculations are included as equality constraints in the solver. No equations are optimized and no inequality constraints are included.

An initial gas-phase-composition calculation is performed only if the gas phase is defined to have a constant volume and is defined to be initially in equilibrium with a specified solution. The distribution of species for this solution has already been calculated, either by an initial solution calculation or by a batch-reaction or transport calculation. Thus, the values of all master unknowns related to the aqueous phase are known and are used as initial estimates for the initial gas-phase-composition calculation.

For data input to PHREEQC, definition of the initial gas-phase-composition calculation is made with the **GAS_PHASE** data block (see "Description of Data Input").

Batch-Reaction and Transport Calculations

Batch-reaction and transport calculations require calculating equilibrium between the aqueous phase and any equilibrium-phase assemblage, surface assemblage, exchanger assemblage, solid-solution assemblage, and gas phase that is defined to be present in a chemical system. Irreversible reactions that occur prior to equilibration include mixing, specified stoichiometric reactions, kinetic reactions, and temperature change. The complete set of Newton-Raphson equations that can be included in batch-reaction and transport calculations contains f_e, f_H, f_{H_2O}, $f_{m'}$, f_O, $f_{P_{total}}$, f_p, $f_{p_{ss}}$, f_{s_k}, f_z, $f_{z,s}$, f_μ, and f_{Ψ_s}.

Equations for mole balance on hydrogen f_H, activity of water f_{H_2O}, mole balance on oxygen f_O, charge balance f_z, and ionic strength f_μ are always included and are associated with the master unknowns $\ln a_{e^-}$, $\ln a_{H_2O}$, W_{aq} (mass of water), $\ln a_{H^+}$, and μ, which are always included as master unknowns.

Mole-balance equations $f_{m'}$ are included for total concentrations of elements, not individual valence states or combinations of individual valence states. A mole-balance equation for alkalinity can not be included; it is used only in initial solution calculations.

The equation $f_{P_{total}}$ is included if a fixed-pressure gas phase is specified and is present at equilibrium. The equations f_e are included if an exchange assemblage is specified. The equations f_{s_k} are included if a surface assemblage is specified. In addition, f_{Ψ_s} is included for each surface for which an implicit diffuse-layer

calculation is specified or $f_{z,s}$ is included for each surface for which an explicit diffuse-layer calculation is specified. An equation f_p is included for each pure phase that is present at equilibrium. An equation $f_{p_{ss}}$ is included for each component of each solid solution that is present at equilibrium.

It is not known at the beginning of the calculation whether a pure phase, solid solution, or fixed-pressure gas phase will be present at equilibrium. Thus, at each iteration, the following logic is used to determine which of the equations should be included in the equilibrium calculations. The equation for a phase is included if it has a positive moles, $n_p > 0$, or if the saturation index is calculated to be greater than the target saturation index. If the equation is not included in the matrix, then all coefficients for the unknown dn_p in the matrix are set to zero.

For an ideal solid solution, the equations $f_{p_{ss}}$ are included if the moles of any of the components are greater than a small number ($1 x 10^{-13}$) or if the sum, $\sum_{p_{ss}} \dfrac{IAP_{p_{ss}}}{K_{p_{ss}}}$, is greater than 1.0. For an ideal solid solution, $\dfrac{IAP_{p_{ss}}}{K_{p_{ss}}} = x_{p_{ss}}$, so the summation determines if the sum of the mole fractions is greater than 1.0. If the equations for a solid solution are not included in the matrix, then all coefficients for the unknowns $dn_{p_{ss}}$ in the matrix are set to zero.

For nonideal, binary solid solutions the following procedure to determine whether to include solid-solution equations is developed from the equations of Glynn and Reardon (1990, equations 37 through 48). If the moles of any of the solid-solution components are greater than a small number ($1 x 10^{-13}$) then all the equations for the solid solution are included. Otherwise, the aqueous activity fractions of the components are calculated from

$$x_{1,aq} = \frac{IAP_1}{IAP_1 + IAP_2} \text{ and } x_{2,aq} = \frac{IAP_2}{IAP_1 + IAP_2}, \tag{89}$$

where IAP is the ion activity product for the pure component. Next the mole fractions of the solids that would be in equilibrium with those aqueous activity fractions are determined by solving the following equation for x_1 and x_2 ($=1-x_1$):

$$x_1 \lambda_1 K_1 + x_2 \lambda_2 K_2 = \frac{1}{\dfrac{x_{1,aq}}{\lambda_1 K_1} + \dfrac{x_{2,aq}}{\lambda_2 K_2}}, \tag{90}$$

where x_1 and x_2 are the mole fractions in the solid phase, K_1 and K_2 are the equilibrium constants for the pure components, λ_1 and λ_2 are the activity coefficients of the components as calculated from the Guggenheim parameters for the excess free energy. This equation is highly nonlinear and is solved by first testing subintervals between 0 and 1 to find one that contains the mole fraction of component 1 that satisfies the equation and then interval halving to refine the estimate of the mole fraction. Once the mole fractions of the solid have been determined, two values of the "total activity product" ($\sum \Pi$) are calculated as follows:

$$\sum \Pi_{aq} = IAP_1 + IAP_2 \tag{91}$$

and $\sum \Pi_{solid} = x_1 \lambda_1 K_1 + x_2 \lambda_2 K_2.$ $\tag{92}$

If $\sum \Pi_{solid} < \sum \Pi_{aq}$, then the equations for the solid solution are included, otherwise, the equations are not included. If the equations for a solid solution are not included in the matrix, all coefficients for the unknowns $dn_{p_{ss}}$ in the matrix are set to zero.

At each iteration, the equation for the sum of partial pressures of gas components in the gas phase is included for a fixed-pressure gas phase if the moles in the gas phase are greater than a small number ($1 x 10^{-14}$), or if the sum of the partial pressures of the gas-phase components, as calculated from the activities of aqueous species, is greater than the total pressure. If the equation for the sum of the partial pressures of gas components in the gas phase is not included in the matrix, then all coefficients of the unknown dN_g are set to zero.

Equations $f_{P_{total}}$, f_p and $f_{p_{ss}}$ are included as optimization equations in the solver. All other equations are included as equality constraints in the solver. In addition, several inequality constraints are included in the solver:

(1) the value of the residual of an optimization equation f_p, which is equal to $b_p - \sum_j a_{p,j} x_j$, is constrained to be nonnegative, which maintains an estimate of saturation or undersaturation for the mineral; (2) the value of the residual of an optimization equation $f_{p_{ss}}$, which is equal to $b_{p_{ss}} - \sum_j a_{p_{ss},j} x_j$, is constrained to be nonnegative, which maintains an estimate of saturation or undersaturation for the component of the solid solution; (3) the residual of the optimization equation for $f_{P_{total}}$ is constrained to be nonnegative, which maintains a nonnegative estimate of the total gas pressure; (4) the decrease in the mass of a pure phase, dn_p, is constrained to be less than or equal to the total moles of the phase present, n_p; (5) the decrease in the mass of a component of a solid solution, $dn_{p_{ss}}$, is constrained to be less than or equal to the total moles of the component present, n_p; and (6) the decrease in the moles in the gas phase, dN_{gas}, is constrained to be less than the moles in the gas phase, N_{gas}.

Initial values for the master unknowns for the aqueous phase are taken from the previous distribution of species for the solution. If mixing of two or more solutions is involved, the initial values are the sums of the values in the solutions, weighted by their mixing factor. If exchangers or surfaces have previously been equilibrated with a solution, initial values are taken from the previous equilibration. If they have not been equilibrated with a solution, the estimates of the master unknowns are the same as those used for initial exchange-composition and initial surface-composition calculations. Initial values for the moles of each phase in the pure-phase assemblage, each component in the solid solutions in the solid-solution assemblage, and each gas component in the gas phase are set equal to the input values or the values from the last simulation in which they were saved.

For data input to PHREEQC, definition of batch-reaction and transport calculations rely on many of the data blocks. Initial conditions are defined with **SOLUTION** or **SOLUTION_SPREAD**, **EXCHANGE**, **SURFACE**, **GAS_PHASE**, **EQUILIBRIUM_PHASES**, **SOLID_SOLUTIONS**, and **USE** data blocks. Batch reactions are defined by initial conditions and with **MIX**, **KINETICS**, **REACTION**, **REACTION_TEMPERATURE**, and **USE** data blocks. Transport calculations are specified with the **ADVECTION** or the **TRANSPORT** data block (see "Description of Data Input").

NUMERICAL METHOD AND RATE EXPRESSIONS FOR CHEMICAL KINETICS

A major deficiency with geochemical equilibrium models is that minerals, organic substances, and other reactants often do not react to equilibrium in the time frame of an experiment or a model period. A kinetically controlled reaction of a solid or a nonequilibrium solute generates concentration changes of aqueous species according to the rate equation:

$$\frac{dm_i}{dt} = c_{i,k}R_k,$$

(93)

where $c_{i,k}$ is the stoichiometric coefficient of species i in the kinetic reaction, and R_k is the overall reaction rate for substance k (mol/kgw/s). In general, reaction rates vary with reaction progress, which leads to a set of ordinary differential equations that must be solved.

Kinetic rates have been published for numerous reactions, and for various conditions of temperature, pressure, and solution composition. However, different researchers applied different rate expressions to fit observed rates, and it is difficult to select rate expressions (which commonly have been hard coded into programs) that have sufficient generality. The problem is circumvented in PHREEQC with an embedded BASIC interpreter that allows definition of rate expressions for kinetic reactions in the input file in a general way, obviating the need for hard-coded rate expressions in the program.

Numerical Method

The rate must be integrated over a time interval, which involves calculating the changes in solution concentrations while accounting for effects on the reaction rate. Many geochemical kinetic reactions result in "stiff" sets of equations in which some rates (the time derivatives of concentration change) are changing rapidly while others are changing slowly as the reactions unfold in time. PHREEQC solves such systems by a Runge-Kutta (RK) algorithm, which integrates the rates over time. An RK scheme by Fehlberg (1969) is used, with up to 6 intermediate evaluations of the derivatives. The scheme includes an RK method of lower order to derive an error estimate. The error estimate is compared with a user-defined error tolerance to automatically decrease or increase the integration time interval to maintain the errors within the given tolerance. Furthermore, if the rates in the first three RK evaluations differ by less than the tolerance, the final rate is calculated directly and checked once more against the required tolerance. The user can specify the number of intermediate RK subintervals which are evaluated before final integration of the interval is attempted (see "Description of Data Input"). The coefficients in the scheme are from Cash and Karp (1990).

Rate Expressions

The overall rate for a kinetic reaction of minerals and other solids is:

$$R_k = r_k \frac{A_0}{V} \left(\frac{m_k}{m_{0k}} \right)^n,$$

(94)

where r_k is the specific rate (mol/m^2/s), A_0 is the initial surface area of the solid (m^2), V is the amount of solution (kgw), m_{0k} is the initial moles of solid, m_k is the moles of solid at a given time, and $(m_k/m_{0k})^n$ is a factor to account for changes in A_0/V during dissolution and also for selective dissolution and aging of the solid. For uniformly dissolving spheres and cubes $n = 2/3$. All calculations in PHREEQC are in moles, and the factor A_0/V must be provided by the user to obtain the appropriate scaling.

The specific rate expressions, r_k, for a selection of substances have been included in the database under keyword **RATES**. These specific rates have various forms, largely depending on the completeness of the experimental information. When information is lacking, a simple rate that is often applied is

$$r_k = k_k \left(1 - \left(\frac{IAP}{K_k} \right)^{\sigma} \right), \tag{95}$$

where k_k is an empirical constant and IAP/K_k is the saturation ratio (SR). This rate equation can be derived from transition-state theory, where the coefficient σ is related to the stoichiometry of the reaction when an activated complex is formed (Aagaard and Helgeson, 1982; Delany and others, 1986). Often, $\sigma = 1$. An advantage of this expression is that the rate equation applies for both supersaturation and undersaturation, and the rate is zero at equilibrium. The rate is constant over a large domain whenever the geochemical reaction is far from equilibrium ($IAP/K < 0.1$), and the rate approaches zero when IAP/K approaches 1.0 (equilibrium).

The rate expression may also be based on the saturation index (SI) in the following form:

$$r_k = k_k \sigma \log \left(\frac{IAP}{K_k} \right). \tag{96}$$

This rate expression has been applied with some success to dissolution of dolomite (Appelo and others, 1984).

Rate expressions often contain concentration-dependent terms. One example is the Monod equation:

$$r_k = r_{max} \left(\frac{C}{K_m + C} \right), \tag{97}$$

where r_{max} is the maximal rate, and K_m is equal to the concentration where the rate is half of the maximal rate. The Monod rate equation is commonly used for simulating the sequential steps in the oxidation of organic matter (Van Cappellen and Wang, 1996). A series of rate expressions can be developed in line with the energy yield of the oxidant; first O_2 is consumed, then NO_3^-, and successively other, more slowly operating oxidants such as Fe(III) oxides and SO_4^{-2}. The coefficients in the Monod equation can be derived from first-order rate equations for the individual processes. For degradation of organic matter (C) in soils the first-order rate equation is

$$\frac{ds_C}{dt} = -k_1 s_C, \tag{98}$$

where s_C is organic carbon content (mol/kg soil), and k_1 is the first-order decay constant (s^{-1}). The value of k_1 is approximately equal to 0.025 yr^{-1} in a temperate climate with aerobic soils (Russell, 1973), whereas in sandy aquifers in The Netherlands, where NO_3^- is the oxidant, $k_1 \approx 5e-4$ yr^{-1}. Concentrations of up to 3 μM O_2 are found in ground water even outside the redox-domain of organic degradation by O_2, and 3 μM O_2 may be taken as the concentration where the (concentration-dependent) rate for aerobic degradation equals the reaction rate for denitrification. First-order decay ($k_1 = 0.025$ yr^{-1} for 0.3 mM O_2 and $k_1 = 5$e-4 yr^{-1} for 3 μM O_2) is obtained with the coefficients $r_{max} = 1.57$e-9 s^{-1} and $K_m = 294$ μM in the Monod equation, and oxygen as the limiting solute. A similar estimate for denitrification is based on $k_1 = 5$e-4 yr^{-1} for $NO_3^- = 3$ mM and $k_1 = 1$e-5 yr^{-1} for $NO_3^- = 3$ μM, which yields $r_{max} = 1.67$e-11 s^{-1} and $K_m = 155$ μM. The combined overall Monod expression for degradation of organic carbon in a fresh-water aquifer is then:

$$R_C = 6 \, s_C \left(\frac{s_C}{s_{C_0}} \right) \left[\frac{1.57 \times 10^{-9} m_{O_2}}{2.94 \times 10^{-4} + m_{O_2}} + \frac{1.67 \times 10^{-11} m_{NO_3^-}}{1.55 \times 10^{-4} + m_{NO_3^-}} \right] \tag{99}$$

where the factor 6 derives from recalculating the concentration of s_C from mol/kg soil to mol/kg pore water.

A further aspect of organic matter decomposition is that a part appears to be refractory and particularly

resistant to degradation. Some models have been proposed to account for the tendency of part of the sedimentary organic carbon to survive; tentatively, a factor $\left(\dfrac{s_C}{s_{C_0}}\right)$ may be assumed, which makes the overall rate second order. This factor implies that a decrease to 1/10 of the original concentration results in a decrease of 1/100 in the rate of further breakdown. It must be noted that this simple factor is used to approximate a very complicated process and a more thorough treatment of the process is needed, but is also possible given the flexibility of defining rates in PHRE-EQC.

Still other rate expressions are based on detailed measurements in solutions with varying concentrations of the aqueous species that influence the rate. For example, Williamson and Rimstidt (1994) give a rate expression for oxidation of pyrite:

$$r_{pyrite} = 10^{-10.19} m_{O_2}^{0.5} m_{H^+}^{-0.11}, \tag{100}$$

which shows a square root dependence on the molality of oxygen, and a small increase of the rate with increase in pH. This rate is applicable for the dissolution reaction only, and only when the solution contains oxygen. It is probably inadequate when the solution approaches equilibrium or when oxygen is depleted.

An example of a more complete rate expression which applies for both dissolution and precipitation is the rate equation for calcite. Plummer and others (1978) have found that the rate can be described by the equation:

$$r_{calcite} = k_1[H^+] + k_2[CO_2] + k_3[H_2O] - k_4[Ca^{2+}][HCO_3^-], \tag{101}$$

where bracketed chemical symbols indicate activity, and the coefficients k_1, k_2 and k_3 have been determined as a function of temperature by Plummer and others (1978) from calcite dissolution experiments in CO_2-charged solutions. The rate contains a forward part r_f (first three terms of equation 101), and a backward part r_b (last term of equation 101), and thus is applicable for both dissolution and precipitation. The backward rate contains a coefficient k_4, the value of which depends on the solution composition. In a pure water-calcite system, bicarbonate concentration is approximately equal to twice the calcium concentration and the backward rate can be approximated as

$$r_b = k_4[Ca^{2+}][HCO_3^-] \approx k_4 2[Ca^{2+}]^2. \tag{102}$$

At equilibrium, $[Ca^{2+}]$ is the activity at saturation $[Ca^{2+}]_s$. Also $r_{calcite} = 0$, and therefore,

$$2k_4 = \frac{r_f}{[Ca^{2+}]_s^2}. \tag{103}$$

Combining equations 101, 102, and 103 gives:

$$r_{calcite} = r_f\left[1 - \left(\frac{[Ca^{2+}]}{[Ca^{2+}]_s^2}\right)^2\right]. \tag{104}$$

In a pure Ca-CO_2 system at constant CO_2 pressure, the ion activity product (IAP) is:

$$IAP_{calcite} = \frac{[Ca^{2+}][HCO_3^-]^2}{P_{CO_2}} \approx 4\frac{[Ca^{2+}]^3}{P_{CO_2}} \text{ and } K_{Calcite} = 4\frac{[Ca^{2+}]_s^3}{P_{CO_2}}. \tag{105}$$

Thus, for a calcite-water system, the rate for calcite can be approximated as:

$$r_{calcite} \approx r_f \left[1 - \left(\frac{IAP}{K_{calcite}} \right)^{\frac{2}{3}} \right] \tag{106}$$

where r_f contains the first three terms given in equation 101.

EQUATIONS AND NUMERICAL METHOD FOR TRANSPORT MODELING

PHREEQC has the capability to model several one-dimensional transport processes including: (1) diffusion, (2) advection, (3) advection and dispersion, and (4) advection and dispersion with diffusion into stagnant zones, which is referred to as dual porosity. All of these processes can be combined with equilibrium and kinetic chemical reactions.

The Advection-Reaction-Dispersion Equation

Conservation of mass for a chemical that is transported (fig. 1) yields the advection-reaction-dispersion (ARD) equation:

$$\frac{\partial C}{\partial t} = -v\frac{\partial C}{\partial x} + D_L\frac{\partial^2 C}{\partial x^2} - \frac{\partial q}{\partial t}, \tag{107}$$

where C is concentration in water (mol/kgw), t is time (s), v is pore water flow velocity (m/s), x is distance (m), D_L is the hydrodynamic dispersion coefficient [m^2/s, $D_L = D_e + \alpha_L v$, with D_e the effective diffusion coefficient, and α_L the dispersivity (m)], and q is concentration in the solid phase (expressed as mol/kgw in the pores). The term $-v\frac{\partial C}{\partial x}$ represents advective transport, $D_L\frac{\partial^2 C}{\partial x^2}$ represents dispersive transport, and $\frac{\partial q}{\partial t}$ is the change in concentration in the solid phase due to reactions (q in the same units as C). The usual assumption is that v and D_L are equal for all solute species, so that C can be the total dissolved concentration for an element, including all redox species.
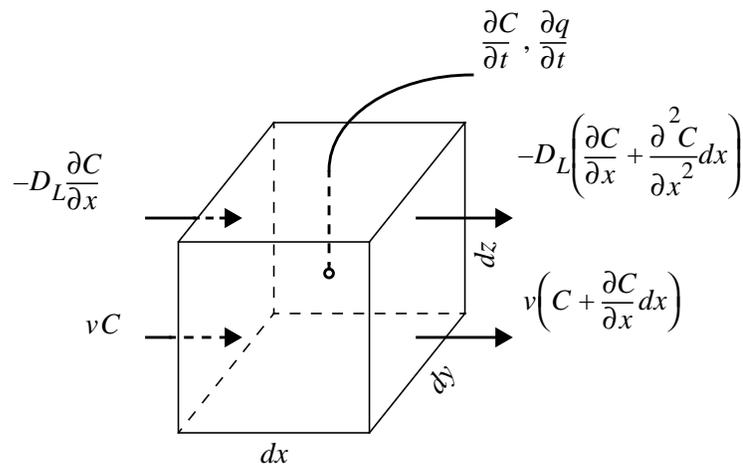

Figure 1.--Terms in the advection-reaction-dispersion equation.

The transport part of equation 107 is solved with an explicit finite difference scheme that is forward in time, central in space for dispersion, and upwind for advective transport. The chemical interaction term $\partial q / \partial t$ for each element is calculated separately from the transport part for each time step and is the sum of all equilibrium and non-equilibrium reaction rates. The numerical approach follows the basic components of the ARD equation in a split-operator scheme (Press and others, 1992; Yanenko, 1971). With each time step, first advective transport is calculated, then all equilibrium and kinetically controlled chemical reactions, thereafter dispersive transport, which is followed again by calculation of all equilibrium and kinetically controlled chemical reactions. The scheme differs from the majority of other hydrogeochemical transport models (Yeh and Tripathi, 1989) in that kinetic and equilibrium chemical reactions are calculated both after the advection step and after the dispersion step. This reduces numerical dispersion and the need to iterate between chemistry and transport.

A major advantage of the split-operator scheme is that numerical accuracy and stability can be obtained by adjusting time step to grid size for the individual parts of the equation. Numerical dispersion is minimized by always having the following relationship between time and distance discretization:

$$(\Delta t)_A = \frac{\Delta x}{v}, \tag{108}$$

where $(\Delta t)_A$ is the time step for advective transport, and Δx is the cell length. Numerical instabilities (oscillations) in the calculation of diffusion/dispersion are eliminated with the constraint:

$$(\Delta t)_D \leq \frac{(\Delta x)^2}{3 D_L}, \tag{109}$$

where $(\Delta t)_D$ is the time step (s) for dispersive/diffusive transport calculations. The two conditions of equation 108 and 109 are the Courant condition for advective transport and the Von Neumann criterion for dispersive transport calculations, respectively (for example, Press and others, 1992). Numerical dispersion is in many cases negligible when $\Delta x \leq \alpha_L$, because physical dispersive transport is then equally or more important than advective transport. When a fine grid is used to reduce numerical dispersion, the time step for dispersive transport calculations (equation 109) may become smaller than the time step for advective calculations (equation 108), because the first has quadratic dependence on grid size. The conflict is solved by multiple dispersion time steps such that $\sum (\Delta t)_D = (\Delta t)_A$, and calculating chemical reactions after each of the dispersion time steps. For input to PHRE-EQC, a time step must be defined which equals the advection time step $(\Delta t)_A$, or, if diffusion is modeled, equals the diffusion period. Furthermore, the number of *shifts* must be defined, which is the number of advection time steps (or diffusion periods) to be calculated.

Dispersive transport in a central difference scheme is essentially mixing of cells. A mixing factor *mixf* is defined as

$$mixf = \frac{D_L (\Delta t)_A}{n (\Delta x)^2}, \tag{110}$$

where n is a positive integer. The restriction is that never more is mixed out of a cell than stays behind, that is, *mixf* must be less than 1/3 as follows from equation 109. When, according to equation 110 with $n = 1$, *mixf* is greater than 1/3, the value of n is increased such that *mixf* is less than or equal to 1/3. The dispersion time step is then

$$(\Delta t)_D = \frac{(\Delta t)_A}{n} \text{ and } n \text{ mixes are performed.}$$

The numerical scheme has been checked by comparison with analytical solutions for simple cases with linear exchange. Linear exchange results when the exchange coefficient for the exchange half-reaction is equal for two homovalent cations. It gives a linear retardation $R = 1 + CEC / C$, where CEC is the cation exchange capacity, expressed in mol/kgw. In the following example, a 130 m flow tube contains water with an initial concentration $C(x,0) = C_i = 0$. The displacing solution has concentration $C = C_0 = 1$ mmol/kgw, and the pore-water flow velocity is $v = 15$ m/year. The dispersivity is $\alpha_L = 5$ m, and the effective diffusion coefficient is $D_e = 0$ m²/s. The profile is given after 4 years for two chemicals, one with $R = 1$ (Cl⁻) and the other with $R = 2.5$ (Na⁺).

Two boundary conditions can be considered for this problem. One entails a flux or third type boundary condition at $x = 0$:

$$C(0, t) = C_0 + \frac{D_L}{v} \frac{\partial C(x_{end}, t)}{\partial x} . \tag{111}$$

This boundary condition is appropriate for laboratory columns with inlet tubing much smaller than the column cross section. The solution for the ARD equation is then (Lindstrom and others, 1967):

$$C(x, t) = C_i + \frac{1}{2}(C_0 - C_i) \, A , \tag{112}$$

where, with $D_L = \alpha_L v$:

$$A = \mathrm{erfc}\left(\frac{x - vt/R}{\sqrt{4\alpha_L vt/R}}\right) + \sqrt{\frac{x}{\pi\alpha_L}} \exp\left[-\frac{(x - vt/R)^2}{4\alpha_L vt/R}\right] - \frac{1}{2}\left(1 + \frac{x}{\alpha_L} + \frac{vt/R}{\alpha_L}\right)\exp\left(\frac{x}{\alpha_L}\right)\mathrm{erfc}\left(\frac{x + vt/R}{\sqrt{4\alpha_L vt/R}}\right) . \tag{113}$$

Figure 2 shows the comparison for three simulations with different grid spacings, $\Delta x = 15$, 5, and 1.67 m, which correspond with $(\Delta t)_A = 1$, 1/3, and 1/9 years, respectively. For Cl⁻, which has $R = 1$, the fronts of the three simulations are indistinguishable and in excellent agreement with the analytical solution. For the retarded ion Na⁺, which has $R = 2.5$, the average location of the breakthrough curve for all grid spacings is correct and is in agreement with the analytical solution. However, the simulations with coarser grids show a more spread-out breakthrough that is due to numerical dispersion. The finest grid gives the closest agreement with the analytical solution, but requires the most computer time.

Computer time is primarily dependent on the number of calls to the geochemical subroutines of PHREEQC, and in the absence of kinetic reactions, the number of calls is proportional to (number of cells) x (number of advection steps) x (1 + number of dispersion steps). In this example, $D_L = D_e + \alpha_L v = 0 + 5$ x 15 m²/yr. Thus, by equation 110, $mixf = 1/3$, 1, and 3, respectively for the progressively smaller cell sizes. For the 15-meter cell-size ($mixf = 1/3$), one dispersion step is taken for each advection step; for the 5-meter cell size ($mixf = 1$), three dispersion steps are taken for each advection step; and for the 1.67-meter cell size ($mixf = 3$), nine dispersion steps are taken for each advection step. Figure 2 shows profiles the advective front of Cl ($C/C_0 = 0.5$) after 4 years of travel, when it has arrived at 60 m; for the 15-meter cell size, this requires 4 advection steps. The flowtube consists of 9 cells for which geochemical calculations are done for each step; therefore, the number of the reaction calculations is 9 x 4 x (1 + 1) = 72. Larger numbers of cells and advection steps apply for the smaller grids. The number of calls to the reaction calculations for the other two cases is 27 x 12 x (1 + 3) = 1,296; and 81 x 36 x (1 + 9) = 29,160.

The examples given here have linear retardation to enable comparison with analytical solutions. However, linear retardation is subject to large numerical dispersion, and the examples are, in a sense, worst cases with respect
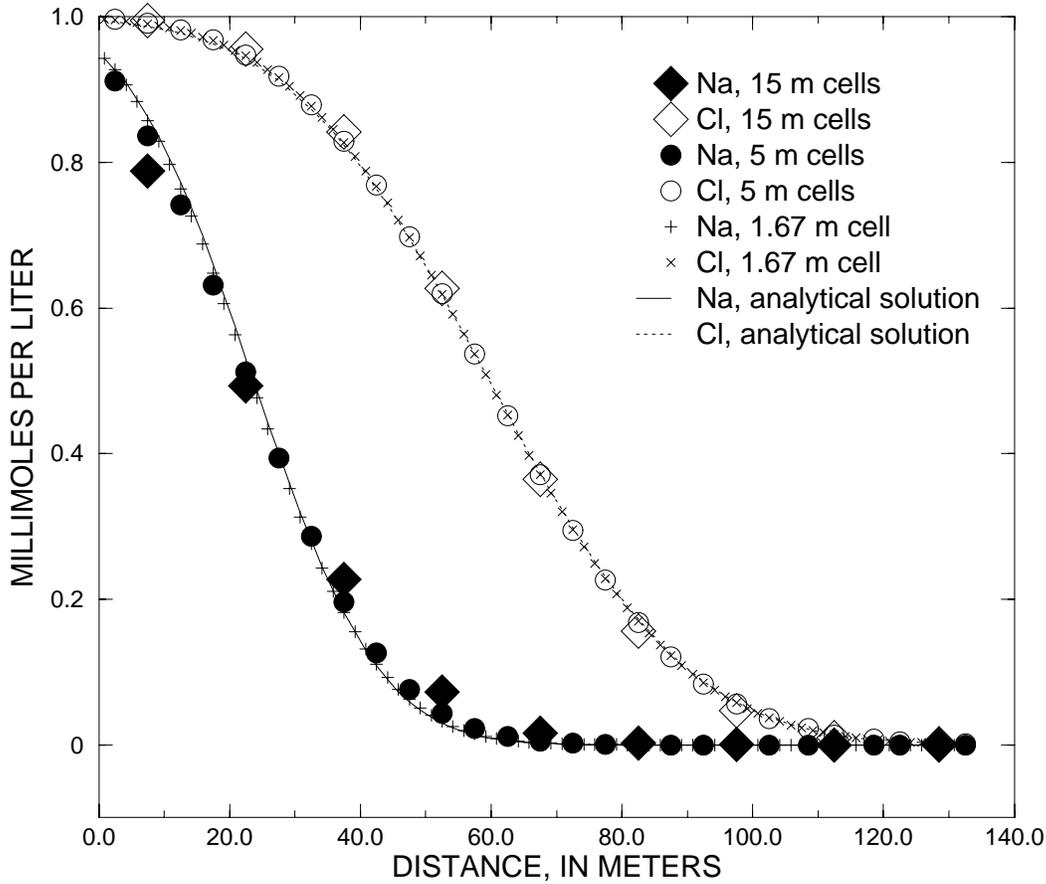
Figure 2.--Analytical solution for 1D transport with ion-exchange reactions and flux boundary condition compared with PHREEQC calculations at various grid spacings.

to numerical dispersion. In many cases of geochemical interest, the chemical reactions help to counteract numerical dispersion because the reactions tend to sharpen fronts, for example with precipitation/dissolution reactions and displacement chromatography. In other cases, exchange with a less favored ion may give a real, chemical dispersion that exceeds the effects of numerical dispersion.

Another boundary condition is the Dirichlet, or first-type, boundary condition, which involves a constant concentration $C(0,t)$ at $x = 0$:

$$C(0, t) = C_0. \tag{114}$$

This boundary condition is valid for water infiltrating from a large reservoir in full contact with the underlying soil, for example infiltration from a pond, or diffusion of seawater into underlying sediment. The solution for the ARD equation is in this case (Lapidus and Amundson, 1952):

$$C(x, t) = C_i + \frac{1}{2}(C_0 - C_i) \, B, \tag{115}$$

where,

$$B = \mathrm{erfc}\left(\frac{x - vt/R}{\sqrt{4\alpha_L vt/R}}\right) + \exp\left(\frac{x}{\alpha_L}\right)\mathrm{erfc}\left(\frac{x + vt/R}{\sqrt{4\alpha_L vt/R}}\right). \tag{116}$$

Figure 3 shows the results of three simulations with the same discretizations as the previous transport example. Again, the conservative solute (Cl⁻ with $R = 1$) is modeled accurately for all three grid sizes. The retarded chemical (Na⁺, $R = 2.5$) shows numerical dispersion for the coarser grids, but again, the average front locations agree. With the constant concentration-boundary condition, the number of dispersion time steps is twice the number for the flux case because of the specified condition at $x = 0$. Also the effect of the first-type boundary condition is to increase diffusion over the contact surface of the column with the outer solution. The flux of chemical over the boundary is correspondingly larger and the fronts have progressed a few meters further than in figure 2. More comparisons of analytical solutions are given in the discussion of example 11 (breakthrough at the outlet of a column) and example 12 (diffusion from a constant source).
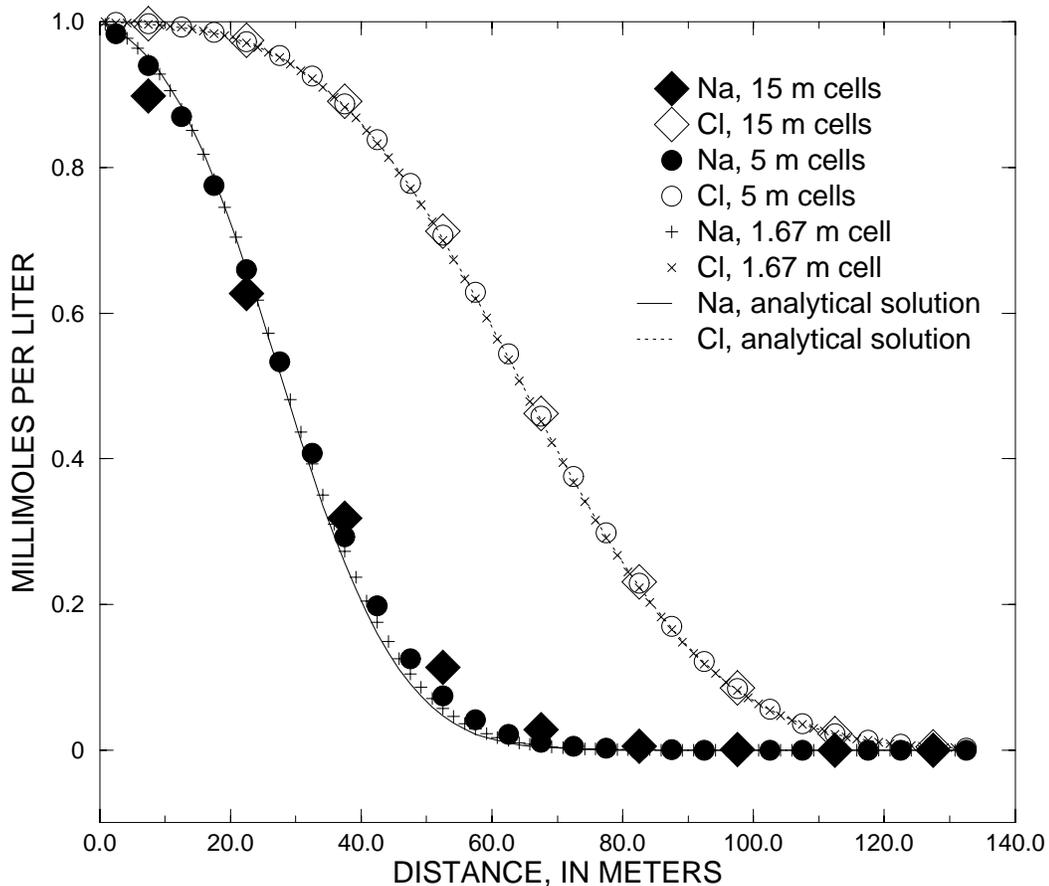

Figure 3.--Analytical solution for 1D transport with ion-exchange reactions and constant boundary condition compared with PHREEQC calculations at various grid spacings.

Transport of Heat

Conservation of heat yields the transport equation for heat, or rather, for the change of temperature. The equation is identical to the advection-reaction-dispersion equation for a chemical substance:

$$(\theta \, \rho_w k_w)\frac{\partial T}{\partial t} \; + \; (1-\theta) \, \rho_s k_s \, \frac{\partial T}{\partial t} \; = \; -(\theta \, \rho_w k_w) \, v \, \frac{\partial T}{\partial x} + \kappa \, \frac{\partial^2 T}{\partial x^2} \; , \tag{117}$$

where T is the temperature (°C), θ is the porosity (a fraction of total volume, unitless), ρ is the density (kg/m^3), k is the specific heat (kJ°C^{-1}kg^{-1}), κ is a term which entails both the dispersion by advective flow and the heat conductivity of the aquifer (kJ°C^{-1}m^{-1}s^{-1}), and subscripts w and s indicate water and solid, respectively. The temperature T is assumed to be uniform over the volume of water and solid.

Dividing equation 117 by $\theta\rho_w k_w$ gives:

$$R_T \; \frac{\partial T}{\partial t} \; = \; -v \; \frac{\partial T}{\partial x} + \kappa_L \frac{\partial^2 T}{\partial x^2} \; , \tag{118}$$

where $R_T \; = \; 1 + \dfrac{(1-\theta)\,\rho_s k_s}{\theta\,\rho_w k_w}$ is the temperature retardation factor (unitless), and $\kappa_L \; = \; \dfrac{\kappa}{\theta\,\rho_w k_w}$ is the thermal dispersion coefficient. The thermal dispersion coefficient contains a component for pure diffusion, and a component for dispersion due to advection: $\kappa_L \; = \; \kappa_e + \beta_L v$, similar to the hydrodynamic dispersion coefficient. The analogy permits the use of the same numerical scheme for both mass and heat transport.

De Marsily (1986) suggests that the thermal dispersivity β_L and the hydrodynamic dispersivity α_L may be equal, whereas the thermal diffusion coefficient κ_e is orders of magnitude larger than D_e. Thus, dispersion due to advection can be calculated in the same algorithm for both mass and heat, while thermal diffusion may require an additional calculation when it exceeds hydrodynamic diffusion. When temperatures are different in the column, and when the thermal diffusion coefficient is larger than the hydrodynamic diffusion coefficient, PHREEQC first calculates, for one time step, the temperature distribution and the chemical reactions due to thermal diffusion in excess of the hydrodynamic diffusion. Subsequently PHREEQC calculates transport for the combination of heat and mass due to hydrodynamic diffusion for the time step. The temperature retardation factor and the thermal diffusion coefficient must be defined in the input file (identifier **-thermal_diffusion** in keyword **TRANSPORT**). Both parameters may vary in time, but are uniform (and temperature independent) over the flow domain.

The similarity between thermal and hydrodynamic transport is an approximation which mainly falls short because diffusion of mass is by orders of magnitude larger in water than in minerals, whereas diffusion of heat is comparable in the two media although often anisotropic in minerals. The (small) difference in thermal diffusivity leads to complicated heat transfer at phase boundaries which is not accounted for by PHREEQC. Also, PHREEQC does not consider the convection that may develop in response to temperature gradients.

Transport in Dual Porosity Media

Water in structured soils and in solid rock has often a dual character with regard to flow: part of the water is mobile and flows along the conduits (continuous joints, fractures, connected porosity), while another part remains immobile or stagnant within the structural units. Exchange of water and solutes between the two parts may occur through diffusion. Dual porosity media can be simulated in PHREEQC either with the first-order exchange approximation or with finite differences for diffusion in the stagnant zone.

First-Order Exchange Approximation

Diffusive exchange between mobile and immobile water can be formulated in terms of a mixing process between mobile and stagnant cells. In the following derivation, one stagnant cell is associated with one mobile cell. The first-order rate expression for diffusive exchange is

$$\frac{dM_{im}}{dt} = \theta_{im} R_{im} \frac{dC_{im}}{dt} = \alpha(C_m - C_{im}), \tag{119}$$

where subscript m indicates mobile and im indicates immobile, M_{im} are moles of chemical in the immobile zone, θ_{im} is porosity of the stagnant (immobile) zone (a fraction of total volume, unitless), R_{im} is retardation in the stagnant zone (unitless), C_{im} is the concentration in stagnant water (mol/kgw), t is time (s), C_m is the concentration in mobile water (mol/kgw), and α is the exchange factor (s^{-1}). The retardation is equal to $R = 1 + dq/dC$, which is calculated implicitly by PHREEQC through the geochemical reactions. The retardation contains the change dq in concentration of the chemical in the solid due to *all* chemical processes including exchange, surface complexation, kinetic and mineral reactions; it may be non-linear with solute concentration and it may vary over time for the same concentration.

The equation can be integrated with the following initial conditions:

$C_{im} = C_{im_0}$ and $C_m = C_{m_0}$, at $t = 0$, and by using the mole-balance condition:

$$C_m = C_{m_0} - (C_{im} - C_{im_0}) \frac{R_{im}\theta_{im}}{R_m\theta_m}.$$

The integrated form of equation 119 is then:

$$C_{im} = \beta f C_{m_0} + (1 - \beta f) C_{im_0}, \tag{120}$$

where $\beta = \dfrac{R_m\theta_m}{R_m\theta_m + R_{im}\theta_{im}}$, $f = 1 - \exp\left(-\dfrac{\alpha t}{\beta\theta_{im}R_{im}}\right)$, θ_m is the water filled porosity of the mobile part (a fraction of total volume, unitless), and R_m is the retardation in the mobile area.

A mixing factor, $mixf_{im}$, can be defined that is a constant for a given time t:

$$mixf_{im} = \beta f. \tag{121}$$

When $mixf_{im}$ is entered in equation 120, the first-order exchange is shown to be a simple mixing process in which fractions of two solutions mix:

$$C_{im} = mixf_{im} C_{m_0} + (1 - mixf_{im}) C_{im_0}. \tag{122}$$

Similarly, an equivalent mixing factor, $mixf_m$, for the mobile zone concentrations is obtained with the mole-balance equation:

$$mixf_m = mixf_{im} \frac{R_{im}\theta_{im}}{R_m\theta_m} \tag{123}$$

and the concentration of C_m at time t is

$$C_m = (1 - mixf_m) C_{m_0} + mixf_m C_{im_0}. \tag{124}$$

The exchange factor α can be related to specific geometries of the stagnant zone (Van Genuchten, 1985). For example, when the geometry is spherical, the relation is

$$\alpha = \frac{D_e \theta_{im}}{(a f_{s \to 1})^2}, \tag{125}$$

where D_e is the diffusion coefficient in the sphere (m^2/s), a is the radius of the sphere (m), and $f_{s \to 1}$ is a shape factor for sphere-to-first-order-model conversion (unitless). Other geometries can likewise be transformed to a value for α using other shape factors (Van Genuchten, 1985). These shape factors are given in table 1.

An analytical solution is known for a pulse input in a medium with first-order mass transfer between mobile and stagnant water (Van Genuchten, 1985; Toride and others, 1993); example 13 defines a simulation that can be compared with the analytical solution. A 2 m column is discretized in 20 cells of 0.1 m. The resident solution is 1 mM KNO$_3$ in both the mobile and the stagnant zone. An exchange complex of 1 mM is defined, and exchange coefficients are adapted to give linear retardation $R = 2$ for Na$^+$. A pulse that lasts for 5 shifts of 1 mM NaCl is followed by 10 shifts of 1 mM KNO$_3$. The Cl ($R = 1$) and Na ($R = 2$) profiles are calculated as a function of depth.

The transport variables are $\theta_m = 0.3$; $\theta_{im} = 0.1$; $v_m = 0.1 / 3600 = 2.778$e-5 m/s; and $\alpha_L = 0.015$ m. The stagnant zone consists of spheres with radius a = 0.01 m, diffusion coefficient $D_e = 3.$e-10 m^2/s, and a shape factor $f_{s \to 1} = 0.21$. This gives an exchange factor $\alpha_L = 6.8$e-6 s^{-1}. In the PHREEQC input file α, θ_m, and θ_{im} must be given; R_m and R_{im} are calculated implicitly by PHREEQC through the geochemical reactions.

Figure 4 shows the comparison of PHREEQC with the analytical solution (obtained with CXTFIT, version 2, Toride and others, 1995). The agreement is excellent for Cl$^-$ ($R = 1$), but the simulation shows numerical dispersion for Na$^+$ ($R = 2$). When the grid is made finer so that Δx is equal to or smaller than α_L (0.015 m), numerical dispersion is much reduced. In the figure, the effect of a stagnant zone is to make the shape of the pulse asymmetrical. The leading edge is steeper than the trailing edge, where a slow release of chemical from the stagnant zone maintains higher concentrations for a longer period of time.

Finite Differences for the Stagnant Zone

As an alternative to first-order exchange of stagnant and mobile zones, a finite difference grid can be laid over the stagnant region. Fick's diffusion equations, $F = -D_e \nabla C$ and $\frac{\partial C}{\partial t} = -\nabla \bullet F$, transform to finite differences for an arbitrarily shaped cell j:

$$C_j^{t2} = C_j^{t1} + D_e \Delta t \sum_{i \neq j}^{n} \frac{A_{ij}}{h_{ij} V_j} (C_i^{t1} - C_j^{t1}) f_{bc}, \tag{126}$$

where C_j^{t1} is the concentration in cell j at the current time, C_j^{t2} is the concentration in cell j after the time step, Δt is the time step [s, equal to $(\Delta t)_D$ in PHREEQC], i is an adjacent cell, A_{ij} is shared surface area of cell i and j (m^2), h_{ij} is the distance between midpoints of cells i and j (m), V_j is the volume of cell j (m^3), and f_{bc} is a factor for boundary cells (-). The summation is for all cells (up to n) adjacent to j. When A_{ij} and h_{ij} are equal for all cells, a central difference algorithm is obtained that has second-order accuracy [$O(h)^2$]. It is therefore advantageous to make the grid regular.

The correction factor f_{bc} depends on the ratio of the volume of the mobile zone, V_m, to the volume of the boundary cell which contacts the mobile zone, V_{bc}. When the two volumes are equal, $f_{bc} = 1$. It can be shown that $f_{bc} \to 2$ when $V_m \to \infty$ (or if the concentration is constant in the mobile region, Appelo and Postma, 1993, p. 376). Likewise, $f_{bc} = 0$ when $V_m = 0$. To a good approximation therefore,

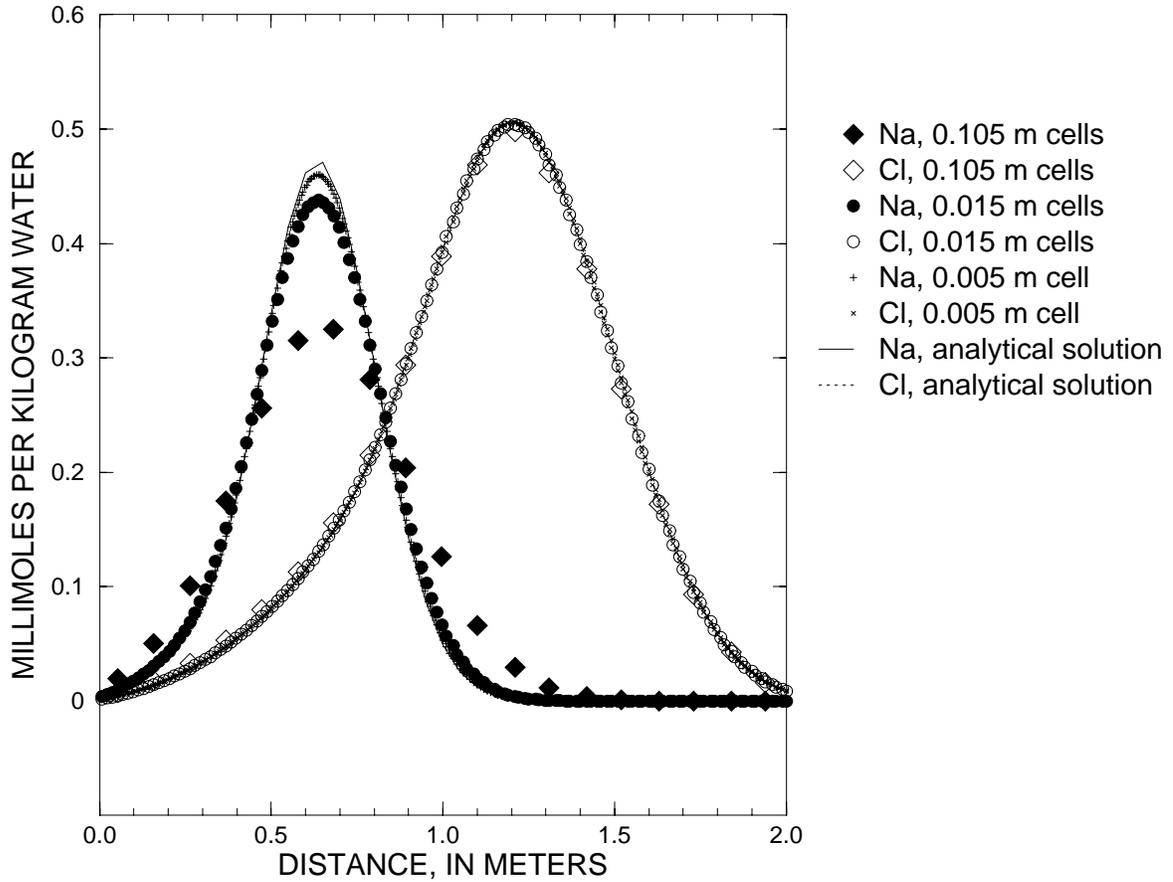$$f_{bc} = 2 \frac{V_m}{V_m + V_{bc}}. \tag{127}$$

Figure 4.--Analytical solution for transport with stagnant zones, a pulse input, and ion-exchange reactions compared with PHREEQC calculations at various grid spacings.

Equation 126 can be restated in terms of mixing factors for combinations of adjacent cells. For an adjacent cell, the mixing factor contains the terms which multiply the concentration difference (C_i - C_j),

$$mixf_{ij} = \frac{D_e \Delta t A_{ij} f_{bc}}{h_{ij} V_j} \tag{128}$$

and for the central cell, the mixing factor is

$$mixf_{jj} = 1 - D_e \Delta t \sum_{i \neq j}^{n} \frac{A_{ij} f_{bc}}{h_{ij} V_j}, \tag{129}$$

which give in equation 126:

$$C_j^{t2} = mixf_{jj} C_j^{t1} + \sum_{i \neq j}^{n} mixf_{ij} C_i^{t1}. \tag{130}$$

It is necessary that $0 < mixf < 1$ to prevent numerical oscillations. If any *mixf* is outside the range, the grid of mobile and stagnant cells must be adapted. Generally, this requires a reduction of Δt, which can be achieved by increasing the number of mobile cells. An example calculation is given in example 13, where the stagnant zone consists of spheres.

Table 1.--Shape factors for diffusive first-order exchange between cells with mobile and immobile water

Shape of stagnant region	Dimensions (*x, y, z*) or (*2r, z*)	First-order equivalent $f_{s \to 1}$	Comments
Sphere	$2a$	0.210	$2a$ = diameter
Plane sheet	$2a, \infty, \infty$	0.533	$2a$ = thickness
Rectangular prism	$2a, 2a, \infty$	0.312	rectangle
	$2a, 2a, 16a$	0.298	
	$2a, 2a, 8a$	0.285	
	$2a, 2a, 6a$	0.277	
	$2a, 2a, 4a$	0.261	
	$2a, 2a, 3a$	0.246	
	$2a, 2a, 2a$	0.220	cube $2a$x$2a$x$2a$
	$2a, 2a, 4a/3$	0.187	
	$2a, 2a, a$	0.162	
	$2a, 2a, 2a/3$	0.126	
	$2a, 2a, 2a/4$	0.103	
	$2a, 2a, 2a/6$	0.0748	
	$2a, 2a, 2a/8$	0.0586	
Solid cylinder	$2a, \infty$	0.302	$2a$ = diameter
	$2a, 16a$	0.289	
	$2a, 8a$	0.277	
	$2a, 6a$	0.270	
	$2a, 4a$	0.255	
	$2a, 3a$	0.241	
	$2a, 2a$	0.216	
	$2a, 4a/3$	0.185	
	$2a, a$	0.161	
	$2a, 2a/3$	0.126	
	$2a, 2a/4$	0.103	
	$2a, 2a/6$	0.0747	
	$2a, 2a/8$	0.0585	
Pipe wall	$2r_i, 2r_o, \infty$		$2r_i$ = pore diameter
(surrounds the	$2a, 4a$	0.657	$2r_o$ = outer diameter
mobile pore)	$2a, 10a$	0.838	of pipe (Enter wall
	$2a, 20a$	0.976	thickness $r_o - r_i = a$ in
	$2a, 40a$	1.11	Equation 125).
	$2a, 100a$	1.28	
	$2a, 200a$	1.40	
	$2a, 400a$	1.51	

For data input to PHREEQC, 1D transport including only advection is accomplished with the **ADVECTION** data block. All other 1D transport calculations, including diffusion, advection and dispersion, and advection and dispersion in a dual porosity medium, require the **TRANSPORT** data block. Initial conditions of the transport column are defined with **SOLUTION** (or **SOLUTION_SPREAD**), **EQUILIBRIUM_PHASES**, **EXCHANGE**, **GAS_PHASE**, **SOLID_SOLUTIONS**, and **SURFACE** data blocks. Kinetic reactions are defined with **KINETICS** data blocks. Infilling solutions are defined with **SOLUTION** (or **SOLUTION_SPREAD**) data blocks (see "Description of Data Input").

EQUATIONS AND NUMERICAL METHOD FOR INVERSE MODELING

PHREEQC has capabilities for geochemical inverse modeling, which attempts to account for the chemical changes that occur as a water evolves along a flow path (Plummer and Back, 1980; Parkhurst and others, 1982; Plummer and others, 1991, Plummer and others, 1994). In inverse modeling, one aqueous solution is assumed to mix with other aqueous solutions and to react with minerals and gases to produce the observed composition of a second aqueous solution. Inverse modeling calculates mixing fractions for the aqueous solutions and mole transfers of the gases and minerals that produce the composition of the second aqueous solution. The basic approach in inverse modeling is to solve a set of linear equalities that account for the changes in the moles of each element by the dissolution or precipitation of minerals (Garrels and Mackenzie, 1967, Parkhurst and others, 1982). Previous approaches have also included equations to account for mixing, conservation of electrons, which forces oxidative reactions to balance reductive reactions, and isotope mole balance (Plummer and Back, 1980; Parkhurst and others, 1982; Plummer and others, 1983; Plummer, 1984; Plummer and others, 1990; Plummer and others, 1991; and Plummer and others, 1994).

Equations and Inequality Constraints

PHREEQC expands on previous approaches by the inclusion of a more complete set of mole-balance equations and the addition of inequality constraints that allow for uncertainties in the analytical data. Mole-balance equations are included for (1) each element or, for a redox-active element, each valence state of the element, (2) alkalinity, (3) electrons, which allows redox processes to be modeled, (4) water, which allows for evaporation and dilution and accounts for water gained or lost from minerals, and (5) each isotope (Parkhurst, 1997). Also included are (6) a charge-balance equation for each aqueous solution, and (7) an equation that relates uncertainty terms for pH, alkalinity, and total dissolved inorganic carbon for each solution. Furthermore, inequalities are used (8) to constrain the size of the uncertainty terms within specified limits, and (9) to constrain the sign of the mole transfer of reactants.

The unknowns for this set of equations and inequalities are (1) the mixing fraction of each aqueous solution α_q, (2) the mole transfers of minerals and gases into or out of the aqueous solution α_p, (3) the aqueous mole transfers between valence states of each redox element α_r (the number of redox reactions for each redox element is the number of valence states minus one), and (4) a set of uncertainty terms that account for uncertainties in the analytical data $\delta_{m,q}$. Unlike previous approaches to inverse modeling, uncertainties are assumed to be present in the analytical data, as evidenced by the charge imbalances found in all water analyses. Thus, the uncertainty terms $\delta_{m,q}$ represent uncertainties due to analytical error and spatial or temporal variability in concentration of each

element, element valence state, or alkalinity, m, in each aqueous solution q. The uncertainty terms can be constrained to be less than specified uncertainty limits, $u_{m, q}$, which allows user-supplied estimates of uncertainty for each element or element valence state to limit the deviation from the analytical data ($T_{m, q}$) of revised element concentrations ($T_{m, q} + \delta_{m, q}$) that are calculated in mole-balance models.

Mole-Balance Equations

The mole-balance equations, including the uncertainty terms and redox reactions, for elements and valence states are defined as

$$\sum_{q}^{Q} c_q \alpha_q (T_{m, q} + \delta_{m, q}) + \sum_{p}^{P} c_{m, p} \alpha_p + \sum_{r}^{R} c_{m, r} \alpha_r = 0, \tag{131}$$

where Q indicates the number of aqueous solutions that are included in the calculation, $T_{m, q}$ is the total moles of element or element valence state m in aqueous solution q, $\delta_{m, q}$ can be positive or negative, $c_{m, p}$ is the coefficient of master species m in the dissolution reaction for phase p (by convention, all chemical reactions for phases are written as dissolution reactions; precipitation in mole-balance models is indicated by negative mole transfers, $\alpha_p < 0$), P is the total number of reactive phases, $c_{m, r}$ is the stoichiometric coefficient of secondary master species m in redox reaction r, and R is the total number of aqueous redox reactions. The last aqueous solution, number Q, is assumed to be formed from mixing the first Q-1 aqueous solutions, or, $c_q = 1$ for $q < Q$ and $c_Q = -1$.

For PHREEQC, redox reactions are taken from the reactions for secondary master species defined in **SOLUTION_SPECIES** input data blocks. Dissolution reactions for the phases are derived from chemical reactions defined in **PHASES** and **EXCHANGE_SPECIES** input data blocks (see "Description of Data Input").

Alkalinity-Balance Equation

The form of the mole-balance equation for alkalinity is identical to the form for other mole-balance equations:

$$\sum_{q}^{Q} c_q \alpha_q (T_{Alk, q} + \delta_{Alk, q}) + \sum_{p}^{P} c_{Alk, p} \alpha_p + \sum_{r}^{R} c_{Alk, r} \alpha_r = 0, \tag{132}$$

where Alk refers to alkalinity. The difference between alkalinity and other mole-balance equations is contained in the meaning of $c_{Alk, r}$ and $c_{Alk, p}$. What is the contribution to the alkalinity of an aqueous solution due to aqueous redox reactions or the dissolution or precipitation of phases? The alkalinity contribution of a reaction is defined by the sum of the alkalinities of the aqueous species in a redox or phase-dissolution reaction. PHREEQC defines $c_{Alk, r}$ and $c_{Alk, p}$ as follows:

$$c_{Alk, r} = \sum_{i}^{N_{aq}} b_{Alk, i} c_{i, r}, \tag{133}$$

and

$$c_{Alk, p} = \sum_{i}^{N_{aq}} b_{Alk, i} c_{i, p}, \tag{134}$$

where $b_{Alk, i}$ is the number of equivalents of alkalinity per mole of species i, $c_{i, r}$ is the stoichiometric coefficient of the species i in the aqueous redox reaction r, and $c_{i, p}$ is the stoichiometric coefficient of the species i in the dissolution reaction for phase p.

Electron-Balance Equation

The mole-balance equation for electrons assumes that no free electrons are present in any of the aqueous solutions. Electrons may enter or leave the system through the aqueous redox reactions or through the phase dissolution reactions. However, the electron-balance equation requires that any electrons entering the system through one reaction be removed from the system by another reaction:

$$\sum_{r}^{R} c_{e^-, r} \alpha_r + \sum_{p}^{P} c_{e^-, p} \alpha_p = 0,$$ (135)

where $c_{e^-, r}$ is the number of electrons released or consumed in aqueous redox reaction r, and $c_{e^-, p}$ is the number of electrons released or consumed in the dissolution reaction for phase p.

Water-Balance Equation

The mole-balance equation for water is

$$\sum_{q}^{Q} \frac{W_{aq, q}}{GFW_{H_2O}} c_q \alpha_q + \sum_{r}^{R} c_{H_2O, r} \alpha_r + \sum_{p}^{P} c_{H_2O, p} \alpha_p = 0,$$ (136)

where GFW_{H_2O} is the gram formula weight for water (approximately 0.018 kg/mol), $W_{aq, q}$ is the mass of water in aqueous solution q, $c_{H_2O, r}$ is the stoichiometric coefficient of water in aqueous redox reaction r, and $c_{H_2O, p}$ is the stoichiometric coefficient of water in the dissolution reaction for phase p.

Charge-Balance Equation

The charge-balance equations for the aqueous solutions constrain the unknown δ's to be such that, when the δ's are added to the original data, charge balance is produced in each aqueous solution. The charge-balance equation for an aqueous solution is

$$\sum_{m}^{M} \tilde{z}_m \delta_{m, q} = -T_{z, q},$$ (137)

where $T_{z, q}$ is the charge imbalance in aqueous solution q calculated by a speciation calculation and $\tilde{z}_m$ is defined to be the charge on the master species plus the alkalinity assigned to the master species, $\tilde{z}_m = z_m + b_{Alk, m}$. For alkalinity, $\tilde{z}_{Alk}$ is defined to be -1.0. The summation ranges over all elements or element valence states and includes a term for alkalinity, just as charge balance is commonly calculated by summing over cationic and anionic elements plus a contribution from alkalinity. In the definition of $\tilde{z}_m$, the alkalinity of the master species is added to the charge for that master species to remove the equivalents for the element or element redox state that are already accounted for in the alkalinity. For example, the contribution of carbonate species in equation 137 is zero with this definition of $\tilde{z}_m$ ($z_{CO_3^{2-}} = -2$, $b_{Alk, CO_3^{2-}} = 2$, $\tilde{z} = 0$); all of the charge contribution of carbonate species is included in the alkalinity term of the summation.

Isotope-Balance Equations

Geochemical mole-balance models must account for the isotopic composition as well as the chemical composition of the final aqueous solution. In general, isotopic evolution requires solving a differential equation that accounts for fractionation processes for precipitating solids and exsolving gases. In the development presented here, only the simpler case of isotopic mole balance, without fractionation, is considered. This approach is correct if aqueous mixing occurs and (or) all isotope-bearing phases dissolve, but is approximate when isotope-bearing phases precipitate or exsolve. The approach does not calculate isotopic compositions of individual redox states within the aqueous phase, only net changes in isotopic composition of the aqueous phase are considered.

Mole balance for an isotope can be written as

$$0 = \sum_q^Q \left(c_q \alpha_q \sum_m^{M_e} \left(R_{m,q}^i + \delta_{R_{m,q}^i} \right) (T_m + \delta_{m,q}) \right) + \sum_p^P c_{e,p} \left(R_{e,p}^i + \delta_{R_{e,p}^i} \right) \alpha_p, \tag{138}$$

where M_e is the number of valence states of element e, $R_{m,q}^i$ is the isotopic ratio [which may be delta notation (for example $\delta^{13}C$ or $\delta^{34}S$), ^{14}C activity in percent modern carbon, or any units that allow linear mixing] for isotope i for valence state m in aqueous solution q, $\delta_{R_{m,q}^i}$ is an uncertainty term for the isotopic ratio for a valence state in the aqueous solution, $R_{e,p}^i$ is the isotopic ratio of element e in phase p, and $\delta_{R_{e,p}^i}$ is an uncertainty term for the isotopic ratio of the element in the phase.

Expanding equation 138 and neglecting the products of δ's gives the following approximation:

$$0 \approx \sum_q^Q \sum_m^{M_e} \left(c_q R_{m,q}^i T_m \alpha_q + c_q R_{m,q}^i \alpha_q \delta_{m,q} + c_q T_m \alpha_q \delta_{R_{m,q}^i} \right) + \sum_p^P \left(c_{e,p} R_{e,p}^i \alpha_p + c_{e,p} \alpha_p \delta_{R_{e,p}^i} \right). \tag{139}$$

Commonly, $\delta_{m,q}$ will be small relative to the concentration of the valence state or $\delta_{R_{m,q}^i}$ for the isotopic ratio will be small relative to the isotopic ratio itself. In either case, the products of δ's that are neglected will be small relative to the other terms and equation 139 will be a good approximation. The approximation in equation 139 will be poor only if the concentration of the valence state and the isotopic ratio have large calculated δ's. In this case, the overall effect is that the true values of the uncertainty terms will be larger than specified uncertainty limits. The neglected terms can be made smaller by decreasing the uncertainty limits on either the valence-state concentrations or the isotopic ratios for each aqueous solution.

Relation Among pH, Alkalinity, and Total Dissolved Inorganic Carbon Uncertainty Terms

One additional equation is added for each aqueous solution to relate the uncertainty terms in pH, alkalinity, and total dissolved inorganic carbon. Unlike all other mole-balance quantities, which are assumed to vary independently, alkalinity, pH, and inorganic carbon are not independent. The following equation is used to relate the uncertainty terms for each of these quantities:

$$\delta_{Alk,q} = \frac{\partial Alk_q}{\partial C_q} \delta_{C_q} + \frac{\partial Alk_q}{\partial pH_q} \delta_{pH,q}, \tag{140}$$

where Alk_q is the alkalinity of solution q, and C_q is the total inorganic carbon of solution q. The partial derivatives are evaluated numerically for each aqueous solution.

Inequality Constraints

This formulation of the inverse problem makes sense only if the values of the δ's are small, meaning that the revised aqueous solution compositions (original plus δ's) do not deviate unreasonably from the original data. A set of inequalities places limits on the magnitudes of the δ's. The absolute value of each δ is constrained to be less than or equal to a specified uncertainty limit, $u_{m,q}$:

$$|\delta_{m,q}| \leq u_{m,q}. \tag{141}$$

Inequality constraints (equation 141) are also included for carbon(+4), alkalinity, and pH for each aqueous solution. In addition, the mixing fractions for the initial aqueous solutions ($q < Q$) are constrained to be nonnegative,

$$\alpha_q \geq 0, \tag{142}$$

and the final aqueous-solution mixing fraction is fixed to -1.0 ($\alpha_Q = -1.0$). If phases are known only to dissolve, or only to precipitate, the mole transfer of the phases may be constrained to be nonnegative or nonpositive:

$$\alpha_p \geq 0, \tag{143}$$

or

$$\alpha_p \leq 0. \tag{144}$$

Change of Variables

The system of equations for inverse modeling, formulated in the previous section, is nonlinear because it includes the product of unknowns of the form $\alpha_q(T_{m,q} + \delta_{m,q})$, where α and δ are unknowns. However, the equations can be linearized with the substitution

$$\varepsilon_{m,q} = \alpha_q \delta_{m,q}. \tag{145}$$

The mole-balance equations now become

$$\sum_q^Q c_q T_{m,q} \alpha_q + \sum_q^Q c_q \varepsilon_{m,q} + \sum_p^P c_{m,p} \alpha_p + \sum_r^R c_{m,r} \alpha_r = 0. \tag{146}$$

The alkalinity balance equation can be written as

$$\sum_q^Q c_q T_{Alk,q} \alpha_q + \sum_q^Q c_q \varepsilon_{Alk,q} + \sum_p^P c_{Alk,p} \alpha_p + \sum_r^R c_{Alk,r} \alpha_r = 0. \tag{147}$$

The electron-balance equation is unchanged. The charge-balance equation can be rewritten into

$$\sum_m^M \tilde{z}_m \varepsilon_{m,q} + \alpha_q T_{z,q} = 0. \tag{148}$$

The water-balance equation is unchanged. The isotope-balance equation 139 is

$$0 \approx \sum_q^Q \sum_m^{M_e} \left(c_q R_{m,q}^i T_m \alpha_q + c_q R_{m,q}^i \varepsilon_{m,q} + c_q T_m \varepsilon_{R_{m,q}^i} \right) + \sum_p^P \left(c_{e,p} R_{e,p}^i \alpha_p + c_{e,p} \varepsilon_{R_{e,p}^i} \right) \tag{149}$$

The relation among carbon(+4), pH, and alkalinity is

$$\varepsilon_{Alk, q} = \frac{\partial Alk_q}{\partial C_q}\varepsilon_{C, q} + \frac{\partial Alk_q}{\partial pH_q}\varepsilon_{pH, q};$$ (150)

and lastly, the inequality constraints become

$$\left|\varepsilon_{m, q}\right| \le \alpha_q u_{m, q}.$$ (151)

All of these equality and inequality equations are linear in the unknowns α and ε, and once the values of all of the α and ε are known, the values of the uncertainty terms δ can be determined.

This formulation of the inverse-modeling problem produces a series of linear equality and inequality constraints, which are solved with the algorithm developed by Barrodale and Roberts (1980). Their algorithm performs an L1 optimization (minimize sum of absolute values) on a set of linear equations subject to equality and inequality constraints. The problem can be posed with the following matrix equations:

$$AX = B$$
$$CX = D$$ (152)
$$EX \le F .$$

The first matrix equation is minimized in the sense that $\sum_i \left| b_i - \sum_j a_{i, j}x_j \right|$ is a minimum, where i is the index of rows and j is the index for columns, subject to the equality constraints of the second matrix equation and the inequality constraints of the third matrix equation. The method will find a solution that minimizes the objective functions ($AX = B$) or it will determine that no feasible model for the problem exists.

Initially, $AX = B$ is set to minimize $\sum_q \sum_m \frac{S\left|\varepsilon_{m, q}\right|}{u_{m, q}}$, where $S = 0.001$ is a scaling factor that limits the size of the coefficients in the A matrix; A is a diagonal matrix with elements $\frac{S}{u}$, and $B = 0$. The equality constraints ($CX = D$) include all mole-balance, alkalinity-balance, charge-balance, electron-balance, and water-balance equations and all inorganic carbon-alkalinity-pH relations. The inequality constraints ($EX \le F$) include two inequalities for each of the ε's, one for positive and one for negative (to account for the absolute values used in the formulation), an inequality relation for each mixing fraction for the aqueous solutions, which forces each mixing fraction to be nonnegative, and an inequality relation for each phase that is specified to dissolve only or precipitate only. Application of the optimization technique will determine whether any inverse models exist that are consistent with the constraints.

Thus, one set of mixing fractions and phase mole transfers (plus associated ε's) that satisfy the constraints may be found. Ignoring the values of the ε's and redox mole transfers (α_r), let the set of nonzero α_q and α_p (mixing fractions and phase mole transfers) uniquely identify an inverse model. The magnitude of the α's is not important in the identity of an inverse model, only the fact that the α's are nonzero in a certain set is considered. (At this point, little significance should be placed on the exact mole transfers that are found, only that it is possible to account for the observations using the aqueous solutions and phases of the inverse model.) But could other sets of aqueous solutions and phases also produce feasible inverse models? An additional algorithm is used to find all of the unique inverse models.

Assuming P phases and Q aqueous solutions, we proceed as follows: If no model is found when all Q aqueous solutions and P phases are included in the equations, we are done and no feasible models exist. If a model is found, then each of the phases in the model is sequentially removed and the remaining set of phases and aqueous solutions

is tested to see if other feasible models exist. If no model is found when a particular phase is removed, the phase is retained in the model; otherwise, the phase is discarded. After each phase has been tested and possibly discarded, the phases that remain constitute a "minimal" model, that is, to obtain a feasible model none of the phases can be removed. Three lists are kept during this process: each feasible model is kept in one list, each infeasible model is kept in another list, and each minimal model is kept in a third list.

Next, each combination of *P-1* phases is tested for feasible models in the following way. If a trial model with *Q* aqueous solutions and *P-1* phases is a subset of a model in the infeasible- or minimal-model list, the trial model is skipped because it must be either infeasible or a previously identified minimal model. If only minimal models are to be found (**-minimal** in **INVERSE_MODELING** data block), the trial model is skipped if it is a superset of a model in the minimal-model list. Otherwise, the inverse problem is formulated for the trial model and solved using the set of aqueous solutions and the *P-1* phases in the same way as described above, maintaining the three lists during the process. Once all sets of *P-1* phases have been tested, the process continues with sets of *P-2* phases, and so on until the set containing no phases is tested or until, for the given number of phases, every trial model is either a subset of a model in the infeasible- or minimal-model list.

At this point, the entire process is repeated using each possible combination of one or more of the *Q* aqueous solutions. Although the process at first appears extremely computer intensive, most sets of phases are rapidly eliminated by subset and superset comparisons with models in the three lists. The number of models that are formulated and solved by the optimization methods are relatively few. Also the process has the useful feature that if no feasible models exist, this is determined immediately when the optimization procedure is invoked the first time. For PHREEQC, during all of the testing, whenever a feasible model is found, it is printed to the output device or optionally, only the minimal models are printed to the output device.

An alternative formulation of the objective functions can be used to determine the range of mole transfer for each aqueous solution and each phase that is consistent with the specified uncertainty limits. For the "range" calculation (**-range** in **INVERSE_MODELING** data block), the equations for a given model are solved twice for each aqueous solution and phase in the model, once to determine the maximum value of the mixing fraction or mole transfer and once to determine the minimum value of the mixing fraction or mole transfer. In these calculations, the $\frac{\varepsilon}{u}$'s are not minimized, but instead, the single objective function for maximization is

$$\alpha = M , \tag{153}$$

and in the minimization case,

$$\alpha = -M , \tag{154}$$

where α refers to either α_q or α_p, and *M* is a large number. By default, the value of *M* is 1000. The optimization method will try to minimize the difference between α and 1000 for maximization and between α and -1000 for minimization. It is possible that the mixing fraction for a solution (α_q) could exceed 1000 in an evaporation problem. In this case, the method would fail to find the true maximum for α_q, and instead find a value closest to 1000. This error can be remedied by choosing a larger value for *M*. The value of *M* may be changed with the **-range** identifier in the **INVERSE_MODELING** data block.

For data input to PHREEQC, identifiers in the **INVERSE_MODELING** data block are used for the selection of aqueous solutions (**-solutions**), uncertainty limits (**-uncertainties** and **-balances**), reactants (**-phases**), mole-balance equations (**-balances**), range calculations (**-range**) and minimal models (**-minimal**).