GEM-Selektor geochemical modeling package
Revised algorithm and GEMS3K numerical kernel for coupled simulation codes

Journal Article

Author(s):
Kulik, Dmitrii A.; Wagner, Thomas; Dmytrieva, Svitlana V.; Kosakowski, Georg; Hingerl, Ferdinand F.; Chudnenko, Konstantin V.; Berner, Urs R.

ORIGINAL PAPER

GEM-Selektor geochemical modeling package: revised algorithm and GEMS3K numerical kernel for coupled simulation codes

Dmitrii A. Kulik · Thomas Wagner · Svitlana V. Dmytrieva · Georg Kosakowski · Ferdinand F. Hingerl · Konstantin V. Chudnenko · Urs R. Berner

Abstract Reactive mass transport (RMT) simulation is a powerful numerical tool to advance our understanding of complex geochemical processes and their feedbacks in relevant subsurface systems. Thermodynamic equilibrium defines the baseline for solubility, chemical kinetics, and RMT in general. Efficient RMT simulations can be based on the operator-splitting approach, where the solver of chemical equilibria is called by the mass transport part for each control volume whose composition, temperature, or pressure has changed. Modeling of complex natural systems requires consideration of multiphase–multicomponent geochemical models that include nonideal solutions (aqueous electrolytes, fluids, gases, solid solutions, and melts). Direct Gibbs energy minimization (GEM) methods have numerous advantages for the realistic geochemical modeling of such fluid–rock systems. Substantial improvements and extensions to the revised GEM interior point method algorithm based on Karpov's convex programming approach are described, as implemented in the GEMS3K C/C++ code, which is also the numerical kernel of GEM-Selektor v.3 package (http://gems.web.psi.ch). GEMS3K is presented in the context of the essential criteria of chemical plausibility, robustness of results, mass balance accuracy, numerical stability, speed, and portability to high-performance computing systems. The stand-alone GEMS3K code can treat very complex chemical systems with many nonideal solution phases accurately. It is fast, delivering chemically plausible and accurate results with the same or better mass balance precision as that of conventional speciation codes. GEMS3K is already used in several coupled RMT codes (e.g., OpenGeoSys-GEMS) capable of high-performance computing.

D. A. Kulik (✉) · G. Kosakowski · F. F. Hingerl · U. R. Berner
Laboratory for Waste Management, Paul Scherrer Institut, 5232 Villigen PSI, Switzerland
e-mail: dmitrii.kulik@psi.ch

T. Wagner · F. F. Hingerl
Institute for Geochemistry and Petrology, ETH Zurich, Clausiusstrasse 25, 8092 Zurich, Switzerland

S. V. Dmytrieva
Institute of Environmental Geochemistry, Palladin ave. 34a, 03680 Kyiv-142, Ukraine

K. V. Chudnenko
Vinogradov Institute of Geochemistry SB RAS, 1a Favorski str., 664033 Irkutsk, Russia

1 Introduction

Numerical geochemical and coupled geochemical–physical modeling is important for understanding the complexity and feedback behavior of natural geosystems, in particular, when combined with field observations and analytical data. The specific advantage of geochemical modeling of complex multicomponent–multiphase–multiprocess systems lies in the possibility to systematically explore the parameter space, to identify the feedback between key parameters, and to

investigate geosystems at pressure–temperature conditions and timescales that are not accessible to direct observation or laboratory experiments. The dramatic increase in computer speed during the last decade has facilitated construction of ever more complex models of geochemical and geophysical processes, which are now closer to the complexity of real natural systems than ever before. Examples include coupled geochemical–geophysical models of geodynamic processes [3, 37, 52], coupled models for biogeochemical cycles [42, 56, 61], and reactive transport models of the shallow to deep subsurface [22, 53].

Simulation of the spatial–temporal–chemical evolution of natural subsurface systems needs to account for transport processes (diffusion, advection) coupled with partial chemical equilibria and their feedback on the pore space and transport properties. Many coupled reactive transport codes are based on the operator-splitting approach (e.g., [57]), in which the chemical solver program is called by the mass transport part for each control volume every time when the composition, temperature, or pressure of the volume has changed. In realistic simulations, up to a million of nodes and time steps have to be specified for a reasonable 3-D spatial resolution and time discretization. Thus, chemical plausibility, robustness of results, mass balance accuracy, numerical stability, speed, and portability become crucial properties of the chemical solver. These can be defined as follows.

Chemical plausibility. Realistic modeling of natural multicomponent–multiphase systems requires that the solver must return not only the aqueous speciation in equilibrium with a set of pure solids but also the amounts and composition of many coexisting nonideal multicomponent solid solutions, sorption, fluid, and melt phases subject to optional metastability restrictions, along with saturation/stability indexes for all phases. Unlike Gibbs energy mimimization (GEM) codes such as ChemApp [2, 18] or HCh [50], law of mass action (LMA) codes such as GWB [4] or PHREEQC [38] seem to partially fail against this criterion, in particular, in chemical systems relevant to hydrothermal geochemistry and petrology.

Robustness of results. The chemical solver must compute the correct equilibrium speciation and activities of components in phases for any input composition of the system, even if the latter is not well buffered. This principal task is not trivial, and it is related to the theoretical existence of single or multiple minima of the total Gibbs energy function, to the degree of non-ideality of mixing, and to possible existence of phase miscibility gaps.

Mass balance precision. Most mass transport codes typically guarantee high precision, but may also cause numerical oscillations into the chemical composition of control volumes. The chemical solver must keep the mass balance residuals even smaller than the mass transport code (typically 10^{-13} relative), in order to avoid the accumulation of errors in simulation over time.

Numerical performance. In a typical coupled reactive mass transport (RMT) simulation of, e.g., 10^6 nodes over 10^6 time steps, the chemical solver will be called at least 10^{12} times, which poses a very high demand to its numerical stability, calculation speed, and portability to high-performance computers (HPC). In this regard, existing LMA and GEM codes appear to behave very differently [2, 25, 27, 31, 34, 38, 50, 58]. From our theoretical estimates in Section 4 below, it follows that, for similar problems, GEM codes are ca. 2 to 10 times slower than LMA codes. The reason is that the GEM algorithm finds unknown phase assemblage by minimizing total Gibbs energy while maintaining mass balance, whereas the LMA algorithm directly minimizes the mass balance residuals, but must do additional cycles if the stable phase assemblage is not known in advance.

Portability. Clearly, the compiled-linked-in code with data exchange in main memory (RAM) achieves the best performance on particular computer architecture. However, most chemical solvers are only available without the source code, either as binary dynamically linked libraries or as executable files. Sometimes, the user is forced to perform the data exchange between the transport and the chemistry parts via input/output files, which slows down the overall performance and makes the parallelization difficult. This problem can only be solved by giving access to the source code of the chemical solver.

The purpose of this paper is to describe the revised GEM interior point method (IPM) algorithm based on the convex programming method [8, 24, 25], with substantial improvements and extensions implemented recently in the GEMS3K code for using it within coupled RMT codes. We provide arguments for the central role of chemical thermodynamic calculations in RMT, and why using this particular algorithm and code would be beneficial for coupling with existing transport codes such as OpenGeoSys (OGS) [29, 30, 46, 60]. Finally, we provide an example related to RMT simulations based on an already implemented coupling in OGS-GEMS code. On the practical side, this paper should provide guidance, sufficient understanding, and necessary detail for those who wish to couple GEMS3K with their own transport code efficiently.

2 Advantages of the GEM method

Solving the phase equilibrium problem for a complex heterogeneous chemical system with multiple nonideal multicomponent solution phases (aqueous, fluid, condensed, melt, etc.) requires an efficient and stable GEM algorithm. LMA algorithms are not well suited for this purpose (see discussion and details in [6]) because some of them cannot handle nonaqueous solutions at all [4] or cannot consider in the mass balance more phases than prescribed by the Gibbs phase rule [38, 62]. The latter limitation is usually circumvented by first computing the aqueous speciation only, then using it for calculation of saturation indexes of all stoichiometrically feasible solid phases [43]. The most oversaturated phase is added to the mass balance, and the equilibrium speciation is computed again, updating also saturation indices of all solid phases. If another oversaturated phase is detected, then it is inserted to mass balance; if an undersaturated phase is found in the mass balance, then it is removed. The whole cycle is repeated until no oversaturated phase can be detected. It may not be possible to calculate accurately the saturation index of a nonideal solution phase, which makes the entire iterative selection procedure unrigorous and unreliable. A related problem is that calculations may be started from aqueous speciation in chemical systems with potentially high solid/water ratio, which would lead to unrealistically high dissolved concentrations, strongly biased activity coefficients, and unrealistic saturation indexes for some solids. Hence, the LMA method can be used with confidence and success only in water-dominated aqueous systems, where the stable mineral phase assemblage is a priori known.

Compared to chemical speciation models based on the LMA method such as PHREEQC [38], MINEQL [62], CHILLER [43–45], or GWB [4], the GEM method has the advantage that it can handle in one run any number of potentially stable phases, including many (highly) nonideal solutions. In addition, GEM calculations can directly account for processes of phase immiscibility such as liquid–vapor equilibria and solid–solid exsolution that are very important for modeling natural fluid–rock systems (e.g., [15]). Volumes of all phases in the reactive system can be found when the standard molar volumes for their components and volumes of mixing are provided. The disadvantage of GEM is a slower convergence compared to that of LMA. The applicability of GEM algorithms is limited only by the availability of consistent standard-state molar properties of end-members and nonideal interaction parameters of mixing in each multicomponent phase (all

corrected to temperature T and pressure P of interest, if necessary).

The GEM method of chemical thermodynamic modeling has been implemented in several computer codes that are used in material science, petrology, and hydrothermal geochemistry, such as ChemApp/FactSage [1, 2, 18, 19], Gibbs/HCh [5, 48–50], Selektor-C [8, 25], Perplex [10, 11], Theriak-Domino [13, 14], MELTS [16, 20], and GEM-Selektor ([31]; http://gems.web.psi.ch).

The convex programming GEM method [23–25] implemented in GEMS3K computes simultaneously the chemical speciation in all stable phases and the chemical potentials of chemical elements and charge. From that, the chemical potentials, activities, and concentrations of dependent components, saturation indices and stability criteria for phases, as well as pH and Eh in aqueous solution, are obtained. If nonideal solution phases are included, then the activity coefficients of their components are adjusted on GEM iterations and are calculated in conjunction with the primal solution. A specific feature of this approach consists in the operator splitting between GEM and mass balance refinement (MBR) steps.

One of the main goals of our work was to improve the suitability of GEM algorithm for reactive mass transport simulations. RMT simulation of the spatial–temporal evolution of a subsurface geochemical system needs to account simultaneously for fluid/tracer transport and for chemical reactions in/between phases. There are several levels of complexity of such models (Fig. 1). Existing coupled code packages (e.g., OpenGeoSys [30, 60]) use the operator-splitting approach, in which a chemical equilibrium solver is called for each volume after every time step of mass transport (Fig. 2). The chemical solver code receives the modified bulk chemical composition, temperature, and pressure of the reactive part of the volume and returns the updated (meta)stable phase speciation (perhaps with changed volumes of some solid phases affecting the porosity), which is then used for performing the next time step of mass transport. Taking into account potentially large numbers of control volumes and time steps and high precision of mass conservation maintained by the hydraulic and transport algorithms, the RMT simulations demand from the chemical equilibria solver very high levels of chemical plausibility, robustness and accuracy of results, mass balance precision, numerical stability, calculation speed, and portability to high-performance computers.

The main motivation for the development of the stand-alone GEMS3K code was to provide a fast, ro-
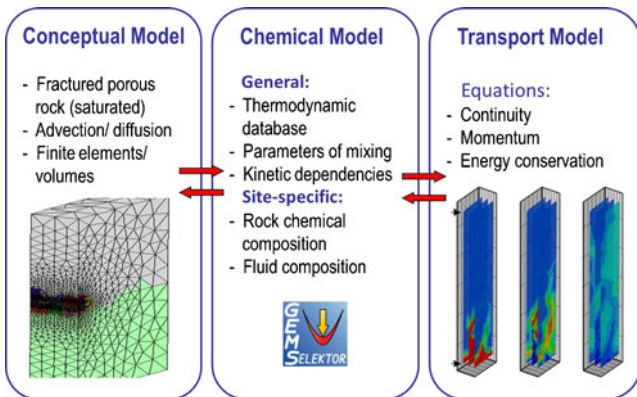
Fig. 1 Levels of complexity in RMT simulations in terms of the number of control volumes (nodes) and the number of time steps

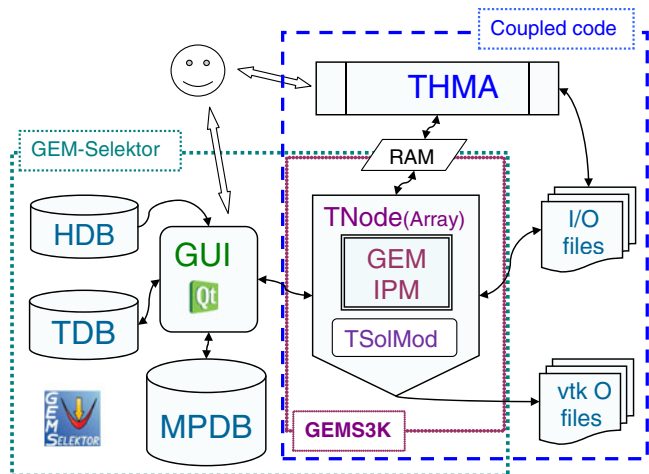| Complexity	N volumes	N steps
Reaction path (titration, T change, metastability change)	1	10 to 1000
Sequential- and flow-through reactors (transfer of fluid or solid)	10 - 100	1 - 10
Box-flux transport models (megasystem)	2 - 100	10 - 1000
1-D (column) transport (advection, diffusion)	100 - 1000	100 - 10^6
2-D and 3-D realisticTH(M)C simulation	> 10^5	> 10^6

bust, and efficient solver of complex geochemical equilibria with a simple data exchange interface that would facilitate the implementation of various operator-splitting coupled modeling codes. To date, GEMS3K has been coupled with the OpenGeoSys [30, 46, 60], CSMP++ [12], and MCOTAC [39] fluid mass transport simulation codes. GEMS3K can be potentially used in generic parameter fitting algorithms, calculation and plotting of phase diagrams, geodynamic modeling, model sensitivity analysis, and other coupled code applications. The second aim of the development of GEMS3K was to prepare it to be released with the source code in order to encourage future development of coupled codes suitable for parallel HPC platforms. The third aim was to maintain a full compatibility between the GEM-Selektor package and any coupled code by sharing the same GEMS3K kernel numerical code. This would allow rigorous benchmarking and verification of different coupled codes using the same chemical system definition and the same reactive mass transport model setup.

Fig. 2 Schematic concept of coupled RMT modeling using the GEMS chemical solver

3 GEMS3K code and its place in the GEM-Selektor framework

The overall structure of GEM-Selektor and stand-alone GEMS3K codes (all written in C/C++) and relationships between them are shown in Fig. 3. The GEMS3K code consists of the GEM IPM numerical algorithm and the TSolMod class library, embedded in the TNode/TNodeArray class interface that can exchange the data in RAM (computer memory) or via input–output (I/O) files. The GEMS3K is also integrated into the GEM-Selektor code with the graphical

Fig. 3 Diagrammatic relations between GEM-Selektor and GEMS3K codes. *MPDB* modeling projects data base; *TDB* default thermodynamic data base (provided); *HDB* help database (provided); *GUI* graphical user's interface and database management code; *GEM IPM* GEM IPM algorithm code implementation; *THMA* thermohydraulic mass transport algorithm code; *TNode* interface connecting GEM IPM kernel with existing MT code and with I/O (input/output) files containing the exported definition of chemical systems, as well as the output in VTK format; *TSolMod* library of classes for mixing models of phases of solutions [59]; *TNodeArray* extension of TNode interface for developing new coupled RMT codes

user interface and databases. In a coupled code, the upper-level thermohydraulic mass transport algorithm (THMA) code can communicate with one or more stand-alone instances of GEMS3K, depending on the number of processors in computer architecture. GEMS3K can optionally write output files in VTK format (http://www.vtk.org) commonly used in RMT simulations. The versatile new TSolMod C++ class library for equations of state and activity models of nonideal phases (described in a companion paper [59]) provides a connection between GEMS3K and many relevant application fields in Earth sciences and chemical engineering.

3.1 GEM convex programming approach

The convex programming method [8, 23–25] simultaneously computes the primal and dual solution of the GEM chemical thermodynamic problem. The primal solution $\widehat{n}^{(x)}$, usually called *speciation*, consists of mole amounts of dependent components (chemical species in phases) and their corresponding activity coefficients in equilibrium state. The dual solution $\widehat{u}$ consists of *chemical potentials* of independent components (usually chemical elements and electric charge) for the same equilibrium state. Both $\widehat{n}^{(x)}$ and $\widehat{u}$ vectors are consistent with the same scalar value of the minimal total Gibbs energy of the system. The dual solution, which cannot be computed in LMA codes, is particularly useful because from the $\widehat{u}$ vector, any activity-based thermodynamic quantities (activities, fugacities, pH, pe) of dependent components, the saturation indexes of phases, and criteria of phase stability can be readily obtained. In GEM codes, if phases with nonideal mixing are included, the activity coefficients of their dependent components (end members) are calculated together with the primal solution at every GEM iteration. The GEMS3K algorithm can handle heterogeneous systems with any number of potentially stable phases, including several (highly) nonideal solutions, and enhanced system complexity slows down the calculations only moderately.

The GEM problem consists in calculation of the (unknown) equilibrium speciation and phase assemblage in the system defined by pressure P, temperature T, bulk composition $n^{(b)}$ (expressed in absolute mole amounts of independent components), the standard state thermodynamic data for dependent components, and, optionally, parameters of models of nonideal mixing in multicomponent phases. Solving this problem means finding mole amounts of dependent components $n^{(x)} = \left\{ n_j^{(x)}, \quad j \in L \right\}$ with indexes belonging to the set

L (vector x in [24]), such that (see full list of symbols used in Appendix 1)

$$G\left(n^{(x)}\right) \Rightarrow \min\left\{G\left(n^{(x)}\right) / n^{(x)} \in M1\right\} \tag{1a}$$

$$M1 = \left\{n^{(x)} / An^{(x)} = n^{(b)}, \ n^{(x)} \in R1\right\} \tag{1b}$$

where $n^{(b)} = \{n_i^{(b)}, \ i \in N\}$ are the input amounts of independent components with indexes belonging to the set N and $A = \left\{a_{ij}, \ i \in N, \ j \in L\right\}$ is a matrix constructed from the stoichiometry coefficients of independent components in formulae of dependent components. Furthermore,

$$R1 = \begin{cases} n^{(x)} / n_j^{(x)} \geq 0, j \in D_0; \\ \underline{n}_j^{(x)} \leq n_j^{(x)}, j \in D_1; \\ n_j^{(x)} \leq \bar{n}_j^{(x)}, j \in D_2; \\ \underline{n}_j^{(x)} \leq n_j^{(x)} \leq \bar{n}_j^{(x)}, j \in D_3 \end{cases}$$

$$L = D_0 \cup D_1 \cup D_2 \cup D_3 \tag{2}$$

is the set of additional constraints on $n_j^{(x)}$ composed of trivial nonnegativity constraints (set D_0), and optional sets of input metastability constraints from below ($\underline{n}_j^{(x)} > 0, \ j \in D_1 \cup D_3$) and from above ($\bar{n}_j^{(x)} > 0, \ j \in D_2 \cup D_3$). Finally, $G(n^{(x)})$ is the total Gibbs energy function of the system:

$$G\left(n^{(x)}\right) = \sum_j n_j^{(x)} \upsilon_j, \quad j \in L. \tag{3}$$

In Eq. 3, υ_j is the normalized chemical potential of the jth dependent component, written in a simplified dimensionless form as

$$\upsilon_j = \frac{g_j}{RT} + \ln C_j + \ln \gamma_j + \Xi, \quad j \in L \tag{4}$$

where g_j is the standard state molar Gibbs energy function of the jth dependent component corrected to temperature and pressure of interest, R is the universal gas constant, and $C_j = f\left(n_j^{(x)}\right)$ is the concentration relative to the chosen standard concentration scale for the respective phase. For the component of the kth condensed nonelectrolyte solution phase and for the water solvent in aqueous electrolyte, the concentration is defined as

$$C_j = x_j = \frac{n_j^{(x)}}{\sum_{j_p} n_{j_p}^{(x)}}, \ j_p \in l_k \tag{5a}$$

where x_j is the mole fraction, and l_k is the subset of all dependent components belonging to kth phase. For aqueous electrolyte species, concentration is defined as

$$C_j = m_j = \frac{1000}{18.0153} \cdot \frac{x_j}{x_{j_w}}, \quad j_w \in l_{aq} \tag{5b}$$

where m is the molality (in moles per kilogram water), and j_w is the index of water solvent. For gas, plasma, or fluid species with indexes belonging to the subset l_g, concentration is taken as partial pressure

$$C_j = x_j P, \quad j \in l_g, \tag{5c}$$

and for any dependent component in a stable pure substance phase,

$$C_j = x_j = 1, \quad j \in l_k \text{ and } n(l_k) = 1. \tag{5d}$$

The activity coefficient γ_j of the jth dependent component in its respective phase is taken according to the chosen model of nonideal mixing (details in [59]). The concentrations and activity coefficients for adsorbed species can also be considered [32, 33]. The nonlogarithmic asymmetry term Ξ is [25]

$$\Xi = 1 - x_{j_w}, \quad \forall \ j \in l_{aq} \setminus j_w \quad \text{for the aqueous species} \tag{6a}$$

$$\Xi = 2 - x_{j_w} - 1/x_{j_w}, \quad j_w \in l_{aq} \text{ for the water solvent} \tag{6b}$$

while $\Xi = 0$ for condensed mixture end-members, gases, and pure-substance phases.

The convex set $M1$ in Eqs. 1a and 1b is called a *feasible domain*, composed of the system of mass balance constraints and of the set of constraints $R1$ (Eq. 2). If only trivial nonnegativity constraints are present in the $R1$ set, i.e., $D_1 = \emptyset, \ D_2 = \emptyset, \ D_3 = \emptyset$, then the speciation vector $\widehat{n}^{(x)}$ will be the primal solution of the problem (1a) and (1b) only when such a dual solution vector $\widehat{u}$ exists that the Karush–Kuhn–Tucker (KKT) necessary and sufficient conditions of equilibrium are satisfied [24]:

$$\upsilon - A^T \widehat{u} \geq 0 \tag{7a}$$

$$A\widehat{n}^{(x)} = n^{(b)}; \quad \widehat{n}^{(x)} \geq 0 \tag{7b}$$

$$\widehat{n}^{(x)^T} \left(\upsilon - A^T \widehat{u} \right) = 0. \tag{7c}$$

Here, T is the transpose operator, and $\wedge$ denotes "optimal," written in vector–matrix notation. Condition (7a), rewritten with indices using Eq. 4,

$$\frac{g_j}{RT} + \ln C_j + \ln \gamma_j + \Xi - \sum_i a_{ij} \widehat{u}_i \geq 0, \quad j \in L, \ i \in N \tag{8}$$

implies that for the jth dependent component present at some equilibrium concentration $C_j > 0$ in its phase, the primal chemical potential υ_j must numerically equal the dual chemical potential

$$\widehat{\eta}_j = \sum_i a_{ij} \widehat{u}_i, \quad j \in L, \ i \in N. \tag{9}$$

From Eqs. 5b, 5c and 8, it follows that (in equilibrium) the GEM dual solution $\widehat{u}_i$ values (Lagrange multipliers for balance constraints based on $n_i^{(b)}$) are the chemical potentials of thermodynamically independent components and that $\widehat{u}_i$ must have the same value in all coexisting phases. The condition (7c) zeroes off the mole amounts of unstable species and phases. The KKT conditions are actually checked over dependent components whose indexes belong to the set L_S defined as [25]

$$L_S = \left\{ j / j \in l_k, \ k \in \Phi, \ n(l_k) = 0 \vee n(l_k) \right.$$
$$\left. = 1 \vee n(l_k) > 0 \wedge n_j^{(x)} > \varepsilon_x \right\}. \tag{10}$$

Here, Φ is the set of indexes of phases, and ε_x is the numerical cutoff operational threshold ($10^{-40} < \varepsilon_x < 10^{-20}$ mol). Amounts of all species below ε_x are zeroed off, and indexes of such species are removed from the L_S set.

In the actual GEM IPM algorithm, the extended KKT conditions [8, 25] are considered for the complete "nontrivial" $R1$ set in order to simulate various metastable or so-called kinetically controlled *partial equilibrium* states:

$$\upsilon_j - \widehat{\eta}_j \geq 0, \ (\upsilon_j - \widehat{\eta}_j)\, \widehat{n}_j^{(x)} = 0, \ \widehat{n}_j^{(x)} \geq 0, \ j \in D_0 \cap L_S \tag{11a}$$

$$\upsilon_j - \widehat{\eta}_j \geq 0, \ (\upsilon_j - \widehat{\eta}_j) \left(\underline{n}_j^{(x)} - \widehat{n}_j^{(x)} \right) = 0, \quad j \in D_1 \tag{11b}$$

$$\upsilon_j - \widehat{\eta}_j \geq 0, \left(\upsilon_j - \widehat{\eta}_j\right)\left(\widehat{n}_j^{(x)} - \overline{n}_j^{(x)}\right) = 0, \quad j \in D_2 \cap L_S$$

$$(11c)$$

$$\upsilon_j - \widehat{\eta}_j + p_j \geq 0, \left(\upsilon_j - \widehat{\eta}_j + \widehat{p}_j\right)\left(\underline{n}_j^{(x)} - \widehat{n}_j^{(x)}\right) = 0,$$

$$\widehat{p}_j \geq 0, \quad \widehat{p}_j\left(\widehat{n}_j^{(x)} - \overline{n}_j^{(x)}\right) = 0, \quad j \in D_3 \qquad (11d)$$

$$\sum_j a_{ij}\widehat{n}_j^{(x)} = n_i^{(b)}, \quad i \in N, \ j \in L \qquad (11e)$$

The parameters $\widehat{p}_j$ in Eq. 11d comprise additional Lagrange multipliers conjugate to dependent components with nontrivial two-side constraints (set D_3), e.g., additional parts of the dual solution of the problem (1a) and (1b) with two-side metastability amount constraints.

3.2 The GEM IPM algorithm as implemented in the GEMS3K code

The original GEM IPM is described elsewhere [24, 25]. Later on, it was considerably modified into the IPM-2 algorithm [8, 9], claimed to be more accurate and stable. However, our attempts to couple IPM-2 (in the form of the stand-alone GEMIPM2K code) with mass transport codes (e.g., [46]) have shown that the accuracy and stability of IPM-2 were still not sufficient. The problem was finally solved by reconsidering potential pathways of error propagation in primal and dual solution approximations within the IPM algorithm, modifying the definition of convergence criteria in the MBR loops of the algorithm, developing a modified method for calculating phase and species stability, and adding internal rescaling of the input bulk composition to a fixed total number of moles of independent components in the system. This revised GEM IPM version 3 algorithm, which constitutes the core of the GEMS3K code, is briefly described below. As in many other numerical methods, the accuracy and feasibility of the initial approximation $n^{(y)}$ of the primal solution are paramount for the efficient convergence and for the quality of results. This is because the GEM IPM algorithm is organized internally in an operator-splitting mode, where the mass balance is improved in one subroutine and the direct minimization of Gibbs energy is performed in another procedure, which maintains the previously obtained mass balance precision. Hence, the mass balance precision in IPM is basically inherited from that of the preceding approximation of the $n^{(y)}$ vector.

3.2.1 Structure and flow chart of the GEM IPM algorithm

The flow chart of the revised GEM IPM algorithm is shown in Fig. 4. A more detailed description of the main steps of GEM IPM, relevant for coupling with the GEMS3K code, is given in Appendix 2.

Briefly, the chemical system definition provided at input always contains thermodynamic data for dependent components and phases plus the bulk chemical composition $n^{(b)}$ of the system. In addition, it might contain the initial chemical speciation $n^{(y)}$, obtained from a previous equilibrium calculation, that will be used as initial approximation. By default, in the automatic initial approximation (AIA) mode, the initial speciation $n^{(y)}$ that is required to proceed with the main IPM descent algorithm is obtained by solving a simplified GEM problem using the linear programming (LP) simplex algorithm, modified to tolerate nontrivial one- or two-side metastability constraints $\underline{n}_j^{(x)} > 0$ and $\overline{n}_j^{(x)} < 10^6$ mol (Appendix 2.1).

The linear programming SolveSimplex() procedure yields at most $n(N)$ nonzero values in the $n_{(s)}^{(y)}$ initial approximation vector. In order to involve all possible phases and components in the main GEM IPM calculation, all components in the $n_{(s)}^{(y)}$ vector that have zero amount after convergence of the LP simplex-method calculation must be filled out with a small value ε_f (ca. 10^{-5} mol for a total amount of 1,000 mol of all independent components in the system). This insertion distorts the mass balance with respect to the input $n_i^{(b)}$ values and increases some balance residuals to about $n \cdot 10^{-5}$ mol. To correct for this distortion, a nonlinear MBR procedure (Appendix 2.2) is used for minimizing the mass balance residuals for all independent components to less than a threshold value (typically 10^{-13}) normalized relative to the absolute input mole amount $n_i^{(b)}$ of any ith independent component. This yields a balanced feasible initial approximation (FIA) vector $n^{(y)}$, suitable as input for the MainIPMDescent() procedure. "Feasible" means that the $n^{(y)}$ vector belongs to the so-called feasible domain, i.e., the set $M1$ in Eqs. 1a and 1b.

Alternatively, if a suitable speciation vector $n^{(y)}$ is already available at input (e.g., from a previous GEM calculation), the procedure for obtaining the FIA as described above can be skipped. Substituting the input $n^{(y)}$ vector as FIA is called the smart initial approximation (SIA) mode, which normally results in much faster GEM calculations relative to the AIA mode. Unlike for the $n^{(y)}$ vector obtained in the AIA mode, it is not guaranteed that the input speciation vector used in the SIA mode will always represent the correct
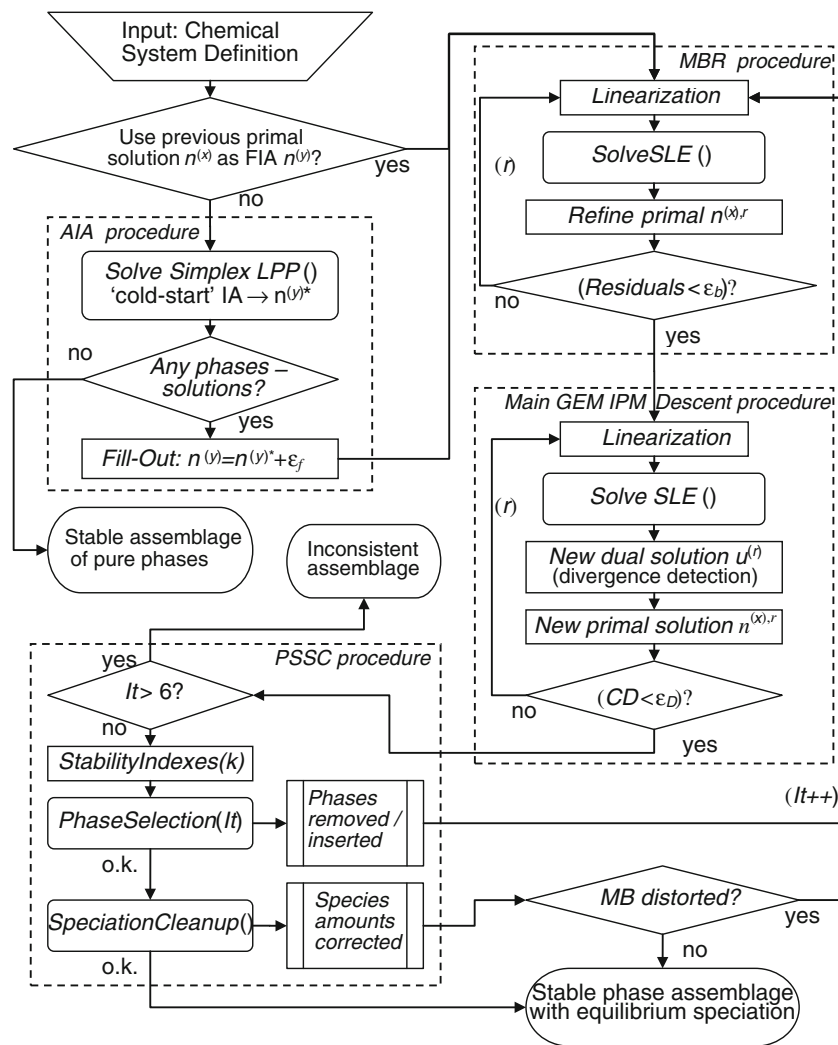
equilibrium phase assemblage. Hence, the speed gain by SIA is associated with a certain risk to obtain an inconsistent set of phases present in the mass balance. For chemical equilibrium problems involving a stable aqueous phase, gas mixture, and only pure solids, any inconsistent phase can be rigorously detected and corrected by the subsequent phase stability analysis (see Appendix 2.4), but this is not always possible if non-ideal solution phases are included in the system but they are not present in SIA. Conversely, for chemical problems with many nonideal solution phases (aqueous phase, solid solutions, fluid mixtures, etc.), the slower AIA mode is preferable.

The vector $n^{(y)}$ is substituted into the main IPM descent algorithm (Appendix 2.3), which adjusts simultaneously primal and dual solution approximations of the GEM problem at every iteration, until the criterion of convergence $C_D < \varepsilon_D$ is satisfied (ε_D is the empirical threshold chosen in the range 10^{-7} to 10^{-5}). Optionally, the new routine for detection of dual solution approx-

imation divergence is used to obtain an approximate but valid solution for cases where the chemical system is ill-defined. The GEM IPM algorithm eliminates any dependent component whose mole amount drops below the numerical threshold ε_x and proceeds to the stable phase assemblage and speciation by successively zeroing off all such unstable dependent components.

In the phase selection–speciation cleanup subroutine (Appendix 2.4), the stability indices (Ω_k, Λ_k, $f_{\alpha,k}$) for all phases are calculated from the IPM dual solution, and the quality of the entire primal solution $\widehat{n}_j^{(x)}$ is checked and (if necessary) corrected within the SpeciationCleanup() algorithm. If any phase is detected that is not present in the mass balance but should be stable or if some correction violates the mass balance residual tolerance, the loop involving the main IPM descent procedure is repeated using the current (cleaned up) primal solution as FIA. Otherwise, the highly accurate and consistent GEM dual and primal solutions are delivered as output, together with activity coefficients

of dependent components in all nonideal solution phases.

3.3 Precision and robustness of the GEM IPM algorithm

Given that consistent thermodynamic data are used, the GEM IPM algorithm must always yield a correct and accurate equilibrium phase assemblage and speciation, regardless of how different are the input total amounts of independent components, and how many pure and solution phases were included into the initial definition of the chemical system. However, in calculation of real systems, this is limited by the achievable level of consistency between the primal and dual solutions of the numerical IPM algorithm, by the mass balance precision, and by the presence or absence of internal redox buffers in the chemical system. In fact, these issues are usually not discussed with reference to chemical speciation codes, with rare exceptions (e.g., PHREEQC [38]).

The analysis of error propagation throughout the flowchart (Fig. 4) can be performed keeping in mind that (1) the GEM IPM algorithm produces two solutions (primal and dual) that are identical only in theory, (2) the GEM and the MBR steps are separated in an operator-splitting mode, (3) steps where species (dependent components) are inserted or eliminated partially degrade the mass balance, (4) correct composition–activity relationships in nonideal solution phases may be partially distorted by applied smoothing procedures (Appendix 2.7), and (5) activity coefficients of dependent components are only adjusted after each IPM iteration.

The requirement of obtaining robust results implies that the GEM IPM solver must compute correct phase assemblage, equilibrium speciation, and activities of components for any feasible input bulk composition of the system and for a wide range of numerical tolerances. This task is not trivial, and it is related to the theoretical existence of single or multiple minima of the total Gibbs energy function, to the extent of nonideality of mixing, and to the presence of miscibility gaps (see Appendix 2.6). We consider the system numerically stiff when chemical potentials of some ICs or DCs are not well constrained by physically significant amounts of relevant chemical species or persistent contributions from activity coefficients. In such a system, the resulting IPM solution may significantly depend on numerical tolerances (e.g., ε_D), including the controls of smoothing (Appendix 2.7). To detect numerically stiff systems, a special new procedure has been added to the GEMS3K IPM algorithm implementation (Appendix 2.8). In most cases, the detected stiffness manifested in

the divergence of dual solution approximation is due to the absence of redox buffering in the system. This then calls for a more careful setup of the system definition.

4 Performance of the GEM IPM kernel code

The GEM IPM algorithm described above uses an operator-splitting approach that minimizes mass balance residuals (MBR stage) and then minimizes the total Gibbs energy of the system while preserving the mass balance (main IPM descent stage). In both procedures, each iteration involves solving $n = n(N)$ linear equations with $n = n(N)$ unknowns using a $(n \cdot n)$ coefficient matrix (where $n(N)$ is the number of independent components). The MBR stage typically converges in a few iterations, whereas the IPM descent stage may need up to few thousand iterations in certain cases (see estimates below). Any GEM algorithm that simultaneously looks for the minimum in total Gibbs energy and minimizes mass balance residuals would have to solve a SLE of at least $(2n \cdot 2n)$ size at each iteration [5, 23], performing at least the same number of iterations as the IPM descent algorithm. This explains qualitatively why the operator splitting used in GEM IPM is fast and numerically efficient. Additional gain in speed can be achieved in most cases through usage of the SIA that reduces the number of IPM descent iterations by 10–20 times.

4.1 Theoretical estimates of GEM IPM computational costs as compared to other methods

It is known that the theoretical number of floating-point operations (flops) of a SLE solver using the Cholesky or LU decomposition is about $n^3/3$ and $2n^3/3$, respectively [55]. Assuming that both steps of the GEM IPM algorithm perform an equal number of iterations $n(r)$ (in practice, the MBR does much less iterations than the IPM) and that the LU decomposition is used in rare cases, the calculation time of the whole GEM IPM operator-splitting procedure will be proportional to $2/3 \cdot n(r) \cdot n^3$, where $n = n(N)$ is the number of independent components. If $n(N)$ is doubled, the calculation time must consequently increase eight times. Conversely, for a hypothetical GEM algorithm that does not use an operator-splitting approach but solves a problem of simultaneous minimization of total Gibbs energy and mass balance residuals, the size of the SLE to solve on $n(r)$ iterations would be at least $2n$, this SLE cannot be guaranteed positively definite, and the calculation time would be theoretically proportional to $1/3 \cdot n(r) \cdot [2n]^3$ (Cholesky) or $2/3 \cdot n(r) \cdot [2n]^3$ (LU). Such a

code would be four to eight times slower than the GEM IPM kernel code for the same number of independent components in the system.

By comparison, LMA algorithms such as MINEQL [62] directly minimize the mass balance deviations while maintaining the system of boundary conditions imposed by the LMA expressions and equilibrium constants for "product species." The commonly used Newton–Raphson algorithm has a quadratic convergence behavior (e.g., [28]), which means that, in order to reach the double precision of 14 decimal digits $\left(\frac{s_i^{(r)}}{n_i^{(b)}} < 10^{-14}\right)$, it needs $n(r) < 8$ iterations. For each iteration, the Jacobian matrix has to be inverted by solving the SLE with the $(n \cdot n)$ coefficient matrix; hence, the LMA algorithm is expected to require a calculation time proportional to between $7/3 \cdot n^3$ and $14/3 \cdot n^3$, i.e., roughly 2–10 times faster than the GEM IPM. However, this advantage of LMA is outweighted in complex chemical thermodynamic systems with many potentially possible solid phases that cannot all be included in the initial approximation. For such systems, many external loops in an operator-splitting mode are necessary for adding oversaturated and removing undersaturated phases in the mass balance (see Section 1, see also [6]).

A combined GEM-LMA algorithm (e.g., in the HCh code, Shvarov, 2011, personal communication) starts with obtaining a rough GEM primal solution, which is then improved by successive LMA runs, followed by application of phase stability criteria to decide which phases to remove and which to insert into the mass balance. Because the computational cost of the GEM calculation is about $n(r) \cdot n^3$ (with $n(r)$ about 100) and that of LMA is about $4l \cdot n^3$ (l is the number of loops for phase stability checking, typically about 10), the total cost in this case would be $(n(r) + 4l) \cdot n^3$, i.e., in multiphase systems definitely greater than that of the operator-splitting GEM IPM algorithm. However, this toll is justified for systems with diverging chemical potentials of some independent components, because the GEM method alone cannot accurately solve such equilibria.

GEM algorithms used mainly for plotting complex phase diagrams in petrology, such as Theriak [13, 14] and Perplex [11], reduce the nonlinear minimization problem (1a) and (1b) to a sequence of LP problems solved by the simplex method. Simplex LP algorithms have the exponential time complexity [55] but show in practice the performance of $O \cdot (m + n)$, so they are in most cases very efficient algorithms when n and m are not very large. In the given problem (1a) and (1b), the sizes are $n = n(N)$ (basis) and $m = n(\Phi) \leq n(L)$,

and the number of dependent components $n(L)$ can be several times that of $n(N)$. Assuming that for a chemical system typically $n(L) = 9 \, n(N)$, the cost of one simplex LP run can thus be estimated as $100 \cdot n$. In the Perplex code [11], any solution phase is compositionally discretized to a large number of phases with fixed composition (termed pseudocompounds), but only one LP step per equilibrium speciation problem is done, so the computational cost can be about $10^4 \cdot n$.

In the Theriak [14] code, the LP run for the whole system is embedded in several l loops where solving the LP is preceded by adjusting equilibrium compositions of all multicomponent phases [13], which is done by a steepest descent nonlinear minimization of Gibbs energy of each phase composed of n_f end members ($n_f = 4 \pm 2$). Assuming the cost of the steepest descent algorithm to be about $n(r) \cdot (n_f)^3$ (where $n(r)$ is typically 15 ± 10) and the number of solution phases equal to n, the total cost can be estimated as $l \cdot (n \cdot 15 \cdot 4^3 + 100 \, n)$ or about $l \cdot 10^3 \cdot n$, i.e., comparable to that of Perplex. For a system with 10 independent components, performance of both LP-based algorithms is slightly slower than that of LMA, but for a system with 30 independent components and many multicomponent phases, it may be significantly faster. However, LP-based algorithms suffer from inefficient calculation of aqueous speciation and from insufficient mass balance precision due to the technical limits of phase composition discretization to pseudocompounds (J. Connolly, personal communication).

The estimated computational costs (summarized in Table 1) show that GEM IPM is significantly slower than LMA or LP-based algorithms, but substantially faster than other GEM or combined GEM + LMA methods. The rather high uncertainty in such estimates comes from the complexity of the chemical system and must be associated with the number of GEM iterations $n(r)$, usually ranging between 1 and 1,000, and the number of phase stability loops l, assumed to be between 1 and 20.

4.2 Impact of nonideal solution phases

Another important factor that may strongly affect the algorithm performance is the presence of strong non-ideality of mixing in some phases, which can either disturb the convergence of the IPM or cause difficulties in the vicinity of the consolute point of miscibility gaps. In this area, the differences in Gibbs energy between exsolving phases, as well as the IPM gradients, become very small, slowing down the convergence. However, GEM IPM always converges correctly in this region, and the number of iterations dramatically decreases

Table 1 Comparative estimates of computational costs in flop (assuming optimal coding)

Code type	Method	SLE dim.	n of loops	Costs, flop
LMA	Newton–Raphson method to minimize MBRs + Jacobian inversion to solve for LMAs	$n \times n$	$l = \max(1, \Phi - n)$	$\sim 4l \cdot n^3$
GEMS3K	Operator splitting between GEM and MBR steps (IPM)	$n \times n$	$n(r)$ avg. 60 (AIA) or 6 (SIA)	$\sim 2/3 \cdot n(r) \cdot n^3$
GEM implicit	Simultaneously minimizes Gibbs energy and MBRs (Newton method)	$2n \times 2n$	$n(r)$ avg. 60	$> 8/3 \cdot n(r) \cdot n^3$
GEM + LMA	Operator splitting between GEM and multiple LMA steps	$n \times n$	$n(r)$ avg. 60, $l = \max(1, \Phi - n)$	$\sim (n(r) + 4l) \cdot n^3$
LP-based (Perplex)	Operator splitting between LP simplex and phase-solution composition discretization	$n \times 10^4 n$	$n(r) = 1, l < 5$	$\sim l \cdot n \cdot 10^4$
LP-based (Theriak)	Operator splitting between LP simplex and phase-solution composition 'equilibration' by GEM	$n \times 9n$	$n(r) = 15, l < 100, n_f = 4$	$\sim l \cdot n \cdot (n(r) \cdot (n_f)^3 + 100)$

MBRs: mass balance residuals

when moving away from the consolute point. In some systems of this type, up to 7,000 IPM iterations may be required (regardless whether AIA or SIA mode), while less than 100 iterations are sufficient in most chemical systems when the AIA mode is used. In the SIA mode, the typical number of iterations may be 10 to 50 times less. The SIA mode has, therefore, a high potential for speeding up coupled reactive transport simulations.

5 Data exchange interface (TNode class), memory structures, and input/output files

The revised GEM IPM algorithm described above is implemented as a C++ class, which is embedded in the TNode C++ class implementing the data exchange interface. Together, both classes comprise the GEMS3K code, compatible with doxygen specifications (see http://www.doxygen.org) for generating the source code documentation. One particular goal of the GEMS3K development was to retain the code implementation and handling of input/output data as simple as possible, as a prerequisite for compiling the coupled RMT-GEM program on any computer architecture, up to parallel HPC clusters. Consequently, the GEMS3K program can only use the data for a compressed GEM problem encapsulated in the IPM work data structure. Because of this, the GEMS3K code has some restrictions compared to the full GEM-Selektor-3 package (Table 2). Based on considerations of data organization and exchange efficiency, the input data

Table 2 Differences in functionality between the stand-alone GEMS3K program and its variant embedded in the GEM-Selektor v.3 code package

Functionality	GEMS3K in stand-alone codes	GEMS3K in GEM-Selektor-3 code
Phase- and process-scripted models	Not possible	Used
Built-in TSolMod class functions for activity coefficients	Used	Used
T–P corrections for input thermodynamic data	Lagrangian interpolation using 2-D lookup arrays (exported into DCH file from GEM-Selektor)	Direct access to default or project thermodynamic data base and built-in thermodynamic and EoS P,T correction functions
Access to input, output, and work data	Batch mode only (data exchange via I/O files or in RAM)	Direct interactive GUI access, including multithread stepwise calculation mode; many dialogs and data access screen forms
Runtime help and tooltips	Not available	Available as runtime local HTML help system, with some enhancements on GEMS web site

of GEMS3K have been subdivided into three main subsets, as shown in Table 3. Only one relatively large DCH and one IPM file contain all the data relevant for the definition of the chemical system, models of mixing, and numerical settings of the GEM IPM algorithm. The node-specific data are exclusively contained in a set of small and compact input–output DBR files. The advantage of having three different types of input files instead of one single file type is that in a typical RMT problem with a large number of nodes, higher performance and flexibility of the coupled code can be achieved at much less amount of data exchange.

The `TNode` class, designed for connecting GEMS3K to upper-level (e.g. mass transport) codes, provides a collection of access methods (function calls) for loading the input chemical system definition and related thermodynamic and composition data into the IPM work structure, running the GEM IPM algorithm for one calculation of equilibrium state, and extracting the calculated speciation results. For the subsequent visualization, the output in VTK format files (http://www.vtk.org) is also supported. The `TNode` class uses two data exchange interface structures, which are `DATACH` and `DATABR`. The structures can be allocated in the computer memory and written into or read from text/binary files (DCH and DBR file, respectively). The parameters of nonideal mixing models in phases, as well as the numerical controls and tolerances of GEM IPM, can be exchanged via the IPM file (loaded directly into the IPM work structure). A set of direct access methods to parameters of mixing models and thermodynamic data kept in IPM work structure is provided in the TNode class for the efficient implementation of coupled codes for GEM input parameter fitting, in particular, GEMSFIT.

Because the `TNode` class is also plugged into the GEM-Selektor package, DCH, IPM, and DBR files can be exported using a specific GUI dialog and then directly used as input in the stand-alone coupled code, as described below. Moreover, results of stand-alone calculations of equilibria in nodes can be written in DBR files that can be imported into a "parent" GEM-Selektor modeling project as additional system-equilibrium (SysEq) records (details in GEM-Selektor runtime help). There, the input and output GEM data can be conveniently explored and compared with internally recomputed speciation. This functionality can save the user a lot of time for tedious work with I/O files. It supports our recommendation first to set up in a GEM-Selektor project the chemical system definitions and initial bulk compositions of node types and test them within relevant temperature and pressure intervals, before exporting them in GEMS3K I/O files and starting expensive calculations with a stand-alone RMT code.

5.1 The DATACH data structure

The `DATACH` structure (mirrored in the DCH input file) contains the input definition of the chemical system visible to both GEMS3K and to the external coupled code. The respective dimensions, lists of components and phases, etc. will be automatically copied into the IPM work data structure (belonging to the `TMulti` class) upon initialization of the `TNode` class instance, which includes reading of one DCH, one IPM, and at least one DBR file.

Representation of thermodynamic data for variable pressure–temperature conditions. Direct access to the GEM-Selektor thermodynamic database is not possible, and extensive functionality for pressure–

Table 3 Main data subsets used in `TNOde` class for I/O file exchange

Data subset	Data structure	Code visibility	I/O file	Comment
DCH	`DATACH`	GEM and MT parts	`*DCH` `.dat`	Input definition of the chemical system (compressed) with selection of components and phases to be exchanged with the mass transport part (common to all nodes of the RMT problem)
DBR	`DATABR`	GEM and MT parts	`*DBR` `.dat`	Data bridge for node-specific chemical input and output properties (bulk composition, speciation, temperature, pressure, optional metastability constraints, etc.)
IPM	`MULTI`	GEM part only	`*IPM` `.dat`	Input information not relevant for the MT part but needed for GEM IPM calculations (IPM numerical settings, parameters of nonideal mixing models, adsorption models, etc.)

DCH and DBR files are automatically (un)packed to/from the `MULTI` (IPM work data) memory structure using functions provided by the `TNode` class

temperature corrections of thermodynamic data is not available in the stand-alone GEMS3K code. Therefore, the retrieval of thermodynamic data for dependent components is implemented in GEMS3K by an interpolation through lookup arrays provided in the DATACH structure and in the DCH file. These lookup arrays can be generated through a GUI dialog in the GEM-Selektor-3 package from the modeling project thermodynamic database when exporting the DCH file. To use the lookup arrays, the TNode class employs a 2-D Lagrangian interpolation routine that extracts thermodynamic data for temperature and pressure of interest and loads them into the IPM work data structure before starting the next GEM IPM calculation.

5.2 The DATABR data structure

This data structure (and DBR file), which contains all information relevant for node data exchange with the external multispecies mass transport code, relies on dimensions provided in the DATACH structure. Hence, the DCH file must be read before IPM and DBR file(s) in order to initialize the IPM work structure and the DATABR instances and to perform the dynamic memory allocation in them.

Modes of interaction between GEMS3K and the upper-level code. For calling GEM IPM calculations, the NodeStatusCH flag sets the GEM initial approxima-

tion mode at input and will be reset to a GEM IPM return code after each GEM calculation. Hence, the upper-level code must reset this flag every time when the GEM calculation should be performed, and check the return value immediately after getting the control back. The most important situations (at TNode level of coupling) are summarized in Table 4.

In the AIA mode, GEM calculations are robust and reproducible but may be comparatively slow (20 to 1,000 IPM iterations, depending on the chemical system definition and composition). The alternative SIA mode uses the current contents of the speciation (xDC) and activity coefficients gam vectors (perhaps modified through the DATABR interface after the mass transport step) as the initial approximation. If the xDC vector for the current node does not violate the mass balance relative to the current bulk composition vector (bIC) and neither the phase assemblage nor the redox state is changing, then GEM IPM is expected to perform only very few iterations (i.e., 10 to 20 times faster than in the AIA mode). This substantial acceleration can be paramount for running large 2-D or 3-D coupled RMT models.

The SIA mode may be preferable over the AIA mode when the mass transport part only moves aqueous species. In this case, any changed chemical speciation in the node is guaranteed to remain mass-balanced because the corrections to the bulk composition vector ($n^{(b)}$ in Eq. 1b, called bIC in DATABR structure and DBR file) are calculated from the corrections to speciation (primal solution) vector $n^{(x)}$ in Eq. 1a, called

Table 4 Modes of GEM IPM initial approximation and required NodeStatusCH actions

|	
AIA mode of GEM calculations: slow, but yields robust results	Before calling GEM run: NodeStatusCH = NEED_GEM_AIA (=1).
	After GEM run: check that NodeStatusCH = OK_GEM_AIA (=2).
	If yes: copy GEM results to MT data structures and proceed with the next node.
	If NodeStatusCH = BAD_GEM_AIA (=3), then check whether the current bulk composition in the reactive subsystem is well balanced/buffered, also with respect to redox state.
	If NodeStatusCH = ERR_GEM_AIA (=4) or = T_ERROR_GEM (=9), then generate a diagnostic message and break the coupled modeling calculations.
	Terminal error code 9 (T_ERROR_GEM) indicates a corruption of the dynamic memory in GEMS3K.
SIA mode of GEM calculations: Fast, but results may be not always robust	Before calling GEM run: set NodeStatusCH = NEED_GEM_SIA (=5).
	After GEM run: check that NodeStatusCH = OK_GEM_SIA (=6). or = OK_GEM_AIA (=2).
	If yes: copy GEM results to MT data structures and proceed with the next node.
	NodeStatusCH = OK_GEM_AIA (=2) implies that the GEM IPM algorithm had found the node chemical speciation inappropriate as initial approximation and has switched to the automatic (simplex) initial approximation (AIA).
	Error codes 3 or 4 have the same meaning as above for the AIA start.
	NodeStatusCH may return BAD_GEM_SIA (=7) or ERR_GEM_SIA (8); if so, try restarting GEM run for this node in the AIA mode.

"`xDC`" in DATABR). The latter then is a good initial approximation at least in the sense of mass balance, and the MBR procedure is not expected to perform any additional iterations.

In meaningful RMT problems, typically the phase assemblage changes or redox transitions occur across some distinct reaction fronts that are represented by only a small fraction of the nodes. In such a node, the current speciation taken in SIA mode might be an invalid GEM initial approximation. The GEM IPM algorithm has several built-in criteria that can detect this condition (see Appendices 2.2, 2.3, and 2.8 for details). In cases where this situation is detected, the program automatically switches to the AIA and obtains a correct new phase assemblage and speciation at more computational cost. However, for the vast majority of nodes, slight changes of speciation and composition as a consequence of mass transport would not result in a phase transformation or a redox transition.

6 Experience from the provisional coupling of GEMS3K with transport codes

6.1 OGS-GEMS

OGS, hosted at the Helmholtz Centre for Environmental Research (UFZ), is a research initiative for the development of computational methods for the solution of thermo–hydro–mechanical–chemical processes in porous-fractured media. Applications are in the fields of groundwater and surface hydrology, water resources, geotechnics, radioactive waste disposal, geothermal energy, and CO_2 storage [7, 30, 35, 46, 47, 51, 54, 60, 63]. The concept of OpenGeoSys is a distributed open source scientific software development that should benefit from joint efforts of all involved groups. The OpenGeoSys modules for simulation of groundwater flow and transport of dissolved species have been coupled to the GEMS3K kernel. The coupling strategy, verification examples, and the first applications were briefly described by Shao et al. [46]. Like for similar coupled codes, application of the OGS-GEMS coupled code is limited by long calculation times and extensive memory requirements. Driven by these shortcomings, we recently parallelized the central geochemical loop in OGS-GEMS that contains independent calls to GEMS3K for each finite element node. We used an existing MPI implementation of OGS [60] and combined this with a multithreaded calculation of the coupling loop [64]. The combination of multithreading and MPI provides an effective and flexible environment to speed
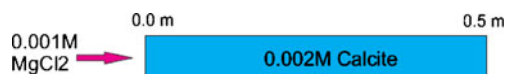

Fig. 5 Model domain, transport boundary, and initial conditions [29] for the RMT example

up OGS-GEMS calculations while limiting the required memory use. We would like to remark that this was only possible because the GEMS3K code is thread-safe, i.e., it is safely possible to run multiple instances of GEMS3K, as these do not share any data structures.

6.2 Benchmark example

The verification example is based on a hypothetical setup of a benchmark test case, which conceptionally considers 1-D advection-dominated mass transport, instantaneous calcite dissolution, and dolomite precipitation processes. This benchmark test was first proposed by Engesgaard and Kipp [17] for model verification of the MST1D code and later used by Prommer [41] for the PHT3D code. It is also used as an official test case for different couplings between OGS and chemical equilibrium solvers. The following description is adapted from [29, 46]. A 1-D column that initially contains calcite and water saturated with calcite in the pore space is continuously infiltrated by water that contains dissolved magnesium chloride (Fig. 5). With the progressive movement of the $MgCl_2$ advection front, calcite dissolves, and dolomite is temporarily formed as a product of incongruent dissolution but eventually dissolves as well. The media properties and the thrrmodynamic setup are detailed in [46]. For OGS-GEMS calculations, all the independent components
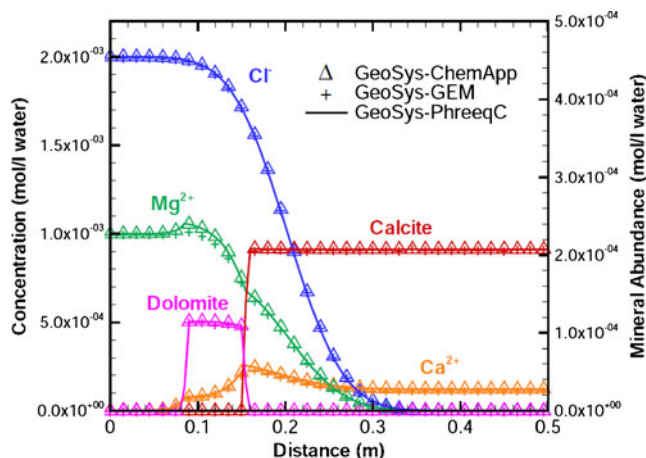
Fig. 6 Benchmark results from OGS-ChemApp (*triangles*), OGS-PHREEQC (*solid lines*), and OGS-GEMS (*crosses*) coupled simulations [29]. Temperature 25 °C, pressure 1 bar

and species need to be explicitly set for initial and boundary conditions. In this example, all concentration values are given in the unit of moles per cubic meter of water. This test has been simulated by OGS-PHREEQC, OGS-ChemApp, and OGS-GEMS. In all these benchmarks, the same GEMS built-in version of Nagra/PSI 01/01 chemical thermodynamic database was used. Figure 6 depicts a comparison of simulation results, which shows excellent agreement between the different coupled RMT code implementations.

7 Outlook

The GEMS3K geochemical equilibrium solver facilitates geochemical modeling of multicomponent–multiphase equilibria, delivering robust and numerically stable results with high mass balance precision that is comparable to conventional speciation codes that use the LMA method. Based on the revised GEM IPM convex programming algorithm, GEMS3K is able to accurately solve complex geochemical systems with many nonideal solution phases, including dilute to concentrated aqueous solutions, solid solutions, gas mixtures, supercritical fluids, melts, and sorption phases. The modular and computationally efficient code design, foreseen to be released an open source under the LGPL v.3 license [21], is favorable for coupling GEMS3K to existing mass transport codes, developing new RMT codes, and implementing advanced geophysical–geochemical models that simulate geodynamic processes and global geochemical cycles. Because GEMS3K is also included into the GUI-driven GEM-Selektor code package, all input data required for coupled reactive transport simulations can be straightforwardly created and tested in GEM-Selektor and then exported interactively into a file set for use in stand-alone codes coupled with GEMS3K.

Future developments will include substantial extension of the built-in selection of equation-of-state and activity models for multicomponent phases available in the TSolMod class library, reimplementation of sorption models into a TSorpMod class library using a object-oriented code architecture similar to TSolMod, implementation of models for mineral dissolution and precipitation kinetics in a TKinMet class library, extensions to enable the GEMSFIT coupled code to access thermodynamic data for dependent components, coupling GEMS3K to codes for calculation and plotting phase diagrams, and implementing methods that would facilitate consistent minimization of thermodynamic potentials other than Gibbs energy [26], such as Helmholtz energy and negative entropy.

Appendix 1

Table 5 List of mathematical symbols and acronyms

Symbol	Explanation
Mathematical symbols	
A	Matrix of stoichiometry coefficients a_{ij} of the ith IC in formula of jth DC
$a_j^{(\widehat{u})}$	Dual thermodynamic activity of the jth DC in its phase
α_γ	Smoothing factor for improving IPM convergence in highly nonideal systems
C_D	Dikin's criterion of convergence of the IPM descent algorithm
c_j	Constant in the LP minimization subproblem
C_j	Concentration of the jth DC in its respective phase
D_0	Set of indexes of dependent components with trivial nonnegativity constraints
D_1	Set of indexes of DCs with additional constraints from below $\underline{n}_j^{(x)}$
D_2	Set of indexes of DCs with additional constraints from above $\overline{n}_j^{(x)}$
D_3	Set of indexes of DCs with additional constraints below $\underline{n}_j^{(x)}$ and above $\overline{n}_j^{(x)}$
Δ	Vector of correction of the $n^{(y)}$ approximation on MBR or IPM iterations
$\delta_j^{(x)}, \overline{\delta}_j, \underline{\delta}_j$	Magnitudes of correction of the jth DC amount in speciation Cleanup()
$\delta_j^{(\mu)}$	Difference between primal and dual DC chemical potential in Cleanup()
ε_α	Empirical uppermost value for the smoothing factor α_γ
ε_δ	Empirical minimum value for the smoothing factor α_γ
ε_b	Relative tolerance for the IC mole balance residuals ς_i
ε_D	Tolerance for checking the Dikin's criterion of IPM convergence
$\varepsilon_{\delta\mu}$	Tolerance for $\delta_j^{(\mu)}$ to perform DC amount correction in Cleanup()
$\varepsilon_{e\Phi}$	Tolerance for detecting the mass balance violation in Cleanup()
ε_f	Empirical fill-out constant to obtain the initial approximation $n^{(y)}$ from $n_{(s)}^{(y)}$

Table 5 (continued)

Symbol	Explanation
ε_{fi}	Empirical constant for insertion of a single-component phase in Cleanup()
$\varepsilon_{n\Phi}$	Tolerance for the amount of stable phase present in the mass balance
$\varepsilon_{\Phi-}$	Tolerance for the stability index Ω_k for kth phase for elimination
$\varepsilon_{\Phi+}$	Tolerance for the stability index Ω_k for kth phase for insertion
ε_u	General tolerance for the detection of diverging dual solution approximation
ε_x	Tolerance for zeroing off the amounts of dependent components $n_j^{(x)}$
$f_{\alpha,k}$	Karpov's (old) phase stability index
f_j^o	Fugacity of the jth pure nonideal gas/fluid component at T,P of interest
Φ	Set of indexes of phases in the chemical system definition
$G(n^{(x)})$	Total Gibbs energy of the chemical system (scalar) in moles
g^{o}	Input vector of standard state Gibbs energy functions (per mole) of dependent components at temperature of interest and reference state pressure in J mol^{-1}
g_j^{DQF}	Darken's parameter for a component of solid solution or melt
γ_j	Activity coefficient of the jth dependent component in its phase
i	Index of the independent component
It	Index of the PhaseSelection() loop iteration
j	Index of the dependent component
j_{w}	Index of H$_2$O solvent in aqueous phase
k	Index of the phase
Λ_k	Logarithmic phase stability index used in PhaseSelection() procedure
L	Set of indexes j of dependent components
L_{S}	Set of indexes of dependent components subject to KKT conditions (L$_{\text{S}} \subseteq$ L)
l_{aq}	Set of indexes of dependent components in aqueous phase
l_{g}	Set of indexes of dependent components in gas/fluid phases
l_k	Set of indexes of dependent components belonging to the kth phase (l$_{\text{k}} \subseteq$ L)
λ_j	Constant (initial) activity coefficient term
$\lambda_b^{(r)}, \lambda_r$	Scalar optimal step length on MBR or IPM descent iterations
m_j	Molality of the jth aqueous species
$M1$	Set of constraints to the $G(n^{(x)})$ minimization problem
μ	Chemical potential
$\widehat{\eta}_j$	Dual-solution chemical potential of the jth dependent component
N	Set of indexes i of independent components
n	Number of unknowns in the internal SLE (MBR and IPM algorithms)
$n^{(b)}$	Input vector of total amounts of independent components $n_i^{(b)}$

Table 5 (continued)

Symbol	Explanation
Mathematical symbols	
$n_k^{(\Phi)}$	Total amount of the kth phase (the sum of mole amounts of all its components)
$n^{(x)}$	Speciation vector of mole amounts of dependent components $n_j^{(x)}$
$n^{(y)}$	Initial approximation of the speciation vector $n^{(x)}$ (for the IPM algorithm)
$n_{(\text{s})}^{(y)}$	Initial approximation of $n^{(x)}$ obtained in the simplex LP step
$\widehat{n}^{(x)}$	Primal (optimal) solution of the GEM IPM problem—DC mole amounts $\widehat{n}_j^{(x)}$
$n_j^{(x,c)}$	Corrected amount of dependent component in speciation Cleanup()
$\underline{n}_j^{(x)}$	Additional constraint from below on the mole amount of jth DC $n_j^{(x)}$
$\overline{n}_j^{(x)}$	Additional constraint from above on the mole amount of jth DC $n_j^{(x)}$
$n(A)$	The number of elements in the set A
$n(r)$	Number of iterations done by MBR() or MainIPMDescent() procedure
Ω_k	Phase stability index used in PhaseSelection() procedure
P	Pressure (bar)
p_r	Lagrange multiplier related to the ellipsoid of constraints in IPM
$\widehat{p}_j$	Lagrange multiplier conjugate to the two-side additional constraint on $n_j^{(x)}$
$q_j^{(r)}$	Weight multipliers used in the IPM descent algorithm
Q^2	Diagonal matrix of weight multipliers q_j used in the MBR() algorithm
R	Universal gas constant, 8.31451 J K^{-1} mol^{-1}
$R1$	Set of additional constraints on the elements $n_j^{(x)}$ of the speciation vector $n^{(x)}$
r_{ik}	Coefficient of the matrix for linear equation solver in MBR() and main IPM descent
$(r), r$	Superscript or subscript index of MBR() or MainIPMDescent() iteration
T	Temperature (in degrees Kelvin)
T	Operator transpose (for vector, matrix)
u	Vector of Lagrange multipliers conjugate to elements of n$^{(b)}$ vector
$\widehat{u}$	Dual (optimal) solution of the GEM IPM problem—IC chemical potentials $\widehat{u}_i$
υ	Vector of approximations of primal chemical potentials υ_j
v	Vector of Lagrange multipliers conjugate to mass balance residuals
x_j	Mole fraction of jth dependent component in its phase
$\omega_j^{(\widehat{u})}$	Activity of jth DC in mole fraction scale, estimated from the dual solution
ϖ_j	Mole fraction of DC in its phase, estimated from the dual activity ω_j

Table 5 (continued)

Symbol	Explanation
Ξ	Term for conversion between internal and external activity coefficients
ς_i	Residual of the mole balance for ith independent component
Acronyms	
AIA	Automatic initial approximation
DC	Dependent component (chemical species)
FIA	Feasible initial approximation
Flop	Floating-point operation
Flops	Floating-point operations per second
HPC	High-performance computer
GEM	Gibbs energy minimization
GUI	Graphical user interface
I/O	Input/output
IC	Independent component (usually chemical element, charge, or ligand)
IPM	Interior points method (of non-linear minimization)
KKT	Karush-Kuhn-Tucker conditions (of equilibrium state or $G(n^{(x)})$ minimum)
LP	Linear programming (minimization)
LMA	Law of mass action
MBR	Mass balance refinement (procedure)
MT	Mass transport (part of the coupled code)
PSSC	Phase selection speciation cleanup (procedure)
RMT	Reactive mass transport (algorithm, code)
SIA	Smart initial approximation
SLE	System of linear equations

Appendix 2

Details on revised GEM IPM algorithm and implementation in the GEMS3K code

2.1 Selection of the initial approximation

If the input chemical system definition contains a speciation which is compatible with the bulk chemical composition and phase assemblage, it can be directly used as FIA in the smart initial approximation mode. In this case, the GEM IPM algorithm usually makes approximately 10 to 20 times less iterations, which may be highly beneficial for the overall performance of large coupled reactive transport calculations, even on parallel architectures. However, there is no guarantee that the previously computed speciation, even for the same node, is in the feasible domain. This situation is usually detected with MBR or PSSC procedures, and the GEM IPM algorithm then automatically switches to the AIA mode. This detection mechanism is not absolutely efficient, particularly in systems with more than one highly nonideal solution phases, and more numerical research is required in this area.

An automatic (simplex) FIA must be used in all cases when the stable phase assemblage or the redox state is different from the previously computed equilibrium state or if the latter is not available at all. Otherwise, the old solution vectors $n^{(y)} = n^{(x),\text{old}}$ and $\gamma = \gamma^{(\text{old})}$ with the new bulk composition $n^{(b)} = n^{(b),\text{new}}$ are simply forwarded to the MBR procedure. In the AIA mode, the generic nonlinear GEM problem is first truncated into a LP subproblem by cutting off the concentration and activity coefficient terms in chemical potential expressions (Eq. 4) for all dependent components:

$$\min \qquad \sum_{j \in L_s} c_j n^{(y)}_{j(s)}$$
$$\text{subject to} \quad A n^{(y)}_{(s)} = n^{(b)}, \; n^{(y)}_{(s)} \geq 0 \tag{12}$$

where the constant c_j is defined as

$$c_j = \begin{cases} \dfrac{g_j}{RT} + \ln \lambda_j, & j \in L \setminus (L_g \cup (L_{aq} \setminus j_w)) \\[2mm] \dfrac{g_j}{RT} + \ln \lambda_j + \ln P, & j \in L_g \\[2mm] \dfrac{g_j}{RT} + \ln \lambda_j + \ln 55.5083734, & j \in L_{aq} \setminus j_w. \end{cases} \tag{13}$$

The λ_j is the optional constant controlling the stability of j-th dependent component at LP initial approximation, when each dependent component (species) is treated as a pure phase. The LP subproblem (Eq. 12) is then solved using a modified simplex algorithm, which yields very small relative mass balance residuals ($\varsigma < 10^{-12}$). The GEMS3K calculation is completed at this point, if no solution phases (having two or more dependent components) are present in the chemical system definition. Otherwise, the original nonlinear minimization problem (Eqs. 1a, 1b to 3) is restored. Initial values of $n^{(y)}_j$ are set equal to simplex-calculated amounts $n^{(y)}_{j(s)}$ if $n^{(y)}_{j(s)} < \varepsilon_f$, or to a small value ($\varepsilon_f = 10^{-5}$ mol) if $n^{(y)}_{j(s)} < \varepsilon_f$. This initial approximation fill-out procedure ensures that thermodynamically stable phases and species will not be eliminated from the final GEM result.

2.2 The MBR procedure

In order to obtain a FIA before starting the main IPM descent loop, the mass balance must be substantially improved. This is done within the MassBalanceRefinement() procedure that iteratively adjusts some $n^{(y)}_j$ values in such a way that any mass balance residual

$$\varsigma^{(r)}_i = n^{(b)}_i - \sum_j a_{ij} n^{(y,r)}_j, \quad i \in N, \; j \in L \tag{14}$$

does not exceed a small prescribed tolerance value $\varepsilon_{\mathrm{b},i}$, and the approximation remains in the feasible domain $M1$ (see Eqs. 1a and 1b). Here and below, (r) stands for the iteration index. In the original IPM algorithm [24, 25], these threshold values ε_b for balance residuals ς_i were taken to be the same for all independent components, in the range of $10^{-8} \leq \varepsilon_{\mathrm{b}} \leq 10^{-6}$ mol. Such accuracy was not sufficient for environmental and (radioactive) waste disposal applications, ore metal solubility calculations, and reactive transport simulations, where total amounts of trace elements (radionuclides, precious ore metals) can be 10^{-9} mol or less per 1 kg of water solvent. Development work was directed towards substantial improvement of the mass balance precision particularly for trace elements, resulting in a modified MBR procedure that satisfies the strictest requirements. MBR iterations (usually a few) are completed when

$$\frac{\varsigma_i^{(r)}}{n_i^{(b)}} < \varepsilon_{\mathrm{b}} \quad \forall \, i \in N \backslash i_{\mathrm{e}} \tag{15}$$

where i_{e} is the index of the electrical charge IC, and $10^{-15} < \varepsilon_{\mathrm{b}} < 10^{-12}$ is the empirical threshold. The new relative scale conditions (Eq. 15) result in excellent mass balance precision for both major and trace independent components. It is different from that in previous versions of GEM IPM code, where an absolute condition $\max_i \left| \varsigma_i^{(r)} \right| \leq \varepsilon_3$ (in moles) has been used, resulting in good accuracy for major elements, but frequently in an unsatisfactory one for trace elements.

In GEMS3K, combined criteria of a relative mass balance threshold for minor/trace (Eq. 15) and an absolute threshold $\varepsilon_{\mathrm{abs}}$ for major independent components (for instance, H and O in aquatic systems) are implemented. The latter can either be equal to the relative value ε_{b} (this is the default) or set separately as $10^{-|\varepsilon_{\mathrm{b,exp}}|}$ using a prescribed decimal exponent $\varepsilon_{\mathrm{b,exp}}$. This combination of both relative and absolute mass balance precision may be beneficial for chemical systems with several nonideal solution phases that contain elements present at both major and trace concentration levels.

The refined vector $n^{(y)}$ is copied into the start primal approximation vector $n^{(x,0)} \in M1$ of the main GEM IPM descent algorithm and also used to compute first approximations of activity coefficients in nonideal phases $\gamma^{(0)}$, as well as those of primal chemical potentials $\upsilon^{(0)}$ (Eq. 4).

2.3 Main IPM descent algorithm

The MainIPMDescent() procedure iteratively adjusts both the primal $n^{(x,y)}$ and dual $u^{(b,y)}$ solution vectors while decreasing the total Gibbs energy function of the system $G(n^{(x)})$. At the rth iteration, first the descent-direction vector $\Delta^{(r)}$ is obtained, then the optimal step length λ_r is estimated, and finally the primal solution approximation is corrected as

$$n^{(x,r+1)} = n^{(x,r)} + \lambda_r \Delta^{(r)}. \tag{16}$$

The direction-of-descent vector $\Delta^{(r)}$ is found by solving a dual nonlinear minimization subproblem

$$\min \sum_{j \in L} \upsilon_j^{(r)} \Delta_j, \text{ s.t. } A\Delta = 0, \ \sum_{j \in L_{\mathrm{S}}} \frac{\Delta_j^2}{q_j^{(r)}} \leq 1 \tag{17}$$

where the constraint $A\Delta = 0$ preserves the mass balance, the ellipsoid $\sum_{j \in L_{\mathrm{S}}} \frac{\Delta_j^2}{q_j^{(r)}} \leq 1$ replaces all nontrivial upper and lower metastability constraints by a single additional constraint, and the corresponding weight multipliers are given as

$$q_j^{(r)} = \begin{cases} n_j^{(x,r)}, & j \in D_0, \\ n_j^{(x,r)} - \underline{n}_j, & j \in D_1, \\ \overline{n}_j - n_j^{(x,r)}, & j \in D_2, \\ \min \left[\left(n_j^{(x,r)} - \underline{n}_j \right), \ \left(\overline{n}_j - n_j^{(x,r)} \right) \right], & j \in D_3. \end{cases} \tag{18}$$

In solving the subproblem (Eq. 17), as well as a similar subproblem in the MBR procedure, the internal linearization procedure is performed, in which the system of linear equations is solved for the Lagrange multipliers $\upsilon_i^{(r)}$ using the Cholesky method. In GEMS3K, the Jama TNT C++ library [40] is used for this purpose.

After calculation of the descent-direction $\Delta^{(r)}$ vector (see Eq. 17), the optimal step length λ_r is found by solving a unidimensional minimization subproblem

$$\lambda_r = \underset{0 \leq \lambda \leq \mu_r}{\arg \min} \left[\sum_{j \in L} \left(n^{(x,r)} + \lambda_r \right) \upsilon_j^{(r)} \right] \tag{19}$$

where the μ_r value is determined from the condition that the point $n^{(x,r)} + \mu_r \Delta^{(r)}$ lies at the boundary of the feasible domain $M1$ [9]. The new primal solution vector $n^{(x,y+1)}$ is then obtained from Eq. 16 and further used for the correction of activity coefficients in solution phases $\gamma_j^{(r+1)}$ to be applied at the next IPM iteration for computing a new approximation of primal chem-

ical potentials $\upsilon_j^{(r+1)}$. Activity coefficients and related configurational entropy terms for various solution and fluid phase models of mixing are computed as described in [59].

The convergence of IPM iteration process is checked using a modified Dikin's criterion C_D, which is defined as

$$C_D = \sum_{j \in L_S} \left| q_j^{(r)} \left(\sum_{i \in N} a_{ij} u_i^{(r)} - \upsilon_j^{(r)} \right) \right|. \tag{20}$$

Iterations are completed when $C_D \leq \varepsilon_D$, where $10^{-7} \leq \varepsilon_D \leq 10^{-3}$ mol is the empirical threshold. For a system having a total amount of 1,000 mol of independent components, the C_D is typically set to about $3 \cdot 10^{-6}$. The number of performed iterations strongly depends on the system complexity (total number of components, presence of strongly nonideal solution phases) and on the mode of initial approximation. In the AIA mode, typically 15 to 1,000 iterations are done, in the SIA mode, about 1 to 100 iterations are required (usually only a few). After the final recalculation of activity coefficients, the main IPM descent run is completed, and the primal and dual solutions are subsequently checked for their consistency.

2.4 PhaseSelection and SpeciationCleanup algorithms

Consistency checks are performed within the PSSC procedure. In GEMS3K, the completely revised PSSC replaces the Select-2 procedure used in older versions of GEM-Selektor and Selektor codes [9, 25]. The latter is, however, retained as an option that might be useful in complex systems with many nonideal solution phases. In such systems, a subtle interplay between small differences in Gibbs energy between different solution phases of similar bulk composition (particularly Fe-Mg silicates) and slight distortion of composition–activity relationships arising from activity coefficient

smoothing procedures (see below) might result in numerical convergence problems that are in some cases better resolved by the old Select-2 procedure. The main reason for introducing the new PSSC procedure was to improve the overall robustness and precision of the primal solution $\widehat{n}^{(x)}$, sometimes found to be insufficient for coupled reactive mass transport calculations. The PSSC procedure has two essential steps, which are PhaseSelection and SpeciationCleanup.

The PhaseSelection run begins with the StabilityIndexes() procedure, where the stability indexes Λ_k, Ω_k, $k \in \Phi$ for all phases in the system are computed. They are defined as

$$\Lambda_k = \log_{10} \Omega_k; \; \Omega_k = \sum_{j \in l_k} \varpi_j \tag{21}$$

where

$$\varpi_j = \frac{\omega_j^{(\widehat{u})}}{\gamma_j^{(\widehat{n})}} \tag{22}$$

is the estimate of mole fraction of the jth component in the kth phase, $\omega_j^{(\widehat{u})}$ is the activity calculated from the dual solution $\widehat{u}$, and $\gamma_j^{(\widehat{n})}$ is the activity coefficient calculated from the primal solution of the GEM IPM minimization problem. Both $\omega_j^{(\widehat{u})}$ and $\gamma_j^{(\widehat{n})}$ are taken in the mole fraction scale and calculated depending on the phase type [59]. The $\omega_j^{(\widehat{u})}$ values are obtained, as derived from Eqs. 4 to 6a, 6b. After calculation of stability indices for all phases, including those eliminated in the IPM run, the primal solution is checked phase by phase for consistency against the dual solution according to phase selection rules. When the loop over all phases is completed, the PSSC procedure proceeds with the SpeciationCleanup algorithm, which is performed in another loop over all phases whose components belong to the L_S set, as obtained from PhaseSelection. After all components of the phase have been checked and corrected, the PSSC procedure analyzes the results and

Table 6 Cases resulting after the PSSC procedure

A	The best case when no phases have been inserted and no mass balance violations occurred upon phase elimination or speciation cleanup. It is then sufficient to run the final MBR step, recalculate activity coefficients, and release the primal $\widehat{n}^{(x,c)}$ and dual $\widehat{u}$ solutions that are highly accurate and consistent with each other
B	If the mass balance violation flag has been activated at least for one species or phase, $n^{(y)} = \widehat{n}^{(x,c)}$ is assigned, and the full main GEM IPM sequence (Fig. 4) MBR(It), IPM Descent(It), and PSSC(It) is repeated after incrementing the global loop counter It. To prevent an infinite loop, maximum six iterations (It < 6) are allowed in GEMS3K. If inconsistencies persist after the sixth attempt in the PSSC procedure, the changes are not accepted, and the IPM primal solution $\widehat{n}^{(x)}$ is returned, marked as incompletely consistent with the primal solution

identifies subsequent actions of the GEM IPM solver (Table 6).

The overall behavior of the GEM IPM algorithm strongly depends on the mode of the initial approximation. In principle, error conditions can be generated at each numerical step shown in Fig. 4. In the SIA mode, any such error leads to a complete restart of GEM IPM calculation using the AIA mode. In the AIA mode, two error conditions do not prevent yielding solutions that are either incompletely consistent (after six PSSC loops have been performed) or insufficiently accurate (when MBR cannot converge at prescribed maximum number of iterations). In these two cases, the respective warning codes are generated. All other errors result in returning an error code but yield no valid primal and dual solution.

2.5 Comparison of new PSSC and old Selekt-2 procedures

The phase stability index Ω_k (Eqs. 21 and 22) is a generalization of the saturation index, broadly used in LMA speciation calculations and in simulation of mineral dissolution and precipitation kinetics [4, 43, 44]. Because Ω_k is directly related to the KKT conditions (Eqs. 7c, 11a, 11b and 11c), it can be compared with the Karpov's phase stability index f_α [23, 24] used in the old Selekt-2 procedure, which in our notation is expressed as

$$f_{\alpha,k} = \sum_{j \in L_k} \left[\exp \left(\eta_j - \upsilon_j + \ln \widehat{x}_j \right) \right] - 1, \quad j \in L_S \cap j \in D_0$$

(23)

where $\widehat{x}_j$ is the mole fraction of the jth component in the kth phase. This definition can be compared with the stability index from Eq. 22:

$$f_{\alpha,k} = \Omega_k - 1 \text{ or } \Lambda_k = \log_{10} \left(f_{\alpha,k} + 1 \right).$$

(24)

There is also a numerical difference in the application of phase selection rules using empirical thresholds. In the Select-2 procedure, $\varepsilon_{\Phi-} = \varepsilon_{\Phi+} = \varepsilon_\Phi$ and the phase is considered stable if $-\varepsilon_\Phi \leq f_{\alpha,k} \leq \varepsilon_\Phi$ [8]. In the PSSC procedure, the phase is considered stable if $-\varepsilon_{\Phi-} < \Lambda_k < \varepsilon_{\Phi+}$ or $-\varepsilon_{\Phi-} < \log_{10} \left(f_{\alpha,k} + 1 \right) < \varepsilon_{\Phi+}$ (Eq. 24). This comparison shows that, for the same ε_Φ threshold values, the linear-scale criterion f_α in Selekt-2 is more sensitive to dependent components present in large mole fractions $\widehat{x}_j$ in the phase and less sensitive to those present in small mole fractions. Conversely, the logarithmic-scale criterion used in PSSC is sensitive to primal activity coefficients and mainly to the dual

chemical potentials, but less to the mole fractions of dependent components. In other words, the index Λ_k is to a less extent based on the primal solution $\widehat{n}^{(x)}$ and uses the dual solution $\widehat{u}$ of the GEM problem to greater extent than the old index $f_{\alpha,k}$.

This notion explains why the introduction of the new phase stability index Eq. 21 and the new PSSC algorithm, together with the modified Dikin's criterion of IPM convergence (Eq. 20), has dramatically improved the robustness and internal consistency of the GEMS3K code, compared to previous versions of GEMIPM2K and Selektor codes. In turn, this success suggests a hypothesis that in the iterative numerical process of the IPM algorithm, the error accumulation mainly occurs in the primal solution approximation, rather than in the dual one. This idea forms the basis of the SpeciationCleanup algorithm. There are, however, infrequent numerically ill-behaved cases, when the dual solution approximation becomes divergent, and dramatic error accumulation or failure in MBR or IPM may occur.

2.6 Calculation of nonideal mixing models for solution phases

The substantial advantage of direct GEM algorithms lies in their ability to potentially handle any number of nonideal solution phases simultaneously in the same chemical thermodynamic system, without any a priori information on which phases are stable and which not. With the current GEM algorithm, it is possible to model very complex nonideal solid-solution aqueous-solution systems (e.g., [15, 36]). The most complex system that we have successfully tested with GEM-Selektor v.3 is composed of 13 independent components (12 chemical elements plus charge), one aqueous solution with 83 species, 44 nonideal mineral solid-solution phases with up to nine end members, and additionally 48 pure solid phases.

The GEMS3K code can compute models of nonideal mixing in two ways, either (1) by interpreting phase model scripts (prepared in phase records stored in the GEM-Selektor project data base) or (2) by executing built-in models of mixing, which are implemented on the basis of the new TSolMod C++ class interface. The stand-alone GEMS3K code can execute any built-in model, but usage of phase model scripts is not supported. The TSolMod class library with its numerous models of mixing, as provided in GEM-Selektor and GEMS3K codes, is described in detail in [59]. Regardless of the type of model (built-in or scripted), the execution occurs within the CalculateAc-

Table 7 Modes of CalculateActivityCoefficients (*LinkMode*) operation

LinkMode	Description	Where called from
TP	Calculations of model parameters that depend on T,P but not on the composition of phases	Start of the GEM IPM algorithm[a]
UX	Calculation of activity coefficients $\gamma_j^{(r)}$ of dependent components within MBR and MainIPMDescent()	After fill-out initial approximation procedure (AIA) or in SIA before MBR(0), in MBR(It) before solving SLE, at the end of each MainIPMDescent(It) iteration[a]
PP[b]	Calculation of integral non-ideal properties of phases after obtaining a valid GEM IPM primal solution	After MainIPMDescent(It) has converged with valid primal and dual solutions[a]

[a] In GEM-Selector, also when reading SysEq records for displaying or sampling results
[b] The PP (phase properties) mode will be operational in GEMS3K v.3.2 and higher

tivityCoefficients() procedure, which has three modes that are summarized in Table 7.

In some highly nonideal phases (for instance, reciprocal sublattice solid solutions), the activities of end members can so strongly change with phase composition that this may disturb the smooth convergence of MBR and main IPM descent procedures. Because of the strong coupling between the chemical potentials of end members of different solid solution phases in such systems, this may result in an oscillating behavior of the GEM IPM algorithm that prevents the descent to the stable minimum Gibbs energy state. In such cases, the application of smoothing procedures may be necessary, although this may slow down the convergence of IPM and the performance of the GEMS3K solver.

2.7 Smoothing in systems with highly nonideal phases

Through extensive numerical tests using systems with highly nonideal phases where the main IPM descent procedure did not properly converge, we found that an efficient procedure for smoothing chemical potential oscillations would depend on the ratio of the actual value of the Dikin's criterion C_D (Eq. 20) and the prescribed IPM convergence threshold ε_D. The newly implemented formula for calculation of the smoothing factor

$$\alpha_\gamma = \alpha_{CD} = \varepsilon_\delta + \exp\left[\ln\left(1 - \varepsilon_\alpha\right) + \ln C_D\right] \quad (25)$$
$$+ \frac{\varepsilon_\alpha - \varepsilon_\delta}{1 + \exp\left(\ln \varepsilon_D - \ln C_D\right)/\varepsilon_\delta}$$

uses three empirical input constants, which are ε_D, $0 < \varepsilon_\alpha < 1$, and $0 < \varepsilon_\delta < 1$. The constant ε_δ sets the minimal value of α_γ that it approaches when C_D becomes very small, and the constant ε_α sets the position of the α_γ plateau at first IPM iterations when C_D values are still high and rather strong corrections of primal chemical potentials are necessary. The smoothing factor is

applied after recalculation of activity coefficients $\gamma_j^{(r+1)}$ subsequent to the rth IPM iteration as follows. The approximations of primal chemical potentials $v_j^{(r+1)}$ are computed from $n^{(x,r+1)}$ and $\gamma_j^{(r+1)}$, using Eq. 4. The deviations are then calculated as

$$\delta_j^{(v,r)} = v_j^{(r+1)} - v_j^{(r)}, \quad j \in L_S \quad (26)$$

and compared with the empirical smoothing sensitivity threshold $\varepsilon_{sm} = 1 \cdot 10^{-3}$:

$$\forall \left| \delta_j^{(v,r)} \right| \geq \varepsilon_{sm}: \quad v_j^{(r+1)} = v_j^{(r)} + \alpha_\gamma \delta_j^{(v,r)}, \quad j \in L_S. \quad (27)$$

As seen from Eq. 27, the smoothing suppresses only increments to chemical potentials of dependent components that are present in the primal solution approximations in nonzero amounts and only when these increments are sufficiently large. The value of ε_{sm} may dramatically affect the GEM IPM performance and the number of iterations. Because the smoothing procedure invariably results in a slight distortion of the activity–composition relationships in nonideal solution phases, it should only be used with great caution. Simple aquatic systems where no nonideal phases are present in addition to the aqueous phase will usually converge without smoothing.

2.8 The procedure for detection of divergence in the dual solution approximation

The revised GEM IPM algorithm and GEMS3K code include a new runtime uDD() procedure for detection of numerical stiffness. When active, the uDD() procedure is called after each IPM iteration r. It starts with comparing IPM dual solution $u_i^{(r)}$ values with generic absolute tolerances. The $u_i^{(r)}$ values are accepted if they fall within an interval $-600 \leq u_i^{(r)} \leq 400$ (for the electrical charge Z, a narrower interval $-50 \leq u_Z^{(r)} \leq$

100 is checked). Note that actual ranges in well-posed systems are not so broad and become much narrower with increasing temperature. For a detailed detection of stiff behavior, for each independent component, a moving average of its u_i value for three consecutive IPM iterations is calculated, and its variation relative to previous value at $(r-1)$ is compared with a prescribed tolerance. If a substantial shift is detected, the warning is issued, the entire calculation is restarted in AIA mode, and the result from IPM iteration $(r-2)$ is returned (the PSSC procedure is skipped in this case). In such a result, the primal and dual IPM solutions are not internally consistent, but cannot be further improved because of the absence of DCs in amounts sufficient to constrain a diverging ith IC chemical potential u_i. If this happens, the system bulk composition has to be redefined such that certain redox buffering for the ith element is provided.

References

1. Bale, C.W., Chartrand, P., Degterov, S.A., Eriksson, G., Hack, K., Ben Mahfoud, R., Melancon, J., Pelton, A.D., Petersen, S.: FactSage thermochemical software and databases. Calphad **26**, 189–228 (2002)
2. Bale, C.W., Belisle, E., Chartrand, P., Degterov, S.A., Eriksson, G., Hack, K., Jung, I.H., Kang, Y.B., Melancon, J., Pelton, A.D., Robelin, C., Petersen, S.: FactSage thermochemical software and databases—recent developments. Calphad **33**, 295–311 (2009)
3. Baumann, C., Gerya, T.V., Connolly, J.A.D.: Numerical modelling of spontaneous slab breakoff dynamics during continental collision. Geol. Soc. Lond. Spec. Publ. **332**, 99–114 (2010)
4. Bethke, C.M.: Geochemical and Biogeochemical Reaction Modeling. Cambridge University Press, New York (2008)
5. Borisov, M.V., Shvarov, Y.V.: Thermodynamics of Geochemical Processes. Moscow State University Publishers, Moscow (in Russian, 1992)
6. Bruno, J., Bosbach, D., Kulik, D., Navrotsky, A.: Chemical Thermodynamics of Solid Solutions of Interest in Radioactive Waste Management. A State-of-the-art Report. OECD NEA, Paris (2007)
7. Centler, F., Shao, H., De Biase, C., et al.: GeoSysBRNS—a flexible multidimensional reactive transport model for simulating biogeochemical subsurface processes. Comput. Geosci. **36**, 397–405 (2010)
8. Chudnenko, K.V.: Thermodynamic Modeling in Geochemistry: The Theory, the Algorithms, the Software, the Applications. Academic Publishing House GEO, Novosibirsk (in Russian, 2010)
9. Chudnenko, K.V., Karpov, I.K., Kulik, D.A.: A High-Precision IPM-2 Non-linear Minimization Module of GEM-Selektor v.2-PSI Program Code for geochemical Thermodynamic Modeling. Technical Report TM-44–02–06. Paul Scherrer Institut, Villigen (2002)
10. Connolly, J.A.D., Petrini, K.: An automated strategy for calculation of phase diagram sections and retrieval of rock properties as a function of physical conditions. J. Metamorph. Geol. **20**, 697–708 (2002)
11. Connolly, J.A.D.: Computation of phase equilibria by linear programming: a tool for geodynamic modeling and its application to subduction zone decarbonation. Earth Planet. Sci. Lett. **236**, 524–541 (2005)
12. Coumou, D., Matthai, S., Geiger, S., Driesner, T.: A parallel FE–FV scheme to solve fluid flow in complex geologic media. Comput. Geosci. **34**, 1697–1707 (2008)
13. de Capitani, C., Brown, T.H.: The computation of chemical equilibrium in complex systems containing non-ideal solutions. Geochim. Cosmochim. Acta **51**, 2639–2652 (1987)
14. de Capitani, C., Petrakakis, K.: The computation of equilibrium assemblage diagrams with Theriak/Domino software. Am. Mineral. **95**, 1006–1016 (2010)
15. Dolejs, D., Wagner, T.: Thermodynamic modeling of non-ideal mineral-fluid equilibria in the system Si-Al-Fe-Mg-Ca-Na-K-H-O-Cl at elevated temperatures and pressures: implications for hydrothermal mass transfer in granitic rocks. Geochim. Cosmochim. Acta **72**, 526–553 (2008)
16. Ebel, D.S., Ghiorso, M.S., Sack, R.O., Grossmann, L.: Gibbs energy minimization in gas + liquid + solid systems. J. Comput. Chem. **21**, 247–256 (2000)
17. Engesgaard, P., Kipp, K.L.: A geochemical transport model for redox-controlled movement of mineral fronts in groundwater-flow systems—a case of nitrate removal by oxidation of pyrite. Water Resour. Res. **28**, 2829–2843 (1992)
18. Eriksson, G., Hack, K.: Chemsage—a computer program for the calculation of complex chemical equilibria. Metall. Mater. Trans. B **21**, 1013–1023 (1990)
19. Eriksson, G., Thompson, W.T.: A procedure to estimate equilibrium concentrations in multicomponent systems and related applications. Calphad **13**, 389–400 (1989)
20. Ghiorso, M.S.: Algorithms for the estimation of phase stability in heterogeneous thermodynamic systems. Geochim. Cosmochim. Acta **58**, 5489–5501 (1994)
21. GNU free software foundation: GNU Lesser General Public License. Free Software Foundation, Boston. http://www.gnu.org/licenses/lgpl.html (2007)
22. Hammond, G., Lichtner, P., Lu, C.: Subsurface multiphase flow and multicomponent reactive transport modeling using high-performance computing. J. Phys. Conf. Ser. **78**(012025), 1–10 (2007)
23. Karpov, I.K.: Computer-Aided Physico-Chemical Modeling in Geochemistry. Nauka Publ., Novosibirsk (in Russian, 1981)
24. Karpov, I.K., Chudnenko, K.V., Kulik, D.A.: Modeling chemical mass-transfer in geochemical processes: thermodynamic relations, conditions of equilibria and numerical algorithms. Am. J. Sci. **297**, 767–806 (1997)
25. Karpov, I.K., Chudnenko, K.V., Kulik, D.A., Avchenko, O.V., Bychinskii, V.A.: Minimization of Gibbs free energy in geochemical systems by convex programming. Geochem. Int. **39**, 1108–1119 (2001)
26. Karpov, I.K., Chudnenko, K.V., Kulik, D.A., Bychinskii, V.A.: The convex programming minimization of five thermodynamic potentials other than Gibbs energy in geochemical modeling. Am. J. Sci. **302**, 281–311 (2002)
27. Keizer, M.G., Van Riemsdijk, W.H.: ECOSAT. Technical Report. Department Soil Science and Plant Nutrition, Wageningen Agricultural University, Wageningen, The Netherlands (1998)
28. Kelley, C.T.: Solving Nonlinear Equations with Newton's Method. SIAM, Philadelphia (2003)

29. Kolditz, O., Görke, U.-J., Shao, H., Wang, W. (eds.): Benchmarks and Examples for Thermo-Hydro-Mechanical/Chemical Processes in Porous Media. Series: Lecture Notes in Computational Science and Engineering, vol. 86. Springer, Berlin (2012)

30. Kolditz, O., Bauer, S., Bilke, L., Böttcher, N., Delfs, J.O., Fischer, T., Görke, U.J., Kalbacher, T., Kosakowski, G., McDermott, C.I., Park, C.H., Radu, F., Rink, K., Shao, H., Shao, H.B., Sun, F., Sun, Y.Y., Singh, A.K., Taron, J., Walther, M., Wang, W., Watanabe, N., Wu, Y., Xie, M., Xu, W., Zehner, B.: OpenGeoSys: an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C) processes in porous media. Environ. Earth Sci. (2012). doi:10.1007/s12665–012–1546-x

31. Kulik, D., Berner, U., Curti, E.: Modelling chemical equilibrium partitioning with the GEMS-PSI code. In: Smith, B., Gschwend, B. (eds.) PSI Scientific Report 2003/Volume IV, Nuclear Energy and Safety, pp. 109–122. Paul Scherrer Institut, Villigen (2004)

32. Kulik, D.A.: Classic adsorption isotherms incorporated in modern surface complexation models: implications for sorption of actinides. Radiochim. Acta **94**, 765–778 (2006)

33. Kulik, D.A.: Standard molar Gibbs energies and activity coefficients of surface complexes (thermodynamic insights). In: Luetzenkirchen, J. (ed.) Surface Complexation Modelling. Interface Science and Technology, vol. 11, pp. 171–250. Elsevier, Amsterdam (2006)

34. Lukas, H.L., Fries, S., Sundman, B.: Computational Thermodynamics: The Calphad Method. Cambridge University Press, Cambridge (2007)

35. McDermott, C.I., Walsh, R., Mettier, R., Kosakowski, G., Kolditz, O.: Hybrid analytical and finite element numerical modeling of mass and heat transport in fractured rocks with matrix diffusion. Comput. Geosci. **13**, 349–361 (2009)

36. Monecke, T., Kempe, U., Trinkler, M., Thomas, R., Dulski, P., Wagner, T.: Unusual rare earth element fractionation in a tin-bearing magmatic-hydrothermal system. Geology **39**, 295–298 (2011)

37. Nakagawa, T., Tackley, P.J., Deschamps, F., Connolly, J.A.D.: Incorporating self-consistently calculated mineral physics into thermochemical mantle convection simulations in a 3-D spherical shell and its influence on seismic anomalies in Earth's mantle. Geochem. Geophys. Geosyst. (G^3) **10**, Q03004 (2009)

38. Parkhurst, D.L., Appelo, C.A.J.: User's guide to PHREEQC (version 2)—a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. U.S.G.S. Water-Resources Investigations Report 99–4259, Denver, Colorado (1999)

39. Pfingsten, W.: Efficient modeling of reactive transport phenomena by a multispecies random walk coupled to chemical equilibrium. Nucl. Technol. **116**, 208–221 (1996)

40. Pozo, R.: Template numerical toolkit: an interface for scientific computing in C++. National Institute of Standards and Trechnology (NIST), Gaithersburg, MD. http://math.nist.gov/tnt (2004)

41. Prommer, H.: A Reactive Multicomponent Transport Model for Saturated Porous Media. User's Manual Version 1.0. Contaminated Land Assessment and Remediation Research Centre. The University of Edinburgh, UK (2002)

42. Rastetter, E.B.: Modeling coupled biogeochemical cycles. Front. Ecol. Environ. **9**, 68–73 (2011)

43. Reed, M.H.: Calculation of multicomponent chemical equilibria and reaction processes in systems involving minerals, gases and an aqueous phase. Geochim. Cosmochim. Acta **46**, 513–528 (1982)

44. Reed, M.H.: Calculation of simultaneous chemical equilibria in aqueous-mineral-gas systems and its application to modeling hydrothermal processes. Rev. Econ. Geol. **10**, 109–124 (1998)

45. Reed, M.H., Spycher, N.F.: Calculation of pH and mineral equilibria in hydrothermal waters with applications to geothermometry and studies of boiling and dilution. Geochim. Cosmochim. Acta **48**, 1479–1492 (1984)

46. Shao, H., Dmytrieva, S.V., Kolditz, O., Kulik, D.A., Pfingsten, W., Kosakowski, G.: Modeling reactive transport in non-ideal aqueous-solid solution system. Appl. Geochem. **24**, 1287–1300 (2009)

47. Shao, H., Kulik D.A., Berner U., Kosakowski G., Kolditz O.: Modeling the competition between solid solution formation and cation exchange on the retardation of aqueous radium in an idealized bentonite column. Geochem. J. **43**, e37–e42 (2009)

48. Shvarov, Y.V.: A general equilibrium criterion for an isobaric-isothermal model of a chemical system. Geochem. Int. **18**, 38–45 (1981)

49. Shvarov, Y.: A numerical criterion for existence of the equilibrium state in an open chemical system. Sci. Geol. Bull. **42**, 365–369 (1989)

50. Shvarov, Y.V.: HCh: New potentialities for the thermodynamic simulation of geochemical systems offered by windows. Geochem. Int. **46**, 834–839 (2008)

51. Singh, A.K., Goerke, U.-J., Kolditz, O.: Numerical simulation of non-isothermal compositional gas flow: application to carbon dioxide injection into gas reservoirs. Energy **36**, 3446–3458 (2011)

52. Siret, D., Poulet, T., Regenauer-Lieb, K., Connolly, J.A.D.: PreMDB, a thermodynamically consistent material database as a key to geodynamic modelling. Acta Geotech. **4**, 107–115 (2009)

53. Steefel, C.I., DePaolo, D.J., Lichtner, P.C.: Reactive transport modeling: an essential tool and a new research approach for the Earth sciences. Earth Planet. Sci. Lett. **240**, 539–558 (2005)

54. Tenzer, H., Park, C.L., Kolditz, O., McDermott, C.I.: Application of the geomechanical facies approach and comparison of exploration and evaluation methods used at Soultz-sous-Forets (France) and Spa Urach (Germany) geothermal sites. Environ. Earth Sci. **61**, 853–880 (2010)

55. Trefethen, L.N., Bau, D.: Numerical Linear Algebra. SIAM, Philadelphia (1997)

56. Tromp, T.K., Van Cappellen, P., Key R.M.: A global model for the early diagenesis of organic carbon and organic phosphorous in marine sediments. Geochim. Cosmochim. Acta **59**, 1259–1284 (1995)

57. Van der Lee, J., De Windt, L.: Present state and future directions of modeling of geochemistry in hydrogeological systems. J. Contam. Hydrol. **47**, 265–282 (2001)

58. Van der Lee, J., De Windt, L., Lagneau, V., Goblet, P.: Module-oriented modeling of reactive transport with HYTEC. Comput. Geosci. **29**, 265–275 (2003)

59. Wagner, T., Kulik, D.A., Hingerl, F.F., Dmytrieva, S.V.: GEM-Selektor geochemical modeling package: TSolMod library and data interface for multicomponent phase models. Canadian Mineralogist **50** (2012, in press)

60. Wang, W., Kosakowski, G., Kolditz, O.: A parallel finite element scheme for thermo-hydro-mechanical (THM) coupled

problems in porous media. Comput. Geosci. **35**, 1631–1641 (2009)

61. Wang, Y., Van Cappellen, P.: A multicomponent reactive transport model of early diagenesis: application to redox cycling in coastal marine sediments. Geochim. Cosmochim. Acta **60**, 2993–3014. (1996)

62. Westall, J.C., Zachary, J.L., Morel, F.M.M.: MINEQL: A Compact Program for Computation of Chemical Equilibria in Aquatic Systems. R.M. Parsons Laboratory for Water Resources and Hydrodynamics, Massachusetts Institute of Technology, Cambridge, MA (1976)

63. Xie, M., Kolditz, O., Moog, H.C.: A geochemical transport model for thermo-hydro-chemical (THC) coupled processes with saline water. Water Resour. Res. **47**, W02545 (2011)

64. Kosakowski, G., Kulik, D.A., Shao, H.: OpenGeoSys-GEMS: Hybrid parallelization of a reactive transport code with MPI and threads. Geophys. Res. Abstr. **14**, EGU2012–2642 (2012)