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

## Dmitrii A. Kulik<sup>1</sup>, Thomas Wagner<sup>2</sup>, Svitlana V. Dmytrieva<sup>3</sup>, Georg Kosakowski<sup>1</sup>, Ferdinand F. Hingerl<sup>1,2</sup>, Konstantin V. Chudnenko<sup>4</sup>, Urs R. Berner<sup>1</sup>

<sup>1</sup>Laboratory for Waste Management, Paul Scherrer Institut, 5232 Villigen PSI, Switzerland <sup>2</sup>Institute for Geochemistry and Petrology, ETH Zurich, Clausiusstrasse 25, CH-8092 Zurich, Switzerland

<sup>3</sup>Institute of Environmental Geochemistry, Palladin ave. 34a, 03680 Kyiv-142, Ukraine <sup>4</sup>Vinogradov Institute of Geochemistry SB RAS, 1a Favorski str., 664033 Irkutsk, Russia

Submission date: 23.11.2011

Revision date: 19.06.2012

Submitted to: Computational Geosciences

#### 2 Abstract

3 Reactive mass transport (RMT) simulation is a powerful numerical tool to advance our under-4 standing of complex geochemical processes and their feedbacks in relevant subsurface systems. 5 Thermodynamic equilibrium defines the baseline for solubility, chemical kinetics, and RMT in 6 general. Efficient RMT simulations can be based on the operator-splitting approach, where the 7 solver of chemical equilibria is called by the mass-transport part for each control volume whose 8 composition, temperature, or pressure has changed. Modeling of complex natural systems re-9 quires consideration of multiphase-multicomponent geochemical models that include nonideal 10 solutions (aqueous electrolytes, fluids, gases, solid solutions and melts). Direct Gibbs energy mi-11 nimization (GEM) methods have numerous advantages for the realistic geochemical modeling of 12 such fluid-rock systems. Substantial improvements and extensions to the revised GEM interior 13 point method (IPM) algorithm based on Karpov's convex programming approach are described, 14 as implemented in the GEMS3K C/C++ code, which is also the numerical kernel of GEM-15 Selektor v.3 package (http://gems.web.psi.ch). GEMS3K is presented in the context of the essen-16 tial criteria of chemical plausibility, robustness of results, mass-balance accuracy, numerical sta-17 bility, speed, and portability to high-performance computing systems. The standalone GEMS3K 18 code can treat very complex chemical systems with many nonideal solution phases accurately. It 19 is fast, delivering chemically plausible and accurate results with the same or better mass balance 20 precision as that of conventional speciation codes. GEMS3K is already used in several coupled 21 RMT codes (e.g. OpenGeoSys-GEMS) capable of high-performance computing.

22

Keywords: geochemical modeling, reactive mass transport, Gibbs energy minimization, non ideal systems, fluid-rock interaction

#### 26 **1. Introduction**

27 Numerical geochemical and coupled geochemical-physical modeling is important for under-28 standing the complexity and feedback behavior of natural geosystems, in particular, when com-29 bined with field observations and analytical data. The specific advantage of geochemical model-30 ing of complex multicomponent-multiphase-multiprocess systems lies in the possibility to sys-31 tematically explore the parameter space, to identify the feedback between key parameters, and to 32 investigate geosystems at pressure-temperature conditions and timescales that are not accessible 33 to direct observation or laboratory experiments. The dramatic increase in computer speed during 34 the last decade has facilitated construction of ever more complex models of geochemical and 35 geophysical processes, which are now closer to the complexity of real natural systems than ever 36 before. Examples include coupled geochemical-geophysical models of geodynamic processes [3, 37 37, 52], coupled models for biogeochemical cycles [42, 56, 61], and reactive transport models of 38 the shallow to deep subsurface [22, 53].

39 Simulation of the spatial-temporal-chemical evolution of natural subsurface systems 40 needs to account for transport processes (diffusion, advection) coupled with partial chemical 41 equilibria and their feedback on the pore space and transport properties. Many coupled reactive 42 transport codes are based on the operator-splitting approach (e.g. [57]), in which the chemical 43 solver program is called by the mass-transport part for each control volume every time when the 44 composition, temperature, or pressure of the volume has changed. In realistic simulations, up to a 45 million of nodes and time steps have to be specified for a reasonable 3D spatial resolution and 46 time discretization. Thus, chemical plausibility, robustness of results, mass-balance accuracy, 47 numerical stability, speed, and portability become crucial properties of the chemical solver. 48 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

51 pure solids, but also the amounts and composition of many co-existing non-ideal multi-52 component solid solutions, sorption, fluid, and melt phases subject to optional metastability re-53 strictions, along with saturation/stability indexes for all phases. Unlike Gibbs energy mimimiza-54 tion (GEM) codes such as ChemApp [2, 18] or HCh [50], law of mass action (LMA) codes such 55 as GWB [4] or PHREEQC [38] seem to partially fail against this criterion, in particular, in chem-56 ical systems relevant to hydrothermal geochemistry and petrology.

57 *Robustness of results.* The chemical solver must compute the correct equilibrium specia-58 tion and activities of components in phases for any input composition of the system, even if the 59 latter is not well buffered. This principal task is not trivial, and it is related to the theoretical exis-60 tence of single- or multiple minima of the total Gibbs energy function, to the degree of non-61 ideality of mixing, and to possible existence of phase miscibility gaps.

62 *Mass-balance precision*. Most mass transport codes typically guarantee high precision, 63 but may also cause numerical oscillations into the chemical composition of control volumes. The 64 chemical solver must keep the mass balance residuals even smaller than the mass transport code 65 (typically 10<sup>-13</sup> relative), in order to avoid the accumulation of errors in simulation over time.

Numerical performance. In a typical coupled RMT simulation of e.g. 10<sup>6</sup> nodes over 10<sup>6</sup> 66 time steps, the chemical solver will be called at least  $10^{12}$  times, which poses a very high demand 67 68 to its numerical stability, calculation speed, and portability to high-performance computers 69 (HPC). In this regard, existing LMA and GEM codes appear to behave very differently [2, 25, 70 27, 31, 34, 38, 50, 58]. From our theoretical estimates in section 4 below, it follows that, for sim-71 ilar problems, GEM codes are ca. 2 to 10 times slower than LMA codes. The reason is that the 72 GEM algorithm finds unknown phase assemblage by minimizing total Gibbs energy while main-73 taining mass balance, whereas the LMA algorithm directly minimizes the mass balance residuals, 74 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.

82 The purpose of this paper is to describe the revised GEM interior point method (IPM) al-83 gorithm based on the convex programming method [8, 24, 25], with substantial improvements 84 and extensions implemented recently in the GEMS3K code for using it within coupled reactive 85 mass transport (RMT) codes. We provide arguments for the central role of chemical thermody-86 namic calculations in RMT, and why using this particular algorithm and code would be benefi-87 cial for coupling with existing transport codes such as OpenGeoSys (OGS) [29, 30, 46, 60]. Fi-88 nally, we provide an example related to RMT simulations based on an already implemented 89 coupling in OGS-GEMS code. On the practical side, this paper should provide guidance, suffi-90 cient understanding, and necessary detail for those who wish to couple GEMS3K with their own 91 transport code efficiently.

92

#### 93 2. Advantages of the GEM method

Solving the phase equilibrium problem for a complex heterogeneous chemical system with multiple non-ideal 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 non-aqueous 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

100 only, then using it for calculation of saturation indexes of all stoichiometrically feasible solid 101 phases [43]. The most over-saturated phase is added to the mass balance, and the equilibrium 102 speciation is computed again, updating also saturation indices of all solid phases. If another over-103 saturated phase is detected then it is inserted to mass balance; if an under-saturated phase is 104 found in the mass balance then it is removed. The whole cycle is repeated until no oversaturated 105 phase can be detected. It may not be possible to calculate accurately the saturation index of a 106 non-ideal solution phase, which makes the entire iterative selection procedure unrigorous and 107 unreliable. A related problem is that calculations are started from aqueous speciation in chemical 108 systems with potentially high solid/water ratio, which would lead to unrealistically high dis-109 solved concentrations, strongly biased activity coefficients, and unrealistic saturation indexes for 110 some solids. Hence, the LMA method can be used with confidence and success only in water-111 dominated aqueous systems, where the stable mineral phase assemblage is a priori known.

112 Compared to chemical speciation models based on the LMA method such as 113 PHREEQC [38], MINEQL [62], CHILLER [43, 44, 45] or GWB [4], the GEM method has the 114 advantage that it can handle in one run any number of potentially stable phases, including many 115 (highly) non-ideal solutions. In addition, GEM calculations can directly account for processes of 116 phase immiscibility such as liquid-vapor equilibria and solid-solid exsolution that are very im-117 portant for modeling natural fluid-rock systems (e.g. [15]). Volumes of all phases in the reactive 118 system can be found when the standard molar volumes for their components and volumes of 119 mixing are provided. The disadvantage of GEM is a slower convergence compared to that of 120 LMA. The applicability of GEM algorithms is limited only by the availability of consistent stan-121 dard-state molar properties of end-members and non-ideal interaction parameters of mixing in 122 each multi-component phase (all corrected to temperature T and pressure P of interest, if neces-123 sary).

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, 49, 50], Selektor-C [8, 25], Perplex [10, 11], Theriak-Domino [13, 14], MELTS [16, 20], and GEM-Selektor ([31]; Kulik et al., 2012 http://gems.web.psi.ch).

129 The convex programming GEM method [23, 24, 25] implemented in GEMS3K computes 130 simultaneously the chemical speciation in all stable phases and the chemical potentials of chemi-131 cal elements and charge. From the chemical potentials, activities and concentrations of depen-132 dent components, saturation indices and stability criteria for phases, as well as pH and Eh in 133 aqueous solution are obtained. If non-ideal solution phases are included then the activity coeffi-134 cients of their components are adjusted on GEM iterations and are calculated in conjunction with 135 the primal solution. A specific feature of this approach consists in the operator splitting between 136 GEM and mass balance refinement (MBR) steps.

137 One of the main goals of our work was to improve the suitability of GEM algorithm for 138 reactive mass transport simulations. RMT simulation of the spatial-temporal evolution of a sub-139 surface geochemical system needs to account simultaneously for fluid/tracer transport and for 140 chemical reactions in/between phases. There are several levels of complexity of such models 141 (Fig. 1). Existing coupled code packages (e.g. OpenGeoSys [30, 60]) use the operator splitting 142 approach, in which a chemical equilibrium solver is called for each volume after every time step 143 of mass transport (Fig. 2). The chemical solver code receives the modified bulk chemical compo-144 sition, temperature, and pressure of the reactive part of the volume, and returns the updated (me-145 ta)stable phase speciation (perhaps with changed volumes of some solid phases affecting the po-146 rosity), which is then used for performing the next time step of mass transport. Taking into ac-147 count potentially large numbers of control volumes and time steps, and high precision of mass 148 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

151 to high performance computers.

152 The main motivation for the development of the standalone GEMS3K code was to pro-153 vide a fast, robust and efficient solver of complex geochemical equilibria with a simple data ex-154 change interface that would facilitate the implementation of various operator-splitting coupled 155 modeling codes. To date, GEMS3K has been coupled with the OpenGeoSys [30, 46, 60], 156 CSMP++ [12], and MCOTAC [39] fluid mass transport simulation codes. GEMS3K can be po-157 tentially used in generic parameter fitting algorithms, calculation and plotting of phase diagrams, 158 geodynamic modeling, model sensitivity analysis, and other coupled code applications. The 159 second aim of the development of GEMS3K was to prepare it to be released with the source code 160 in order to encourage future development of coupled codes suitable for parallel HPC platforms. 161 The third aim was to maintain a full compatibility between the GEM-Selektor package and any 162 coupled code by sharing the same GEMS3K kernel numerical code. This would allow rigorous 163 benchmarking and verification of different coupled codes using the same chemical system defini-164 tion and the same reactive mass transport model setup.

165

#### 166 **3. GEMS3K code and its place in the GEM-Selektor framework**

The overall structure of GEM-Selektor and standalone GEMS3K codes (all written in C/C++) and relationships between them are shown in Figure 3. The GEMS3K code consists of the GEM IPM (interior points method) 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 user interface and databases. In a coupled code, the upper-level thermo-hydraulicmass-transport algorithm (THMA) code can communicate with one or more standalone 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
non-ideal phases (described in a companion paper [59]) provides a connection between
GEMS3K and many relevant application fields in Earth sciences and chemical engineering.

179

#### 180 *3.1. GEM convex programming approach*

181 The convex programming method [8, 23, 24, 25] simultaneously computes the primal and dual 182 solution of the GEM chemical thermodynamic problem. The primal solution  $\hat{n}^{(x)}$ , usually called 183 speciation, consists of mole amounts of dependent components (chemical species in phases) and 184 their corresponding activity coefficients in equilibrium state. The dual solution  $\hat{u}$  consists of 185 chemical potentials of independent components (usually chemical elements and electric charge) 186 for the same equilibrium state. Both  $\hat{n}^{(x)}$  and  $\hat{u}$  vectors are consistent with the same scalar value 187 of the minimal total Gibbs energy of the system. The dual solution, which cannot be computed in 188 LMA codes, is particularly useful because from the  $\hat{u}$  vector, any activity-based thermodynamic quantities (activities, fugacities, pH, pe) of dependent components, the saturation indexes of 189 190 phases and criteria of phase stability can be readily obtained. In GEM codes, if phases with non-191 ideal mixing are included, the activity coefficients of their dependent components (end-192 members) are calculated together with the primal solution at every GEM iteration. The GEMS3K 193 algorithm can handle heterogeneous systems with any number of potentially stable phases, in-194 cluding several (highly) non-ideal solutions, and enhanced system complexity slows down the 195 calculations only moderately.

196 The GEM problem consists in calculation of the (unknown) equilibrium speciation and 197 phase assemblage in the system defined by pressure *P*, temperature *T*, bulk composition  $n^{(b)}$  (ex-198 pressed in absolute mole amounts of independent components), the standard state thermodynam-

- 199 ic data for dependent components, and, optionally, parameters of models of non-ideal mixing in
- 200 multi-component phases. Solving this problem means finding mole amounts of dependent com-
- 201 ponents  $n^{(x)} = \{n_j^{(x)}, j \in L\}$  with indexes belonging to the set L (vector x in [24]), such that (see
- 202 full list of symbols used in Annex A)

203 
$$G(n^{(x)}) \Rightarrow \min \left\{ G(n^{(x)}) / n^{(x)} \in M1 \right\}$$
(1a)

204 
$$M1 = \{n^{(x)} / An^{(x)} = n^{(b)}, n^{(x)} \in R1\}$$
 (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  $A = \{a_{ij}, i \in N, j \in L\}$  is a matrix constructed from the stoichiometry coeffi-

207 cients of independent components in formulae of dependent components. Furthermore

208 
$$R1 = \begin{cases} n^{(x)} / n_j^{(x)} \ge 0, j \in D_0; \\ \underline{n}_j^{(x)} \le n_j^{(x)}, j \in D_1; \\ n_j^{(x)} \le \overline{n}_j^{(x)}, j \in D_2; \\ \underline{n}_j^{(x)} \le n_j^{(x)} \le \overline{n}_j^{(x)}, j \in D_3 \end{cases} \quad L = D_o \cup D_1 \cup D_2 \cup D_3$$
(2)

is the set of additional constraints on  $n_j^{(x)}$  composed of trivial non-negativity 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 ( $\overline{n}_j^{(x)} > 0, j \in D_2 \cup D_3$ ). Finally,  $G(n^{(x)})$  is the total Gibbs energy function of the system:

212 
$$G(n^{(x)}) = \sum_{j} n_{j}^{(x)} \upsilon_{j}, \quad j \in L$$
 (3)

213 In Eq. (3),  $v_j$  is the normalized chemical potential of the j-th dependent component, written in a 214 simplified dimensionless form as

215 
$$\upsilon_j = \frac{g_j}{RT} + \ln C_j + \ln \gamma_j + \Xi, \quad j \in L$$
(4)

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

221 
$$C_j = x_j = \frac{n_j^{(x)}}{\sum_{jp} n_{jp}^{(x)}}, jp \in l_k$$
 (5a)

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

224 
$$C_j = m_j = \frac{1000}{18.0153} \cdot \frac{x_j}{x_{jw}}, \ jw \in l_{aq}$$
 (5b)

where *m* is the molality (mol kg<sup>-1</sup> water), and *jw* 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

227 
$$C_j = x_j P, \quad j \in l_g$$
 (5c),

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

229 
$$C_j = x_j = 1, \quad j \in l_k \text{ and } n(l_k) = 1$$
 (5d).

The activity coefficient  $\gamma_j$  of the *j*-th dependent component in its respective phase is taken according to the chosen model of non-ideal mixing (details in [59]). The concentrations and activity coefficients for adsorbed species can also be considered [32, 33]. The non-logarithmic asymmetry term  $\Xi$  is [25]

234 
$$\Xi = 1 - x_{jw}, \forall j \in l_{aq} \setminus jw$$
 for the aqueous species (6a)

235 
$$\Xi = 2 - x_{jw} - \frac{1}{x_{jw}}, \quad jw \in l_{aq} \text{ for the water-solvent}$$
 (6b)

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

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

$$\nu - \mathbf{A}^T \hat{\boldsymbol{u}} \ge 0 \tag{7a}$$

244 
$$A\hat{n}^{(x)} = n^{(b)}; \quad \hat{n}^{(x)} \ge 0$$
 (7b)

245 
$$\hat{n}^{(x)^{T}}(\upsilon - \mathbf{A}^{T}\hat{u}) = 0$$
 (7c)

Here  $^{T}$  is the transpose operator, and  $^{\circ}$  denotes 'optimal', written in vector-matrix notation. Condition (7a), re-written with indices using eq. (4),

248 
$$\frac{g_j}{RT} + \ln C_j + \ln \gamma_j + \Xi - \sum_i a_{ij} \hat{u}_i \ge 0, \quad j \in L, i \in N$$
(8)

implies that for the *j*-th dependent component present at some equilibrium concentration  $C_j > 0$ in its phase, the primal chemical potential  $v_j$  must numerically equal the dual chemical potential

251 
$$\hat{\eta}_j = \sum_i a_{ij} \hat{u}_i, \quad j \in L, i \in N$$
(9)

From eqs. (7) and (8) it follows that (in equilibrium) the GEM dual solution  $\hat{u}_i$  values (Lagrange multipliers for balance constraints based on  $n^{(b)}_i$ ) are the chemical potentials of thermodynamically independent components, and that  $\hat{u}_i$  must have the same value in all co-existing 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

258 
$$L_{s} = \left\{ j \mid j \in l_{k}, k \in \Phi, n(l_{k}) = 0 \lor n(l_{k}) = 1 \lor n(l_{k}) > 0 \land n_{j}^{(x)} > \varepsilon_{x} \right\}$$
(10)

[25]. Here,  $\Phi$  is the set of indexes of phases, and  $\varepsilon_x$  is the numerical cut-off operational threshold (10<sup>-40</sup> <  $\varepsilon_x$  < 10<sup>-20</sup> mol). Amounts of all species below  $\varepsilon_x$  are zeroed off, and indexes of such species are removed from the  $L_s$  set.

263

264

In the actual GEM IPM algorithm, the extended KKT conditions [8, 25] are considered for the complete 'non-trivial' *R*1 set in order to simulate various metastable or so-called kinetically-controlled *partial equilibrium* states:

265 
$$\upsilon_j - \hat{\eta}_j \ge 0, (\upsilon_j - \hat{\eta}_j) \hat{n}_j^{(x)} = 0, \, \hat{n}_j^{(x)} \ge 0, \, j \in D_0 \cap L_S$$
 (11a)

266 
$$\upsilon_j - \hat{\eta}_j \ge 0, \ (\upsilon_j - \hat{\eta}_j)(\underline{n}_j^{(x)} - \hat{n}_j^{(x)}) = 0, \quad j \in D_1$$
 (11b)

267 
$$\upsilon_j - \hat{\eta}_j \ge 0, \ (\upsilon_j - \hat{\eta}_j)(\hat{n}_j^{(x)} - \overline{n}_j^{(x)}) = 0, \quad j \in D_2 \cap L_s$$
 (11c)

268 
$$\begin{array}{c} \nu_{j} - \hat{\eta}_{j} + p_{j} \geq 0, \ (\nu_{j} - \hat{\eta}_{j} + \hat{p}_{j})(\underline{n}_{j}^{(x)} - \hat{n}_{j}^{(x)}) = 0, \\ \hat{p}_{j} \geq 0, \quad \hat{p}_{j}(\hat{n}_{j}^{(x)} - \overline{n}_{j}^{(x)}) = 0, \quad j \in D_{3} \end{array}$$
(11d)

269 
$$\sum_{j} a_{ij} \hat{n}_{j}^{(x)} = n_{i}^{(b)}, \quad i \in N, \ j \in L$$
 (11e)

The Lagrange multipliers  $\hat{p}_j$  in eqs (11d) comprise additional Lagrange multipliers conjugate to dependent components with non-trivial two-side constraints (set  $D_3$ ), e.g. additional parts of the dual solution of the problem eq. (1) with two-side metastability amount constraints.

273

#### 274 3.2. The GEM IPM algorithm as implemented in the GEMS3K code

275 The original GEM IPM is described elsewhere [24, 25]. Later on, it was considerably modified 276 into the IPM-2 algorithm [8, 9], claimed to be more accurate and stable. However, our attempts 277 to couple IPM-2 (in the form of the standalone GEMIPM2K code) with mass transport codes 278 (e.g., [46]) have shown that the accuracy and stability of IPM-2 were still not sufficient. The 279 problem was finally solved by re-considering potential pathways of error propagation in primal 280 and dual solution approximations within the IPM algorithm, modifying the definition of conver-281 gence criteria in the mass balance refinement (MBR) loops of the algorithm, developing a mod-282 ified method for calculating phase and species stability, and adding internal re-scaling of the in-283 put 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 284 285 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 286 for the quality of results. This is because the GEM IPM algorithm is organized internally in an 287 288 operator splitting mode, where the mass balance is improved in one subroutine, and the direct 289 minimization of Gibbs energy is performed in another procedure, which maintains the previously 290 obtained mass balance precision. Hence, the mass balance precision in IPM is basically inherited 291 from that of the preceding approximation of the  $n^{(y)}$  vector.

292

293 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 Annex B.

297 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. 298 In addition, it might contain the initial chemical speciation  $n^{(y)}$ , obtained from a previous equili-299 300 brium 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 301 302 IPM descent algorithm is obtained by solving a simplified GEM problem using the linear pro-303 gramming (LP) simplex algorithm, modified to tolerate non-trivial one- or two-side metastability constraints  $\underline{n}_{j}^{(x)} > 0$  and  $\overline{n}_{j}^{(x)} < 10^{6} \text{ mol (Annex B.1)}.$ 304

305 The linear programming SolveSimplex() procedure yields at most n(N) non-zero values 306 in the  $n_{(s)}^{(y)}$  initial approximation vector. In order to involve all possible phases and components in 307 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  $\varepsilon_f$  (ca. 308 10<sup>-5</sup> mol for a total amount of 1000 moles of all independent components in the system). This 309 insertion distorts the mass balance with respect to the input  $n^{(b)}_{i}$  values, and increases the balance 310 residuals to about  $n \cdot 10^{-5}$  mol. To correct for this distortion, a non-linear mass-balance refinement 311 (MBR) procedure (Annex B.2) is used for minimizing the mass balance residuals for all inde-312 pendent components to less than a threshold value (typically 10<sup>-13</sup>) normalized relative to the ab-313 solute input mole amount  $n^{(b)}_{i}$  of any *i*-th independent component. This yields a balanced feasible 314 initial approximation (FIA) vector  $n^{(y)}$ , suitable as input for the MainIPMDescent() procedure. 315 "Feasible" means that the  $n^{(y)}$  vector belongs to the so-called feasible domain, i.e. the set M1 in 316 317 eq. (1).

Alternatively, if a suitable speciation vector  $n^{(y)}$  is already available at input (e.g. from a 318 319 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) 320 mode, which normally results in much faster GEM calculations relative to the AIA mode. Unlike 321 322 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 323 324 speed gain by SIA is associated with a certain risk to obtain an inconsistent set of phases present 325 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 326 327 the subsequent phase stability analysis (see Annex A.4), but this is not always possible if non-328 ideal solution phases are included in the system but they are not present in SIA. Conversely, for 329 chemical problems with many non-ideal solution phases (aqueous phase, solid solutions, fluid 330 mixtures etc.), the slower AIA mode is preferable.

331 The vector  $n^{(y)}$  is substituted into the main IPM descent algorithm (Annex B.3), which ad-332 justs simultaneously primal and dual solution approximations of the GEM problem at every iteration, until the criterion of convergence  $C_D < \varepsilon_D$  is satisfied ( $\varepsilon_D$  is the empirical threshold chosen in the range 10<sup>-7</sup> to 10<sup>-5</sup>). Optionally, the new routine for detection of dual solution approximation 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  $\varepsilon_x$ , and proceeds to the stable phase assemblage and speciation by successively zeroing off all such unstable dependent components.

339 In the phase selection - speciation cleanup (PSSC) subroutine (Annex A.4), the stability indices  $(\Omega_k, \Lambda_k, f_{\alpha,k})$  for all phases are calculated from the IPM dual solution, and the quality of 340 the entire primal solution  $\hat{n}_{i}^{(x)}$  is checked and (if necessary) corrected within the SpeciationClea-341 342 nup() algorithm. If any phase is detected that is not present in the mass balance but should be 343 stable, or if some correction violates the mass balance residual tolerance, the loop involving the 344 main IPM descent procedure is repeated using the current (cleaned-up) primal solution as FIA. 345 Otherwise, the highly accurate and consistent GEM dual and primal solutions are delivered as 346 output, together with activity coefficients of dependent components in all non-ideal solution 347 phases.

348

#### 349 3.3. Precision and robustness of the GEM IPM algorithm

350 Given that consistent thermodynamic data are used, the GEM IPM algorithm must always yield a 351 correct and accurate equilibrium phase assemblage and speciation, regardless of how different 352 are the input total amounts of independent components, and how many pure- and solution phases 353 were included into the initial definition of the chemical system. However, in calculation of real 354 systems this is limited by the achievable level of consistency between the primal and dual solu-355 tions of the numerical IPM algorithm, by the mass balance precision, and by the presence or ab-356 sence 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]). 357

The analysis of error propagation throughout the flowchart (Fig. 4) can be performed keeping in mind that (i) the GEM IPM algorithm produces two solutions (primal and dual) that are identical only in theory, (ii) the GEM and the MBR steps are separated in an operatorsplitting mode, (iii) steps where species (dependent components) are inserted or eliminated partially degrade the mass balance, (iv) correct composition-activity relationships in non-ideal solution phases may be partially distorted by applied smoothing procedures (Annex B.7), and (v) activity coefficients of dependent components are only adjusted after each IPM iteration.

365 The requirement of obtaining robust results implies that the GEM IPM solver must com-366 pute correct phase assemblage, equilibrium speciation, and activities of components for any feas-367 ible input bulk composition of the system and for a wide range of numerical tolerances. This task 368 is not trivial, and it is related to the theoretical existence of single or multiple minima of the total 369 Gibbs energy function, to the extent of non-ideality of mixing, and to the presence of miscibility 370 gaps (see Annex B.6). We consider the system numerically stiff when chemical potentials of 371 some ICs or DCs are not well constrained by physically significant amounts of relevant chemical 372 species or persistent contributions from activity coefficients. In such a system, the resulting IPM solution may significantly depend on numerical tolerances (e.g.  $\varepsilon_D$ ), including the controls of 373 374 smoothing (Annex B.7). To detect numerically stiff systems, a special new procedure has been 375 added to the GEMS3K IPM algorithm implementation (Annex B.8). In most cases, the detected 376 stiffness manifested in the divergence of dual solution approximation is due to the absence of 377 redox buffering in the system. This then calls for a more careful setup of the system definition.

378

#### 379 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 383 involves solving n = n(N) linear equations with n = n(N) unknowns using a (n·n) coefficient ma-384 trix (where n(N) is the number of independent components). The MBR stage typically converges 385 in a few iterations, whereas the IPM Descent stage may need up to few thousand iterations in 386 certain cases (see estimates below). Any GEM algorithm that simultaneously looks for the mini-387 mum in total Gibbs energy and minimizes mass balance residuals would have to solve a SLE of 388 at least (2n·2n) size at each iteration [5, 23], performing at least the same number of iterations as 389 the IPM Descent algorithm. This explains qualitatively why the operator splitting used in GEM 390 IPM is fast and numerically efficient. Additional gain in speed can be achieved in most cases 391 through usage of the SIA that reduces the number of IPM Descent iterations by 10-20 times.

392 Theoretical estimates of GEM IPM computational costs as compared to other methods. It 393 is known that the theoretical number of floating-point operations (flops) of a SLE solver using 394 the Cholesky or LU decomposition is about  $n^3/3$  and  $2n^3/3$ , respectively [55]. Assuming that both 395 steps of the GEM IPM algorithm perform an equal number of iterations n(r) (in practice, the 396 MBR does much less iterations than the IPM), and that the LU decomposition is used in rare cas-397 es, 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 398 399 calculation time must consequently increase 8 times. Conversely, for a hypothetical GEM algo-400 rithm that does not use an operator-splitting approach but solves a problem of simultaneous mi-401 nimization of total Gibbs energy and mass balance residuals, the size of the SLE to solve on n(r)402 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  $2/3 \cdot n(r) \cdot [2n]^3$ 403 (LU). Such a code would be 4-8 times slower than the GEM IPM kernel code for the same num-404 405 ber of independent components in the system.

406By comparison, LMA algorithms such as MINEQL [62] directly minimize the mass bal-407ance deviations, while maintaining the system of boundary conditions imposed by the LMA ex-

408 pressions and equilibrium constants for 'product species'. The commonly used Newton-Raphson 409 algorithm has a quadratic convergence behavior (e.g., [28]), which means that, in order to reach 410 the double precision of 14 decimal digits ( $\frac{\zeta_i^{(r)}}{n_i^{(b)}} < 10^{-14}$ ), it needs n(r) < 8 iterations. For each ite-

411 ration, the Jacobian matrix has to be inverted by solving the SLE with the  $(n \cdot n)$  coefficient ma-412 trix, hence the LMA algorithm is expected to require a calculation time proportional to between 413  $7/3 \cdot n^3$  and  $14/3 \cdot n^3$ , i.e. roughly 2-10 times faster than the GEM IPM. However, this advantage of 414 LMA is out-weighted in complex chemical thermodynamic systems with many potentially poss-415 ible solid phases that cannot all be included in the initial approximation. For such systems, many 416 external loops in an operator-splitting mode are necessary for adding over-saturated and remov-417 ing under-saturated phases in the mass balance (see Introduction; see also [6]).

418 A combined GEM-LMA algorithm (e.g. in the HCh code, Shvarov, 2011, personal com-419 munication) starts with obtaining a rough GEM primal solution, which is then improved by suc-420 cessive LMA runs, followed by application of phase stability criteria to decide which phases to 421 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 422 of loops for phase stability checking, typically about 10), the total cost in this case would be 423  $(n(r)+4l)\cdot n^3$ , i.e. in multi-phase systems definitely greater than that of the operator-splitting GEM 424 425 IPM algorithm. However, this toll is justified for systems with diverging chemical potentials of 426 some independent components, because the GEM method alone cannot accurately solve such 427 equilibria.

428 GEM algorithms used mainly for plotting complex phase diagrams in petrology, such as 429 THERIAK [13, 14] and Perplex [11], reduce the non-linear minimization problem (eq. 1a and 430 1b) to a sequence of LP problems solved by the simplex method. Simplex LP algorithms have 431 the exponential time complexity [55], but show in practice the performance of O(m + n), so they are in most cases very efficient algorithms when *n* and *m* are not very large. In the given problem (eq. 1) the sizes are n = n(N) (basis) and  $m = n(\Phi) \le n(L)$ , 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·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  $10^4$ ·n.

439 In the THERIAK [14] code, the LP run for the whole system is embedded in several l440 loops where solving the LP is preceded by adjusting equilibrium compositions of all multi-441 component phases [13], which is done by a steepest-descent non-linear minimization of Gibbs 442 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  $n(r) \cdot (n_t)^3$  (where n(r) is typically  $15 \pm 10$ ) and the number of 443 solution phases equal to n, the total cost can be estimated as  $l \cdot (n \cdot 15 \cdot 4^3 + 100 \cdot n)$ , or about  $l \cdot 10^3 \cdot n$ . 444 445 i.e. comparable to that of PERPLEX. For a system with 10 independent components, perfor-446 mance of both LP-based algorithms is slightly slower than that of LMA, but for a system with 30 447 independent components and many multi-component phases it may be significantly faster. How-448 ever, LP based algorithms suffer from inefficient calculation of aqueous speciation, and from 449 insufficient mass balance precision due to the technical limits of phase composition discretiza-450 tion 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 1000, and the number of phase stability loops *l*, assumed to be between 1 and 20.

458

#### 459 4.2. Impact of non-ideal solution phases

460 Another important factor that may strongly affect the algorithm performance is the presence of 461 strong non-ideality of mixing in some phases, which can either disturb the convergence of the 462 IPM, or cause difficulties in the vicinity of the consolute point of miscibility gaps. In this area, 463 the differences in Gibbs energy between exsolving phases, as well as the IPM gradients, become 464 very small, slowing down the convergence. However, GEM IPM always converges correctly in 465 this region, and the number of iterations dramatically decreases when moving away from the 466 consolute point. In some systems of this type, up to 7000 IPM iterations may be required (regard-467 less whether AIA or SIA mode) are used, while less than 100 iterations are sufficient in most 468 chemical systems when the AIA mode is used. In the SIA mode, the typical number of iterations 469 may be 10 to 50 times less. The SIA mode has, therefore, a high potential for speeding up 470 coupled reactive transport simulations.

471

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

473 The revised GEM IPM algorithm described above is implemented as a C++ class, which is em-474 bedded in the TNode C++ class implementing the data exchange interface. Together, both 475 classes comprise the GEMS3K code, compatible with doxygen specifications (see 476 http://www.doxygen.org) for generating the source code documentation. One particular goal of 477 the GEMS3K development was to retain the code implementation and handling of input/output 478 data as simple as possible, as a prerequisite for compiling the coupled RMT-GEM program on 479 any computer architecture, up to parallel HPC clusters. Consequently, the GEMS3K program can 480 only use the data for a compressed GEM problem encapsulated in the IPM work data structure. 481 Because of this, the GEMS3K code has some restrictions compared to the full GEM-Selektor-3

482 package (Table 2). Based on considerations of data organization and exchange efficiency, the 483 input data of GEMS3K have been subdivided into three main subsets, as shown in Table 3. Only 484 one relatively large DCH and one IPM file contain all the data relevant for the definition of the 485 chemical system, models of mixing, and numerical settings of the GEM IPM algorithm. The 486 node-specific data are exclusively contained in a set of small and compact input-output DBR 487 files. The advantage of having three different types of input files instead of one single file type is 488 that in a typical RMT problem with a large number of nodes, higher performance and flexibility 489 of the coupled code can be achieved at much less amount of data exchange.

490 The TNode class, designed for connecting GEMS3K to upper-level (mass transport) 491 codes, provides a collection of access methods (function calls) for loading the input chemical 492 system definition and related thermodynamic and composition data into the IPM work structure, 493 running the GEM IPM algorithm for one calculation of equilibrium state, and extracting the cal-494 culated speciation results. For the subsequent visualization, the output in VTK format files 495 (http://www.vtk.org) is also supported. The TNode class uses two data exchange interface struc-496 tures, which are DATACH and DATABR. The structures can be allocated in the computer memory 497 and written into or read from text/binary files (DCH and DBR file, respectively). The parameters 498 of non-ideal mixing models in phases, as well as the numerical controls and tolerances of GEM 499 IPM, can be exchanged via the IPM file (loaded directly into the IPM work structure). A set of 500 direct access methods to parameters of mixing models and thermodynamic data kept in IPM 501 work structure is provided in the TNode class for the efficient implementation of coupled codes 502 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 standalone coupled code, as described below. Moreover, results of standalone calculations of equilibria in nodes can be written in DBR files that can be imported into a 'parent' GEM- 507 Selektor modeling project as additional system-equilibrium (SysEq) records (details in GEM-508 Selektor runtime help). There, the input and output GEM data can be conveniently explored, and 509 compared with internally re-computed speciation. This functionality can save the user a lot of 510 time for tedious work with I/O files. It supports our recommendation first to set up in a GEM-511 Selektor project the chemical system definitions and initial bulk compositions of node types and 512 test them within relevant temperature and pressure intervals, before exporting them in GEMS3K 513 I/O files and starting expensive calculations with a standalone RMT code.

514

#### 515 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.

521 Representation of thermodynamic data for variable pressure-temperature conditions. Di-522 rect access to the GEM-Selektor thermodynamic database is not possible, and extensive functio-523 nality for pressure-temperature corrections of thermodynamic data is not available in the standa-524 lone GEMS3K code. Therefore, the retrieval of thermodynamic data for dependent components 525 is implemented in GEMS3K by an interpolation through look-up arrays provided in the DATACH 526 structure and in the DCH file. These look-up arrays can be generated through a GUI dialog in the 527 GEM-Selektor-3 package from the modeling project thermodynamic database when exporting 528 the DCH file. To use the lookup arrays, the TNode class employs a 2-D Lagrangian interpolation 529 routine that extracts thermodynamic data for temperature and pressure of interest and loads them 530 into the IPM work data structure before starting the next GEM IPM calculation.

#### 532 5.2. The DATABR data structure

This data structure (and DBR file), which contains all information relevant for node data exchange with the external multi-species 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 approximation 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.

544 In the AIA mode, GEM calculations are robust and reproducible, but may be compara-545 tively slow (20 to 1000 IPM iterations, depending on the chemical system definition and compo-546 sition). The alternative SIA mode uses the current contents of the speciation (xDC) and activity 547 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 548 549 mass balance relative to the current bulk composition vector (bIC), and neither the phase assem-550 blage nor the redox state are changing, then GEM IPM is expected to perform only very few ite-551 rations (i.e. 10 to 20 times faster than in the AIA mode). This substantial acceleration can be pa-552 ramount 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 "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.

560 In meaningful RMT problems, typically the phase assemblage changes or redox transi-561 tions occur across some distinct reaction fronts that are represented by only a small fraction of 562 the nodes. In such a node, the current speciation taken in SIA mode might be an invalid GEM 563 initial approximation. The GEM IPM algorithm has several built-in criteria that can detect this 564 condition (see Annexes B2, B3, B8 for details). In cases where this situation is detected, the pro-565 gram automatically switches to the automatic initial approximation (AIA) and obtains a correct 566 new phase assemblage and speciation at more computational cost. However, for the vast majority 567 of nodes, slight changes of speciation and composition as a consequence of mass transport would 568 not result in a phase transformation or a redox transition.

569

#### 570 6. Experience from the provisional coupling of GEMS3K with transport codes

#### 571 *6.1. OGS-GEMS3K*

572 OpenGeoSys (OGS), hosted at the Helmholtz Centre for Environmental Research (UFZ), is a 573 research initiative for the development of computational methods for the solution of thermo-574 hydro-mechanical-chemical (THMC) processes in porous-fractured media. Applications are in 575 the fields of groundwater and surface hydrology, water resources, geotechnics, radioactive waste disposal, geothermal energy, and CO<sub>2</sub> storage [7, 30, 35, 46, 47, 51, 54, 60, 63]. The concept of 576 577 OpenGeoSys is a distributed open source scientific software development that should benefit 578 from joint efforts of all involved groups. The OpenGeoSys modules for simulation of groundwa-579 ter flow and transport of dissolved species have been coupled to the GEMS3K kernel. The coupl-580 ing strategy, verification examples, and the first applications were briefly described by Shao et

581 al. [46]. Like for similar coupled codes, application of the OGS-GEMS3K coupled code is lim-582 ited by long calculation times and extensive memory requirements. Driven by these shortcom-583 ings, we recently parallelized the central geochemical loop in OGS-GEMS3K that contains inde-584 pendent calls to GEMS3K for each finite element node. We used an existing MPI implementa-585 tion of OGS [60] and combined this with a multi-threaded calculation of the coupling loop [64]. 586 The combination of multi-threading and MPI provides an effective and flexible environment to 587 speed up OGS-GEMS3K calculations, while limiting the required memory use. We would like to 588 remark that this was only possible because the GEMS3K code is thread-safe, i.e. it is safely pos-589 sible to run multiple instances of GEMS3K, as these do not share any data structures.

590 6.2. Benchmark example

591 The verification example is based on a hypothetical setup of a benchmark test case, which con-592 ceptionally considers one-dimensional advection-dominated mass transport, instantaneous calcite 593 dissolution, and dolomite precipitation processes. This benchmark test was first proposed by En-594 gesgaard and Kipp [17] for model verification of the MST1D code, and later used by Prommer 595 [41] for the PHT3D code. It is also used as an official test case for different couplings between 596 OGS and chemical equilibrium solvers. The following description is adapted from [29, 46]. A 597 one-dimensional column that initially contains calcite and water saturated with calcite in the pore 598 space is continuously infiltrated by water that contains dissolved magnesium chloride (Fig. 5). 599 With the progressive movement of the MgCl<sub>2</sub> advection front, calcite dissolves, and dolomite is 600 temporarily formed as a product of incongruent dissolution, but eventually dissolves as well. The 601 media properties and the threemodynamic setup are detailed in [46]. For OGS-GEMS calcula-602 tions, all the independent components and species need to be explicitly set for initial and boundary conditions. In this example, all concentration values are given in the unit of mol·m<sup>-3</sup> water. 603 This test has been simulated by OGS-PHREEQC, OGS-ChemApp, and OGS-GEMS. In all these 604 benchmarks, the same GEMS built-in version of Nagra/PSI 01/01 chemical thermodynamic da-605

tabase was used. Figure 6 depicts a comparison of simulation results, which shows excellentagreement between the different coupled RMT code implementations.

608

609 **7. Outlook** 

610 The GEMS3K geochemical equilibrium solver facilitates geochemical modeling of multicompo-611 nent-multiphase equilibria, delivering robust and numerically stable results with high mass bal-612 ance precision that is comparable to conventional speciation codes that use the LMA method. 613 Based on the revised GEM IPM convex programming algorithm, GEMS3K is able to accurately 614 solve complex geochemical systems with many nonideal solution phases, including dilute to 615 concentrated aqueous solutions, solid solutions, gas mixtures, supercritical fluids, melts, and 616 sorption phases. The modular and computationally efficient code design, foreseen to be released 617 open-source under the LGPL v.3 license [21], is favorable for coupling GEMS3K to existing 618 mass transport codes, developing new RMT codes, and implementing advanced geophysical-619 geochemical models that simulate geodynamic processes and global geochemical cycles. Be-620 cause GEMS3K is also included into the GUI-driven GEM-Selektor code package, all input data 621 required for coupled reactive transport simulations can be straightforwardly created and tested in 622 GEM-Selektor, and then exported interactively into a file set for use in standalone codes coupled 623 with GEMS3K.

Future developments will include substantial extension of the built-in selection of equation-of-state and activity models for multi-component phases available in the TSolMod class library, re-implementation of sorption models into a TSorpMod class library using a objectoriented 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 consis-

- tent minimization of thermodynamic potentials other than Gibbs energy [26], such as Helmholtzenergy and negative entropy.
- 633

#### 634 Acknowledgments

The GEM-Selektor code (http://gems.web.psi.ch) is the latest variant of a Selektor code family developed over 30 years by the effort of numerous people. We thank Frieder Enzmann and Wilfried Pfingsten for helpful early discussions on the GEMS3K data and file exchange interface, and Gillian Grün for designing the GEM-Selektor icons used in Figs.2 and 3. Nagra, Wettingen, and the Swiss National Science Foundation (SNSF) are thanked for contributing partial funding to this long-term research effort.

## 642 Annex A

## 643 List of mathematical symbols and acronyms

Symbol	Explanation				
Mathematical symbols					
${f A} a_i^{(\hat{u})}$	Matrix of stoichiometry coefficients $a_{ij}$ of the <i>i</i> -th IC in formula of <i>j</i> -th DC Dual-thermodynamic activity of the j-th DC in its phase				
$egin{array}{c} lpha_{\gamma} & \ C_D & \ c_j & \ C_j & \ D_0 & \ D_1 & \end{array}$	Smoothing factor for improving IPM convergence in highly non-ideal systems Dikin's criterion of convergence of the IPM descent algorithm Constant in the LP minimization sub-problem Concentration of the <i>j</i> -th DC in its respective phase Set of indexes of dependent components with trivial non-negativity constraints 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 $\overline{n}_{j}^{(x)}$				
$\Delta \ \delta^{(x)}_{j}, ar{\delta}_{j}, ar{\delta}_{j}$	Vector of correction of the $n^{(y)}$ approximation on MBR or IPM iterations Magnitudes of correction of the <i>j</i> -th DC amount in speciation Cleanup()				
$\delta^{(\mu)}_{j}$	Difference between primal and dual DC chemical potential in Cleanup()				
$egin{array}{c} arepsilon_lpha & arepsilon \\ arepsilon_\delta & arepsilon_b & $	Empirical uppermost value for the smoothing factor $\alpha_{\gamma}$ Empirical minimum value for the smoothing factor $\alpha_{\gamma}$ Relative tolerance for the IC mole balance residuals $\varsigma_i$ Tolerance for checking the Dikin's criterion of IPM convergence				
$\mathcal{E}_{\delta\mu}$	Tolerance for $\delta_j^{(\mu)}$ to perform DC amount correction in Cleanup()				
E <sub>eΦ</sub> Ef	Tolerance for detecting the mass balance violation in Cleanup() Empirical fill-out constant to obtain the initial approximation $n^{(y)}$ from $n^{(y)}$				
$egin{array}{ccc} \mathcal{E}_{fi} & & & \\ \mathcal{E}_{n\Phi} & & & \\ \mathcal{E}_{\Phi} & & & \\ \mathcal{E}_{\Phi+} & & & \\ \mathcal{E}_{u} & & & \\ \mathcal{E}_{r} & & & \end{array}$	Empirical fini-out constant to obtain the initial approximation $n^{-1}$ from $n_{(s)}^{(s)}$ Empirical constant for insertion of a single-component phase in Cleanup() Tolerance for the amount of stable phase present in the mass balance Tolerance for the stability index $\Omega_k$ for <i>k</i> -th phase for elimination Tolerance for the stability index $\Omega_k$ for <i>k</i> -th phase for insertion General tolerance for the detection of diverging dual solution approximation Tolerance for zeroing-off the amounts of dependent components $n_{(x)}^{(x)}$				
f., h	Karpov's (old) phase stability index				
$\int a_{,\kappa}^{o}$	Fugacity of the <i>j</i> -th pure non-ideal gas/fluid component at <i>T</i> , <i>P</i> of interest				
	Set of indexes of phases in the chemical system definition Total Gibbs energy of the chemical system (scalar) in moles Input vector of standard state Gibbs energy functions (per mole) of dependent components at temperature of interest and reference state pressure in J mol <sup>-1</sup> Darken's parameter for a component of solid solution or melt				
$\gamma_j$	Activity coefficient of the <i>j</i> -th dependent component in its phase				

i	Index of the independent component
It	Index of the PhaseSelection() loop iteration
j	Index of the dependent component
jw	Index of H <sub>2</sub> O solvent in aqueous phase
k	Index of the phase
$\Lambda_k$	Logarithmic phase stability index used in PhaseSelection() procedure
L	Set of indexes <i>j</i> of dependent components
$L_S$	Set of indexes of dependent components subject to KKT conditions ( $L_S \subseteq L$ )
	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 k-th phase $(l_k \subseteq L)$
$\lambda_j$	Scalar antimal stop length on MDD or IDM descent iterations
$\Lambda_b^{(r)}$ , $\Lambda_r$	Scalar optimal step length on MBK of IFM descent iterations
$m_j$	Molality of the <i>j</i> -th aqueous species $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{$
M1	Set of constraints to the $G(n^{(3)})$ minimization problem
$\hat{\mu}$	Chemical potential
$\eta_{j}$	Dual-solution chemical potential of the <i>j</i> -th dependent component
Ν	Set of indexes <i>i</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 $n_i^{(0)}$
$n_k^{(\Phi)}$	Total amount of the <i>k</i> -th 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)
$n_{(s)}^{(y)}$	Initial approximation of $n^{(x)}$ obtained in the Simplex LP step
$\hat{n}^{(x)}$	Primal (optimal) solution of the GEM IPM problem – DC mole amounts $\hat{n}_j^{(x)}$
$n_j^{(x,c)}$	Corrected amount of dependent component in SpeciationCleanup()
$\underline{n}_{j}^{(x)}$	Additional constraint from below on the mole amount of <i>j</i> -th DC $n_j^{(x)}$
$\overline{n}_{j}^{(x)}$	Additional constraints from above on the mole amount of <i>j</i> -th DC $n_j^{(x)}$
n(A)	The number of elements in the set <i>A</i>
n(r)	Number of iterations done by MBR() or MainIPMDescent() procedure
$\Omega_k$	Phase stability index used in PhaseSelection() procedure
P	Pressure (bar)
$p_r$	Lagrange multiplier conjugate to the two side additional constraints in $\mathbf{n}^{(x)}$
$P_j$	Lagrange multiplier conjugate to the two-side additional constraint on $n_j$
$q_j^{(r)}$	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, $8.31451 \text{ J K}^{-1} \text{ mol}^{-1}$
RI	Set of additional constraints on the elements $n_j^{(x)}$ of the speciation vector $n^{(x)}$
<b>r</b> <sub>ik</sub>	Coefficient of the matrix for linear equation solver in MBR() and main IPM descent
(r), r T	Superscript or subscript index of MBR() or MainIPMDescent() iteration Temperature (K)

Т	Operator transpose(for vector, matrix)
и	Vector of Lagrange multipliers conjugate to elements of $n^{(o)}$ vector
û	Dual (optimal) solution of the GEM IPM problem – IC chemical potentials $\hat{u}_i$
υ	Vector of approximations of primal chemical potentials $v_i$
v	Vector of Lagrange multipliers conjugate to mass balance residuals
$x_i$	Mole fraction of <i>j</i> -th dependent component in its phase
$\omega_{i}^{(\hat{u})}$	Activity of <i>j</i> -th DC in mole fraction scale, estimated from the dual solution
$\overline{\omega}_{j}$	Mole fraction of DC in its phase, estimated from the dual activity $\omega_j$
Ξ	Term for conversion between internal and external activity coefficients
Si	Residual of the mole balance for <i>i</i> -th 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

#### 651 Annex B

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

#### 653 *B.1.* Selection of the initial approximation

654 If the input chemical system definition contains a speciation which is compatible with the bulk 655 chemical composition and phase assemblage, it can be directly used as FIA in the Smart Initial 656 Approximation mode. In this case, the GEM IPM algorithm usually makes approximately 10 to 657 20 times less iterations, which may be highly beneficial for the overall performance of large 658 coupled reactive transport calculations, even on parallel architectures. However, there is no guar-659 antee that the previously computed speciation, even for the same node, is in the feasible domain. 660 This situation is usually detected within MBR or PSSC procedures and the GEM IPM algorithm 661 then automatically switches to the AIA mode. This detection mechanism is not absolutely effi-662 cient, particularly in systems with more than one highly non-ideal solution phases, and more 663 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  $n^{(y)} = n^{(x),old}$  and  $\gamma = \gamma^{(old)}$  with the new bulk composition  $n^{(b)} = n^{(b),new}$  are simply forwarded to the MBR procedure. In the AIA mode, the generic non-linear GEM problem is first truncated into a LP sub-problem by cutting off the concentration and activity coefficient terms in chemical potential expressions (eq. 4) for all dependent components:

671 
$$\min_{\substack{j \in L_{s} \\ subject \ to \ An_{(s)}^{(y)} = n^{(b)}, \ n_{(s)}^{(y)} \ge 0}$$
(B-1)

672 where the constant  $c_j$  is defined as

673 
$$c_{j} = \begin{cases} \frac{g_{j}}{RT} + \ln \lambda_{j}, j \in L \setminus (L_{g} \cup (L_{aq} \setminus jw)) \\ \frac{g_{j}}{RT} + \ln \lambda_{j} + \ln P, \ j \in L_{g} \\ \frac{g_{j}}{RT} + \ln \lambda_{j} + \ln 55.5083734, \ j \in L_{aq} \setminus jw \end{cases}$$
(B-2)

674 The  $\lambda_i$  are the optional constants controlling the stability of dependent components at LP initial 675 approximation, when each dependent component (species) is treated as a pure phase. The LP 676 sub-problem (eq. 12) is then solved using a modified simplex algorithm, which yields very small relative mass-balance residuals ( $\zeta < 10^{-12}$ ). The GEMS3K calculation is completed at this point, 677 678 if no solution phases (having two or more dependent components) are present in the chemical 679 system definition. Otherwise, the original non-linear minimization problem (eqs. 1 to 3) is restored. Initial values of  $n_j^{(y)}$  are set equal to simplex-calculated amounts  $n_{j(s)}^{(y)}$  if  $n_{j(s)}^{(y)} > \varepsilon_{f}$ , or to a 680 small value ( $\varepsilon_{\rm f} = 10^{-5}$  mol) if  $n_{j(s)}^{(y)} < \varepsilon_{\rm f}$ . This initial approximation fill-out procedure ensures that 681 682 thermodynamically stable phases and species will not be eliminated from the final GEM result. 683

#### 684 B.2. The mass-balance refinement (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  $n_i^{(y)}$  values in such a way that any mass balance residual

688 
$$\zeta_i^{(r)} = n_i^{(b)} - \sum_j a_{ij} n_j^{(y,r)}, \ i \in N, \ j \in L$$
 (B-3)

does not exceed a small prescribed tolerance value  $\varepsilon_{b,i}$ , and the approximation remains in the feasible domain M1 (see eq. 1). Here and below, (*r*) stands for the iteration index. In the original IPM algorithm [24, 25], these threshold values  $\varepsilon_b$  for balance residuals  $\varsigma_i$  were taken to be the same for all independent components, in the range of  $10^{-8} \le \varepsilon_b \le 10^{-6}$  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

$$699 \qquad \frac{\varsigma_i^{(r)}}{n_i^{(b)}} < \varepsilon_b \ \forall i \in N \setminus ie \tag{B-4}$$

where *ie* is the index of the electrical charge IC, and  $10^{-15} < \varepsilon_b < 10^{-12}$  is the empirical threshold. The new relative scale conditions (eq. B-4) 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  $\max_i |\varsigma_i^{(r)}| \le \varepsilon_3$  (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  $\varepsilon_{abs}$  for major independent components (for instance, H and O in aquatic systems) are implemented. The latter can either be equal to the relative value  $\varepsilon_b$  (this is the default), or set separately as  $10^{-|\varepsilon_{b,exp}|}$  using a prescribed decimal exponent  $\varepsilon_{b,exp}$ . This combination of both relative and absolute mass balance precision may be beneficial for chemical systems with several non-ideal 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  $n^{(x,0)} \in M_1$  of the main GEM IPM descent algorithm, and also used to compute first approximations of activity coefficients in non-ideal phases  $\gamma^{(0)}$ , as well as those of primal chemical potentials  $\upsilon^{(0)}$  (eq. 4).

#### 717 B.3. Main IPM descent algorithm

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

722 
$$n^{(x,r+1)} = n^{(x,r)} + \lambda_r \Delta^{(r)}$$
 (B-5)

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

725 
$$\min\sum_{j\in L} \upsilon_j^{(r)} \Delta_j, \quad st. \quad A\Delta = 0, \sum_{j\in L_s} \frac{\Delta_j^2}{q_j^{(r)}} \le 1$$
 (B-6)

726 where the constraint  $A\Delta = 0$  preserves the mass balance, the ellipsoid  $\sum_{j \in L_s} \frac{\Delta_j^2}{q_j^{(r)}} \le 1$  replaces all

non-trivial upper and lower metastability constraints by a single additional constraint, and the
 corresponding weight multipliers are given as

729 
$$q_{j}^{(r)} = \begin{cases} n_{j}^{(x,r)}, & j \in D_{0}, \\ n_{j}^{(x,r)} - \underline{n}_{j}, & j \in D_{1}, \\ \overline{n}_{j} - n_{j}^{(x,r)}, & j \in D_{2}, \\ \min\left[(n_{j}^{(x,r)} - \underline{n}_{j}), (\overline{n}_{j} - n_{j}^{(x,r)})\right], & j \in D_{3}. \end{cases}$$
(B-7)

In solving the sub-problem (B-6), as well as a similar sub-problem in the MBR procedure, the internal linearization procedure is performed, in which the system of linear equations is solved for the Lagrange multipliers  $v_i^{(r)}$  using the Cholesky method. In GEMS3K, the Jama TNT C++ library [40] from NIST is used for this purpose.

After calculation of the descent-direction  $\Delta^{(r)}$  vector (see eq. B-6), the optimal step length  $\lambda_r$  is found by solving a uni-dimensional minimization sub-problem

736 
$$\lambda_r = \underset{0 \le \lambda \le \mu_r}{\operatorname{arg\,min}} \left[ \sum_{j \in L} (n^{(x,r)} + \lambda_r) \upsilon_j^{(r)} \right]$$
(B-8)

where the  $\mu_r$  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  $n^{(x,r+1)}$  is then obtained from eq. (B-5), 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 chemical potentials  $v_j^{(r+1)}$ . Activity coefficients and related configurational entropy terms for various solution and fluid phase models of mixing are computed as described in [59].

743 The convergence of IPM iteration process is checked using a modified Dikin's criterion 744  $C_D$ , which is defined as:

745 
$$C_D = \sum_{j \in L_S} \left| q_j^{(r)} \left( \sum_{i \in N} \mathbf{a}_{ij} u_i^{(r)} - \upsilon_j^{(r)} \right) \right|$$
 (B-9)

Iterations are completed when  $C_D \le \varepsilon_D$ , where  $10^{-7} \le \varepsilon_D \le 10^{-3}$  mol is the empirical threshold. 746 For a system having a total amount of 1000 moles of independent components, the  $C_D$  is typical-747 ly set to about  $3 \cdot 10^{-6}$ . The number of performed iterations strongly depends on the system com-748 749 plexity (total number of components, presence of strongly non-ideal solution phases) and on the 750 mode of initial approximation. In the AIA mode, typically 15 to 1000 iterations are done, in the 751 SIA mode about 1 to 100 iterations are required (usually only a few). After the final recalculation 752 of activity coefficients, the main IPM descent run is completed, and the primal and dual solutions 753 are subsequently checked for their consistency.

754

#### 755 B.4. PhaseSelection and SpeciationCleanup algorithms

756 Consistency checks are performed within the Phase Selection - Speciation Cleanup (PSSC) pro-

cedure. In GEMS3K, the completely revised PSSC replaces the Select-2 procedure used in older

758 versions of GEM-Selektor and Selektor codes [9, 25]. The latter is, however, retained as an op-759 tion that might be useful in complex systems with many non-ideal solution phases. In such sys-760 tems, a subtle interplay between small differences in Gibbs energy between different solution 761 phases of similar bulk composition (particularly Fe-Mg silicates) and slight distortion of compo-762 sition-activity relationships arising from activity coefficient smoothing procedures (see below) 763 might result in numerical convergence problems that are in some cases better resolved by the old 764 Select-2 procedure. The main reason for introducing the new PSSC procedure was to improve the overall robustness and precision of the primal solution  $\hat{n}^{(x)}$ , sometimes found to be insuffi-765 766 cient for coupled reactive mass transport calculations. The PSSC procedure has two essential steps, which are PhaseSelection and SpeciationCleanup. 767

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

770 
$$\Lambda_k = \log_{10} \Omega_k; \ \Omega_k = \sum_{j \in l_k} \overline{\sigma}_j$$
(B-10)

771 where

772 
$$\varpi_j = \frac{\omega_j^{(\hat{a})}}{\gamma_i^{(\hat{n})}}$$
(B-11)

is the estimate of mole fraction of the *j*-th component in the *k*-th phase,  $\omega_j^{(\hat{u})}$  is the activity calculated from the dual solution  $\hat{u}$ , and  $\gamma_j^{(\hat{n})}$  is the activity coefficient calculated from the primal solution of the GEM IPM minimization problem. Both  $\omega_j^{(\hat{u})}$  and  $\gamma_j^{(\hat{n})}$  are taken in the mole fraction scale, and calculated depending on the phase type [59]. The  $\omega_j^{(\hat{u})}$  values are obtained, as derived from eqs. (4) to (6). 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 identifies subsequent actions of the GEM IPM solver (Table B.1).

784 The overall behavior of the GEM IPM algorithm strongly depends on the mode of the ini-785 tial approximation. In principle, error conditions can be generated at each numerical step shown 786 in Fig. 4. In the SIA mode, any such error leads to a complete re-start of GEM IPM calculation 787 using the AIA mode. In the AIA mode, two error conditions do not prevent yielding solutions 788 that are either incompletely consistent (at step K when six loops have been performed) or insuf-789 ficiently accurate (when MBR cannot converge at prescribed maximum number of iterations). In 790 these two cases, the respective warning codes are generated. All other errors result in returning 791 an error code, but yield no valid primal and dual solution.

792

#### 793 B.5. Comparison of new PSSC and old Selekt-2 procedures

The phase stability index  $\Omega_k$  (eqs. B-10, B-11) 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  $\Omega_k$  is directly related to the KKT conditions (eqs. 7c, 11), it can be compared with the Karpov's phase stability index  $f_{\alpha}$  [23, 24] used in the old Selekt-2 procedure, which in our notation is expressed as

799 
$$f_{\alpha,k} = \sum_{j \in L_k} \left[ \exp\left(\eta_j - \upsilon_j + \ln \hat{x}_j\right) \right] - 1, \quad j \in L_S \cap j \in D_0$$
(B-12)

800 where  $\hat{x}_j$  is the mole fraction of the *j*-th component in the *k*-th phase. This definition can be 801 compared with the stability index from eq. (B-11)

802 
$$f_{\alpha,k} = \Omega_k - 1$$
, or  $\Lambda_k = \log_{10}(f_{\alpha,k} + 1)$  (B-13)

There is also a numerical difference in the application of phase selection rules using empirical 803 804 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  $-\varepsilon_{\Phi^{-}} < \Lambda_{k} < \varepsilon_{\Phi^{+}}$ 805 or (eq. B-13)  $-\varepsilon_{\Phi_{-}} < \log_{10}(f_{\alpha,k} + 1) < \varepsilon_{\Phi_{+}}$ . This comparison shows that, for the same  $\varepsilon_{\Phi}$  threshold 806 807 values, the linear-scale criterion  $f_{\alpha}$  in Selekt-2 is more sensitive to dependent components present in large mole fractions  $\hat{x}_i$  in the phase, and less sensitive to those present in small mole fractions. 808 809 Conversely, the logarithmic-scale criterion used in PSSC is sensitive to primal activity coeffi-810 cients and mainly to the dual chemical potentials, but less to the mole fractions of dependent 811 components. In other words, the index  $\Lambda_k$  is to a less extent based on the primal solution  $\hat{n}^{(x)}$  and uses the dual solution  $\hat{u}$  of the GEM problem to greater extent than the old index  $f_{\alpha,k}$ . 812

813 This notion explains why the introduction of the new phase stability index eq. (B-10) and 814 the new PSSC algorithm, together with the modified Dikin's criterion of IPM convergence (eq. 815 B-9), has dramatically improved the robustness and internal consistency of the GEMS3K code, 816 compared to previous versions of GEMIPM2K and Selektor codes. In turn, this success suggests 817 a hypothesis that in the iterative numerical process of the IPM algorithm the error accumulation 818 mainly occurs in the primal solution approximation, rather than in the dual one. This idea forms 819 the basis of the SpeciationCleanup algorithm. There are, however, infrequent numerically ill-820 behaved cases, when the dual solution approximation becomes divergent, and dramatic error ac-821 cumulation or failure in MBR or IPM may occur.

822

#### 823 B.6. Calculation of non-ideal mixing models for solution phases

The substantial advantage of direct GEM algorithms lies in their ability to potentially handle any number of non-ideal 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 non-ideal 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 non-ideal mineral solid-solution phases with up to 9 endmembers, and additionally 48 pure solid phases.

832 The GEMS3K code can compute models of non-ideal mixing in two ways, either (i) by 833 interpreting phase model scripts (prepared in Phase records stored in the GEM-Selektor project 834 data base), or (ii) by executing built-in models of mixing, which are implemented on the basis of 835 the new TSolMod C++ class interface. The standalone GEMS3K code can execute any built-in 836 model, but usage of phase model scripts is not supported. The TSolMod class library with its 837 numerous models of mixing, as provided in GEM-Selektor and GEMS3K codes, is described in 838 detail in [59]. Regardless of the type of model (built-in or scripted), the execution occurs within 839 the CalculateActivityCoefficients() procedure, which has three modes that are summarized in 840 Table B.2.

841 In some highly non-ideal phases (for instance, reciprocal sublattice solid solutions), the 842 activities of end members can so strongly change with phase composition that this may disturb 843 the smooth convergence of MBR and Main IPM descent procedures. Because of the strong 844 coupling between the chemical potentials of end-members of different solid solution phases in 845 such systems, this may result in an oscillating behavior of the GEM IPM algorithm that prevents 846 the descent to the stable minimum Gibbs energy state. In such cases, the application of smooth-847 ing procedures may be necessary, although this may slow down the convergence of IPM and the 848 performance of the GEMS3K solver.

- 849
- 850
- 851

#### 852 B.7. Smoothing in systems with highly non-ideal phases

Through extensive numerical tests using systems with highly non-ideal 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. B-9) and the prescribed IPM convergence threshold  $\varepsilon_D$ . The newly implemented formula for calculation of the smoothing factor

858 
$$\alpha_{\gamma} = \alpha_{CD} = \varepsilon_{\delta} + \exp\left[\ln\left(1 - \varepsilon_{\alpha}\right) + \ln C_{D}\right] + \frac{\varepsilon_{\alpha} - \varepsilon_{\delta}}{1 + \exp\left(\ln \varepsilon_{D} - \ln C_{D}\right)/\varepsilon_{\delta}}$$
(B-14)

uses three empirical input constants, which are  $\varepsilon_D$ ,  $0 < \varepsilon_\alpha < 1$ , and  $0 < \varepsilon_\delta < 1$ . The constant  $\varepsilon_\delta$  sets the minimal value of  $\alpha_\gamma$  that it approaches when  $C_D$  becomes very small, and the constant  $\varepsilon_\alpha$  sets the position of the  $\alpha_\gamma$  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 re-calculation of activity coefficients  $\gamma_j^{(r+1)}$  subsequent to the *r*-th IPM iteration as follows. The approximations of primal chemical potentials  $\upsilon_j^{(r+1)}$  are computed from  $n^{(x,r+1)}$  and  $\gamma_j^{(r+1)}$ , using eq. (4). The deviations are then calculated as

866 
$$\delta_{j}^{(\nu,r)} = v_{j}^{(r+1)} - v_{j}^{(r)}, \ j \in L_{s}$$
 (B-15)

867 and compared with the empirical smoothing sensitivity threshold  $\varepsilon_{sm} = 1 \cdot 10^{-5}$ :

868 
$$\forall \left| \delta_{j}^{(\nu,r)} \right| \ge \varepsilon_{sm} : \upsilon_{j}^{(r+1)} = \upsilon_{j}^{(r)} + \alpha_{\gamma} \delta_{j}^{(\nu,r)}, \ j \in L_{S}$$
(B-16)

As seen from eq. (B-16), the smoothing suppresses only increments to chemical potentials of dependent components that are present in the primal solution approximations in non-zero amounts, and only when these increments are sufficiently large. The value of  $\varepsilon_{sm}$  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.

875 876

#### 877 B.8. The procedure for detection of divergence in the dual solution approximation

878 The revised GEM IPM algorithm and GEMS3K code include a new runtime uDD() procedure 879 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 toler-880 ances. The  $u_i^{(r)}$  values are accepted if they fall within an interval  $-600 \le u_i^{(r)} \le 400$  (for the elec-881 trical charge Z, a narrower interval  $-50 \le u_Z^{(r)} \le 100$  is checked). Note that actual ranges in well-882 883 posed systems are not so broad, and become much narrower with increasing temperature. For a 884 detailed detection of stiff behaviour, for each independent component, a moving average of its  $u_i$ 885 value for three consecutive IPM iterations is calculated, and its variation relative to previous 886 value at (r-1) is compared with a prescribed tolerance. If a substantial shift is detected, the warn-887 ing is issued, the entire calculation is re-started in AIA mode, and the result from IPM iteration 888 (r-2) is returned (the PSSC procedure is skipped in this case). In such a result, the primal and 889 dual IPM solutions are not internally consistent, but cannot be further improved because of the 890 absence of DCs in amounts sufficient to constrain a diverging *i*-th IC chemical potential  $u_i$ . If this 891 happens, the system bulk composition has to be re-defined such that certain redox buffering for 892 the *i*-th element is provided.

- 893
- 894
- 895
- 896

897

Tables to the Annex B.

#### 900

#### 901

902

Α

 Table B.1. Cases resulting after the PSSC procedure.

The best case when no phases have been inserted and no mass balance violations occurred upon phase elimination or speciation clean-up. It is then sufficient to run the final MBR step; recalculate activity coefficients; and release the primal  $\hat{n}^{(x,c)}$  and dual  $\hat{u}$  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,  $n^{(y)} = \hat{n}^{(x,c)}$  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, a maximum number of 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  $\hat{n}^{(x)}$  is returned, marked as incompletely consistent with the primal solution.

908 
 Table B.2. Modes of CalculateActivityCoefficients (LinkMode) operation.

LinkMode	Description	Where called from
ТР	Calculations of model parameters that depend on $T,P$ but not on the composition of phases	Start of the GEM IPM algorithm <sup>1</sup>
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 MainIPMDes- cent( $It$ ) iteration <sup>1</sup>
PP <sup>2</sup>	Calculation of integral non-ideal prop- erties of phases after obtaining a valid GEM IPM primal solution	After MainIPMDescent( $It$ ) has converged with valid primal and dual solutions <sup>1</sup>

<sup>1</sup>In GEM-Selector, also when reading SysEq records for displaying or sampling results;

<sup>2</sup>The PP (phase properties) mode will be operational in GEMS3K v.3.2 and higher. 

#### 914 **References**

- 915 1. Bale, C.W., Chartrand, P., Degterov, S.A., Eriksson, G., Hack, K., Ben Mahfoud, R., Melancon, J., Pelton,
  916 A.D., Petersen, S.: FactSage thermochemical software and databases. Calphad 26, 189-228 (2002)
- 917 2. Bale, C.W., Belisle, E., Chartrand, P., Degterov, S.A., Eriksson, G., Hack, K., Jung, I.H., Kang, Y.B., Me-
- 918lancon, J., Pelton, A.D., Robelin, C., Petersen, S.: FactSage thermochemical software and databases -919recent developments. Calphad 33, 295-311 (2009)
- Baumann, C., Gerya, T.V., Connolly, J.A.D.: Numerical modelling of spontaneous slab breakoff dynamics
   during continental collision. Geological Society of London Special Publications 332, 99-114 (2010)
- 922 4. Bethke, C. M.: Geochemical and Biogeochemical Reaction Modeling. Cambridge University Press, New
   923 York (2008)
- 5. Borisov, M.V., Shvarov, Y.V.: Thermodynamics of geochemical processes. Moscow State University Pub lishers, Moscow (1992), in Russian
- 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. Computers and Geosciences 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 (2010), in Russian
- 933 9. Chudnenko, K.V., Karpov, I.K., Kulik, D.A.: A high-precision IPM-2 non-linear minimization module of
   934 GEM-Selektor v.2-PSI program code for geochemical thermodynamic modeling. Technical Report
   935 TM-44-02-06, Paul Scherrer Institut, Villigen (2002)
- 936 10. Connolly, J.A.D., Petrini, K.: An automated strategy for calculation of phase diagram sections and retrieval
   937 of rock properties as a function of physical conditions. Journal of Metamorphic Geology 20, 697-708
   938 (2002)
- 11. Connolly, J.A.D.: Computation of phase equilibria by linear programming: A tool for geodynamic model ing 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 com plex geologic media. Computers and Geosciences 34, 1697-1707 (2008)
- 943 13. de Capitani, C., Brown, T.H.: The computation of chemical equilibrium in complex systems containing
  944 non-ideal solutions. Geochimica et Cosmochimica Acta 51, 2639-2652 (1987)
- 945 14. de Capitani, C., Petrakakis, K.: The computation of equilibrium assemblage diagrams with The 946 riak/Domino software. American Mineralogist **95**, 1006-1016 (2010)
- 947 15. Dolejs, D., Wagner, T.: Thermodynamic modeling of non-ideal mineral-fluid equilibria in the system Si 948 Al-Fe-Mg-Ca-Na-K-H-O-Cl at elevated temperatures and pressures: Implications for hydrothermal
   949 mass transfer in granitic rocks. Geochimica et Cosmochimica 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. Journal of Computational Chemistry 21, 247-256 (2000)
- 952 17. Engesgaard, P., Kipp, K.L.: A geochemical transport model for redox-controlled movement of mineral
  953 fronts in groundwater-flow systems a case of nitrate removal by oxidation of pyrite. Water Resour.
  954 Res. 28, 2829–2843 (1992)
- 18. Eriksson, G., Hack, K.: Chemsage A computer program for the calculation of complex chemical equili bria. Metallurgical and Materials Transactions **B21**, 1013-1023 (1990)
- 957 19. Eriksson, G., Thompson, W.T.: A procedure to estimate equilibrium concentrations in multicomponent
  958 systems and related applications. Calphad 13, 389-400 (1989)
- 959 20. Ghiorso, M.S.: Algorithms for the estimation of phase stability in heterogeneous thermodynamic systems.
  960 Geochimica et Cosmochimica Acta 58, 5489-5501 (1994)
- 961 21. GNU free software foundation: GNU Lesser General Public License. Free Software Foundation, Boston.
   962 http://www.gnu.org/licenses/lgpl.html (2007)
- 963 22. Hammond, G., Lichtner, P., Lu, C.: Subsurface multiphase flow and multicomponent reactive transport
   964 modeling using high-performance computing. Journal of Physics: Conference Series 78, 012025, p.1 965 10 (2007)
- 966 23. Karpov, I.K.: Computer-aided physico-chemical modeling in geochemistry. Nauka Publ., Novosibirsk
   967 (1981), in Russian
- 968 24. Karpov, I.K., Chudnenko, K.V., Kulik, D.A.: Modeling chemical mass-transfer in geochemical processes:
   969 Thermodynamic relations, conditions of equilibria and numerical algorithms. American Journal of
   970 Science 297, 767-806 (1997)
- Skarpov, 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. Geochemistry International **39**, 11081119 (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. American Journal
  of Science 302, 281-311 (2002)
- 27. Keizer, M.G., Van Riemsdijk, W.H.: ECOSAT. Technical report. Department Soil Science and Plant Nu trition, Wageningen Agricultural University, Wageningen, The Netherlands (1998)
- 979 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 Heidelberg (2012)
- 30. Kolditz O, Bauer S, Bilke L, N. Böttcher, J.O. Delfs, T. Fischer, U.J. Görke, T. Kalbacher, G. Kosakowski,
  C.I. McDermott, C.H. Park, F. Radu, K. Rink, H. Shao, H.B. Shao, F. Sun, Y.Y. Sun, A.K. Singh, J.
  Taron, M. Walther, W. Wang, N. Watanabe, Y. Wu, M. Xie, W. Xu, B. Zehner (2012): OpenGeoSys:
  an open-source initiative for numerical simulation of thermo-hydro-mechanical/chemical (THM/C)
  processes in porous media, Environ. Earth Sci., 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.
  Paul Scherrer Institut, Villigen, pp. 109-122 (2004)
- 32. Kulik, D.A.: Classic adsorption isotherms incorporated in modern surface complexation models: Implica tions for sorption of actinides. Radiochimica 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, v.11, Elsevier, Amsterdam, 171-250 (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 ele ment numerical modeling of mass and heat transport in fractured rocks with matrix diffusion. Compu tational Geosciences 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 in fluence on seismic anomalies in Earth's mantle. Geochemistry Geophysics Geosystems (G<sup>3</sup>) 10,
   Q03004 (2009)
- 1007 38. Parkhurst, D.L., Appelo, C.A.J.: User's guide to PHREEQC (version 2) A computer program for specia 1008 tion, batch-reaction, one-dimensional transport, and inverse geochemical calculations. U.S.G.S. Wa 1009 ter-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. Nuclear Technology 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)
- 1014 41. Prommer, H.: A reactive multicomponent transport model for saturated porous media, User's manual Ver1015 sion 1.0, Contaminated land assessment and remediation research centre. The University of Edin1016 burgh, UK (2002)
- 42. Rastetter, E.B.: Modeling coupled biogeochemical cycles. Frontiers in Ecology and the Environment 9, 6873 (2011)
- 43. Reed, M.H.: Calculation of multicomponent chemical equilibria and reaction processes in systems involving minerals, gases and an aqueous phase. Geochimica et Cosmochimica Acta 46, 513-528 (1982)
- 44. Reed, M.H., Calculation of simultaneous chemical equilibria in aqueous-mineral-gas systems and its appli cation to modeling hydrothermal processes. Reviews in Economic Geology 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. Geochimica et Cosmochimica Acta 48,
  1025 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. Applied Geochemistry 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. Geochemical Journal 43, e37-e42 (2009)
- 48. Shvarov, Y.V.: A general equilibrium criterion for an isobaric-isothermal model of a chemical system.
  Geochemistry International 18, 38-45 (1981)
- 49. Shvarov, Y.: A numerical criterion for existence of the equilibrium state in an open chemical system.
  Sciences Geologique Bulletin 42, 365-369 (1989)
- 50. Shvarov, Y.V.: HCh: New potentialities for the thermodynamic simulation of geochemical systems offered
  by windows. Geochemistry International 46, 834-839 (2008)
- 1037 51. Singh A. K., Goerke U. -J., Kolditz O.: Numerical simulation of non-isothermal compositional gas flow:
  1038 Application to carbon dioxide injection into gas reservoirs. Energy 36, 3446-3458 (2011)
- 1039 52. Siret, D., Poulet, T., Regenauer-Lieb, K., Connolly, J.A.D.: PreMDB, a thermodynamically consistent ma 1040 terial database as a key to geodynamic modelling. Acta Geotechnica 4, 107-115 (2009)
- 1041 53. Steefel, C.I., DePaolo, D.J., Lichtner, P.C.: Reactive transport modeling: An essential tool and a new re-1042 search approach for the Earth sciences. Earth and Planetary Science Letters **240**, 539-558 (2005)
- 1043 54. Tenzer, H., Park C.L., Kolditz O., McDermott, C.I.: Application of the geomechanical facies approach and
   1044 comparison of exploration and evaluation methods used at Soultz-sous-Forets (France) and Spa Urach
   1045 (Germany) geothermal sites. Environmental Earth Sciences 61, 853-880 (2010)
- 1046 55. Trefethen, L.N., Bau, D.: Numerical linear algebra. SIAM (1997)
- 1047 56. Tromp, T. K., Van Cappellen, P., Key R. M.: A global model for the early diagenesis of organic carbon
  1048 and organic phosphorous in marine sediments. Geochimica et Cosmochimica Acta 59, 1259–1284
  1049 (1995)
- 57. Van der Lee, J., De Windt, L.: Present state and future directions of modeling of geochemistry in hydro geological systems. Journal of Contaminant Hydrology 47, 265-282 (2001)
- 1052 58. Van der Lee, J., De Windt, L., Lagneau, V., Goblet, P.: Module-oriented modeling of reactive transport
  1053 with HYTEC. Computers and Geosciences 29, 265-275 (2003)
- 1054 59. Wagner, T., Kulik, D.A., Hingerl, F.F., Dmytrieva, S.V.: GEM-Selektor geochemical modeling package:
  1055 TSolMod library and data interface for multicomponent phase models. Canadian Mineralogist (2012),
  1056 to be accepted with minor revisions.
- 1057 60. Wang, W., Kosakowski, G., Kolditz, O.: A parallel finite element scheme for thermo-hydro-mechanical
  1058 (THM) coupled problems in porous media. Computers and Geosciences 35, 1631-1641 (2009)
- 1059 61. Wang, Y., Van Cappellen, P.: A multicomponent reactive transport model of early diagenesis: Application
   1060 to redox cycling in coastal marine sediments. Geochimica et Cosmochimica Acta 60, 2993-3014.
   1061 (1996)

- 1062 62. Westall, J.C., Zachary, J.L., Morel, F.M.M.: MINEQL: A compact program for computation of chemical
  1063 equilibria in aquatic systems. R.M. Parsons Laboratory for Water Resources and Hydrodynamics,
  1064 Massachusetts Institute of Technology, Cambridge, MA (1976)
- 1065 63. Xie, M., Kolditz, O., Moog, H.C.: A geochemical transport model for thermo-hydro-chemical (THC) cou 1066 pled processes with saline water. Water Resources Research 47, W02545 (2011)
- 64. Kosakowski, G., Kulik, D.A., Shao, H.: OpenGeoSys-GEMS: Hybrid parallelization of a reactive transport
  code with MPI and threads, Geophysical Research Abstracts, 14, EGU2012-2642, (2012)
- 1069

### 1074 Tables

# 1076

Code type	Method	SLE dim.	N of loops	Costs, Flop
LMA	Newton-Raphson method to mi- nimize MBRs + Jacobian inversion to solve for LMAs	n×n	$l = \max(1, \Phi - \mathbf{n})$	$\sim 4l \cdot n^3$
GEMS3K	Operator splitting between GEM and MBR steps (IPM)	n×n	<i>n</i> ( <i>r</i> ) avg. 60 (AIA) or 6 (SIA)	$\sim 2/3 \cdot n(r) \cdot n^3$
GEM Implicit	Simultaneously minimizes Gibbs energy and MBRs (Newton me- thod)	2n×2n	<i>n</i> ( <i>r</i> ) avg. 60	$> 8/3 \cdot n(r) \cdot n^3$
GEM +LMA	Operator splitting between GEM and multiple LMA steps	n×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 compo- sition discretization	$n \times 10^4 n$	n(r) = 1, l < 5	$\sim l \cdot n \cdot 10^4$
LP-based (THERIAC)	Operator splitting between LP simplex and phase-solution compo- sition 'equilibration' by GEM	n×9n	n(r) = 15, l < 100, $n_f = 4$	$\frac{\sim l \cdot \mathbf{n} \cdot}{(n(r) \cdot (\mathbf{n}_{\rm f})^3 + 100)}$

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

MBRs: mass balance residuals.

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

1082

Functionality GEMS3K in standalone codes GEMS3K in GEM-Selektor-3 Phase- and Process Not possible Used scripted models Built-in TSolMod Used Used class functions for activity coefficients *T*-*P* corrections for Lagrangian interpolation using 2-D Direct access to default or project input thermodynamic lookup arrays (exported in DCH thermodynamic data base and builtfile from GEMS3) in thermodynamic and EoS P,T cordata rection functions Access to input, output Batch mode only (data exchange Direct interactive GUI access, inand work data via I/O files or in RAM) cluding multithread stepwise calculation mode; many dialogs and data access screen forms Runtime help and tool-Not available Available as runtime local HTML tips help system, with some enhancements on GEMS web site

1083 1084

1085

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 ex- changed 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 set- tings, parameters of non-ideal mixing models, adsorption models, etc.)

1089 Note: DCH and DBR files are automatically (un)packed to/from the MULTI (IPM work data) memory

1090 structure using functions provided by the TNode class.

## $\textbf{Table 4. Modes of GEM IPM initial approximation and required \texttt{NodeStatusCH} actions.}$

1094

Before calling GEM run: set 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 sub-system 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.			
<b>Before calling GEM run: set</b> 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 re-starting GEM run for this node in the AIA mode.			

#### 1099 Figure Captions

1100

1101 Figure 1. Levels of complexity in RMT simulations in terms of the number of control volumes

1102 (nodes) and the number of time steps.

1103 **Figure 2.** Schematic concept of coupled RMT modeling using the GEMS chemical solver.

1104 Figure 3. Diagrammatic relations between GEM-Selektor and GEMS3K codes. Acronyms:

1105 MPDB: modeling projects data base; TDB: default thermodynamic data base (provided); HDB:

1106 help data base (provided); GUI: graphical user's interface and database management code; GEM

- 1107 IPM: GEM IPM algorithm code implementation; THMA: thermo-hydraulic-mass-transport algo-
- 1108 rithm code; TNode: interface connecting GEM IPM kernel with existing MT code and with I/O

1109 (input/output) files containing the exported definition of chemical systems, as well as the output

1110 in VTK format; TSolMod: library of classes for mixing models of phases-solutions [59]; TNo-

1111 deArray: extension of TNode interface for developing new coupled RMT codes.

1112 Figure 4. Simplified flow chart of the revised GEM IPM algorithm, as implemented in the

- 1113 GEMS3K code. See text for explanations.
- 1114 **Figure 5**. Model domain, transport boundary, and initial conditions [29] for the RMT example.
- 1115 Figure 6. Benchmark results from OGS-ChemApp (triangles), OGS- PHREEQC (solid lines),
- and OGS-GEMS (crosses) coupled simulations [29]. Temperature: 25 °C, pressure: 1 bar.

	Complexity	N volumes	N steps
1	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 <sup>6</sup>
	2-D and 3-D realistic TH(M)C simulation	> 105	> 106

Figure 1.



Figure 2.



Figure 3.



Figure 4.







Figure 6.