ThermoCodegen: a python/C++ package for the generation of custom thermodynamic models

Marc Spiegelman [1,2,¶], **Owen Evans**[*1,3], **Mark Ghiorso** [4], **Lucy Tweed** [2], and **Cian Wilson** [5]

**1** Dept. of Applied Physics and Applied Math, Columbia University **2** Dept. of Earth and Environmental Science, Columbia University **3** Now at Google **4** OFM Research **5** Earth and Planets Laboratory, Carnegie Institution for Science ¶ Corresponding author

Summary

Many important problems in solid-Earth science require a consistent coupling between thermodynamics and geodynamics. Examples include global mantle convection with thermodynamically consistent phase changes, the generation and flow of magma in the mantle, reactive fluid flow and geological carbon sequestration in the crust and even the coupled flow of water and ice in glaciers and ice sheets. Given the complexity of these systems, it is important that the individual user/scientist have considerable control over the formulation and computational choices involved for both the thermodynamic and geodynamic components of models while maintaining transparency and reproducibility. Access to flexible open-source software for the creation and exploration of thermodynamic models and their incorporation into large scale reactive flow simulations is essential for these endeavors.

Statement of need

ThermoCodegen is a software package for the reproducible generation of custom thermodynamic and kinetic reaction models for minerals, melts and some aqueous fluids commonly encountered in the solid-Earth sciences. It is part of the larger ENKI project for thermodynamic modeling whose flagship code ThermoEngine (Ghiorso & Wolf, 2019; Wolf & Ghiorso, 2019) provides a JupyterHub, cloud-based environment for the interactive generation and exploration of thermodynamic models in Earth sciences. In particular, ThermoEngine provides a critical code-generation module, `coder.py` for generating high-performance C code from SymPy expressions for the free energies of thermodynamic endmembers and phases.

ThermoCodegen shares the same underlying code-generation infrastructure as ThermoEngine however its design is driven more by the need for ready incorporation of thermodynamics into large scale geodynamics simulation codes for both equilibrium and disequilibrium transport. Key additional features of ThermoCodegen include:

- *Portable model storage*: Thermodynamic model generation for endmembers and phases are described in detailed Jupyter notebooks using coder.py and SymPy, however, the final models are stored in structured xml files that contain all the information required to describe and auto-generate code for thermodynamic databases of endmembers and phases.
- *Kinetic modeling*: ThermoCodegen also provides code generation and storage of custom kinetic reaction models built on the phase databases.

---

*co-first author

- *C++ support*: The code also generates C++ libraries for inclusion in other modeling systems such as TerraFERMA (Wilson et al., 2017; Wilson & Spiegelman, 2015) or ASPECT (Bangerth et al., 2021; Heister et al., 2017; Kronbichler et al., 2012).
- *Python bindings*: Python bindings of the C++ classes are generated using pybind11 (Wenkel et al., 2017) for direct import into python applications.
- *ThermoEngine support*: Python bindings of ThermoCodegen objects can also be imported back into ThermoEngine as a custom database for use in equilibration or calibration calculations (requires installing the `tcg_model_integration` branch of ThermoEngine).

Relationship to other Software packages for Solid-Earth Thermodynamic modeling

Computational thermodynamic software has been a central tool for solid-Earth science for over 40 years and has focused primarily on calculating phase equilibrium for several distinct self-consistent "Thermodynamic Datasets" such as those for solid mineral phases (Berman, 1988; Holland & Powell, 1998, 2011; Stixrude & Lithgow-Bertelloni, 2005, 2011), silicate liquids (Ghiorso et al., 2002; Ghiorso & Sack, 1995; Holland et al., 2018), and ionic species in aqueous solutions (Shock et al., 1992; Tanger & Helgeson, 1988).

Many of these datasets are associated with specific software for calculating equilibrium assemblages using a range of Gibbs free energy minimization algorithms. Some of the most commonly used software includes the MELTs family of models (Asimow & Ghiorso, 1998; Ghiorso et al., 2002; Ghiorso & Gualda, 2015; Ghiorso & Sack, 1995; Gualda et al., 2012) for equilibration of silicate minerals and melts, which has also been repackaged as alphaMELTS (Antoshechkina & Ghiorso, 2018; Smith & Asimow, 2005) as well as in its current incarnation within the ENKI project. Primarily used for metamorphic petrology, THERMOCALC[1] (Powell et al., 1998) is closely associated with the thermodynamic models of Holland and Powell (Holland & Powell, 1998, 2011), while HeFESTo is used for the calculation of phase equilibria of the high pressure/temperature solid-state models of Stixrude and Lithgow-Bertelloni (Stixrude & Lithgow-Bertelloni, 2005, 2011). Note that not all of these packages have open-source licenses. In addition, there are several more general Gibbs free energy minimization packages that can be used with multiple data sets. Perple_X (Connolly, 1990, 2005, 2009) is one of the most widely used. More recent additions include THERIAK-DOMINO (Capitani & Petrakakis, 2010) which also supports the models of Holland and Powell, Thermolab (Vrijmoed & Podladchikov, 2022) a MATLAB based toolkit, and MAGEMin (Riel et al., 2022) an open-source Gibbs energy minimization package written in C (with julia and MATLAB bindings) for adaptive phase-diagram generation with a range of modern optimization techniques. In particular, MAGEMin includes the igneous thermodynamic database of Holland et al. (2018).

These packages provide critical thermodynamic modeling infrastructure for solid-Earth science but are complementary and distinct in design and applicability to ThermoCodegen. In particular, the primary features added by ThermoCodegen are:

- the ability to more easily create and explore custom, consistent models of all thermodynamic parameters through the use of symbolic differentiation and automatic code generation
- the ability to generate code for disequilibrium reaction kinetics
- the ability to create custom C++ and python libraries for direct inclusion in geodynamics models.

Given the complexity of these high-dimensional open-system problems, having effective open-source software for reproducible exploration of both equilibrium phase diagrams and disequilibrium transport remains a central challenge and goal in geodynamics and ThermoCodegen is designed to address several of these critical challenges.

---

[1]Not to be confused with the commercial package Thermo-Calc

Software design/overview

The design of ThermoCodegen follows the hierarchical structure of thermodynamic modeling commonly used in Earth Sciences and illustrated in Figure 1.
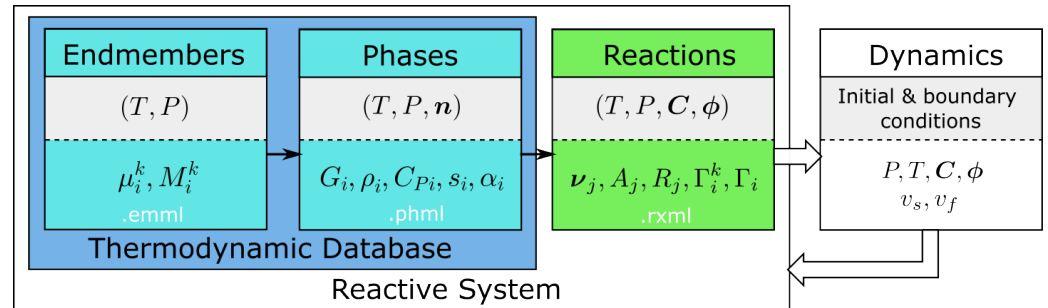
Figure 1: The hierarchical structure of thermodynamic modeling. μ_i^k is an endmember's chemical potential model and M_i^k its mass. G_i is a phase's Gibbs free energy model and ρ_i, C_{Pi}, s_i and α_i are (for example) its density, heat capacity, specific entropy, and thermal expansivity, derived consistently by symbolic differentiation of G_i. ν_j describes the stoichiometry of reactions between the endmembers of the phases, A_j the resulting reaction affinities, R_j their rates, and Γ_i^k and Γ_i the resulting endmember and phase source terms. The Thermodynamic Database and Reactive System are functions of the temperature, T, pressure P, the number of moles, n, or mass fractions, C, of the endmembers, and presence (e.g. volume fraction) of the phases, ϕ. These can be supplied by coupling to a dynamic model that may additionally calculate phase velocities, e.g., v_s and v_f. Image adapted from Tweed (2021).

At the lowest level are thermodynamic **Endmembers** which provide chemical potentials, $\mu_i^k(T, P)$, for stoichiometric chemical components. The actual components can be arbitrary (e.g., oxides, minerals, ionic species) but have fixed chemical composition and thermodynamic properties that only depend on temperature and pressure (or temperature and volume, for models of the Helmholtz free energy).

Given a space of thermodynamic endmembers, we can construct homogeneous **Phases** as combinations of endmembers and develop thermodynamic models for their Gibbs free energy, $G_i(T, P, n_i)$, that depend on temperature, pressure and the number of moles of each endmember $n_i = n_i^k$.

Given $G_i(T, P, n_i)$ for each phase, all of the other thermodynamic properties (and their derivatives) can be derived by partial differentiation of G_i. For example, entropy S_i, volume V_i, chemical potentials $\bar{\mu}_i^k$, heat capacity C_{Pi} and thermal expansivity α_i are all related to G_i by:

$$S_i = -\frac{\partial G_i}{\partial T} \qquad V_i = \frac{\partial G_i}{\partial P} \qquad \bar{\mu}_i^k = \frac{\partial G_i}{\partial n_i^k}$$
$$C_{Pi} = T\frac{\partial S_i}{\partial T} = -T\frac{\partial^2 G_i}{\partial T^2} \qquad \alpha_i = \frac{1}{V_i}\frac{\partial V_i}{\partial T} = \frac{1}{V_i}\frac{\partial^2 G_i}{\partial P^2}$$

Thus if we can define an analytic model for $G_i(T, P, n_i)$ for any phase, we can use symbolic differentiation and code-generation to provide custom routines to generate all thermodynamic properties consistently.

We define a set of phases, together with their constituent thermodynamic endmembers as a **Thermodynamic Database** corresponding to the blue box in Figure 1. Standard databases available in the literature include the mineral database of Berman (1988) and its extension to silicate liquids through the MELTs family of models (Ghiorso et al., 2002; Ghiorso & Sack, 1995), as well as the high-pressure mineral phases of Stixrude and Lithgow-Bertelloni (Stixrude & Lithgow-Bertelloni, 2005, 2011) or metamorphic phases of, for example, Holland & Powell (2011). All of these are currently available through python bindings to ObjectiveC code in

ThermoEngine. However, the ObjectiveC objects are not readily incorporated within other codes.

ThermoEngine also provides support for generating custom thermodynamic models of endmembers and phases through the `thermoengine.coder` module written by Mark Ghiorso. `coder` uses SymPy to describe the free energies of endmembers and phases and then uses automatic differentiation and code-generation to write efficient C source code (and wrap it in Cython) for all thermodynamic properties of phases. Our extensions in `thermocodegen.coder` generate the same underlying C code and interfaces as `thermoengine.coder` but extend its functionality in several critical ways for better integration with large-scale geodynamic modeling.

ThermoCodegen

ThermoCodegen is designed to produce code for just the colored boxes in Figure 1.

Custom Thermodynamic Databases

For **Thermodynamic Databases** of endmembers and phases, ThermoCodegen extends `thermoengine.coder` to record the internal state of each thermodynamic model as a python dictionary in order to store and generate code from xml files that completely describe all variables, parameters and SymPy models for endmembers and phases. These xml files are parsed using the science neutral options package Spud (Ham et al., 2009). Endmembers are stored with the extension `.emml` and phases are stored as `.phml` files, which can be viewed with the Spud GUI `diamond`. Given a set of consistent endmember and phase description files, the ThermoCodegen script `tcg_builddb` can be used to inspect, validate and autogenerate code for a custom Thermodynamic Database as well as provide C++ wrappers and python bindings using pybind11.

At this point, all source code, together with `.emml` and `.phml` files describing the full thermodynamic models, and optionally compiled C++ and python libraries are made available as a compressed tar-ball. This object is considered a Thermodynamic Database and can be used locally or made available as a web-accessible URL, for example with a zenodo DOI.

Given a custom thermodynamic database, these can then be used in optimization codes such as `thermoengine.equilibrate` to find equilibrium assemblages that minimize the global Gibbs free energy.

Reaction Kinetic Objects

For better integration with open-system geodynamics codes that do not assume equilibrium, ThermoCodegen also provides code for generating C++ libraries and python bindings for custom kinetics models that describe a set of reactions between endmembers and phases within a given thermodynamic database. The software provides interfaces for calculating affinities, reaction rates, and thermodynamic properties of phases as well as their exact derivatives for use in non-linear solvers that take Jacobians. These reaction kinetic models are stored as xml files with the extension `.rxml` from which custom C++ libraries and python bindings are generated using the build script `tcg_buildrx`.

These reaction libraries, along with the underlying database objects can then be incorporated in external modeling software such as ThermoEngine, TerraFERMA, ASPECT, Jupyter notebooks and general python or C++ codes. Inclusion in dynamic models is considered external to ThermoCodegen, however, the ThermoCodegen distribution also provides containers that include a complete installation of TerraFERMA along with some example test cases.

Detailed documentation for ThermoCodegen is available that includes a more detailed description of the software design as well as instructions for using the package through the provided docker and singularity containers (local installation guidelines are also available). Most importantly, the documentation also includes fully worked out examples illustrating the

work flow for using ThermoCodegen including a set of Jupyter notebooks for constructing and exploring a range of thermodynamic models.

Acknowledgements

References

Antoshechkina, P. M., & Ghiorso, M. S. (2018). *MELTS for MATLAB: A new Educational and Research Tool for Computational Thermodynamics*. *2018*, ED44B–23. https://ui.adsabs.harvard.edu/abs/2018AGUFMED44B..23A

Asimow, P. D., & Ghiorso, M. S. (1998). Algorithmic modifications extending MELTS to calculate subsolidus phase relations. *American Mineralogist*, *83*(9-10), 1127–1132. https://doi.org/10.2138/am-1998-9-1022

Bangerth, W., Dannberg, J., Fraters, M., Gassmoeller, R., Glerum, A., Heister, T., & Naliboff, J. (2021). *ASPECT v2.3.0*. Zenodo. https://doi.org/10.5281/ZENODO.5131909

Berman, R. G. (1988). Internally-Consistent Thermodynamic Data for Minerals in the System Na2O-K2O-CaO-MgO-FeO-Fe2O3-Al2O3-SiO2-TiO2-H2O-CO2. *Journal of Petrology*, *29*(2), 445–522. https://doi.org/10.1093/petrology/29.2.445

Capitani, C. de, & Petrakakis, K. (2010). The computation of equilibrium assemblage diagrams with Theriak/Domino software. *American Mineralogist*, *95*(7), 1006–1016. https://doi.org/10.2138/am.2010.3354

Connolly, J. A. D. (1990). Multivariable phase diagrams; an algorithm based on generalized thermodynamics. *American Journal of Science*, *290*(6), 666–718. https://doi.org/10.2475/ajs.290.6.666

Connolly, J. A. D. (2005). Computation of phase equilibria by linear programming: A tool for geodynamic modeling and its application to subduction zone decarbonation. *Earth and Planetary Science Letters*, *236*(1-2), 524–541. https://doi.org/10.1016/j.epsl.2005.04.033

Connolly, J. A. D. (2009). The geodynamic equation of state: What and how. *Geochemistry, Geophysics, Geosystems*, *10*(10), Q10014. https://doi.org/10.1029/2009GC002540

Ghiorso, M. S., & Gualda, G. A. R. (2015). An H2O–CO2 mixed fluid saturation model compatible with rhyolite-MELTS. *Contributions to Mineralogy and Petrology*, *169*(6), 53. https://doi.org/10.1007/s00410-015-1141-8

Ghiorso, M. S., Hirschmann, M. M., Reiners, P. W., & Kress, V. C. (2002). The pMELTS: A revision of MELTS for improved calculation of phase relations and major element partitioning related to partial melting of the mantle to 3 GPa. *Geochemistry, Geophysics, Geosystems*, *3*(5), 1–35. https://doi.org/10.1029/2001GC000217

Ghiorso, M. S., & Sack, R. O. (1995). Chemical mass transfer in magmatic processes IV. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid-solid equilibria in magmatic systems at elevated temperatures and pressures. *Contributions to Mineralogy and Petrology*, *119*(2-3), 197–212. https://doi.org/10.1007/BF00307281

Ghiorso, M. S., & Wolf, A. S. (2019). *Thermodynamic Modeling Using ENKI: 1. Overview and Phase Equilibrium Applications*. *2019*, ED53F–0897. https://ui.adsabs.harvard.edu/abs/2019AGUFMED53F0897G

Gualda, G. A. R., Ghiorso, M. S., Lemons, R. V., & Carley, T. L. (2012). Rhyolite-MELTS: A Modified Calibration of MELTS Optimized for Silica-rich, Fluid-bearing Magmatic Systems. *Journal of Petrology*, *53*(5), 875–890. https://doi.org/10.1093/petrology/egr080

Ham, D. A., Farrell, P. E., Gorman, G. J., Maddison, J. R., Wilson, C. R., Kramer, S. C., Shipton, J., Collins, G. S., Cotter, C. J., & Piggott, M. D. (2009). Spud 1.0: Generalising and automating the user interfaces of scientific computer models. *Geoscientific Model Development*, *2*(1), 33–42. https://doi.org/10.5194/gmd-2-33-2009

Heister, T., Dannberg, J., Gassmöller, R., & Bangerth, W. (2017). High accuracy mantle convection simulation through modern numerical methods – II: Realistic models and problems. *Geophysical Journal International*, *210*(2), 833–851. https://doi.org/10.1093/gji/ggx195

Holland, T. J. B., Green, E. C. R., & Powell, R. (2018). Melting of Peridotites through to Granites: A Simple Thermodynamic Model in the System KNCFMASHTOCr. *Journal of Petrology*, *59*(5), 881–900. https://doi.org/10.1093/petrology/egy048

Holland, T. J. B., & Powell, R. (1998). An internally consistent thermodynamic data set for phases of petrological interest. *Journal of Metamorphic Geology*, *16*(3), 309–343. https://doi.org/10.1111/j.1525-1314.1998.00140.x

Holland, T. J. B., & Powell, R. (2011). An improved and extended internally consistent thermodynamic dataset for phases of petrological interest, involving a new equation of state for solids. *Journal of Metamorphic Geology*, *29*(3), 333–383. https://doi.org/10.1111/j.1525-1314.2010.00923.x

Kronbichler, M., Heister, T., & Bangerth, W. (2012). High accuracy mantle convection simulation through modern numerical methods: High accuracy mantle convection simulation. *Geophysical Journal International*, *191*(1), 12–29. https://doi.org/10.1111/j.1365-246X.2012.05609.x

Powell, R., Holland, T., & Worley, B. (1998). Calculating phase diagrams involving solid solutions via non-linear equations, with examples using THERMOCALC. *Journal of Metamorphic Geology*, *16*(4), 577–588. https://doi.org/10.1111/j.1525-1314.1998.00157.x

Riel, N., Kaus, B. J. P., Green, E. C. R., & Berlie, N. (2022). MAGEMin, an Efficient Gibbs Energy Minimizer: Application to Igneous Systems. *Geochemistry, Geophysics, Geosystems*, *23*(7). https://doi.org/10.1029/2022GC010427

Shock, E. L., Oelkers, E. H., Johnson, J. W., Sverjensky, D. A., & Helgeson, H. C. (1992). Calculation of the thermodynamic properties of aqueous species at high pressures and temperatures. Effective electrostatic radii, dissociation constants and standard partial molal properties to 1000°C and 5 kbar. *Journal of the Chemical Society, Faraday Transactions*, *88*(6), 803–826. https://doi.org/10.1039/FT9928800803

Smith, P. M., & Asimow, P. D. (2005). Adiabat_1ph: A new public front-end to the MELTS, pMELTS, and pHMELTS models. *Geochemistry, Geophysics, Geosystems*, *6*(2). https://doi.org/10.1029/2004GC000816

Stixrude, L., & Lithgow-Bertelloni, C. (2005). Thermodynamics of mantle minerals - I. Physical properties. *Geophysical Journal International*, *162*(2), 610–632. https://doi.org/10.1111/j.1365-246X.2005.02642.x

Stixrude, L., & Lithgow-Bertelloni, C. (2011). Thermodynamics of mantle minerals - II. Phase equilibria. *Geophysical Journal International*, *184*(3), 1180–1213. https://doi.org/10.1111/j.1365-246X.2010.04890.x

Tanger, J. C., & Helgeson, H. C. (1988). Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures; revised equations of

state for the standard partial molal properties of ions and electrolytes. *American Journal of Science*, *288*(1), 19–98. https://doi.org/10.2475/ajs.288.1.19

Tweed, L. E. L. (2021). *Coupling the Thermodynamics, Kinetics and Geodynamics of Multiphase Reactive Transport in Earth's Interior* [PhD thesis, Columbia University]. https://doi.org/10.7916/d8-wh72-vb08

Vrijmoed, J. C., & Podladchikov, Y. Y. (2022). Thermolab: A Thermodynamics Laboratory for Nonlinear Transport Processes in Open Systems. *Geochemistry, Geophysics, Geosystems*, *23*(4). https://doi.org/10.1029/2021GC010303

Wenkel, J., Rhinelander, J., & Moldovan, D. (2017). *pybind11 – Seamless operability between C++11 and Python*. https://github.com/pybind/pybind11

Wilson, C. R., & Spiegelman, M. (2015). *Software: TerraFERMA v1.0 — The Transparent Finite Element Rapid Model Assembler, open-source repository: terraferma.github.io*. terraferma.github.io

Wilson, C. R., Spiegelman, M., & Keken, P. E. van. (2017). TerraFERMA: The Transparent Finite Element Rapid Model Assembler for multiphysics problems in Earth sciences. *Geochemistry, Geophysics, Geosystems*, *18*(2), 769–810. https://doi.org/10.1002/2016GC006702

Wolf, A. S., & Ghiorso, M. S. (2019). *Thermodynamic Modeling Using ENKI: 2. Extending Existing Models*. *2019*, ED52A–05. https://ui.adsabs.harvard.edu/abs/2019AGUFMED52A.05W