

2 概要

GeSbTe は“相変化材料”として知られている。低温でアモルファス状態の GeSbTe は、温度を上げることによって結晶化させることができる。その後さらに温度を上げ、急冷させるとアモルファス状態に戻すことができる。このような特性から、記憶媒体など様々な用途で応用がなされている。

本件では GeSbTe のニューラルネットワークポテンシャル (neural-network potential, NNP) を作成した。液相・アモルファス相・立方晶の GeSbTe 系が記述可能で, melt quench シミュレーションを行うことができ、アモルファス相から結晶に転移する様子がとらえられるポテンシャルが得られることを目標とした。

本稿ではどのようにニューラルネットワークポテンシャルを作成したか説明する。また、ニューラルネットワークポテンシャルを用いた分子動力学シミュレーションを実行する方法について具体的に解説する。

ニューラルネットワークポテンシャルについて

本件で作成したニューラルネットワークポテンシャルについて簡単に説明する。詳しくは文献[1], [2]などを参照されたい。

ニューラルネットワークポテンシャルとは、その名が示す通り原子間ポテンシャルをニューラルネットワークによって記述するものである。ニューラルネットワークとは入力層・任意の数の隠れ層・出力層からなるネットワークである。各層は任意の数のノードからなり、さらにバイアス層が紐づけられている。 i 番目の層における j 番目のノードの出力 y_i^j は、 $i-1$ 番目の層の k 番目のノードの出力 y_k^{j-1} を入力として以下のように計算される。

$$y_i^j = f_i^j \left(b_i^j + \sum_{k=1}^{N_{j-1}} w_{k,i}^{j-1,j} y_k^{j-1} \right)$$

ここで $w_{k,i}^{j-1,j}$ は $j-1$ 番目の層の k 番目のノードと j 番目の層の i 番目のノードとの間の重み係数、 b_i^j は j 番目の層の i 番目のノードに対するバイアス層からの寄与である。 $w_{k,i}^{j-1,j}$ や b_i^j が教師データを再現するよう学習によって決める量である。 f_i^j は j 番目の層が i 番目のノードにもたらす作用を表す activation function と呼ばれる関数である。通常 activation function としては何らかの非線形性を持つ関数が用いられる。また、非常に小さいか大きい値に対しては $-1, 1, 0$ などの値に収束する性質をもつ。具体的にはシグモイド関数やガウス関数などが用いられる。出力層においては線形の関数が用いられる場合もある。なお、「0 番目の層の出力」とはすなわち入力ベクトルのことである。出力層における出力はただ一つの値であり、

そのニューラルネットワークポテンシャルに紐づいている原子のエネルギーに対応する。

ニューラルネットワークによって原子間ポテンシャルを表現する場合、その入力ベクトルは原子配置に関連するものであるが、原子配置の xyz 座標をそのまま利用することはできない。これは、原子座標はたとえば原子の入れ替えや全体の回転などの操作に対して不変ではないからである。そこで、保たれるべき不変性が保たれる symmetry function という関数が用いられる場合が多い。たとえば以下のような関数が用いられる。

$$G_i^2 = \sum_{j=1}^{N_{\text{atom}}} \exp[-\eta_2(r_{ij} - R_s)^2] f_c(r_{ij})$$

$$G_i^3 = 2^{1-\zeta} \sum_{j \neq i} \sum_{k \neq i, j} (1 + \lambda \cos \theta_{ijk})^\zeta \exp[-\eta_3(r_{ij}^2 + r_{ik}^2 + r_{jk}^2)] f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk})$$

G_i^2 は i 番目の原子に対する 2 体の symmetry function である。 r_{ij} は原子 ij 間の距離であり、 η_2 や R_s はこの symmetry function の特性を決める係数である。 $f_c(r_{ij})$ はカットオフ関数であり、カットオフ距離に近づくにつれ 0 へなめらかに減衰する関数が用いられる。 G_i^3 は i 番目の原子に対する 3 体の symmetry function である。 θ_{ijk} は原子 ijk の成す角度であり、 η_3, λ, ζ がこの symmetry function の特性を決める係数である。Symmetry function は異なる $\eta_2, \eta_3, R_s, \lambda, \zeta$ を持つものをいくつ用意してもよい。多く用意すればそれだけ多様な原子配置を表現することができるが、当然計算負荷は高くなる。記述したい系の違いをなるべくうまく差別化できるように選べるのが望ましい。Symmetry function の選び方の指針は、たとえば文献[3], [4]などにおいて提案されている。

ニューラルネットワークは通常一つ一つの原子に与えられ、それによってその原子のエネルギーが計算される。このようにして求められた各原子のエネルギーの和がその系のエネルギーとなる。第一原理計算においては通常エネルギーは全系のものがただ一つ求まるのでこの分け方は便宜的なものととらえることもできるが、原子ごと（元素ごと）にニューラルネットワークを割り当てることによって教師データとして用いた系とは異なる原子数（場合によっては異なる化学量論比）の系に対しても適用できるようになるというメリットがある。

ニューラルネットワークポテンシャルはうまく作成すれば第一原理計算の結果を原子当たり 1 meV の精度で再現できる高精度な手法である。他方なんらかの物理的なモデルに基づくものではないため、教師データの“外挿”は不得手としている。排除体積効果さえも含まれないので、うまくいかない場合は極端に短い結合が現れるなど分子動力学シミュレーションとして破綻してしまうような場合もある点がデメリットである。

ニューラルネットワークポテンシャルの作成

ニューラルネットワークポテンシャルの作成方法

ニューラルネットワークポテンシャルを作成するためには、「教師データ」が必要となる。これは原子座標データと対応する全エネルギーと場合によっては原子間力である。このデータは第一原理分子動力学シミュレーションの軌跡から得ることが多い。ニューラルネットワークポテンシャルは教師データの外挿は苦手なので、適用したい領域を包含する原子配置を教師データとして採用する必要がある。

教師データからニューラルネットワークポテンシャルを作成するソフトウェアとして n2p2 (<https://github.com/CompPhysVienna/n2p2>), ænet (<http://ann.atomistic.net/>), deepmd (<https://github.com/deepmodeling/deepmd-kit>) などがある。本件では n2p2 [5] を用いた。

教師データの作成

ニューラルネットワークポテンシャル作成に必要な教師データは PHASE/0 コード[6]による第一原理分子動力学シミュレーションの軌跡より得た。

第一原理分子動力学シミュレーションの計算条件

第一原理分子動力学シミュレーションは以下のような条件で行った。

| 計算条件 | 設定値 |
|-------------|--|
| カットオフエネルギー | 25 Rydberg |
| k 点サンプリング | cubic 系 : $2 \times 2 \times 1$ アモルファス, 液体 : Γ 点のみ |
| 収束判定条件 | $1e-8$ hartree/atom |
| 時間刻み | 4 fs |
| アンサンブル | NVT アンサンブル |

cubic $\text{Ge}_2\text{Sb}_4\text{Te}_5$ の第一原理分子動力学シミュレーション

GeSbTe 系の典型例として $\text{Ge}_2\text{Sb}_4\text{Te}_5$ があげられる。 $\text{Ge}_2\text{Sb}_4\text{Te}_5$ は結晶としては立方晶系と六方晶系があり、六方晶の方がエネルギー的には安定であると言われている。他方アモルファス相とのスイッチングに使いやすいのはエネルギー的には近いであろう立方晶の方ではないかと考えられる。そこで、ここでは結晶の教師データとして立方晶の $\text{Ge}_2\text{Sb}_4\text{Te}_5$ を用いることにした。

立方晶 $\text{Ge}_2\text{Sb}_4\text{Te}_5$ は、面心立方格子の $\langle 111 \rangle$ 方向に原子が Te/Sb ないし Ge ないし空孔 / Te/Sb ないし Ge ないし空孔 / ... という具合にスタックした構造である。その典型例は次

に示すような構造である。

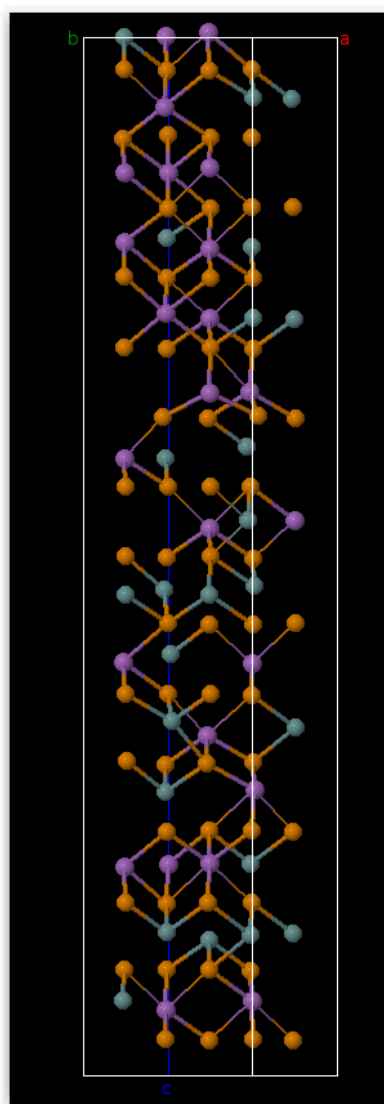


図 1 $\text{Ge}_2\text{Sb}_4\text{Te}_5$ の結晶構造の例

図において緑色で描画されている球が Ge, 紫色で描画されている球が Sb, 橙色で描画されている球が Te である。 Ge:Sb:Te を 2:2:5 という比率で成立させるため細長い単位胞となっている。Te の位置は固定であるのに対し, Ge と Sb は空孔を含めランダムになっている。そこで, Ge/Sb/空孔位置をランダムに埋めた構造を 20 通り用意し, 分子動力学シミュレーションの初期原子配置とした。温度は 500K と 1500 K を採用し, 各々 1 万ステップの *NVT* 第一原理分子動力学シミュレーションを実行した。

液体 $\text{Ge}_2\text{Sb}_2\text{Te}_5$ の第一原理分子動力学シミュレーション
立方体の単位胞に 10 fu 分の原子をランダムに敷き詰めた構造を初期構造とした。温度

3000K で *NVT* 第一原理分子動力学シミュレーションを行った。 $\text{Ge}_2\text{Sb}_2\text{Te}_5$ の融点は 900K 程度なので非常に高温のシミュレーションであるが、多様な構造のサンプリングを狙いこの温度を採用した。Ge 格子定数を二種類用意し、それぞれに対して 4 万ステップ程度のシミュレーションを行った。

メルトクエンチシミュレーション

上述の液体のシミュレーションから 20 通りの初期配置をサンプルし、急冷のシミュレーションを実施した。温度 3000K から 300K まで 3 万ステップかけて冷却した。アンサンブルは *NVT* を採用した。

教師データサンプリング

第一原理分子動力学シミュレーションのステップ間のデータは相関が強く、毎ステップサンプルして教師データとするとデータ数に対して十分多様なデータが得られない可能性が高い。そこで、得られた軌跡から 25 ステップに一度サンプルすることによって最終的な教師データとした。データ点数は合計で 43,000 ほどとなった。

GeSb_2Te_4 の教師データ追加

後述の手続きで作成した NNP を用いて GeSb_2Te_4 の急冷シミュレーションを実施し、教師データに追加した。合計で 4000 点ほど追加した。

n2p2 コードを用いたニューラルネットワークポテンシャルの作成

ニューラルネットワークポテンシャルは n2p2 コード[5]を用いて作成した。Symmetry function は文献[4]の方針に従い定義した。二体の symmetry function は元素の組み合わせあたり 5 通りで計 45 通り、三体の symmetry function は元素の組み合わせあたり 20 通りで計 360 通り用意した。カットオフ距離は 6 Å とした。Activation function としては隠れ層に対しては hyperbolic tangent を、出力層に対しては線形の関数を採用した。ニューラルネットワークとしては隠れ層二層、各層あたり 20 ノードとした。エネルギーだけでなく原子間力も学習に用いる設定を採用した。教師データの 9 割を学習に、1 割をテストに用いた。そのほか学習の仕方の詳細などについては n2p2 のデフォルト設定を採用した。結果得られたニューラルネットワークポテンシャルを用いて作成した parity plot を次に示す。

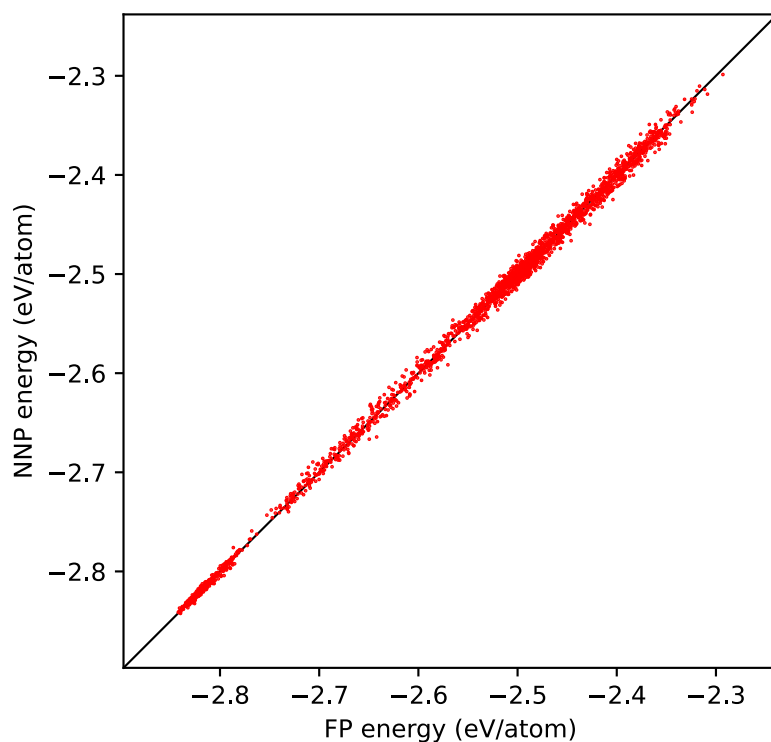


図 2 得られたニューラルネットワークポテンシャルの parity plot.

得られたニューラルネットワークポテンシャルは、学習に用いたデータに対してもテストデータに対してもおおよそ root mean square error がエネルギーに対して 6 meV/atom, 原子間力に対して 0.22 eV/Å の精度で教師データを再現することができた。ニューラルネットワークポテンシャルは問題によってはエネルギーに対して 1 meV/atom 以下の精度を得ることもできる場合があるが、ここで得られた精度は第一原理計算そのものの精度と比較できるオーダーであり、実用上十分であると考えている。

ニューラルネットワークポテンシャルの使い方

ポテンシャルファイル

ポテンシャルファイルは n2p2 ディレクトリー以下に配置されている。以下のファイルが存在する。

input.nn scaling.data weights.032.data weights.051.data weights.052.data

input.nn ファイルには NNP に用いられているニューラルネットワークに関する情報などが記録されている。scaling.data ファイルには symmetry function のスケーリングに関

する情報が記録されている。 `weights.032.data`, `weights.051.data`, `weights.052.data` ファイルにはそれぞれ Ge, Sb, Te のニューラルネットワークの重みが記録されている。このディレクトリーをインプットスクリプトにおいて指定することによって利用することができる。

プログラムのコンパイル方法

n2p2 をインストールする方法

n2p2 をインストールする方法について説明する。Linux 環境 (WSL 上の Linux でも可) を想定しているが、macOS にも似た手続きによってインストールできるかもしれない。

必要なライブラリーのインストール

n2p2 は

GSL (<https://www.gnu.org/software/gsl/>) と

Eigen (<https://eigen.tuxfamily.org/>)

を必要とする。利用したい環境にこれらがインストールされていない場合上述のウェブサイトからアーカイブをダウンロードする。バージョンにそれほど依存しないはずなので、最新版をダウンロードすればよい。

また、BLAS (<https://www.netlib.org/blas/>) も必要である。BLAS はすでにインストールされている場合も多いと思われるが、インストールされていない場合は何らかの BLAS の実装を入手し、インストールする。

GSL はコンパイルする必要がある。通常の GNU/Linux アプリケーションの手続きによってコンパイルすることができる。すなわち

```
./configure
make
sudo make install
```

のような手続きである。管理者権限のないアカウントで作業をする場合は `./configure` のあとに `--prefix=INSTALL_DIRECTORY` をつけることによって書き込み権限のあるディレクトリーをインストールするディレクトリーとして指定する。この場合環境変数 `LD_LIBRARY_PATH` に `INSTALL_DIRECTORY/lib` を加える必要がある。

Eigen はテンプレートライブラリーなのでコンパイルする必要はないが、解凍後得られるトップディレクトリーの下にある `INSTALL` ファイルに記述されている方法でインストールしておくことを推奨する。

n2p2 のコンパイル

n2p2 のアーカイブを配布元 (<https://github.com/CompPhysVienna/n2p2/releases>) からダウンロードし、展開する。バージョンは分子動力学シミュレーターLAMMPS (<https://www.lammps.org/>)に正式に取り入れられた 2.14 以降を用いる。

ソースコードは `src` ディレクトリー以下に配置されている。`src` ディレクトリーに配置されているファイルやディレクトリーは下記の通り。

```
application/  doc/  interface/  libnnp/  libnnpif/  libnnptrain/  makefile
makefile.gnu  makefile.intel  makefile.llvm  pynnp/
```

`makefile.gnu` は `gcc` 用の設定が記述されたファイル、`makefile.intel` には Intel oneAPI 用の設定が記述されたファイル、`makefile.llvm` は `clang` コンパイラー用の設定が記述されたファイルである。これらの内利用したいコンパイラーに対応したファイルをエディターなどで開き、以下の赤色で示した部分を編集する。

```
PROJECT_GSL=/usr/local/include/gsl
PROJECT_EIGEN=/usr/local/include/eigen3
```

`PROJECT_GSL` には `GSL` のインクルードファイルのパスを指定する。`gsl` のインストール先がたとえば `/home/user/gsl` だった場合この指定は `/home/user/gsl/include/gsl` となる。`PROJECT_EIGEN` には `Eigen` のインストールディレクトリーを指定する。`Eigen` のインストール先がたとえば `/home/user/eigen3` だった場合この指定は `/home/user/eigen3/include/eigen3` となる。

この編集ができれば、たとえば `makefile.gnu` を使う場合

```
make COMP=gnu
```

というコマンドを実行すると `n2p2` のコンパイルが始まる。問題が発生しなければ `NNP` を作成するためのプログラムが `bin` 以下に、`LAMMPS` などとリンクするためのライブラリーが `lib` 以下に生成される。

n2p2 の `NNP` が利用できる `LAMMPS` をビルドする方法

`LAMMPS` のアーカイブを配布元 (<https://www.lammps.org/>)からダウンロードする。バージョンは `n2p2` を正式に取り込んだ 29 September 2021 版以降を利用する。

まずは `n2p2` を取り込むための準備として、`lib/hdnnp` ディレクトリーへ移り、そこに配置されている `Install.py` という Python スクリプトを実行する。

```
cd lib/hdnnp
python Install.py -p /home/user/n2p2
```


ここで n2p2 は/home/user/n2p2 以下にインストールされていると仮定した。この操作によって

```
Makefile.lammps includelink liblink
```

という三つのシンボリックリンクが生成されていればこの段は成功である。

つぎに **src** ディレクトリーに移動し、n2p2 の NNP を利用できるようにするパッケージ **ml-hdnp** を有効にする。

```
cd ../../src  
make yes-ml-hdnp
```

ほかにも有効にしたいパッケージがあれば同様の手続きで有効化しておく。さらに必要に応じて **src/MAKE** の下にある **Makefile.mpi** もしくは **Makefile.serial** を編集する。前者は mpi 並列用の Makefile, 後者はシリアル実行用の Makefile である。ここまで準備ができたら **make** コマンドによって LAMMPS をコンパイルすればよい。

```
make mpi  
もしくは  
make serial
```

問題が発生しなければ **src** ディレクトリーの下に **lmp_mpi** (MPI 並列版の場合) もしくは **lmp_serial** (MPI 非並列版の場合) が得られる。

LAMMPS の実行方法

インプットスクリプトの書き方

LAMMPS のインプットスクリプトはニューラルネットワークポテンシャル指定部分以外は通常の設定を採用すればよい。作成したニューラルネットワークポテンシャルはエネルギーの単位としては eV, 長さの単位としては Å を採用しているので, **units metal** を利用し, 初期原子配置作成の際も留意する (後述のようにスケーリングファクターによって調整することもできる)

インプットスクリプトのポテンシャル指定部分は以下のように記述する。

```
pair_style hdnp 6.0 dir ../n2p2 showew no showewsum 1000 resetew yes maxew  
10000 cflength 1.0 cfenergy 1.0  
pair_coeff * * Ge Sb Te
```

一行目

- **pair_style hdnp** によって n2p2 の NNP を使うことを指定している。

- 続く `6.0` によってカットオフが 6.0 \AA であることを指定している。`6.0` は NNP 作成時に採用した値であり、これより短くしてはならない。
- 続く `dir ../n2p2` によって一階層上の `n2p2` というディレクトリーに NNP のファイルが配置されていることを指定している。この部分は計算の実行ディレクトリーの位置によって変わってくる。相対パスではなく絶対パスを用いて指定してもよい。
- 続く `showew no` とすることによって `error warning` を出力しないという設定を行っている。
- 続く `showewsum 1000` によって `1000` ステップに一度積算エラーを出力するように設定している。
- 続く `resetew yes` によって `error warning` は出力する度にリセットすることを設定している。
- 続く `maxew 10000` によって `error warning` が `10000` を超えた場合に計算を終了させることを設定している。
- 続く `cflength 1.0` によって長さに関するスケーリングファクターを `1.0` にしている。用いる `units` が `metal` の場合の設定であり、別の `units` を用いる場合は相応の値を採用する。
- 続く `cfenergy 1.0` によってエネルギーに関するスケーリングファクターを `1.0` にしている。用いる `units` が `metal` の場合の設定であり、別の `units` を用いる場合は相応の値を採用する。

二行目

`pair_coeff` コマンドによって一番目の元素が `Ge`、二番目の元素が `Sb`、三番目の元素が `Te` に対応することを設定している。

計算実行

通常通り LAMMPS を起動すればよい。たとえば以下のようなコマンドを実行する。

```
mpiexec -n N ~/lammps-2Aug2023/src/lmp_mpi -in in
```

`N` は MPI 並列数である。利用しているコンピュータのコア数を超えない数値を指定する。

ログファイル

ヘッダー部分では、通常の LAMMPS の計算のログに加え、用いているニューラルネットワークポテンシャルの様々な情報が出力される。

MD 計算中は上述の `showewsum` の設定に応じてエラーの積算値が出力される。

```
### NNP EW SUMMARY ### TS:      1000 EW      203 EWPERSTEP 2.030e-01
```

`1000` ステップに一度エラーの積算値が出力されており、総計 `203` 回、`1` MD ステップあた

り $2.03\text{e-}01$ 回エラーが発生したことが分かる。この例では初期配置として教師データに含まれない構造を採用したのでエラーが発生してしまったが、以降シミュレーションが進行するに従いエラーは減っていく傾向だった。

エラーは原子配置が教師データに対して外挿となる領域に至った場合に報告される。このエラーが発生したからといって即座に計算を棄却しなければならないとは限らないものの壊滅的な結果が得られてしまう場合もあるので、エラー発生時はその後の結果をよく吟味することが推奨される。

計算例

急冷のシミュレーションによってあらかじめ作成しておいたアモルファス構造を初期配置とし、複数の温度で分子動力学シミュレーションを行った。

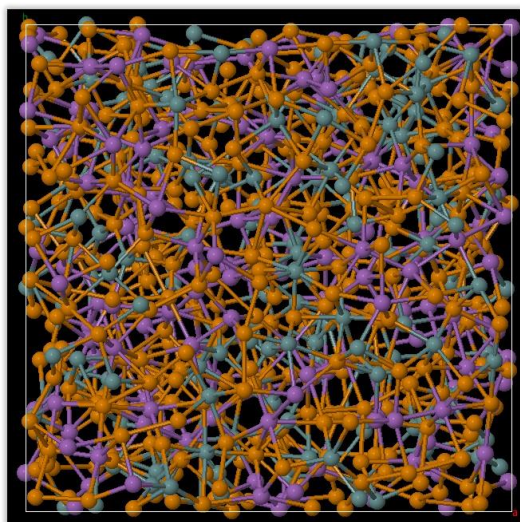
計算条件

分子動力学シミュレーションの主な計算条件は下記の通り。

| 計算条件 | 設定値 |
|--------|----------------------------|
| 時間刻み | 4 fs |
| カットオフ長 | 6 Å |
| アンサンブル | <i>NPT</i> アンサンブル |
| 温度 | 100K から 1000K まで, 100K きざみ |
| ステップ数 | 1000000 |

初期原子配置としてはすべての温度で同じものを利用した。まずは第一原理計算メルトクエンチシミュレーションによって得たアモルファス構造の $2 \times 2 \times 2$ スーパーセルを作成し、1000K から 300K まで急冷のシミュレーションを実施した。得られた最後のスナップショットを初期原子配置とした。その構造を図 3 に示す。

$\text{Ge}_2\text{Sb}_2\text{Te}_5$ 80 fu 720原子



GeSb_2Te_4 96 fu 672原子

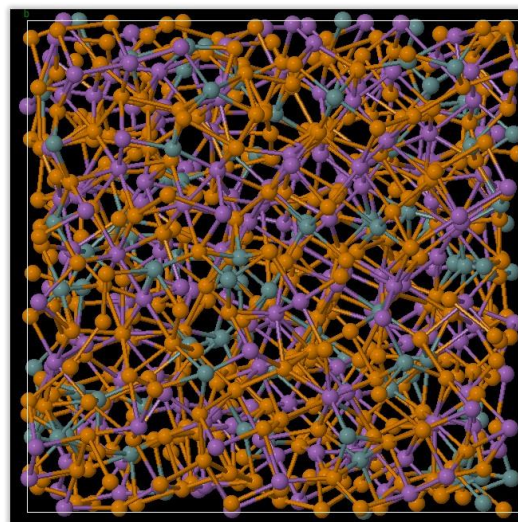
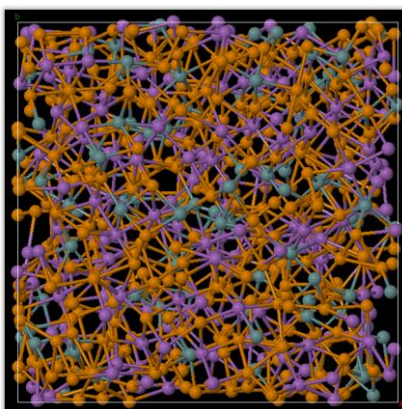


図 3 初期原子配置

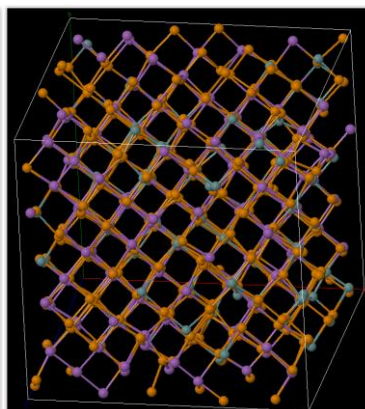
計算結果

得られた結果を紹介する。図 4 と図 5 はそれぞれ 300K, 600K, 900K における $\text{Ge}_2\text{Sb}_2\text{Te}_5$ と GeSb_2Te_4 のスナップショットである。いずれの場合も 300K においてはアモルファス構造が維持されているのに対し 600K では構造が大きく変わり、周期的な構造が得られている。900K の場合は周期構造は維持できず、液体構造が得られている。

300K



600K



900K

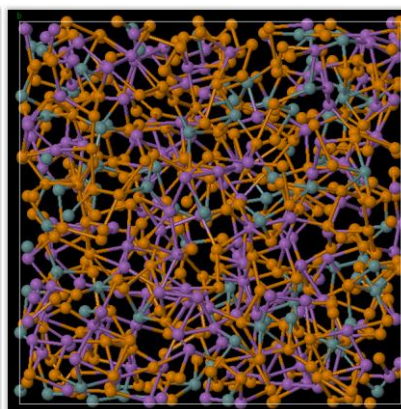


図 4 $\text{Ge}_2\text{Sb}_2\text{Te}_5$ の分子動力学シミュレーションにおいて得られた原子配置のスナップショット

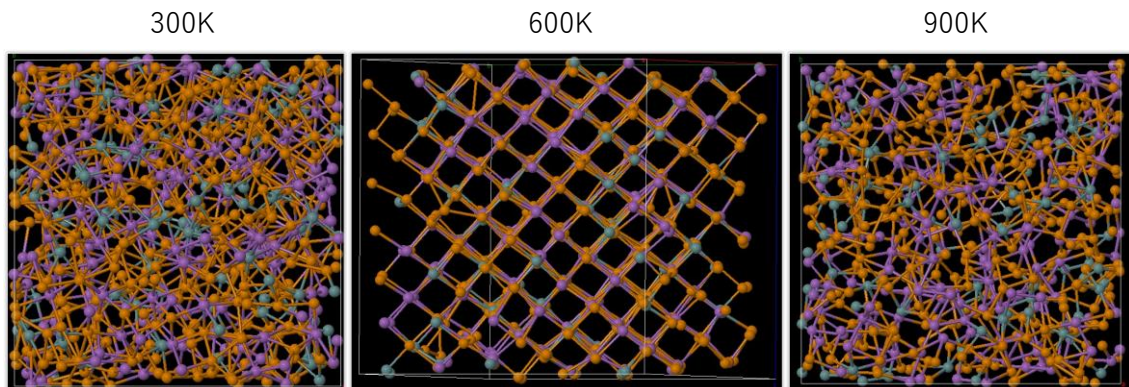


図 5 GeSb_2Te_4 の分子動力学シミュレーションにおいて得られた原子配置のスナップショット

より定量的に判断するため、図 6 に温度と平均エネルギーの関係をプロットする。エネルギーは後半 500000 ステップのエネルギーを平均することによって求めた。通常温度を上げるとそれに伴い系のポテンシャルエネルギーも上昇する。しかしながら今の場合 400-500K あたりにおいていったんエネルギーが下がる傾向がみられる。これはアモルファス構造が周期的な構造に転移したためである。さらに温度が上昇し、800K になると液体に転移するためエネルギーは大幅に上昇する。

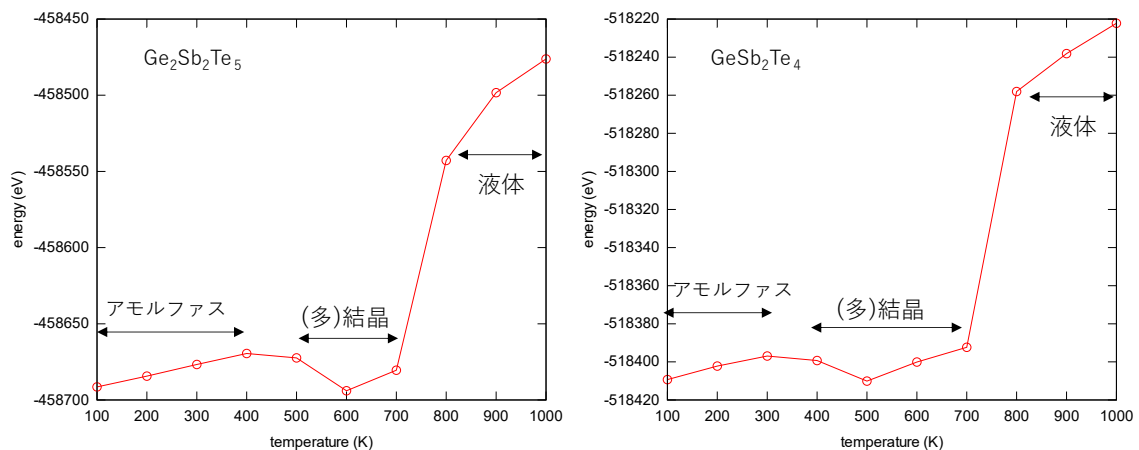


図 6 温度とエネルギーの関係

図 7 には結晶化前後における時間とエネルギーの関係を示す。1.35 ns 以降急激にエネルギーが下がり、1.4 ns からは一定値に振動するような振る舞いになっている。このことは結晶化しはじめてから 100 ps かからず全系が結晶化したことをあらわしている。

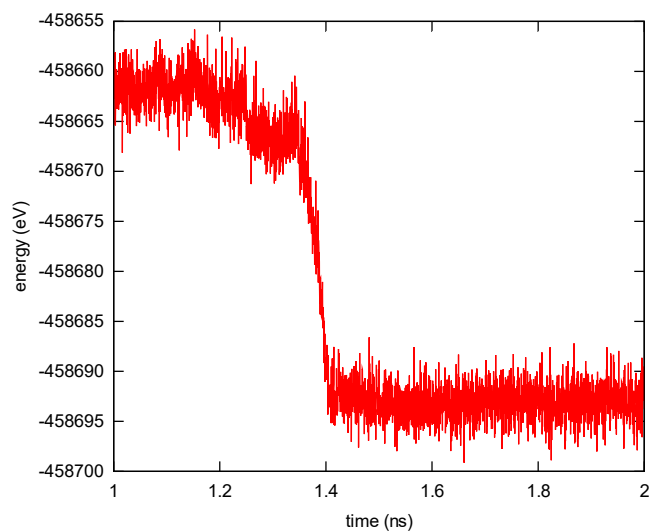


図 7 結晶化前後のエネルギーの履歴

最後に、アモルファス構造から結晶に転移する様子の動画を紹介したい。この動画は $\text{Ge}_2\text{Sb}_2\text{Te}_5$ の 600K のシミュレーションから得られた軌跡から作成したものである。コマ間の間隔は実時間では 0.1 s, シミュレーション時間では 40 ps である。開始数秒で結晶化の兆しが見え始め、その後非常に短い時間で結晶に転移している様子がわかる。

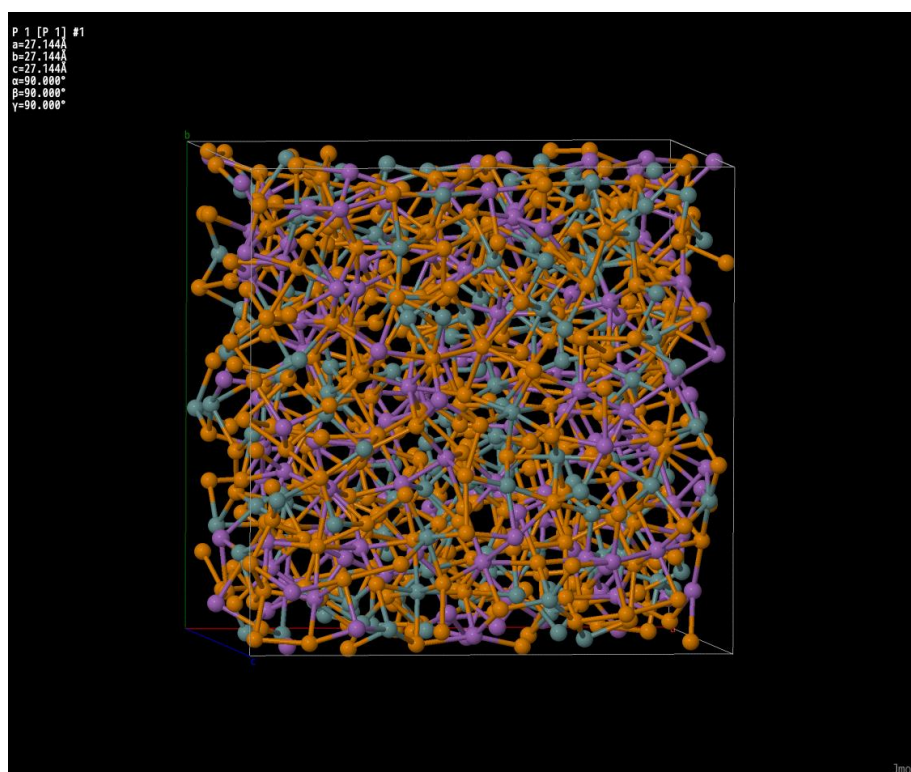


図 8 アモルファス構造から結晶に転移する様子。

参考文献

- [1] J. Behler and M. Parrinello, Phys. Rev. Lett. 98, 146401 (2007).
- [2] J. Behler, International Journal of Quantum Chemistry 115, 1032 (2015)
- [3] Giulio Imbalzano; Andrea Anelli, Daniele Giofr  , Sinja Klees; J rg Behler, Michele Ceriotti, J. Chem. Phys. 148, 241730 (2018).
- [4] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, P. Marquetand, J. Chem. Phys. 148, 241709 (2018).
- [5] A. Singraber, T. Morawietz, J. Behler and C. Dellago, J. Chem. Theory Comput. 2019, 15 (5), 3075–3092.
- [6] T. Yamasaki, A. Kuroda, T. Kato, J. Nara, J. Koga, T. Uda, K. Minami, and T. Ohno, Computer Physics Communications 244, 264-276 (2019).