マテリアルズ・インフォマティクス実例紹介: 半導体ゲート絶縁膜高誘電率相安定化ドーパントの探索 Materials informatics of phase-stabilization dopants for semiconductor gate dielectrics

高度情報科学技術研究機構 古賀 裕明

材料開発におけるビッグデータ活用の取り組みとして、マテリアルズ・インフォマティクス (MI)が注目を集めている。ここではMIの実例として、ハイスループットDFT計算による半導 体ゲート絶縁膜高誘電率相安定化ドーパント探索[Harashima et al., IEEE Trans. Semicond. Manuf. 36, 543 (2023)]について紹介する。特に、ドープ系におけるドーパント原子のすべて の可能な配置を生成する手順や、計算により得られた大量の出力ファイルから必要なデータを抽 出し集計する手順など、MIの基本的技術について重点的に解説する。ジルコニア・ハフニアの様々 なドープ組成に対するハイスループットDFT計算により得られたデータは、SiやGeの相安定化 効果が大きいことや、これらと同周期のIII族・V族元素の等量ドープによりGeと同等の効果が 得られることを示唆している。

1. はじめに

1.1 マテリアルズ・インフォマティクス

近年では、人手に頼る従来の材料開発手法 にかわり、実験や計算の自動化により得られ た膨大なデータから隠れた法則性を見出し、 新材料の発見に結びつけるマテリアルズ・イ ンフォマティクス (MI) が台頭している。 MIの基本的な流れを図1に示す。まず、試 すべき材料の組成や構造など、探索範囲を定 義する。次に、探索範囲全体に対し試料を作 成し測定にかける(実験の場合)か、初期構 造を生成し計算を実行する(計算の場合)。 このステップでは、大量のデータを獲得する ために自動化技術が不可欠であり、特に大量 データ生成に重点をおいた計算をハイスルー プット計算と呼んでいる。また、データ点を 順番に計算するのではなく、機械学習手法を 活用して探索範囲を絞り込んだり、優先順位 を決める取り組みも重要である。この後、得 られた膨大なデータを集計し、ミスがないか チェックする。ミスの原因は、物質本来の不

安定性や計算上の不具合など様々であるの で、実験・計算技術や対象物質に関する知識・ 経験が要求される。こうして用意したデータ セットに対し可視化や機械学習を適用するこ とで、新たな規則性・法則性を見出したり、 予測モデルを構築して材料開発に取り組むこ



とができる。実際には一回のフローで期待し た成果が得られるとは限らない。探索範囲や モデルを修正しつつ、これらの手順を繰り返 すことになる。

情報科学技術に係る研究・技術開発及び科 学技術分野の情報の調査収集等を総合的に推 進することにより、学術および科学技術の発 展に寄与すること [1] をミッションとする RISTにとって、MIは重大な関心事である。 特にRIST計算科学技術部(東京都港区浜松 町)は、公的な研究プロジェクトに参画 [2] する一方で、大学・国研・企業からの個別の 依頼によるコード開発・シミュレーションに 取り組んでおり、今後はMIに関する依頼が 一層増えていくものと予期している。これま で産業界との共同研究として、新規有機光機 能材料探索のための機械学習活用 [3] や、 ケモインフォマティクスのための量子機械学 習法の開発 [4] に取り組んできた。

本稿ではMIの実例として、筑波大学重田 育照教授らによる高誘電率(high-k)物質相安 定化ドーパント探索のためのハイスループッ トDFT計算 [5]を紹介する。特に、この課 題に関してRISTが担当した、supercellツー ル [6]による初期構造生成や、pymatgen [7]による出力データ集計など、MIの前工程・ 後工程にかかる基本的技術を中心に解説する [8]。ただし、得られたデータセットに対す る機械学習については、未公表のため本稿で は記していない。また、本稿はMI全般の動 向に関する記事ではないので、興味のある方 は参考文献 [9]を参照されたい。

1.2 DFT計算

密度汎関数法 (DFT) [10] は MI 登場以前 から存在するメジャーな電子状態計算手法で あり、本稿の主題ではないが、MI において 物性値を計算するためによく使われるので、 簡単に説明する。

一般に物質は原子の集まりでできており、

原子は原子核と電子からできている。原子核 の引力に捉えられた電子の集団の状態を計算 することで、エネルギーなどの物性値を得る ことができる。一般にエネルギーが低い状態 ほど実現しやすいので、エネルギーが低くな るように原子核の座標を最適化すれば、物質 の構造を決めることができる(構造最適化と いう)。構造とエネルギーが求まれば、「正方 晶と単斜晶のどちらが安定か」といった問い に答えることができる。

以上のような理由で、電子状態を計算する 手法が物質科学では重要な役割を果たしてい る。なかでもDFTは、物質科学では標準的 な電子状態計算手法として広く用いられてい る。DFTは、多電子系の電子状態(正確には 基底状態)が電子密度の汎関数[11]になっ ているというHohenberg-Kohn定理[12] に基づいている。全エネルギーを低くするよ うに電子密度を最適化することで、電子状態 を求めることができる。実際には汎関数の形 が完全にわかっているわけではなく、DFT の登場以来、改良が重ねられてきた。固体系 では、GGA-PBE汎関数[13]やその改良版 が用いられる。

DFTの実装は、汎関数のほかにも様々な 近似方法やアルゴリズムにより支えられてい る。まず、電子の軌道を数値データ化するた め、単純な基底関数(ガウス関数や平面波) の重ね合わせとして表現する。また、全ての 電子を考慮すると計算が大変になるので、原 子の外側の電子(価電子)のみ考慮し、内側 の電子(内殻電子)と原子核はイオンコアと して固定する方法もある。この場合、価電子 とイオンコアの相互作用を記述するため、擬 ポテンシャル法 [14] やPAW法 [15] が用い られる。固体物質では、平面波基底、擬ポテ ンシャル法またはPAW法、GGA-PBE汎関数 の組み合わせを用いていることが多い。本稿 で紹介する事例も同様である。固体向けの代 表的なDFT計算コードとしては、Quantum ESPRESSO [16] (QE) や Vienna Ab initio Simulation Package [17] (VASP) が挙げ られる。

1.3 半導体微細化とhigh-k材料

High-k材料 [18] は、大きな誘電率 κ と 絶縁性を兼ね備えた物質で、半導体ゲート絶 縁膜やDRAMキャパシタ絶縁膜に用いられ るものを指す。ムーアの法則に象徴される半 導体IC・LSIの高速化・高集積化は、素子の 微細化によって進められてきた。半導体トラ ンジスタの代名詞であるMOSFETは、ゲー ト電極に電圧を印加し、チャネルに電荷を誘 起することによりソース・ドレイン間電流の ON/OFF 制御を行う電子的スイッチである (図2)。ここでゲート電極/絶縁膜/チャネル はキャパシタとみなすことができ、その容量 をゲート容量という。通常のキャパシタと同 じく、ゲート容量は絶縁膜の誘電率に比例し、 その厚さに反比例する。微細化により絶縁膜 が薄くなると、ゲート容量が増して駆動電流 が大きくなるメリットがある。しかし、従来



図2 MOSFET模式図

から用いられていた二酸化ケイ素膜では、あ まりに薄くするとトンネル効果により絶縁性 が低下し、消費電力が大きくなる問題が生じ た。誘電率の高い物質(high-k物質)を絶縁 膜に使うと、膜厚を保ったままゲート容量を 大きくできる。このような理由から、高い誘 電率と絶縁性をもつ物質が絶縁膜材料として 求められてきた。

本稿で取りあげるハフニア [19] (酸化ハ フニウム、HfO₂)とジルコニア [20] (酸化 ジルコニウム、ZrO₂) は代表的なhigh-k材 料であり、ハフニアベースのゲート絶縁膜 は2007年頃 [18] から使用されている。Hf はZrと同じ4族元素であり、ランタノイド 収縮[21]の影響を受けるので、一周期前 のZrとほぼ同じ原子サイズになる。そのた め、ジルコニアとハフニアはかなり類似した 性質を持つ。これらは常温常圧では蛍石構 造類似の単斜晶となるが、ジルコニア「22] は1050~1170°Cで正方晶、2370°C以上で 立方晶に転移し、ハフニア [23] は1475~ 1600°C以上で正方晶、2700~2900°Cで立方 品に転移する(図3)[24]。相転移により体 積変化すると耐熱材料としては具合が悪いの で、CaやYを添加して高温相を安定化して いる。相安定化したジルコニアは、宝石や酸 素イオン伝導体など様々な用途に用いられて いる。誘電率は単斜晶(κ=20)より正方晶(κ =47) で高い「20] ので、ゲート絶縁膜用途 でも異元素添加などの正方晶安定化技術が重



図3 ジルコニアの結晶相

要になる。

ここで紹介する実例 [5] では、以上の観 点から、様々な添加元素のジルコニア・ハフ ニア正方晶安定化効果についてハイスルー プットDFT計算による系統的評価が行われ た。なお、半導体の電気伝導性を制御するた め異元素(ドーパント)を添加することを ドーピングと呼ぶのにならい、ここでも同様 の呼称を用いているが、半導体のドープ濃度 が100万~1億分の1と極めて小さいのに対 し、相安定化では数%~10%程度の異元素を 添加していることに注意されたい。

2. MI 前工程

2.1 相安定化効果の評価方法

ジルコニアの6.25%Siドープ系を例として、 正方晶相安定化効果を評価する流れを説明す る(図4)。ジルコニアの正方晶および単斜 晶のモデルとして、96原子モデル(2×2×2 セル)を用いた。32個の陽イオン(Zr⁴⁺)の うち2個をSiに置換すると6.25%ドープにな るが、このようなドーパントの配置は正方晶 で9通り、単斜晶で31通り存在する(結晶 の対称性を考慮)。Supercellツール(後述) を用いると全ての可能な配置を生成すること ができる。これら全てについてDFT計算に より体積と構造を最適化し、全エネルギーを 求めた。正方晶と単斜晶それぞれについて、 最もエネルギーが低くなるドーパント配置を 選び、その差を正方晶安定化効果の指標とし た。エネルギーが低い方が安定なので、単斜 晶から正方晶のエネルギーを引いた値が大き いほど、正方晶が安定化されていることを意 味する。

2.2 ドープ元素の選択

計算の対象となった主なドープ組成を表1 に示す。実際には約200の組成、2.5万超の配 置についてデータが得られているが、ここで は主要なものに限って示している。これらの



図4 相安定化効果評価の流れ Zrは黒、Oは白、Siは大きめの球として表示。

族	組成	mol%ª	正方晶 配置数	単斜晶 配置数	対象とした元素	
 ²⁺	$II_1Zr_{31}O_{63}$	3.125	12	64	Mg Ca Sr Ba Zn	
III ³⁺	$III_2Zr_{30}O_{63}$	6.25	148	992	Al Y La	
IV ⁴⁺	IV 2Zr30O64	6.25	9	23	$Si Co Sp Ti 7r/ Hf^{b} Co$	
	$IV_3Zr_{29}O_{64}$	9.375	32	155	SIGE SITTZI/TT CE	
III ³⁺ -V ⁵⁺	$III_1V_1Zr_{30}O_{64}$	各 3.125	9	31	III 族:AI Ga In Sc Y La	
					V 族:P As Sb Bi	

表1 主なドープモデル組成と可能な配置の数

a) 陽イオンに対する比率。

b) ジルコニアに対してはHf、ハフニアに対してはZrをドープ。

元素は次の理由により選択した。まず、Ca やYが既存のジルコニア材料では相安定化に 用いられることから、II族及びIII族元素を検 討対象に含めた。II族元素からは、毒性が懸 念されるBe、Cd、Hgと放射性元素のRaを除 いてMg、Ca、Sr、Ba、Znを選んだ。III族元 素からは、よく用いられているYの他、Alと Laを試した。ZrやHfが+4価の陽イオンであ るのに対し、II族は+2価なので、電荷のバ ランスをとるため酸素イオン(-2価)を一 つ取り除く必要がある。同様にIII族は+3価 になるので、二個のIII族イオンにつき一個 の酸素欠陥を導入する。II族元素の場合、96 原子モデルに1原子導入しただけなら正方晶 12通り、単斜晶64通りであるが、2個導入 すると2個の酸素欠陥を伴うので、急激に 配置の数が増える。よって一個導入の場合 (3.125%)のみ対象とした。III族も同様の理 由で、二個導入(6.25%)のみ対象とした。

先行研究 [25] では IV族ドープ系のDFT 計算により、12.5%Siドープのとき正方晶が 単斜晶より安定になることが報告されている (ただし24原子モデルを使用)。ドープ濃度 を下げたときも同様の効果が得られるのか興 味があるので、IV族元素のSi、Ge、Sn、Ti、 Zr、Hf、Ce [26] の6.25%と9.375%ドープ系 を検討対象に含めた。毒性を考慮しPb は除 外した。またCは原子が小さく、カーボネー トCO₃を作るので、他のIV族とは分けて考 える必要がある。なので今回は検討対象から 除外している。

GaAsなどのIII-V族化合物が半導体として 使われているように、III族元素とV族元素を 1:1で組み合わせるとIV族元素に似た性質を 示す。III族とV族を同数ドープするとIV族 と同様の安定化効果得られると期待し、計算 対象に含めた。III族としてAl、Ga、In、Sc、 Y、Laを、V族としてP、As、Sb、Biを計算対 象とした。BとNはCと同様の理由により除 外した。Tlは毒性を考慮し除外した。VA族 のV、Nb、Taは遷移金属として価数変化する ので、典型元素のV族とはわけて考える必要 がある。よって今回は除外している。

なお、本稿では慣用的にドーピングと呼ん でいるが、表1に示すケースでは添加元素 を陽イオンと置換しているので、複合酸化 物という方が実態に近い。例えばSi₂Zr₃₀O₆₄ なら、シリカとジルコニアの複合酸化物 2SiO₂・30ZrO₂とみなすことができる。また、 AlPZr₃₀O₆₄なら、リン酸アルミニウムとジル コニアの複合物AlPO₄・30ZrO₂とみなすこと ができる。

2.3. ドーパント配置の全生成

ここではドーパント原子の配置の数え方に ついて、模式的な例を用いて説明する。図5 のように二次元平面に原子A(白丸)が正方 格子状に並んでいるときに、半分を原子B(黒 丸)に置き換えることを考える。この格子の 最小構成パターン(基本単位胞)は、図の点 線で示すように一原子のみを含む領域となる が、一原子の半分を別の元素に置き換えるこ とはできないので、最低でも倍の周期を考え る必要がある。例えば図の実線で示した、基



図5 二次元正方格子の単位胞(点線)と2× 2セル(実線)



図6 2×2セル中の2原子を置換して得られた配置a-fと、eとfを4×4セルで表示したものe'、f'(対称性考慮前)

本単位胞を縦横2倍した領域(2×2スーパー セル)には原子4個が含まれているので、こ のうち2個をBに置き換えると50%置換した ことになる。

4個の原子を区別して考えると、4個から 2個を取り出す組合せは、 $_{4}C_{2}=(4\times3)/(2\times1)=6通りとなる。これら6通りの配置を図$ 6のa-fに示す。しかし結晶には対称性、つまり平行移動や回転を施しても元の配列と一致する性質があるので、6通り全てを区別する必要はない。図6のeとfを4×4セルに拡大してe'とf'のように表示するとわかるように、これらは表示位置がずれているだけで、同じ配列である。同様に、a-dも同じ配列を表す。よって対称性を考慮すると、ドーパントの配置は二通りとなる。

ここで考えた2×2セルよりも大きなスー パーセルを取ることもできるが、大きくしす ぎると配置の数が爆発的に増えて計算が困難 になる。実際には計算量と、得られる情報量 (さらに安定な配置が見つかるかどうか等) のバランスを考えて、適当なセルサイズを選 択する。

2.4 Supercell ツールによる全配置生成

次に、supercell [6] ツールによりドーパン ト配置を全生成する手順について説明する。 例として、立方晶ジルコニアを構成する Zr 原子のうち、6.25%をSi に置き換えるケース を考える。まず、単位胞の CIF ファイルを 用意する。CIF [27] は結晶構造を表す代表 的なファイル形式で、Materials Project [28] や Crystallography Open Database [29] などのデータベース [30] からダウンロード できる。また、Open BABEL [31] を用いる と、物質の計算で用いられている様々なファ イル形式からCIFに変換できる。Materials Projectから取得した立方晶ジルコニア(mp-1565)の立方セルのCIFファイルは以下のよ うになっている。 # generated using pymatgen data ZrO2 _symmetry_space_group_name_H-M 'P 1' _cell_length_a 5.09014541 _cell_length_b 5.09014541 cell length c 5.09014541 cell angle alpha 90.0000000 _cell_angle_beta 90.00000000 _cell_angle_gamma 90.00000000 _symmetry_Int_Tables_number 1 _chemical_formula_structural ZrO2 _chemical_formula_sum 'Zr4 O8' _cell_volume 131.88353134 cell formula units Z 4 loop symmetry equiv pos site id _symmetry_equiv_pos_as_xyz 1 'x, y, z' loop_ _atom_type_symbol _atom_type_oxidation_number Zr4+ 4.0 02- -2.0 loop_ _atom_site_type_symbol atom site label _atom_site_symmetry_multiplicity _atom_site_fract_x atom site fract y _atom_site_fract_z _atom_site_occupancy Zr4+ Zr0 1 0.0000000 0.0000000 0.0000000 1 Zr4+ Zr1 1 0.0000000 0.5000000 0.5000000 1 Zr4+ Zr2 1 0.5000000 0.0000000 0.5000000 1 Zr4+ Zr3 1 0.5000000 0.5000000 0.0000000 1 O2- O4 1 0.75000000 0.25000000 0.75000000 1 O2- O5 1 0.75000000 0.75000000 0.75000000 1 O2- O6 1 0.75000000 0.75000000 0.25000000 1 O2- O7 1 0.75000000 0.25000000 0.25000000 1 O2- O8 1 0.25000000 0.25000000 0.25000000 1 O2- O9 1 0.25000000 0.75000000 0.25000000 1 O2- O10 1 0.25000000 0.75000000 0.75000000 1 O2- O11 1 0.25000000 0.25000000 0.75000000 1

冒頭にセルや空間群、続いて対称操作や酸 化数、最後に原子座標が与えられている。 _symmetry_で始まるタグは対称性に関す るデータ、_cell_で始まるタグはセルに関す るデータで、_atom_で始まるタグは原子に 関するデータである。loop_に続くブロック はテーブル形式でデータが与えられており、 loop_に続くタグの並びがテーブルの列の並 びに対応する。原子座標のブロックなら、一 列目が原子種、二列目が原子サイトのラベル、 三列目が原子サイトの多重度、4~6列目が スケールしたxyz座標、7列目がサイト占有 率となっている。占有率は、完全で純粋な物 質なら1であるが、異元素が混じっていたり 原子が欠けていると、1より小さくなる。な お、上に示した例では空間群がP1となって いるが、supercellが対称性を判定しなおす のでこのままでも問題ない。

Zr原子の一部をSi原子に置き換えたいと きは、原子座標テーブルの部分を以下のよう に書き換える(書き換えたファイルの名前を base.cifとする)。

Zr4+ Zr 1 0.0000000 0.0000000 0.0000000 1 Zr4+ Zr 1 0.0000000 0.5000000 0.5000000 1 Zr4+ Zr 1 0.5000000 0.0000000 0.5000000 1 Zr4+ Zr 1 0.5000000 0.5000000 0.0000000 1 Si4+ Si 1 0.0000000 0.0000000 0.0000000 1 Si4+ Si 1 0.0000000 0.5000000 0.50000000 1 Si4+ Si 1 0.50000000 0.00000000 0.50000000 1 Si4+ Si 1 0.5000000 0.5000000 0.00000000 1 O2- O 1 0.75000000 0.25000000 0.75000000 1 O2- O 1 0.75000000 0.75000000 0.75000000 1 O2- O 1 0.75000000 0.75000000 0.25000000 1 O2- O 1 0.75000000 0.25000000 0.25000000 1 O2- O 1 0.25000000 0.25000000 0.25000000 1 O2- O 1 0.25000000 0.75000000 0.25000000 1 O2- O 1 0.25000000 0.75000000 0.75000000 1 O2- O 1 0.25000000 0.25000000 0.75000000 1

変更箇所を太字で示した。まず、Zr原子 はZr、O原子はOというように、サイトラベ ルを統一する。こうすると、supercellにより 同じグループに属するサイトとして認識され る。加えて、Zr原子4個と同じ位置にSi原 子を配置する。占有数はsupercell実行時に 指定できるので、1のままで問題ない。

立方単位胞を各軸に2倍して得られる2× 2×2セル(96原子)の2原子をSiに置き換 えると、6.25%Siドープ系になる。この場合 のSiの可能な配置をsupercellで数え上げる には、 \$ supercell -i base.cif -s 2x2x2 -p "Si:p=2" -m -d

とする。ここで、-i base.cifが入力CIFファ イルの指定を、-s 2x2x2が与えたCIFファイ ルのセルに対し2×2×2倍のスーパーセルを 作ることを意味する。-p "Si:p=2"はSiサイ トのうち2個が占有されていることを意味 する(残りの30サイトはZrで占有される)。 -mは対称性を考慮して同じ構造をマージす ることを意味する。最後の-dはドライラン、 つまり配置の数を数え上げるだけで、配置そ のものは出力しないことを意味する。

上のコマンドを実行すると、以下のような 出力が得られるはずである(重要箇所を太字 で示した)。

Supercell program (v2.1)

(中略)

Group 1 (64 atomic positions in supercell): * Site #1: O (occ. 1) -> FIXED with occupancy 1.000.

Group 2 (32 atomic positions in supercell):

- * Site #1: Zr (occ. 1) -> distributed over 30 positions out of 32 (actual occ.: 0.938).
- * Site #2: Si (occ. 1) -> distributed over 2 positions out of 32 (actual occ.: 0.062). Number of combinations for the group is 496

Minimal distance between atoms of two distinct groups: 2.2041 A.

The total number of combinations is 496

1536 symmetry operation found for supercell. Total enumeration time: 0:00:0.00155212 **Combinations after merge: 5**

実際にドーパント配置を出力するには、-d を外して実行する:

\$ supercell -i base.cif -s 2x2x2 -p "Si:p=2" -m

これを実行すると、supercell_i [番号]_w [多重度].cifというファイルが出力される:

\$ Is supercell_i*_w*.cif supercell_i000_w192.cif supercell_i002_w192.cif supercell_i004_w16.cif supercell_i001_w48.cif supercell_i003_w48.cif

ここで多重度は、対称性により同じ構造と みなせる配置の数を意味する。得られた構造 を図7に示す。

こうして得られたCIFファイルは、Open BABELを使用すれば、DFT計算プログラム で使用可能なファイル形式に変換することが できる。例えば、

\$ obabel supercell_i000_w192.cif -O i000_w192.vasp\ -xw --title i000_w192

を実行すれば、VASP POSCAR形式のファ イルが得られる。ここで--titleはPOSCAR先 頭のタイトル行の文字列の指定である。また、 -xwを指定すると、原子番号順で並び替えて から出力される。

3. MI後工程

3.1 Pymatgenによる DFT 計算データの抽出

Pymatgen (Python Materials Genomics) [7] は、物質系シミュレーションの前工程・ 後工程のためのオープンソースpythonライ ブラリである。Pymatgenにはさまざまな機 能が実装されているが、物質構造を扱うため のpymatgen.core.structureや、DFT など物 質系シミュレーションの計算結果を読み込む ためのpymatgen.ioをおさえておけば、ここ



図7 2×2×2セルに導入したSi原子2個(大きめの球)の可能な配置
 黒い球がZr、白い球がOを表す。

で説明する処理を実行するには十分である。 ここでは、VASPのXMLファイルからデー タを抽出し保存する手順について説明する。 まず、XMLファイルを読み込むには、

from pymatgen.io.vasp.outputs import Vasprun # 必要なモジュールをインポート vrun = Vasprun('vasprun.xml', parse_potcar_file=False, \ parse_eigen=False, parse_dos=False) # 読み込み

とする。引数parse_potcar_fileはポテンシャ ルファイルの解析をするかどうか、parse_ eigenは固有値を読み込むかどうか、parse_ dosは電子状態密度を読み込むかどうかの指 定である。

こうして得られたvrunは、pymatgen.io. vasp.outputs.Vasprunオブジェクトとなっ ている。主な属性を以下に挙げる。

<計算条件設定>

incar:

計算条件設定ファイル INCAR の内容を辞書 形式で保存したもの parameters: 実際に使用された INCAR パラメータの辞 書。INCARで明示的に定義されてないパラ メータに対しても、割り当てられたデフォル ト値がわかる。 kpoints: k 点設定ファイル KPOINTSの中身 actual_kpoints: 実際の計算に用いたk点座標 actual_kpoints_weights: 各k点の重み付け potcar symbols: 計算に用いた擬ポテンシャルファイル POTCARの名前のリスト is hubbard: Hubbard Uが使われているかどうか hubbards: Hubbard Uの値(辞書形式) is_spin: スピン偏極が許されているかどうか run_type: "PBE+U"など、DFT計算タイプを表す文字 列

<計算の終了状態> converged_electronic: 電子状態が収束したがどうか converged_ionic: 構造最適化が収束したかどうか converged: 電子状態と構造最適化が共に収束したかどう か

<計算結果> final_energy: 最終ステップにおける全エネルギー final_structure: 最終ステップにおける構造(Structureオブ ジェクト) structures: 各ステップにおける構造(Structureオブ ジェクトのリスト) efermi: フェルミ準位のエネルギー eigenvalues: 固有値データ nionic_steps: 構造最適化にかかったステップ数 ionic steps: 構造最適化の各ステップの状態(辞書形式) を格納したリスト

ここで辞書(dict)というのは、{'key1':'value1', 'key2':'value2'} のように、キーと値のペアの セットからなる pythonのデータ形式である。 ここで示した属性のうちよく用いられる ものとして、計算後の全エネルギーfinal_ energyと、構造final_structureがある。また、 convergedの値をチェックすれば、計算が正 常に完了しなかったときの結果をデータセッ トから除外することができる。

Vasprunオブジェクトから必要なデータを 一つ一つ取得してもよいが、属性からアクセ スできないときや、必要なデータが決まっ ていないときは不便である。実はpymatgen のオブジェクトの多くはas_dict()というメ ソッドにより辞書化し、これをJSON形式 [32] で保存することができる。vrun.as_ dict()を実行して得られる辞書データのキー と値の例を表2に示す。このうち、inputと outputはそれ自身が辞書データであり、そ の中身は表3、表4のようになっている。例 えば、計算完了時の全エネルギーと構造は以 下のように取得できる。

vdict = vrun.as_dict() # 全体を辞書化 E = vdict['output']['final_energy'] # 全エネルギー sdict = vdict['output']['crystal'] # 構造(辞書データ)

Pymatgenで構造データを処理するには、 Structureオブジェクトに変換する必要があ る。as_dict()で辞書化したデータは、from_ dict()によりStructureオブジェクトに戻すこ とができる。

import pymatgen.core as mg
pymatgenの基本機能をインポート
s = mg.Structure.from_dict(sdict)
辞書データを構造データに戻す

3.2 多数のDFT計算結果の集計

次に、複数の実行結果をまとめて読み込ん で、目的としていた正方晶・単斜晶エネルギー 差を算出するまでの手順を説明する。例とし て、96原子モデルにSiまたはGeを二原子ドー プしたときの正方晶・単斜晶全配置に対する 実行結果が、以下のようなディレクトリーツ リーに収められている状況を考える。

ZrGe2	
├─── mono	
│	
│	
└─── tetragonal	
i000	
i001	

キー	值	説明
vasp_version	5.4.4	コードバージョン
has_vasp_completed	TRUE	計算が完了したかどうか
nsites	12	原子数
unit_cell_formula	{'Zr': 4.0, 'O': 8.0}	組成
reduced_cell_formula	{'Zr': 1.0, 'O': 2.0}	組成 (最小比)
pretty_formula	ZrO2	組成式
is_hubbard	TRUE	Hubbard Uの有無
hubbards	{'Zr': 4.0, 'O': 0.0}	Hubbard Uの値
elements	['O', 'Zr']	元素のリスト
nelements	2	元素の数
run_type	PBE+U	DFT計算タイプ
input	{'incar':	入力データ
output	{'ionic_steps':	出力データ

表 2 Vasprun.as_dict()の出力例

表 3 Vasprun.as_dict()['input']の出力物	表 3	表	Vasprun.as	_dict()['input'] の出力例
------------------------------------	-----	---	------------	-----------------------

キー	値	説明
incar	{'SYSTEM': 'ZrO2',	INCARパラメータの辞書
onuctal	{'@module': 'pymatgen.core.structure',	計算開始時の構造(辞書データ)
crystar	'@class': 'Structure',	
knointe	{ '@module': 'pymatgen.io.vasp.inputs',	k点データ
kpoints	'@class': 'Kpoints',	
nkpoints	4	実際のk点数
potcar	['Zr_sv', 'O']	POTCAR略称
notoor choo	[{'titel': 'PAW_PBE Zr_sv 04Jan2005',	POTCAR詳細名
potcar_spec	'hash': None},]	
potcar_type	['PAW_PBE', 'PAW_PBE']	POTCARタイプ
parameters	{'SYSTEM': 'ZrO2',	INCARパラメータの辞書(詳細版)
latting rag	{'@module': 'pymatgen.core.lattice',	逆格子
lattice_rec	'@class': 'Lattice',	

表 4	Vasprun.as_	_dict()['output']の出力例	(一部省略)
-----	-------------	-----------------------	--------

キー	値	説明
ionic_steps	[{},{]	各ステップの状態(辞書データのリスト)
final_energy	-105.999 eV	計算終了時の全エネルギー
final_energy_per_atom	-8.833 eV	計算終了時の一原子たりエネルギー
	{'@module':	計算終了時の構造(辞書データ)
crystal	'pymatgen.core.structure',	
	'@class': 'Structure',	
efermi	6.94	フェルミエネルギー
eigenvalues	(省略)	全固有値データ
bandgap	4.62	バンドギャップ
cbm	11.24	伝導帯下端
vbm	6.62	価電子帯上端
is_gap_direct	TRUE	直接遷移型かどうか

ZrSi2	
mono	
i000	
i001	
└─── tetragonal	
├─── i000	
i001	

ここで ZrSi2 がSiドープ系、ZrGe2 がGe ドープ系、mono が単斜晶、tetragonal が正 方晶を意味する。i000、i001、...が個々のドー パント配置に対する実行結果を収めたディレ クトリである。

まず、XMLファイルのパスをglobにより 取得する。

import glob

files = sorted(glob.glob('Zr*2/*/i???/vasprun.xml')) # XMLファイルパスのリスト

globで指定するパターン文字列は、シェル で使用するパターンと同じである。?が任意 の一文字、*が任意の0文字以上にマッチす る。globで得られたリストは名前順になって いないので、sorted()によりソートして、変 数filesに格納する。

次に、全てのXMLファイルを読み込み、 辞書化したものをリストvdictsに格納する。

vdicts = [Vasprun(file).as_dict() for file in files]

vdictsの各要素は、一つの配置に対する計 算実行結果を辞書化したものになっている。 このリストは、そのままJSON形式で書き出 すことができる。

```
with open('results.json','w') as fp:
# 出力先をresults.jsonとする
json.dump(vdicts, fp)
```

一般にはJSON内の各データが同じデー タ構造を持っているとは限らないが、ここで は各実行結果がどれも同じデータ構造で保 存されているので、pandasの表形式データ (pandas.DataFrame)として読み込むこと ができる。

import pandas as pd # pandasをインポート df = pd.read_json('results.json', orient='record') # 読み込み

もちろん、vdictsからデータフレームに直 接変換することもできる。

df = pd.DataFrame(vdicts)

各実行結果のドープ組成 (class)・結晶相 (phase)・ドーパント配置 (layout) の情報 も含める必要がある。ここでは、ファイルパ スからこれらのデータを取得する。

df['file'] = files	#globで取得したファイルパスの
リストを列 fileに	収める
df[['class','phase',	'layout']] = df['file'].str.split
('/',expand=True).	loc[:,:2]

二行目のstr.split()により、列fileに収め られているファイルパスを''を区切り文字と して分解する。オプション expand=Trueは、 分解した結果を別々の列として返すことを意 味する。これから最初の三列を loc[:,:2] で取 り出して、列class、phase、layoutに収める。

構造最適化後のエネルギーは列outputの 中にあるので、これを取り出して列 final_ energyにセットする。

df['final_energy'] = df ['output'].apply(lambda x: x['final_energy'])

dfのうち、エネルギー差の算出に必要な

	class	phase	layout	energy
0	ZrGe2	mono	i000	-834.204
1	ZrGe2	mono	i001	-833.859
23	ZrGe2	tetragonal	i000	-833.644
24	ZrGe2	tetragonal	i001	-831.858
32	ZrSi2	mono	i000	-841.252
33	ZrSi2	mono	i001	-840.481
55	ZrSi2	tetragonal	i000	-841.320
56	ZrSi2	tetragonal	i001	-838.133

このdfに対し、

df_min = df.groupby(['class','phase'])['energy'].min(). unstack()

を実行すると、組成・結晶相を同じくするグ ループに分類されて、それぞれにおける最安 定配置のエネルギーが求められる。さらに unstack()により、縦軸が組成、横軸が結晶 相の表に変換されて、変数 df_min に代入さ れる。df_min の中身は、

phase	mono	tetragonal
class		
ZrGe2	-834.344	-834.297
ZrSi2	-841.352	-841.812

のようになる。正方晶・単斜晶エネルギー差 を求めるには、

df_min['mono'] - df_min['tetragonal']

を実行する。結果は以下のようになる。

class ZrGe2 -0.048 ZrSi2 0.460 この値が大きいほど、正方晶が安定化され ていることを意味する。

4. 計算結果の可視化

ハイスループット計算では膨大なデータが 得られるが、数値を眺めているだけでは有用 な知見は得られない。まずはデータ点を2~ 3次元空間にプロットして、大まかな分布を 把握することが重要である。可視化すること で計算ミスがみつかったり、意外な規則性・ 法則性に気づくこともある。

プロットの軸として使用する変数は、物質 に対する知識をもとに選ぶこともあるし、主 成分分析などの機械学習手法で合成すること もある。一例として、正方晶安定化効果(単 斜晶・正方晶エネルギー差)を縦軸、各組成 の最小エネルギー配置における正方晶単位胞 体積を横軸としたときのプロットを図8に示 す。このプロットの上に位置する点ほど、正 方晶安定化効果が大きい組成であることを意 味する(説明のため原論文[5]とは上下を 逆にしている)。

ジルコニアの場合、以下のような傾向と なっている。まずSiとGeでは、顕著な正方 晶安定化効果が得られる。これらはドープ濃 度が上がるほど正方晶安定化効果が大きく なっている。また、単位胞体積はドープ濃度 上昇に伴い減少しており、これらが第3~4 周期の比較的小さい原子であることと関係し ていると見られる。一方、同じIV族元素で も、Ti、Sn、Ceの正方晶安定化効果はSiや Geに比べると小さく、ドープ濃度を上げて もそれほど改善しない。単位胞体積はこの順 に大きくなっており、原子サイズから予期さ れるとおりである。Hfドープでは正方晶安 定性・単位胞体積ともほとんど変わっておら ず、HfとZrの類似性から予想されるとおり である。II族ドープ (Mg、Ca、Sr、Ba、Zn) では、原子サイズ増大に伴い単位胞体積が増 加し、正方晶安定性が緩やかに改善する傾向



図8 ジルコニア(左)とハフニア(右)に対する正方晶安定化効果と単位胞体積 灰色の十字線はアンドープ系の値を示す。Si2やGe3などの数字は、2×2×2セルに導入したドー パント原子数に対応する。×はIII-Vドープのデータ点。

となっている。III族ドープ(Al、Y、La)で は、元素により単位胞体積は大きくかわるが、 正方晶安定性はそれほど差がない。III-Vドー プ(図の×)では、データ点が広く分布して おり、明確な傾向が見られない。Ge二原子 ドープに匹敵する安定化効果を示すのは、第 3~4周期元素の組み合わせ、つまり、Al、 GaとP、Asの組み合わせである(グラフの Ge2のまわり4点)。以上の傾向から、原子 サイズと価数の効果が支配的であることがわ かる。

ハフニアもほぼ同じ傾向となっている。Si とGeで正方晶安定化効果が大きく、ドープ 濃度を上げると更に大きくなる。Ti、Sn、Ce はSiやGeに比べると効果が薄い。Zrはほと んど効果がない。II族では原子サイズが大き くなるほど単位胞体積が増加し、正方晶安定 性が緩やかに改善する。Ge二原子ドープ近 傍の4点はAl、GaとIn、Pの組み合わせであ る。ただし、Si二原子ドープから三原子ドー プに増やしたときに単位胞体積が増加した り、III族のYとLaで正方晶安定性に違いが あるなど、幾つか相違点が見られる。原子サ イズや価数だけでは説明できない要素がある ことを示唆している。

5. おわりに

本稿では、マテリアルズ・インフォマティ クスの実例として、high-k物質であるジル コニア・ハフニアの相安定化ドーパントに 関するハイスループットDFT計算[5]を紹 介した。特に、supercellツールによるドー パント配置の全生成や、DFT計算出力の pymatgenによる読み込み、pandasを用いた データ集計など、MIの前工程・後工程の基 本的技術について重点的に解説した。

DFT 計算により得られたデータから、Si では低濃度でも正方晶安定化効果があること や、第3~4周期のIII-V族元素同時ドープ によりGeと同等の効果が得られることが示 唆された。もっとも、III-VドープではSiに 匹敵する効果は得られていない。価数や原子 サイズだけでは説明できない効果を考慮しな ければ、Siを超えるドーパントは見つけるこ とができないと思われる。

謝辞

筑波大学重田先生、奈良先端大学原嶋先生、 東京エレクトロン守屋様、Zeyuan Ni様をは じめ、本解説記事の掲載を許諾くださいまし た関係者の方々に心よりお礼申し上げます。

参考文献

- [1] 一般財団法人高度情報科学技術研究機構定款;https://www.rist.or.jp/public_ info/teikan
- [2] 防衛装備庁安全保障技術研究推進制度 研究課題「高強度CNTを母材とした 耐衝撃緩和機構の解明と超耐衝撃材の 創出」(代表・筑波大学藤田淳一教授、 令和1~5年度);同研究課題「実験・ 計算科学の融合による革新的塗膜創成 と機序解明の基礎研究」(代表・GSI クレオス柳澤隆氏、令和5年度~)
- 【3】 大越孝洋、大越昌樹、長尾宣明、四 橋聡「ハイスループット計算と機械 学習を活用した光機能性有機材料の 新規構造探索」Panasonic Technical Journal 67, 78 (2021); https:// holdings.panasonic/jp/corporate/ technology/technology-journal/pdf/ v6702/p0116.pdf (パナソニック株式 会社との共同研究)
- [4] 河東田道夫「ケモインフォマティクス 実課題のための量子機械学習法の共
 同研究」RIST NEWS 67, 16 (2021);
 https://www.rist.or.jp/rnews/67/67s3.
 pdf (Quemix Inc.との共同研究)
- [5] Y. Harashima, H. Koga, Z. Ni, T. Yonehara, M. Katouda, A. Notake, H. Matsui, T. Moriya, M.K. Si, R. Hasunuma, A. Uedono, Y. Shigeta, "Systematic Search for Stabilizing Dopants in ZrO₂ and HfO₂ Using First-Principles Calculations," IEEE Trans. Semicond. Manuf. 36,

543 (2023). https://doi.org/10.1109/ TSM.2023.3265658

- [6] https://orex.github.io/supercell/;
 K. Okhotnikov, T. Charpentier,
 S. Cadars, "Supercell program: a combinatorial structure-generation approach for the local-level modeling of atomic substitutions and partial occupancies in crystals," J. Cheminform. 8, 17 (2016). https://doi.org/10.1186/s13321-016-0129-3
- [7] https://pymatgen.org; S.P. Ong,
 W.D. Richards, A. Jain, G. Hautier,
 M. Kocher, S. Cholia, D. Gunter,
 V.L. Chevrier, K.A. Persson, G.
 Ceder, "Python Materials Genomics
 (pymatgen) : A Robust, OpenSource Python Library for Materials
 Analysis," Comp. Mater. Sci. 68,
 314 (2013). https://doi.org/10.1016/
 j.commatsci.2012.10.028
- [8]本稿では手順の説明のためにいくつか コマンドやスクリプトを掲載してい る。筆者の環境で動作確認は行なって いるが、幅広い条件下での動作を保証 するものではない。もし使用される場 合は、十分にテストすることを推奨す る。
- [9] 岩崎悠真「マテリアルズ・インフォマ ティクス―材料開発のための機械学習 超入門」日刊工業新聞社、2019。
- [10] W. Kohn, "Nobel Lecture: Electronic structure of matter-wave functions and density functionals," Rev. Mod. Phys. 71, 1253 (1999). https://doi. org/10.1103/RevModPhys.71.1253
- [11] 通常の関数が入力された値に対し値を 出力するのに対し、汎関数は関数を入 力として値を出力とする。関数の関数 と呼ばれることもある。

- [12] P. Hohenberg, W. Kohn, "Inhomogeneous electron gas," Phys. Rev. 136, B864 (1964). https:// doi.org/10.1103/PhysRev.136.B864
- J.P.Perdew, K.Burke, M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. 77, 3865 (1996). https:// doi.org/10.1103/PhysRevLett.77.3865
- [14] D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism," Phys. Rev. B 41, 7892 (1990). https://doi. org/10.1103/PhysRevB.41.7892
- P.E. Blöchl, "Projector augmentedwave method," Phys. Rev. B 50, 17953 (1994). https://doi.org/10.1103/ PhysRevB.50.17953
- [16] https://www.quantum-espresso. org; Giannozzi et al., "Advanced capabilities for materials modelling with Quantum ESPRESSO," J. Phys.: Condens. Matter 29, 465901 (2017). https://doi.org/10.1088/1361-648X/aa8f79
- [17] https://www.vasp.at; G. Kresse, J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," Phys. Rev. B. 54, 11169 (1996). https://doi.org/10.1103/ PhysRevB.54.11169
- [18] M.T. Bohr, R.S. Chau, T. Ghani, K. Mistry, "The high-k solution," IEEE Spectrum, Oct 2007; https:// spectrum.ieee.org/the-highk-solution
- [19] S. Kol, A.Y. Oral, "Hf-based high- κ dielectrics: A review," Acta Phys. Pol. A. 136, 873 (2019). https://doi. org/10.12693/APhysPolA.136.873

- [20] J. Xie, Z. Zhu, H. Tao, S. Zhou, Z. Liang, Z. Li, R. Yao, Y. Wang, H. Ning, J. Peng, "Research progress of high dielectric constant zirconiabased materials for gate dielectric application," Coatings 10, 698 (2020). https://doi.org/10.3390/ coatings10070698
- [21] ランタノイドのf電子は内殻電子による遮蔽効果を受けにくいため、核に強く引きつけられる。その結果、ランタノイドの後にくるHfなどは、その前の 周期の元素と原子サイズが近くなる。
- [22] R.H. Nielsen, G. Wilfing, "Zirconium and zirconium compounds," Ullmann's Encyclopedia of Industrial Chemistry, Wiley, 2010. https://doi.org/10.1002/14356007. a28_543.pub2
- [23] R.H. Nielsen, G. Wilfing, "Hafnium and hafnium compounds," Ullmann's Encyclopedia of Industrial Chemistry, Wiley, 2010. https://doi.org/10.1002/14356007. a12_559.pub2
- [24] 本稿では結晶構造の表示にVESTAを 用いている。K. Momma, F. Izumi, "VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data," J. Appl. Crystallogr. 44, 1272 (2011). https:// doi.org/10.1107/S0021889811038970
- [25] D. Fischer, A. Kersch, "The effect of dopants on the dielectric constant of HfO₂ and ZrO₂ from first principles," Appl. Phys. Lett. 92, 012908 (2008). https://doi.org/10. 1063/1.2828696
- [26] Ceはランタノイドであるが、ZrO₂と 類似の4価の酸化物を作るので、ここ

ではIV族に含めている。

- [27] 松下能孝「結晶構造データベースと 結晶学共通データ・フォーマット CIFについて」J. Surf. Analysis 21, 71 (2014). https://doi.org/10.1384/ jsa.21.71
- [28] https://next-gen.materialsproject. org; Jain et al., "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," APL Mater. 1, 011002 (2013). https://doi.org/10. 1063/1.4812323
- [29] http://www.crystallography.net/cod/
- [30] データベースによりCIFの記法に若干

の相違があるので、要注意である。泉 富士夫、宮崎晃平「CIFを出発点と する第一原理計算支援用ユーティリ ティー」セラミックス54、473 (2019).

- [31] http://openbabel.org/index.html; N.M.
 O'Boyle, M. Banck, C.A. James, C.
 Morley, T. Vandermeersch, G.R.
 Hutchison, "Open Babel: An open chemical toolbox," J. Cheminf. 3, 33 (2011). https://doi.org/10.1186/1758-2946-3-33
- [32] 辞書形式のデータをテキストデータとして保存したもの。標準的なデータ交換形式として様々な言語でサポートされている。