Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.

Supporting Information

Gapped metals as thermoelectric materials revealed by high-throughput screening.

Francesco Ricci, Alexander Dunn, Anubhav Jain, Gian-Marco Rignanese, Geoffroy Hautier

July 20, 2020

1 Full dataset description

The full dataset of all gapped metals found in this screening is divided in four files according to the number of elements in the compounds: data_binaries.xls, data_ternaries.xls, data_quaternaries.xls, data_quinaries.xls. Each file contains 4 sheets with data relative to different temperatures (i.e. 300 K, 600 K, 900 K, 1200 K). The list of the labels used in those file and their meaning is explain in Table 1. Another file is also provided: data_groupped_binaries_ternaries_600K.xls. This file contains similar data at only 600 K for binaries and ternaries gapped metals, but the compounds are groupped by same composition and same space group. We remind that complete transport data for all these materials can be found in Ref. [1], the band strucuture and other details in the MP website.

2 Automatminer

Models for predicting minimum lattice thermal conductivities in the high temperature limit (κ_{min}) were trained and validated using Automatminer v2019.3.27b0 [2, 3] to predict the elastic moduli from crystal structure. Separate Automatminer pipelines were fit for shear moduli and bulk moduli indepedently. The datasets consisted of elastic moduli sets for 10,987 compounds computed with DFT-GGA gathered from the Materials Project (MP) on 2019/04/19; the dataset was cleaned to remove unphysical (i.e., negative) elastic moduli and compounds containing noble gas elements. The datasets are also freely accessible through the Matminer [4] data retrieval tools under the names "matbench_log_gvrh" and "matbench_log_kvrh". All elastic moduli gathered from MP use the Voigt-Reuss-Hill (VRH) approximation to determine elastic constants from the full elastic tensor. The Automatminer v2019.3.27b0 "express" preset was used for both pipelines; after instantiation of the preset, the pipeline training and configuration process is entirely automatic. Composition features were automatically generated from matminer featurizers [4] including elemental statistics from MagPie [5, 6], ionic properties, electron affinities, and oxidation states, and structure features were automatically generated from the Sine Coulomb Matrix [7], global symmetry features, Global Instability Index [8], Ewald energy, and crystal density. Feature matrices were imputed and then had the overall dimensionality reduced. Feature reduction was

Label	Description	Units
#mp_id	Materials Project identifier	
#formula	Chemical formula	
#spg	Crystal space group	
#eah	Energy above hull	eV
#zt_T	Optimal zT calculated at temperature T (K)	
#opt_pf_T	Optimal PF calculated at temperature T (K)	μW/K^2 m
#sbk@pf_T	Seebeck coefficient value at optimal PF	μV/K
#cond@pf_T	Conductivity value at optimal PF	$(\Omega\ \mathrm{m})^{-1}$
#conc@pf_T	Carrier concentration at optimal PF	cm^{-3}
#kappa@pf_T	Electrical contribution to thermal conductivity at optimal PF	W/K m
#L_factor	Lorenz factor value at optimal PF	W Ω/K^2
#k_min	Minimal thermal conductivity	W/K m
#gap	Closest energy gap to the Fermi level	eV
#mu_gap	Distance of the Fermi level from VBM (-) or CBM (+)	eV
#mu@pf_T	Distance of the Fermi level from VBM (-) or CBM (+) at optimal PF	eV
#n_el	Number of electrons to add (-) or remove (+) to have a semiconductor	
#conc	n_el divided by the volume of the primitive cell	
#spin_pol	If the band structure is spin polarized	
#U	If the U correction has been used	
#Composition	Element composition	
#type	n- or p-type behavior according to the position of the Fermi level	

Table 1: Labels used in the data files and their description.

done first by removing redundant features having more than 0.95 Pearson correlation with existing features; next, a tree-ensemble method retained features at or higher than the 99th percentile of feature importance. Finally, TPOT (see [9]) searched the Automatminer "express" model space, which contains internal pipelines of several normalization techniques and the following models and corresponding hyperparameter grids: Gradient Boosted trees, Random Forest, Extra Trees, k-Nearest Neighbors, LinearSVR, ElasticNet, and Lasso-Least Angle Regression (implemented in scikit-learn [10]). The TPOT optimizer searched the model space using a genetic algorithm subject to a 24 hour total time constraint. The Automatminer pipelines determine relevant features and the optimal internal machine learning pipelines automatically; the features (derived from matminer featrizer feature names) and model parameters (from scikit-learn) are given in Tables 2, 3, 4, and 5. The models' training/testing split was 80%/20%, with an internal 20% (16% of total) validation set utilized by TPOT for all model selection. After model selection, the best found TPOT models were retrained on the entire training dataset before evaluation. The final models resulted in an mean absolute errors on the elastic moduli test sets of 7.88 GPa for K_{VRH} and 11.49 GPa for G_{VRH}. Propagating the predicted elastic moduli through the κ_{min} calculation, we obtained a MAE of 0.0612 W/m·K; the test set mean average deviation is 0.370 W/m·K. Approximately 93% of predictions for the test set are within 20% of the true κ_{min} values. Further details on the ML pipelines can be found in the open source Automatminer code repository [2].

Feature name	**Description**
`MagpieData mean MeltingT`	Mean melting temperature among elements in composition
`MagpieData maximum MendeleevNumber`	Maximum Mendeleev number among elements in composition
`MagpieData mean NUnfilled`	Mean number of unfilled valence orbitals among elements
`MagpieData minimum NValence`	Minimum number of valence electrons among elements
`MagpieData avg_dev NUnfilled`	Average deviation of unfilled valence orbitals among elements
`MagpieData mean Electronegativity`	Mean electronegativity among elements in composition
`MagpieData maximum GSvolume_pa`	Maximum DFT-computed volume of elemental solid among elements in composition
`MagpieData mode MeltingT`	Mode melting temperature among elements in composition
`MagpieData minimum MendeleevNumber`	Minimum Mendeleev number among elements in composition
`MagpieData mode NUnfilled`	Mode number of unfilled valence orbitals among elements
`MagpieData maximum MeltingT`	Maximum melting temperature among elements in composition
`MagpieData mean CovalentRadius`	Mean covalent radius among elements in composition
`MagpieData avg_dev MeltingT`	Average deviation of melting temperature among elements in composition
`MagpieData mode MendeleevNumber`	Mode Mendeleev number among elements in composition
`MagpieData mean NpValence`	Mean number of filled valence p orbitals among elements in composition
`MagpieData mean GSbandgap`	Mean DFT bandgap of elemental solid among elements in composition
`MagpieData mean Row`	Maximum periodic table row among elements in composition
`MagpieData minimum Electronegativity`	Minimum electronegativity among elements in composition
`MagpieData mean GSvolume_pa`	Mean DFT-computed volume of elemental solid among elements in composition
`MagpieData minimum Column`	Minimum periodic table column among elements in composition
`MagpieData avg_dev MendeleevNumber`	Average deviation of Mendeleev number among elements in composition
`packing fraction`	Packing fraction derived from crystal structure
`spacegroup_num`	Spacegroup number of the crystal structure
`vpa`	Voiume per atom of the crystal structure
`sine coulomb matrix eig 0`	Eigenvalue 0 derived from the Sine Coulomb Matrix from crystal structure
`density`	Density derived from crystal structure

Table 2: Features retained in the final automatminer pipeline for predicting bulk modulus.

Operation	Hyperparameters	Description
`SelectFwe`	$\alpha = 0.032$, `f_classif` scoring	Select p values according to Family-wise error
`MinMaxScaler`	n/a	Scale input linearly according to minimum/maximum of training set
`ExtraTreesRegressor`	MSE split criterion, 0.85 max split features, $n_{\text{estimators}} = 200$	Extra Trees regression model

Table 3: Internal TPOT learning pipeline used by Automatminer for predicting bulk modulus.

Feature name	**Description**
`MagpieData mode Electronegativity`	Mode electronegativity among elements in composition
`MagpieData mean CovalentRadius`	Mean covalent radius among elements in composition
`MagpieData mean NsUnfilled`	Mean number of unfilled s valence orbitals among elements
`MagpieData mean AtomicWeight`	Mean atomic weight among elements
`MagpieData avg_dev GSvolume_pa`	Average deviation of DFT-computed volume of elemental solids per composition
`MagpieData mean GSvolume_pa`	Mean DFT-computed volume of elemental solid among elements in composition
`MagpieData maximum MeltingT`	Maximum melting temperature among elements in composition
`MagpieData mean MeltingT`	Mean melting temperature among elements in composition
`MagpieData minimum NUnfilled`	Minimum number of unfilled valence orbitals among elements in composition
`MagpieData avg_dev MendeleevNumber`	Average deviation of Mendeleev numbers among elements in composition
`MagpieData mode MeltingT`	Mode melting temperature among elements in composition
`MagpieData mean MendeleevNumber`	Mean Mendeleev numbers among elements in composition
`MagpieData mode MendeleevNumber`	Mode Mendeleev numbers among elements in composition
`MagpieData maximum MendeleevNumber`	Maximum Mendeleev numbers among elements in composition
`MagpieData mean NUnfilled`	Mean number of unfilled valence orbitals among elements in composition
`MagpieData mean NpValence`	Mean number of filled valence p orbitals among elements in composition
`MagpieData avg_dev MeltingT`	Average deviation of melting temperature among elements in composition
`MagpieData mean NsValence`	Mean number of filled s orbitals among elements in composition
`MagpieData avg_dev NUnfilled`	Average deviation of number of unfilled valence orbitals among elements
`MagpieData mean Electronegativity`	Mean electronegativity among elements in composition
`MagpieData minimum MendeleevNumber`	Minimum Mendeleev numbers among elements in composition
`MagpieData avg_dev NpUnfilled`	Average deviation of number of unfilled p orbitals among elements in composition
`MagpieData avg_dev CovalentRadius`	Average deviation of covalent radii among elements in composition
`vpa`	Volume per atom of the crystal structure
`sine coulomb matrix eig 0`	Eigenvalue 0 derived from the Sine Coulomb Matrix from crystal structure
`sine coulomb matrix eig 2`	Eigenvalue 2 derived from the Sine Coulomb Matrix from crystal structure
`sine coulomb matrix eig 3`	Eigenvalue 3 derived from the Sine Coulomb Matrix from crystal structure
`density`	Density derived from crystal structure
`packing fraction`	Packing fraction of crystal structure
`spacegroup_num`	Spacegroup number derived from crystal structure

Table 4: Features retained in the final automatminer pipeline for predicting shear modulus.

Operation	Hyperparameters	Description
`VarianceThreshold`	$\sigma_{\text{thresh}} = 0.2$	Remove low variance features according to σ_{thresh}
`ZeroCount` [TPOT]	n/a	Add count of zeros and nonzeros (among features) per sample
`GradientBoostingRegressor`	$\alpha = 0.99$, Friedman MSE criterion, $r_{\text{learning}} = 0.1$, max depth of 7, 500 estimators, least squares loss, 40% features per split, 65% subsampling	Gradient Boosted Trees regression model

Table 5: Internal TPOT learning pipeline used by Automatminer for predicting shear modulus.

3 Fermi level distributions in gapped metals

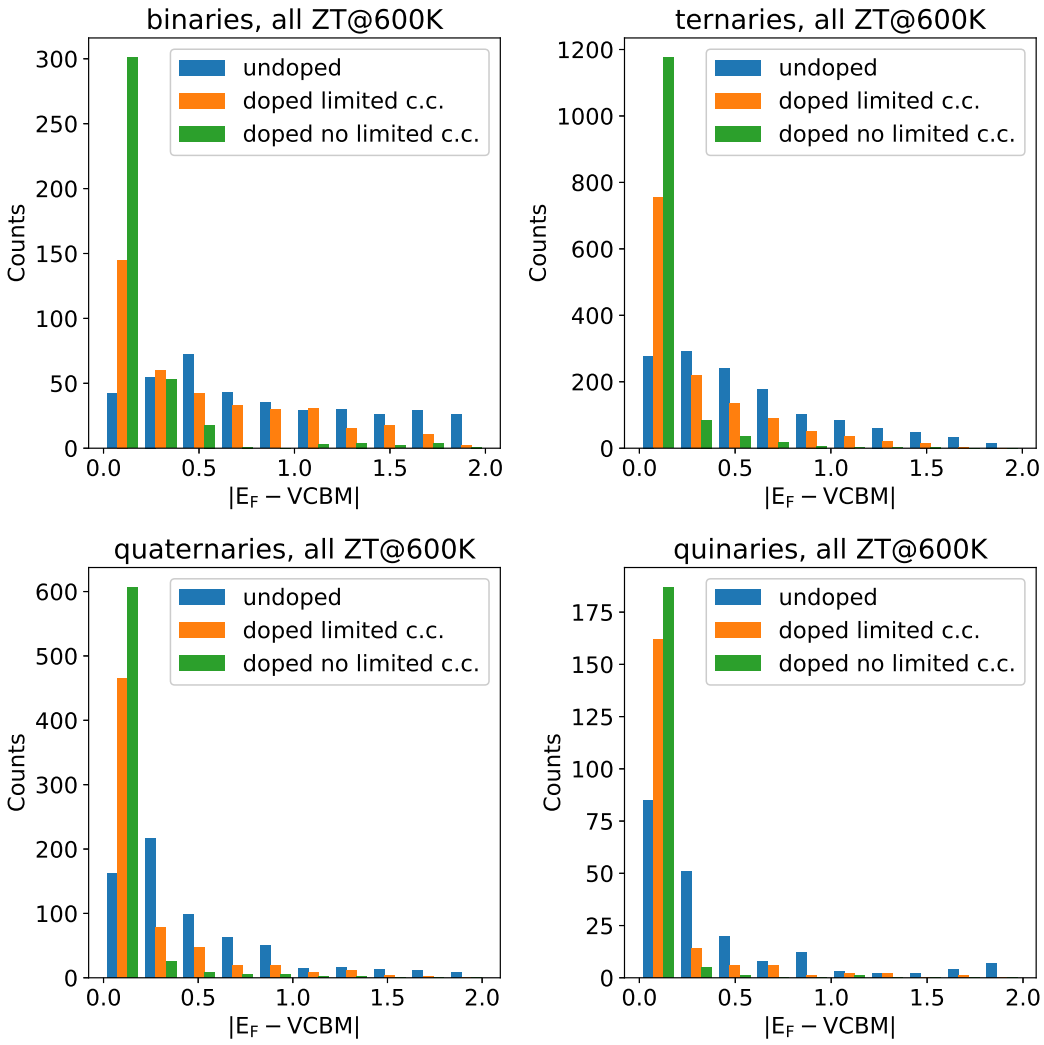

Figure 1:

Distribution of the Fermi level in the gapped metals with no filter on the optimal ZT applied. Blue bars represent the undoped state, the orange bars the optimized case with the limit of the carrier concentration at 10^{22} cm^{-3}, and green bars the optimized case without that limit.

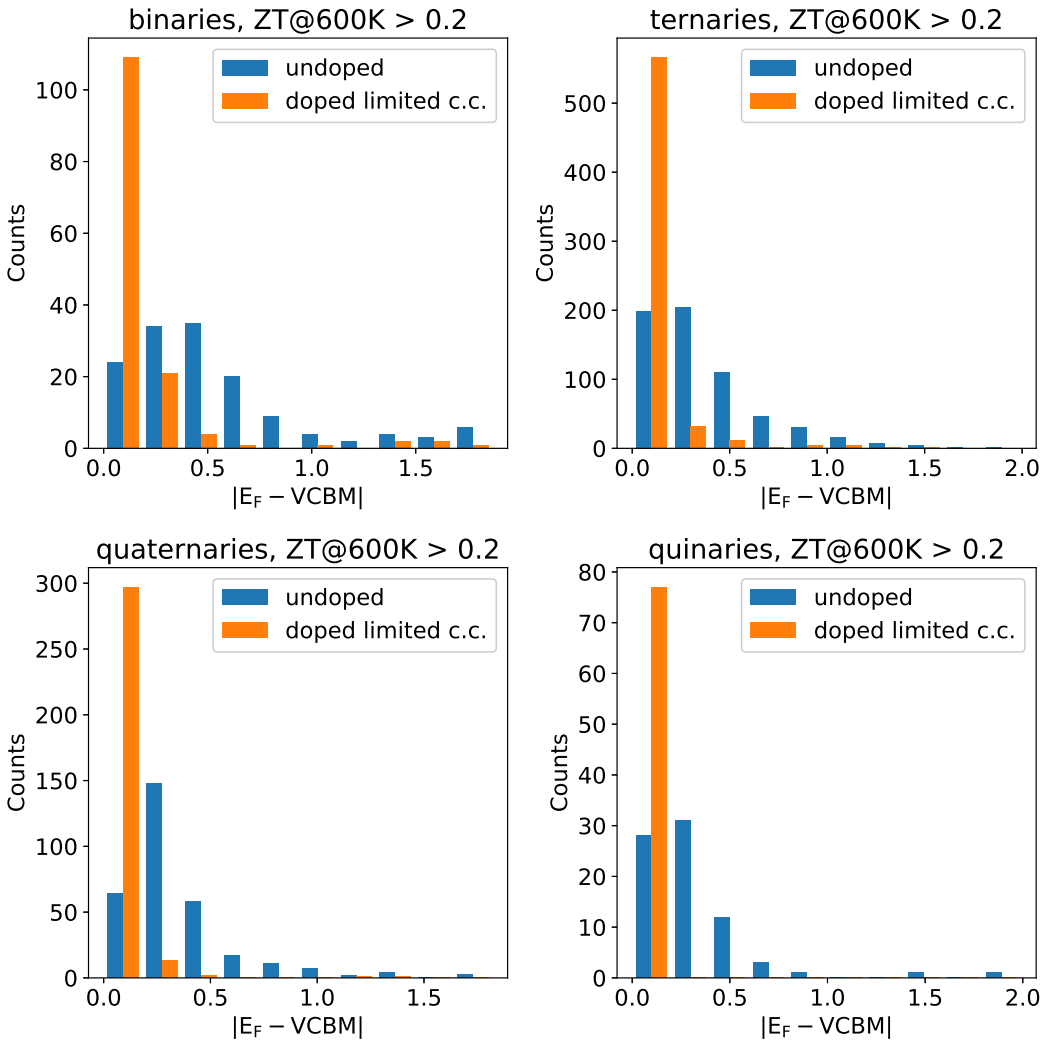

Figure 2:

Distribution of the Fermi level in the gapped metals with a ZT $>$ 0.2 at 600 K. Blue bars represent the undoped state, the orange bars the optimized case with the limit of the carrier concentration at 10^{22} cm^{-3}.

4 Optimal ZT vs temperature in gapped metals

T	Binaries	Ternaries	Quaternaries	Quinaries
300	2	2	1	0
600	17	13	8	9
900	13	13	14	13
1200	68	72	77	78

Table 6: Percentage of compounds that have the maximum optimal ZT at the different temperatures considered in this screening. Only materials with a ZT $>$ 0.2 at 600 K are considered.

5 Plots related to gapped metals taken as reference.

We report here the the crystal structure, the band structure, and the transport properties of La_3Te_4, Mo_3Sb_7, $Yb_{14}MnSb_{11}$, and NbCoSb.

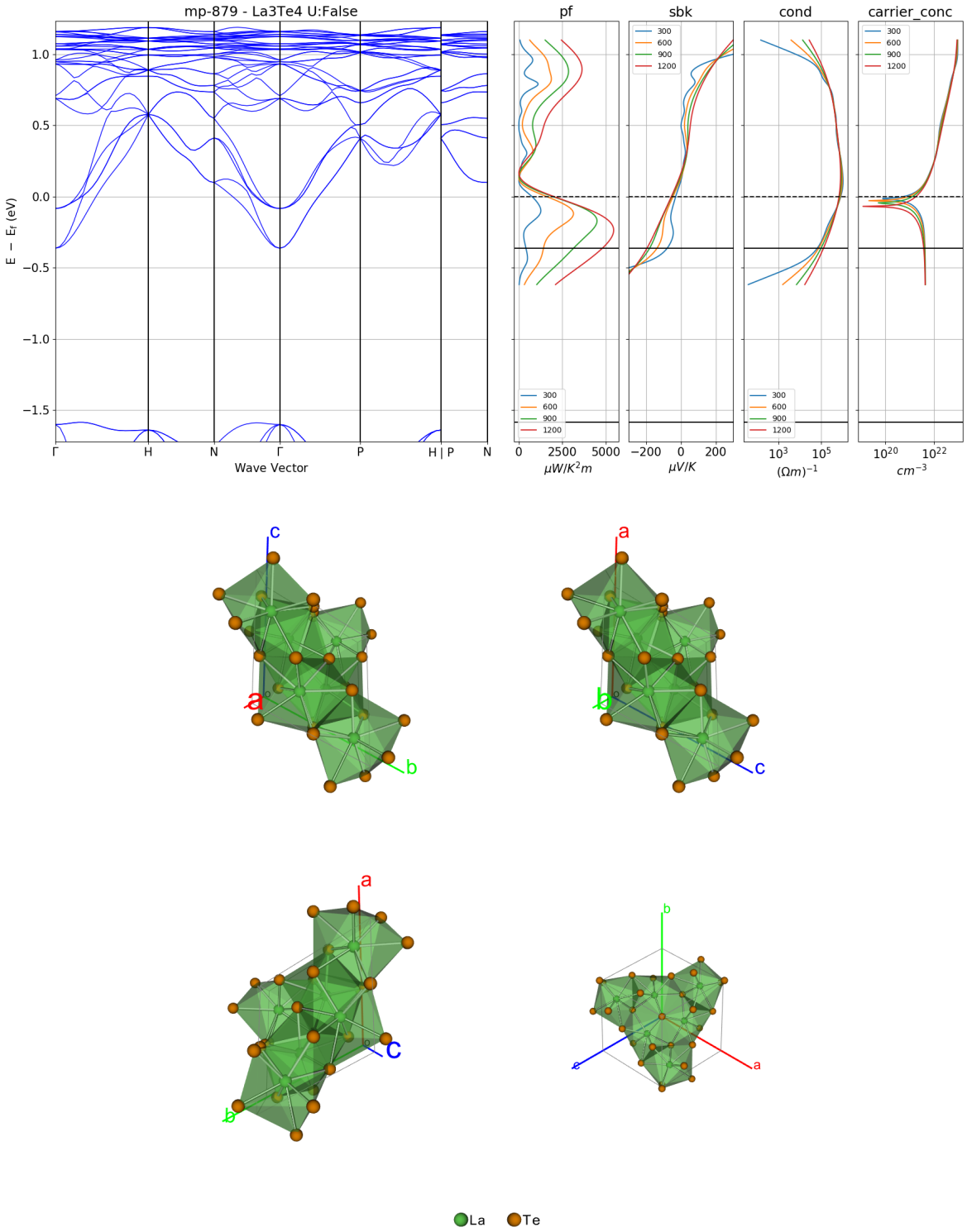

Figure 3: Structure (bottom), band structure, and calculated transport properties (top) of La_3Te_4 (mp-879).

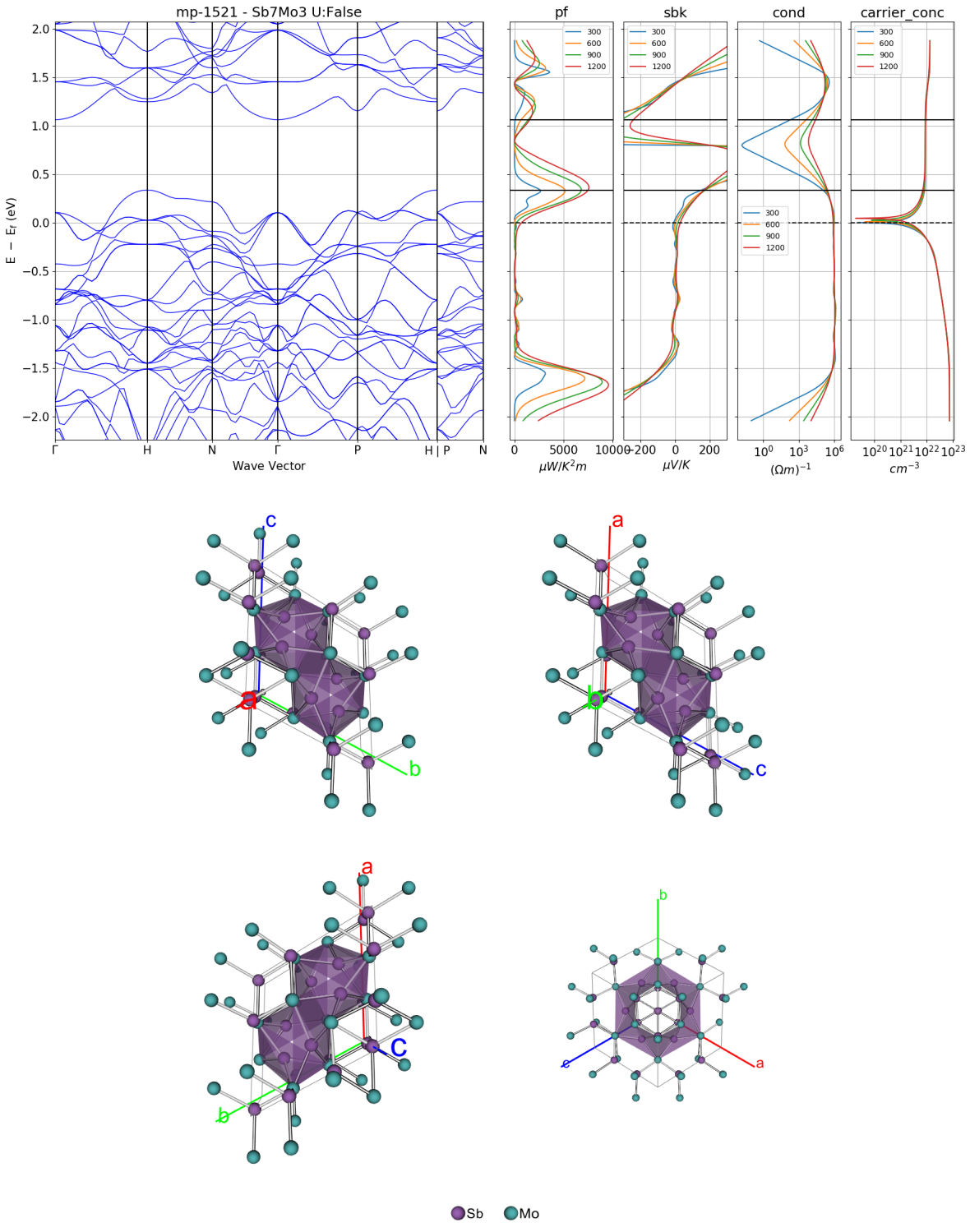

Figure 4: Structure (bottom), band structure, and calculated transport properties (top) of Sb_7Mo_3 (mp-1521).

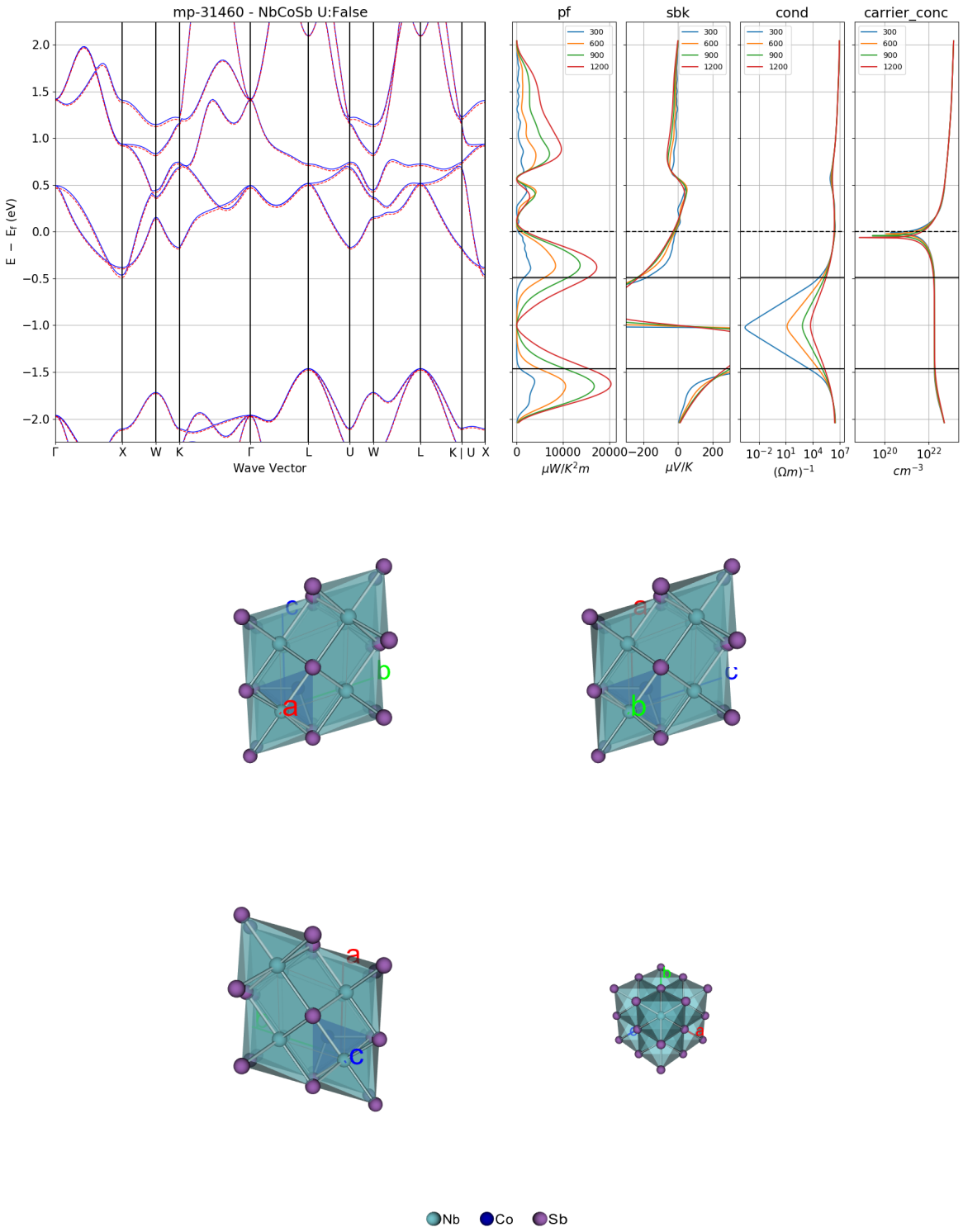

Figure 5: Structure (bottom), band structure, and calculated transport properties (top) of NbCoSb (mp-31460).

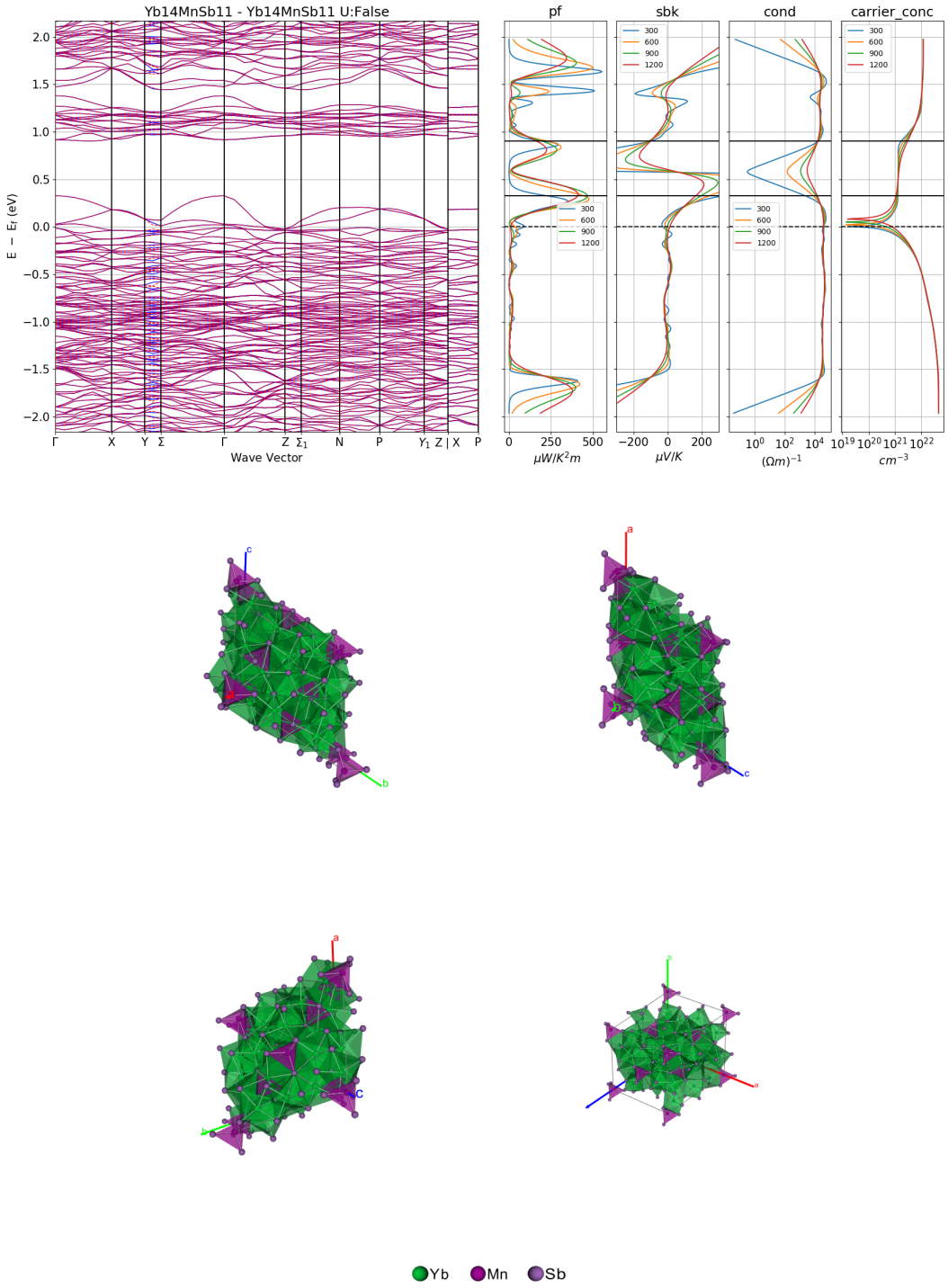

Figure 6: Structure (bottom), band structure, and calculated transport properties (top) of $Yb_{14}MnSb_{11}$.

6 Plots relate to gapped metals with complex band structure.

We report here the crystal structure, the band structure, and the transport properties of $RuBr_3$, an example of gapped metals we excluded from our discussion because of its complex band structure.

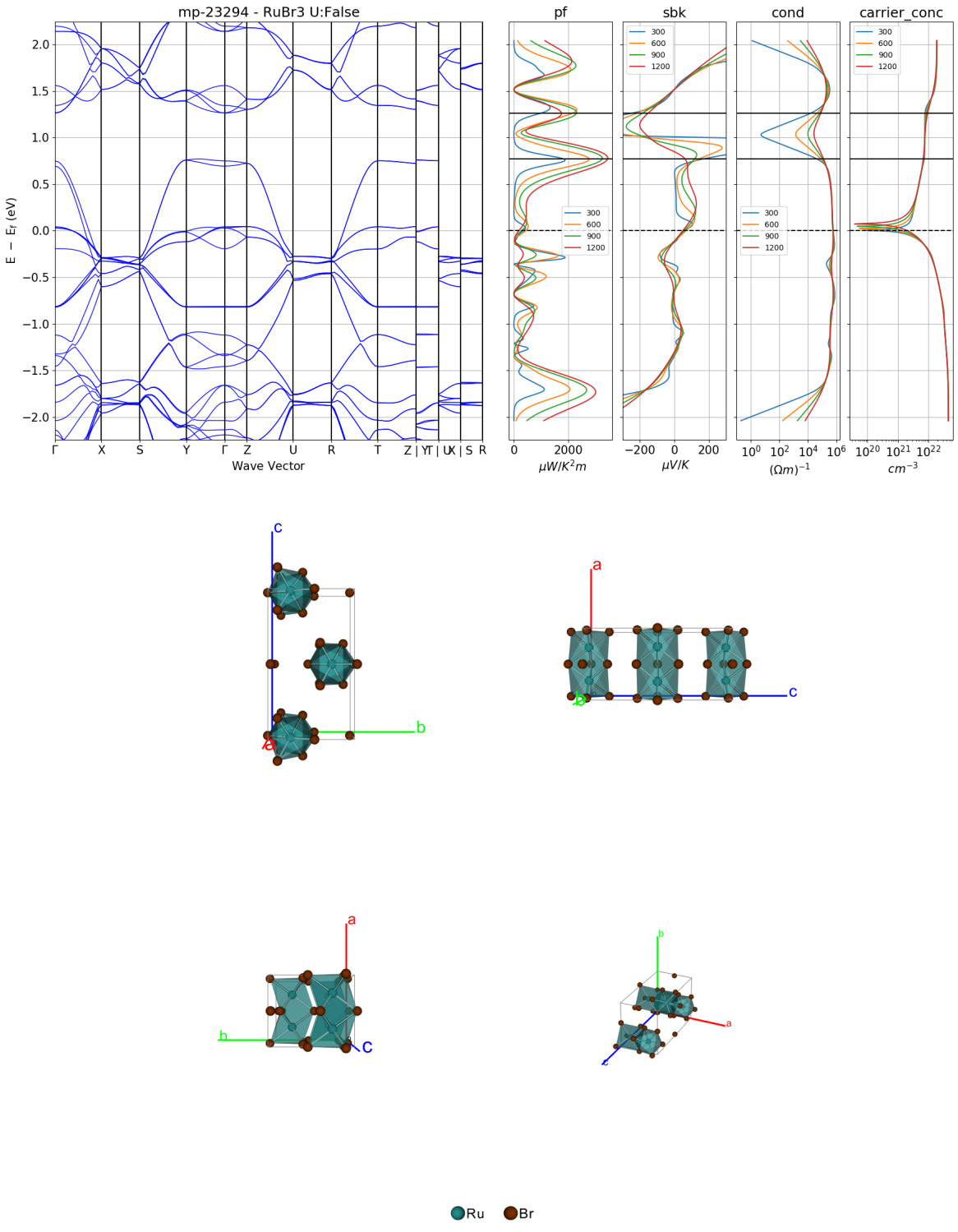

Figure 7: Structure (bottom), band structure, and calculated transport properties (top) of $RuBr_3$ (mp-23294).

7 Plots related to gapped metals in Tables 3 and 4

In the case of materials belonging to the same group that share the same space group, we report here the crystal structure, the band structure, and the transport properties of a single representative of that group.

7.1 Binaries

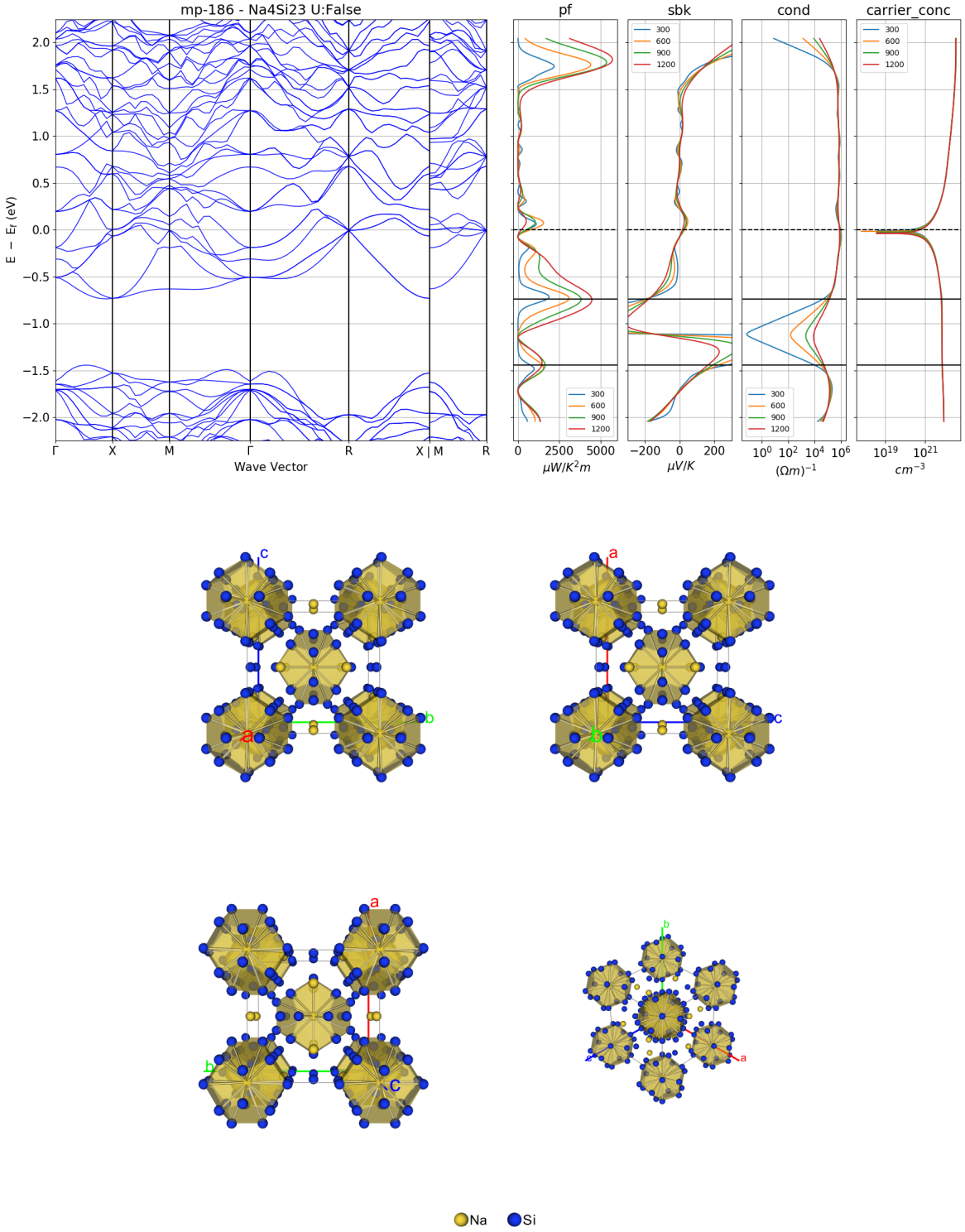

Figure 8: Structure (bottom), band structure, and calculated transport properties (top) of Na_4Si_{23} (mp-186).

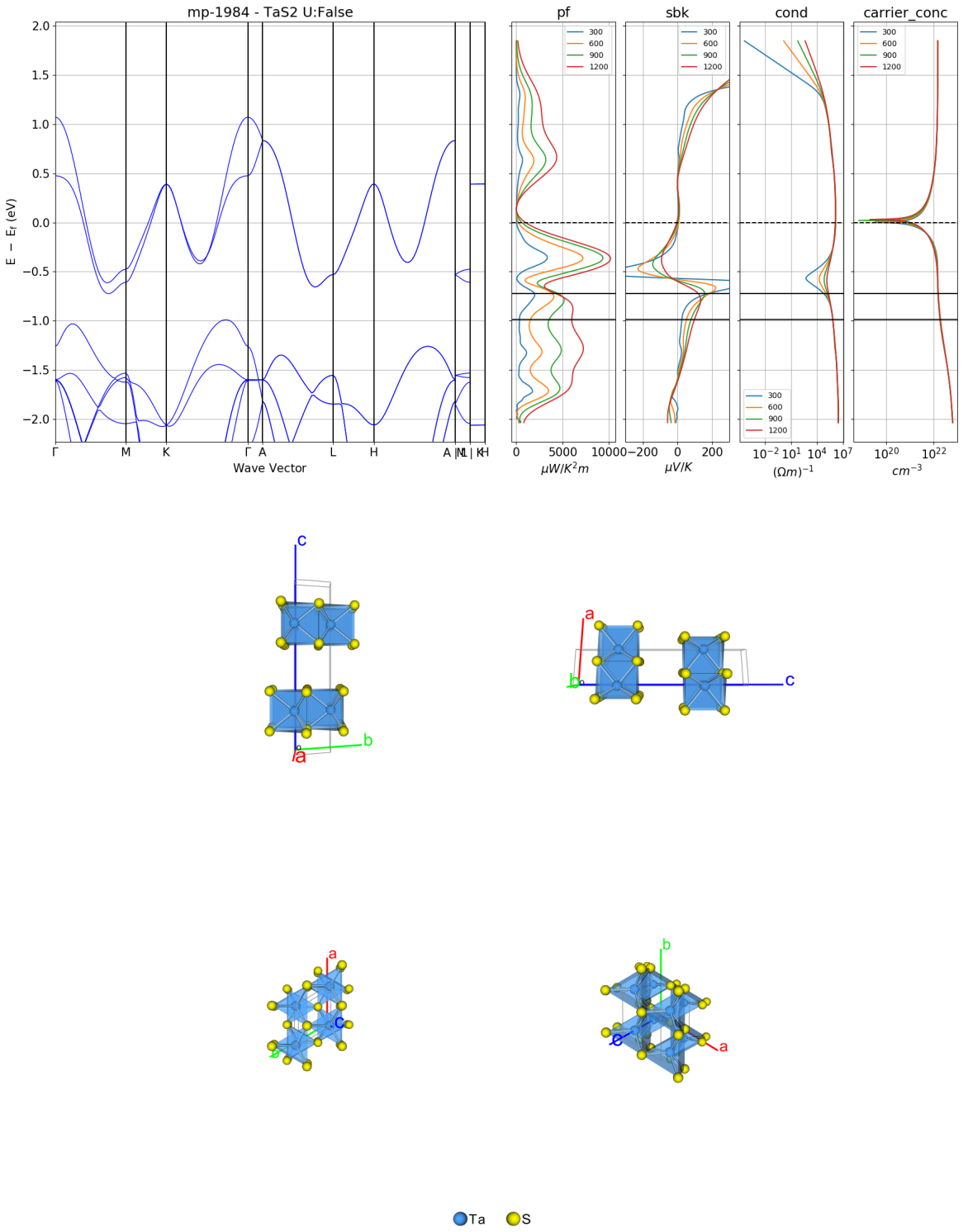

Figure 9: Structure (bottom), band structure, and calculated transport properties (top) of TaS_2 (mp-1984).

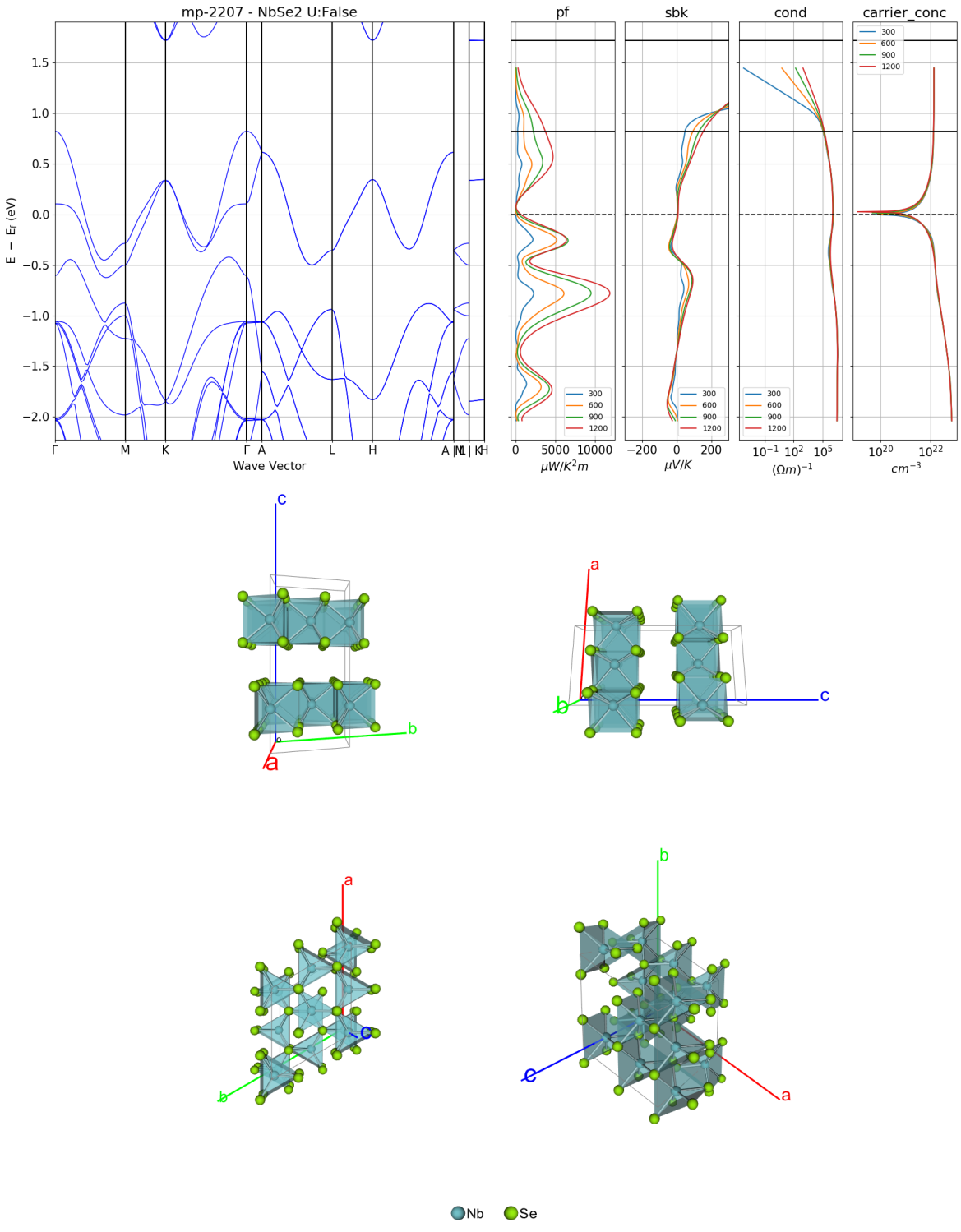

Figure 10: Structure (bottom), band structure, and calculated transport properties (top) of $NbSe_2$ (mp-2207).

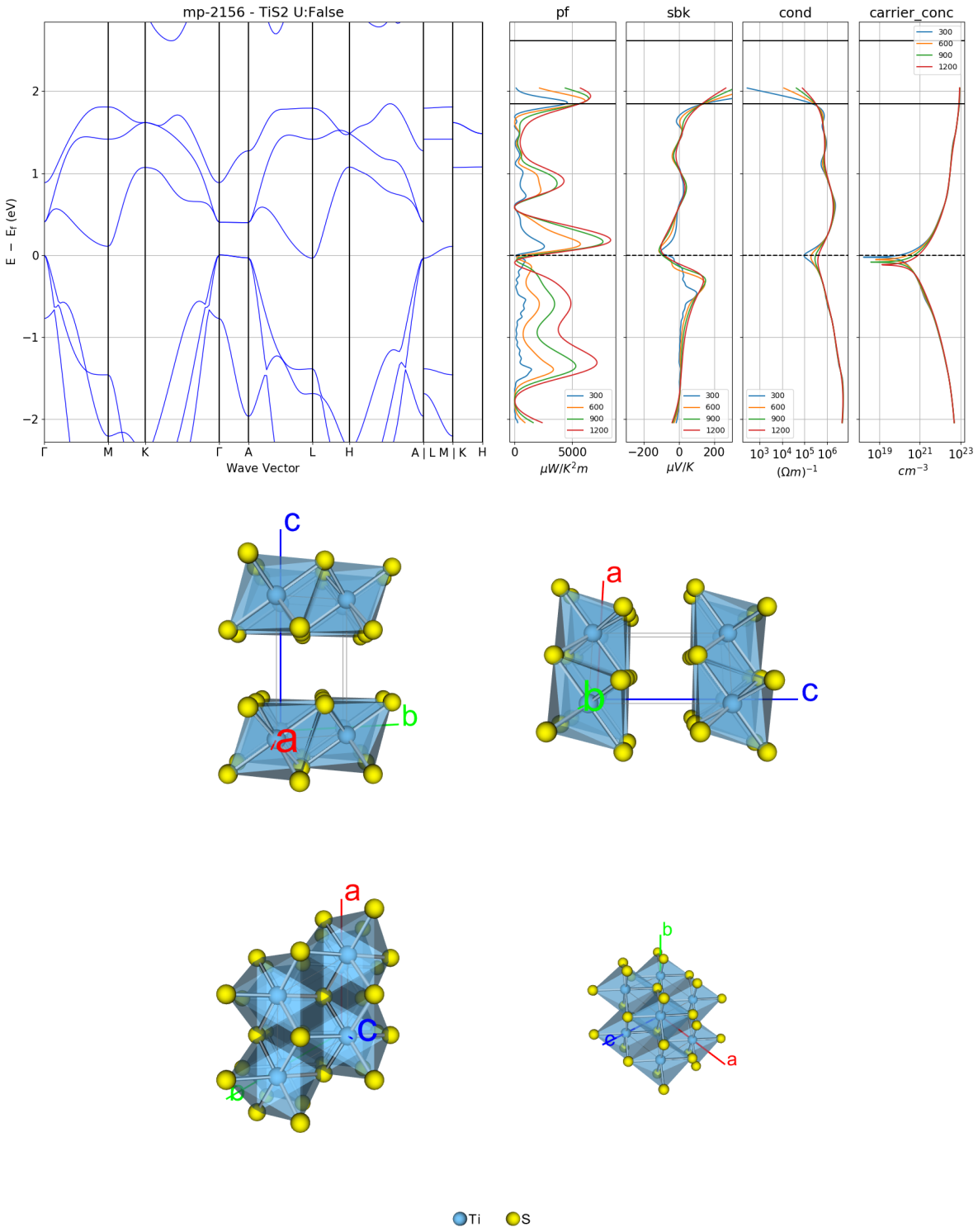

Figure 11: Structure (bottom), band structure, and calculated transport properties (top) of TiS_2 (mp-2156).

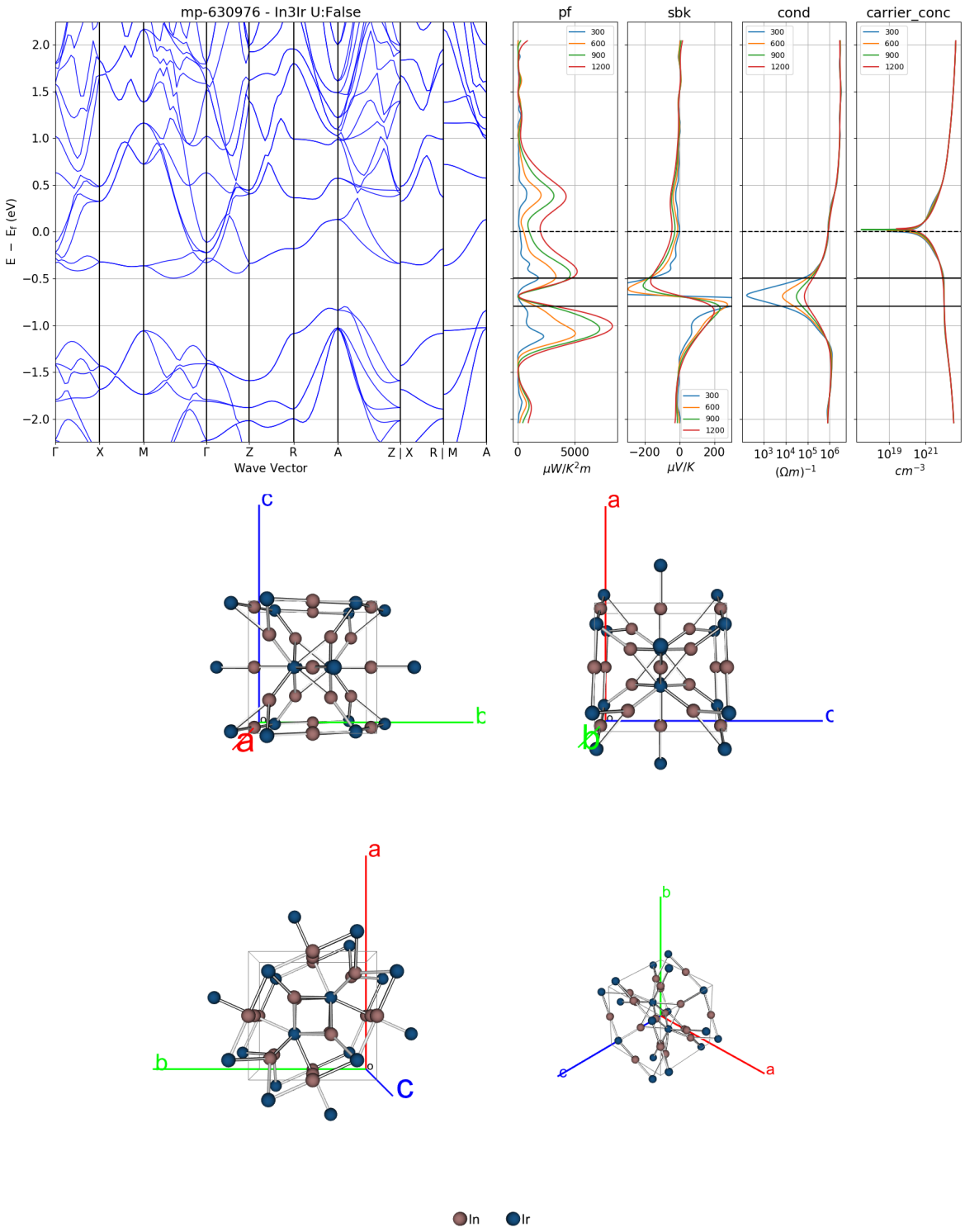

Figure 12: Structure (bottom), band structure, and calculated transport properties (top) of In_3Ir (mp-630976).

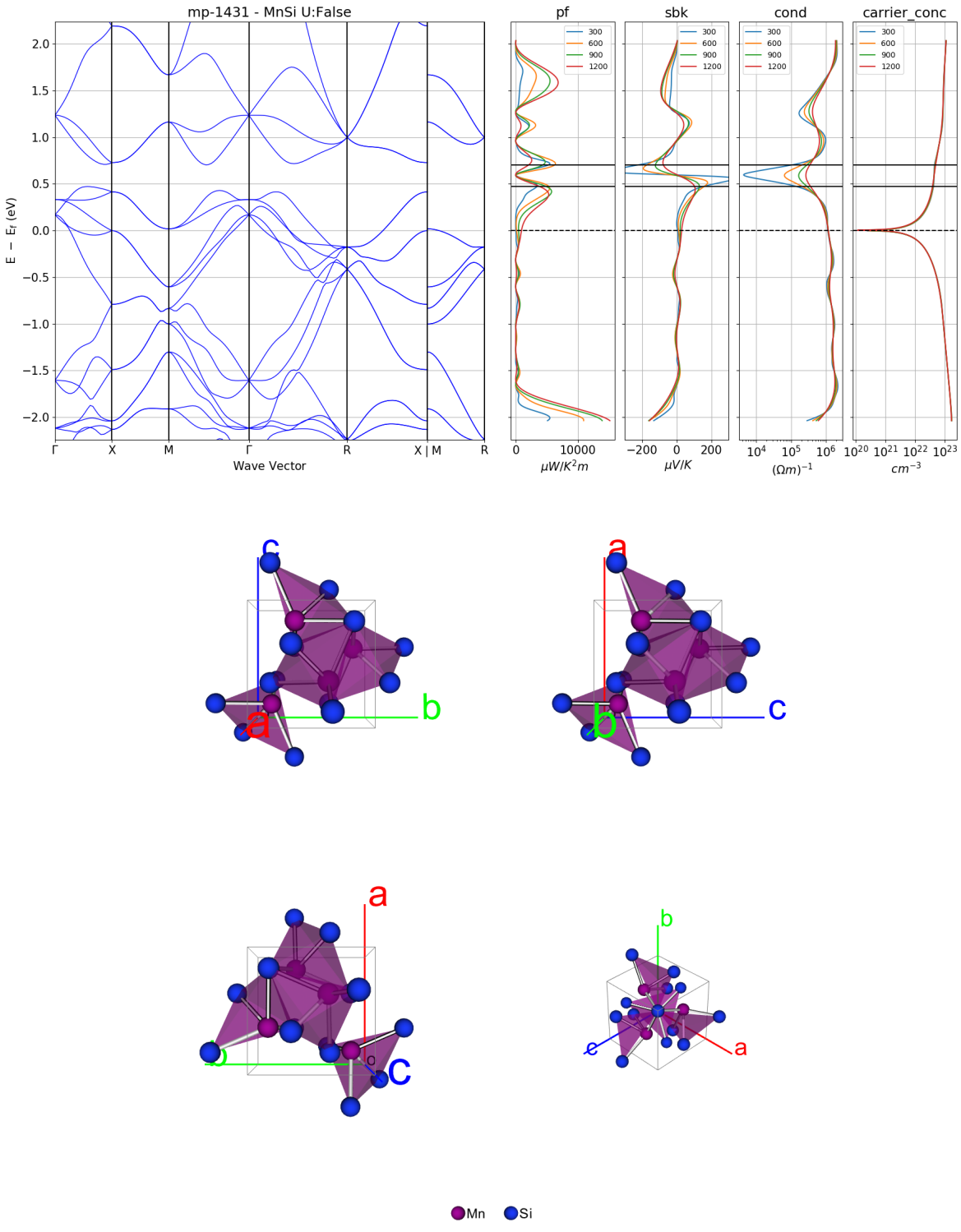

Figure 13: Structure (bottom), band structure, and calculated transport properties (top) of MnSi (mp-1431).

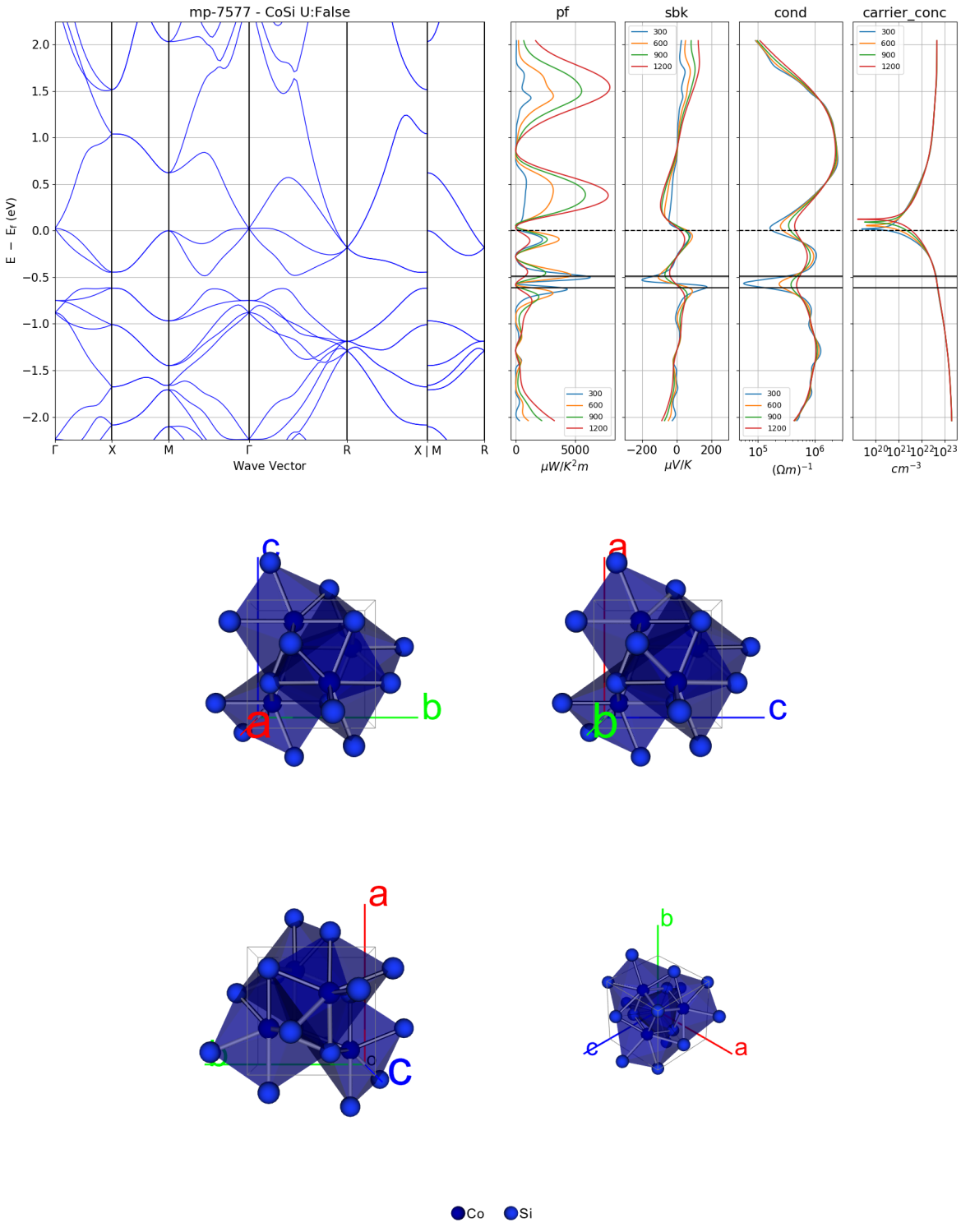

Figure 14: Structure (bottom), band structure, and calculated transport properties (top) of CoSi (mp-7577).

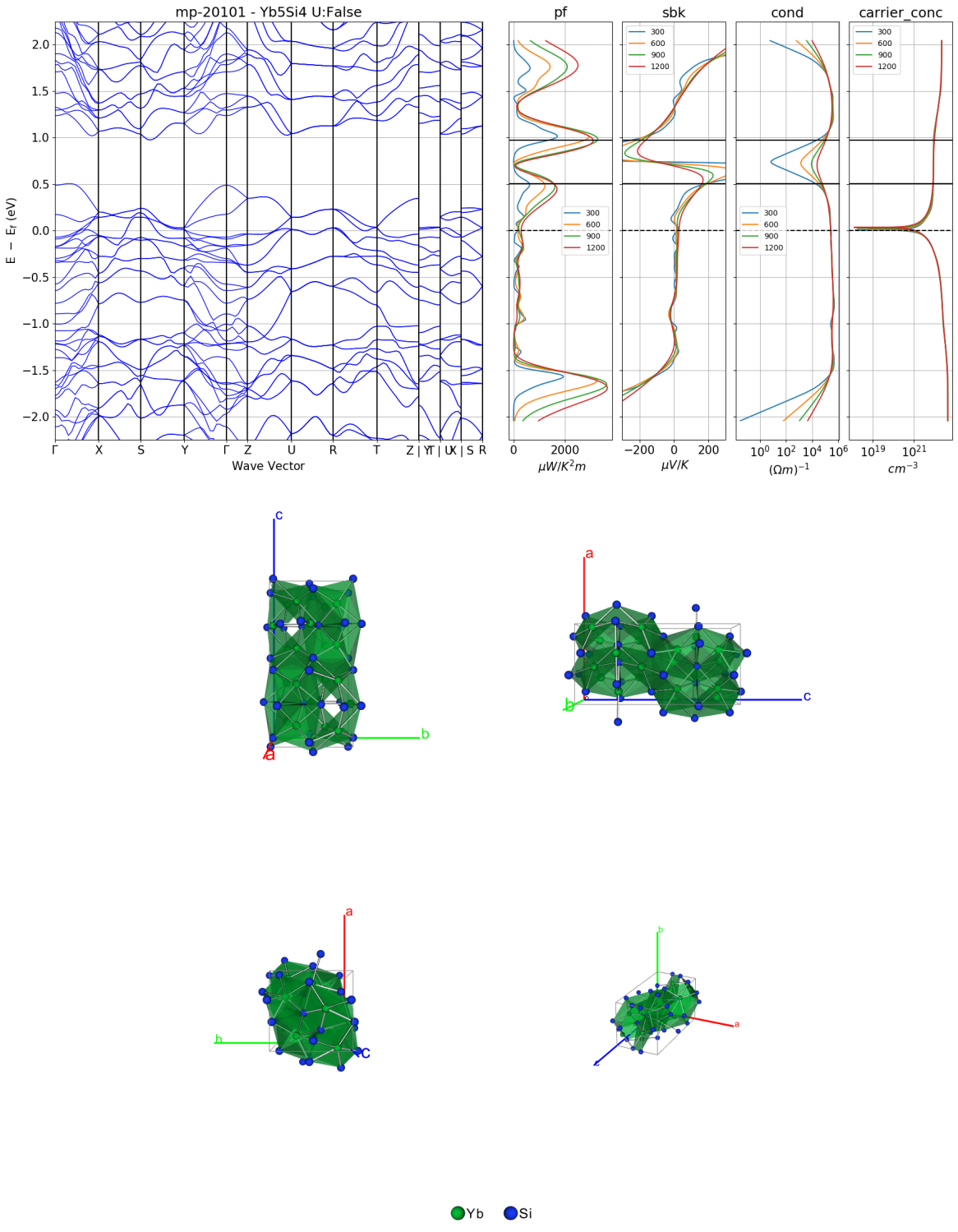

Figure 15: Structure (bottom), band structure, and calculated transport properties (top) of Yb_5Si_4 (mp-20101).

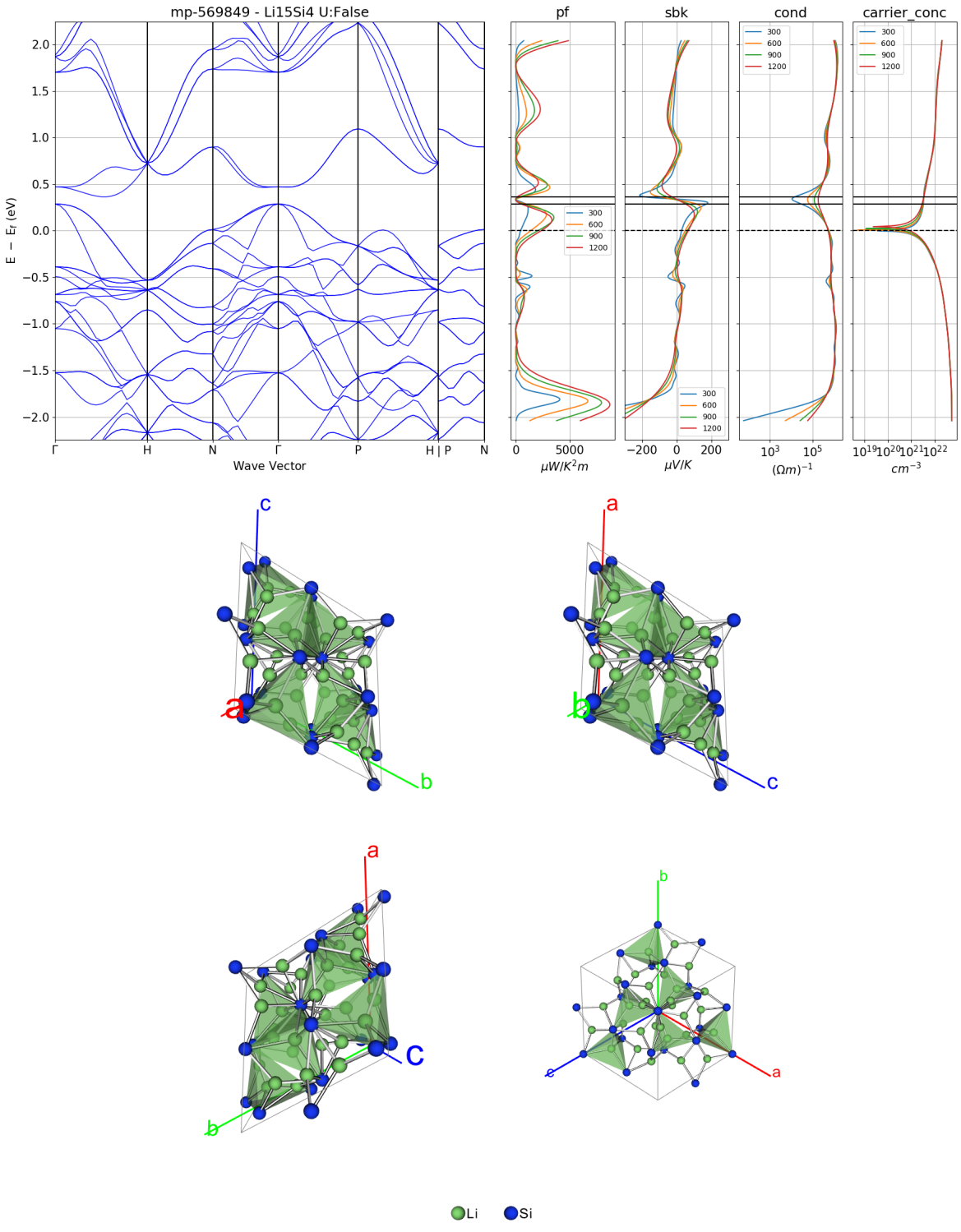

Figure 16: Structure (bottom), band structure, and calculated transport properties (top) of $Li_{15}Si_4$ (mp-569849).

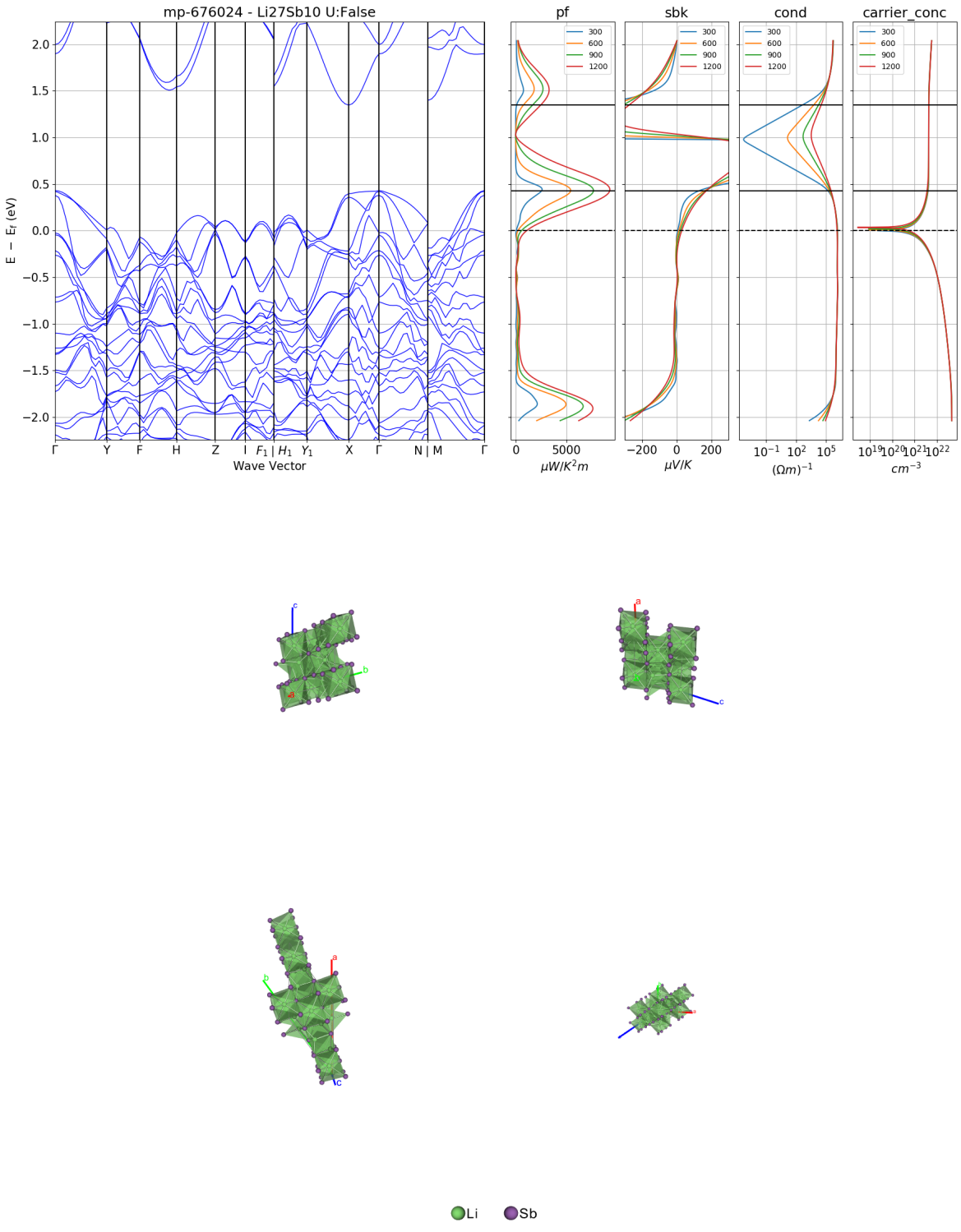

Figure 17: Structure (bottom), band structure, and calculated transport properties (top) of $Li_{27}Sb_{10}$ (mp-676024).

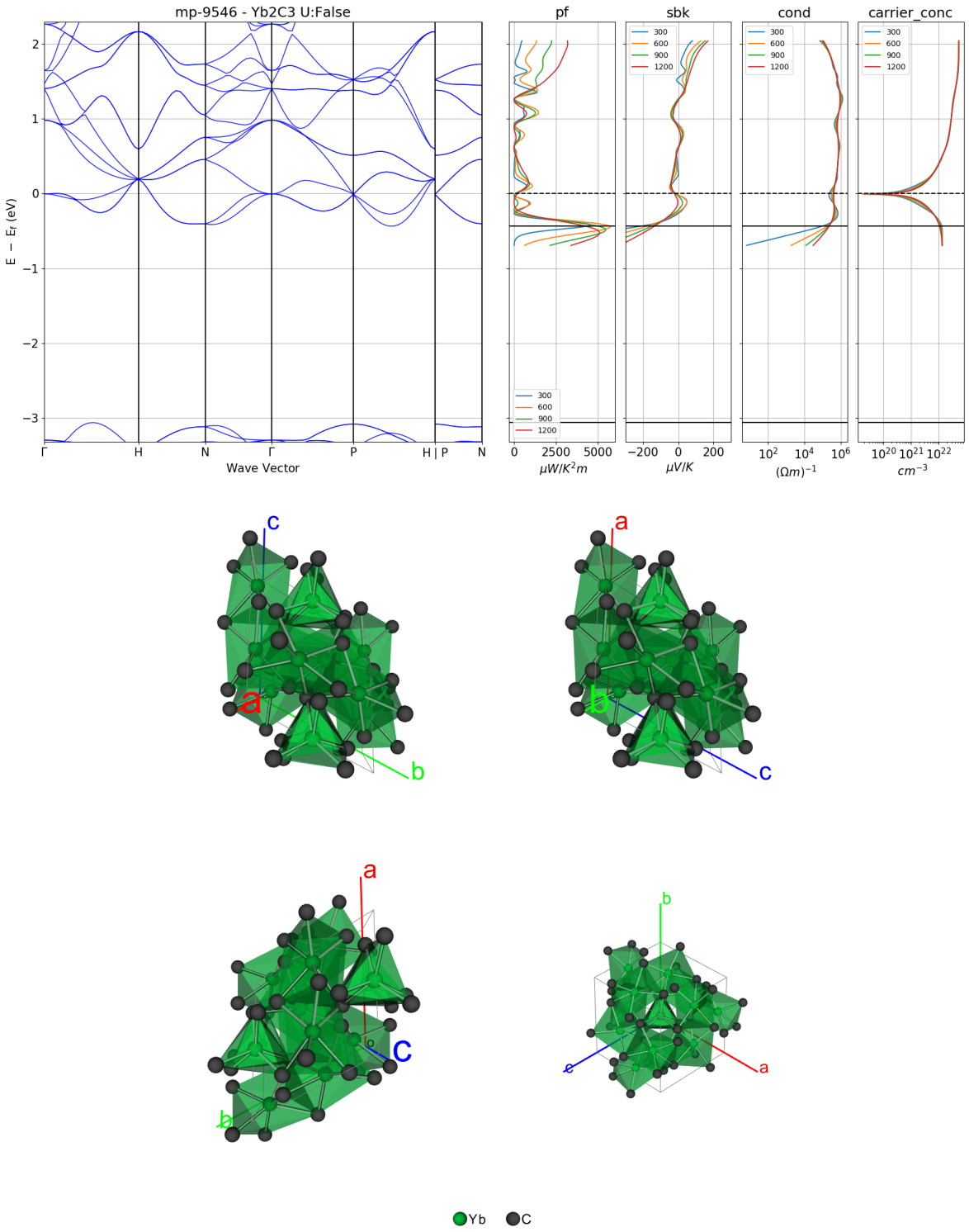

Figure 18: Structure (bottom), band structure, and calculated transport properties (top) of Yb_2C_3 (mp-9546).

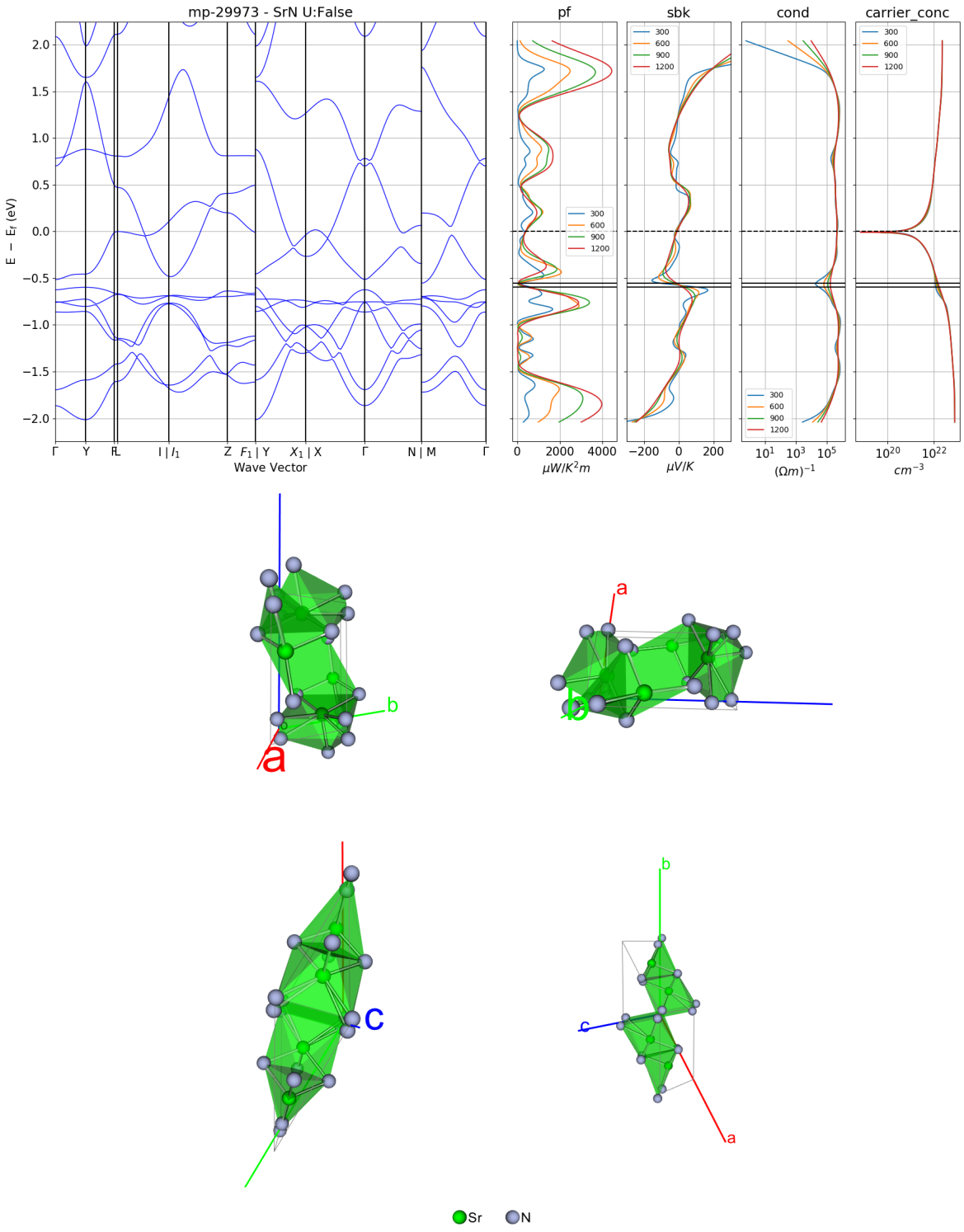

Figure 19: Structure (bottom), band structure, and calculated transport properties (top) of SrN (mp-29973).

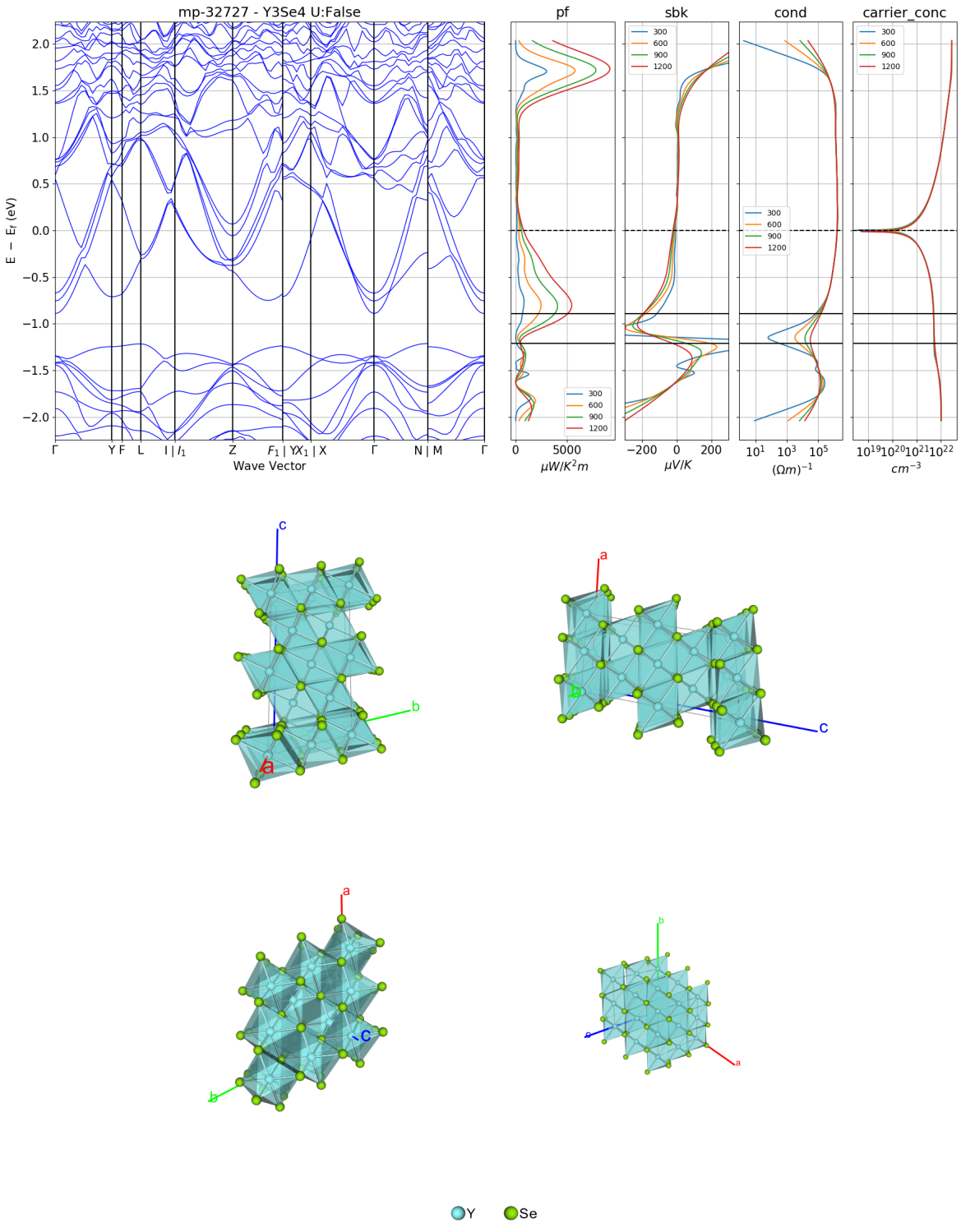

Figure 20: Structure (bottom), band structure, and calculated transport properties (top) of Y_3Se_4 (mp-32727).

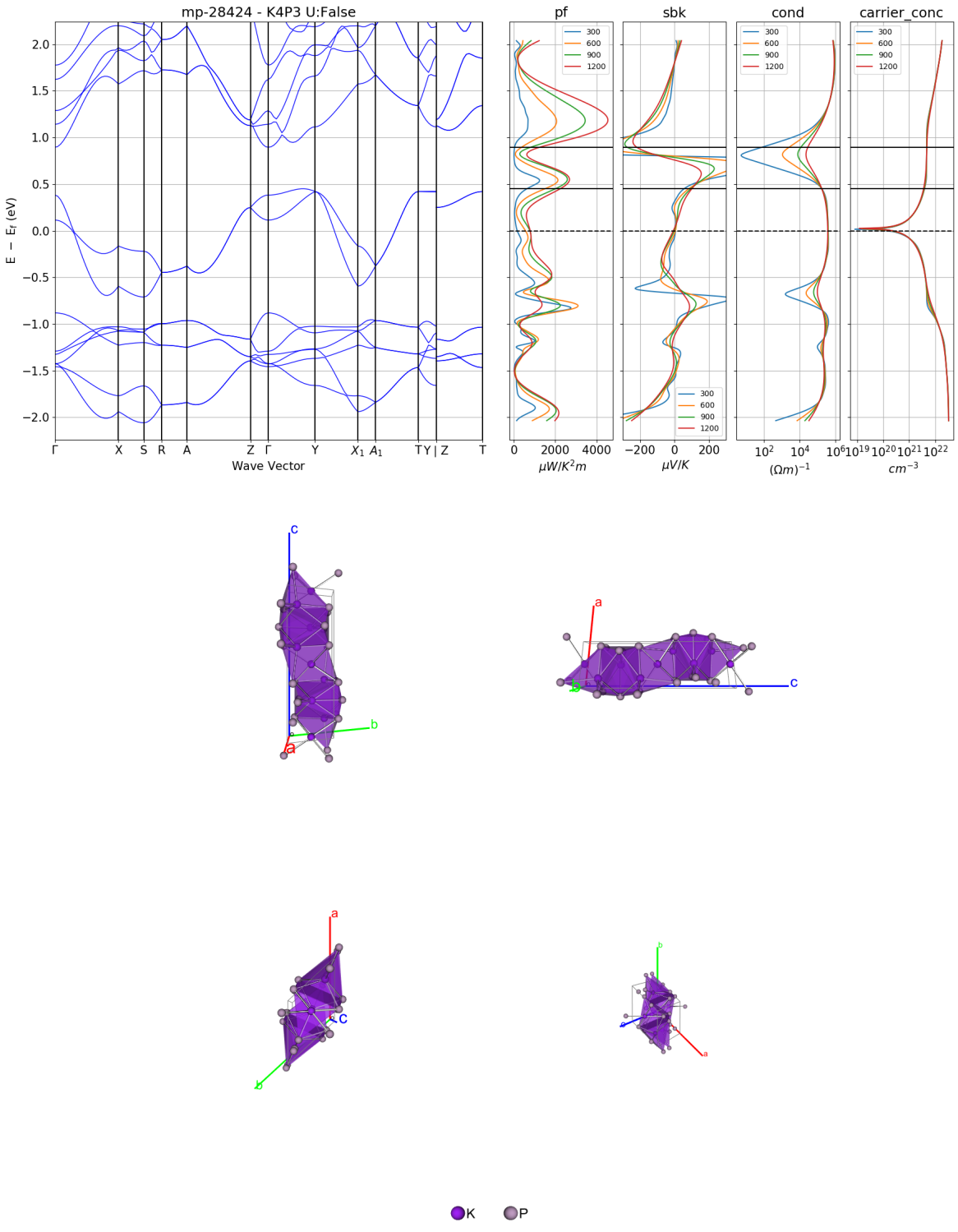

Figure 21: Structure (bottom), band structure, and calculated transport properties (top) of K_4P_3 (mp-28424).

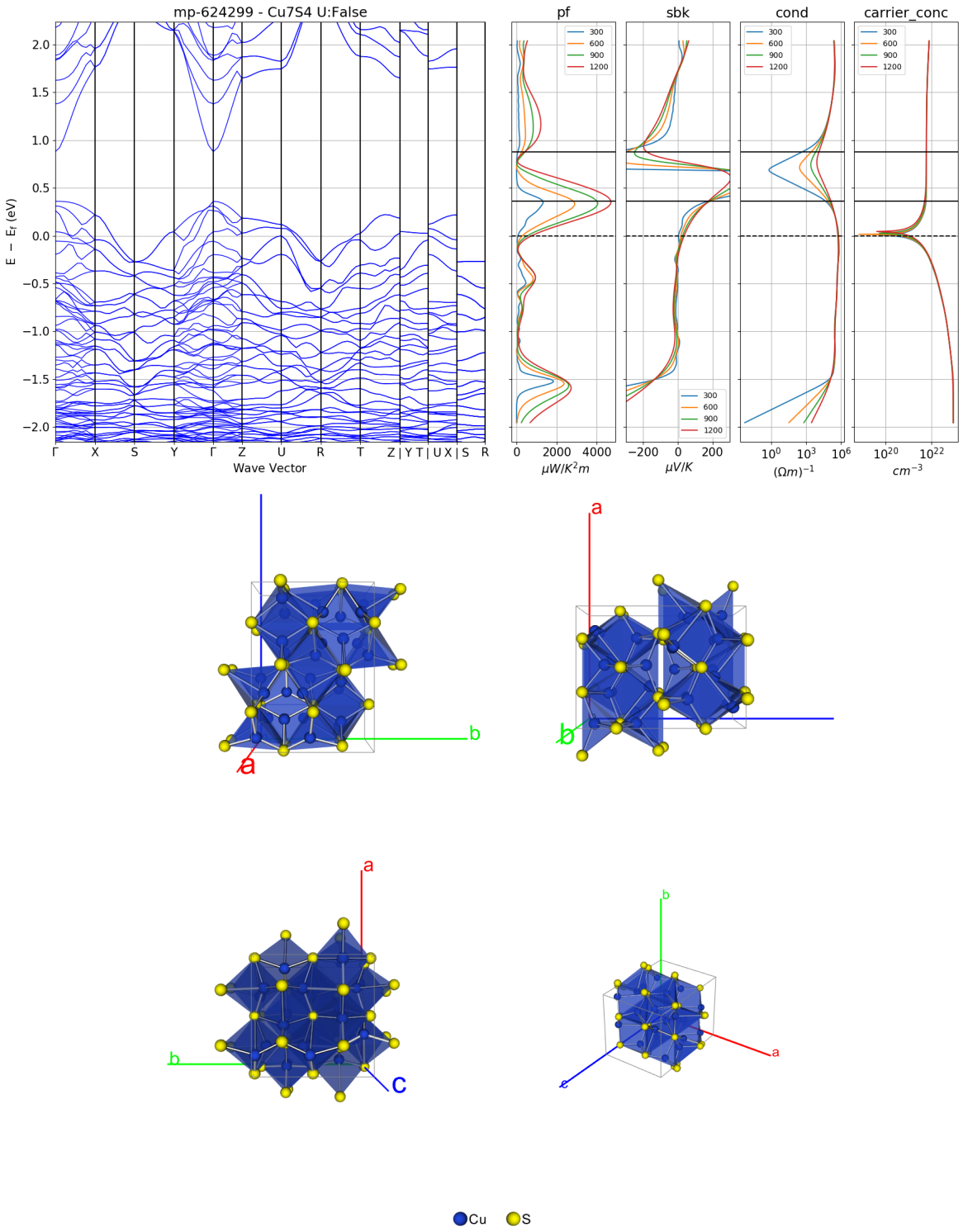

Figure 22: Structure (bottom), band structure, and calculated transport properties (top) of Cu_7S_4 (mp-624299).

7.2 Ternaries

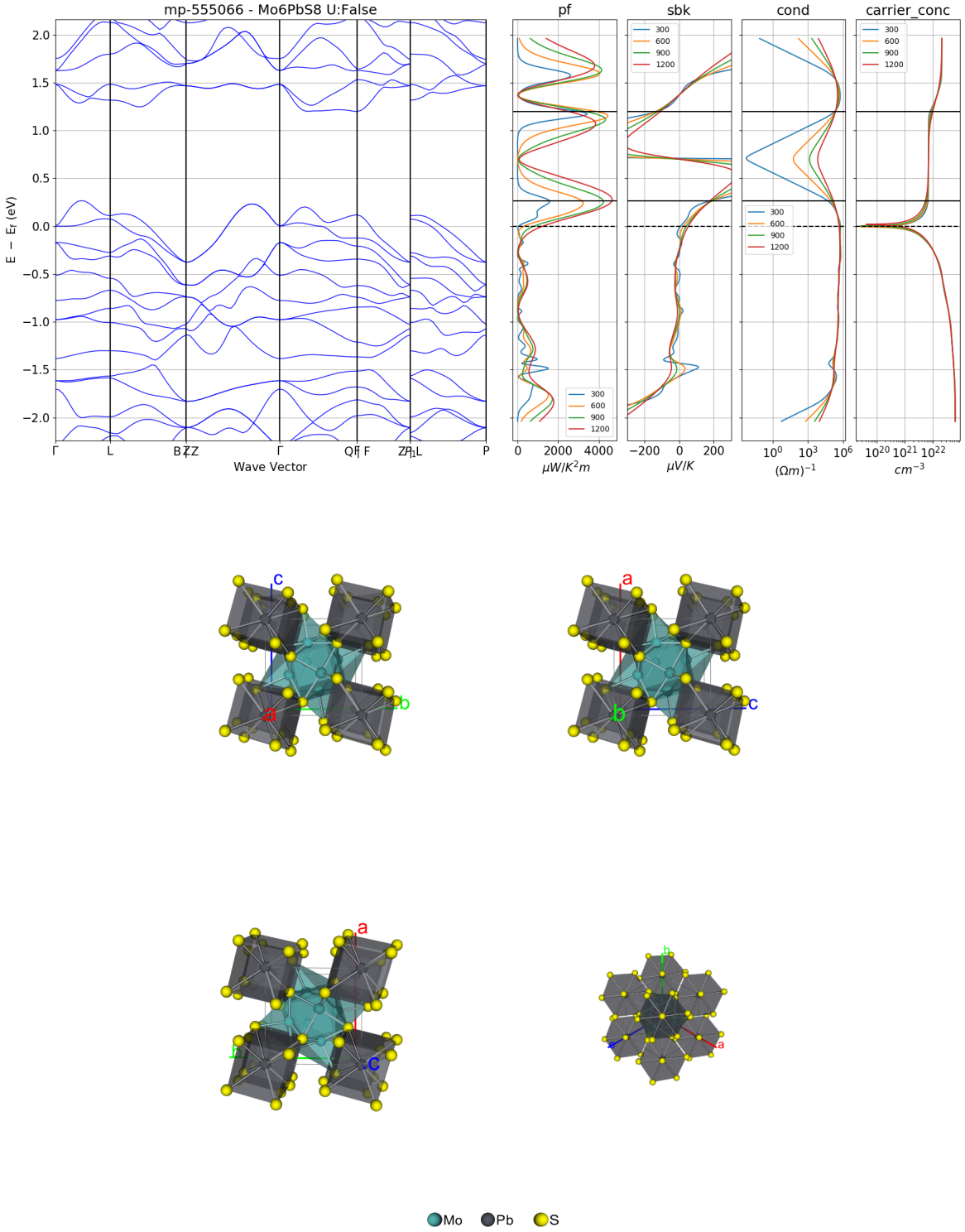

Figure 23: Structure (bottom), band structure, and calculated transport properties (top) of Mo_6PbS_8 (mp-555066).

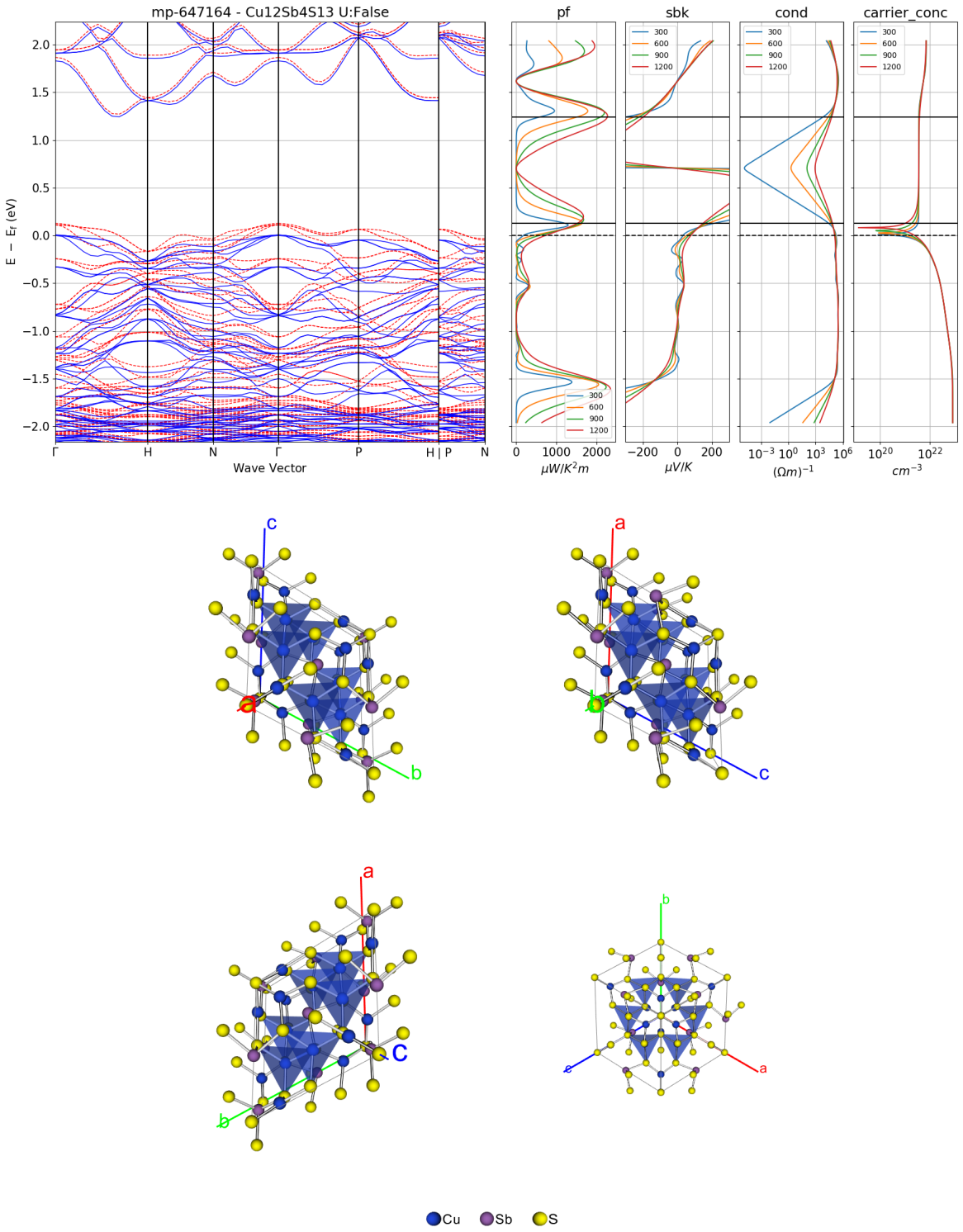

Figure 24: Structure (bottom), band structure, and calculated transport properties (top) of $Cu_{12}Sb_4S_{13}$ (mp-647164).

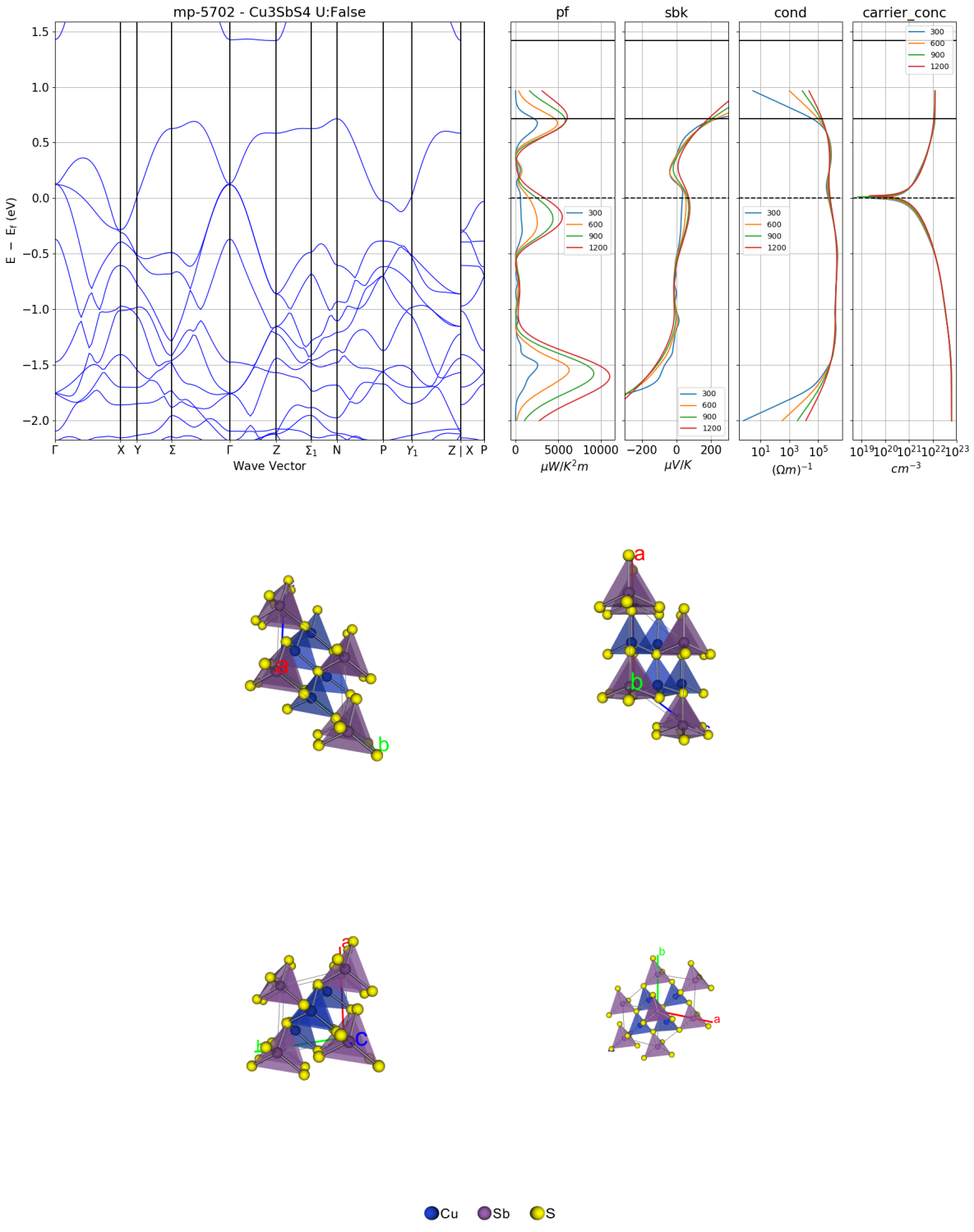

Figure 25: Structure (bottom), band structure, and calculated transport properties (top) of Cu_3SbS_4 (mp-5702).

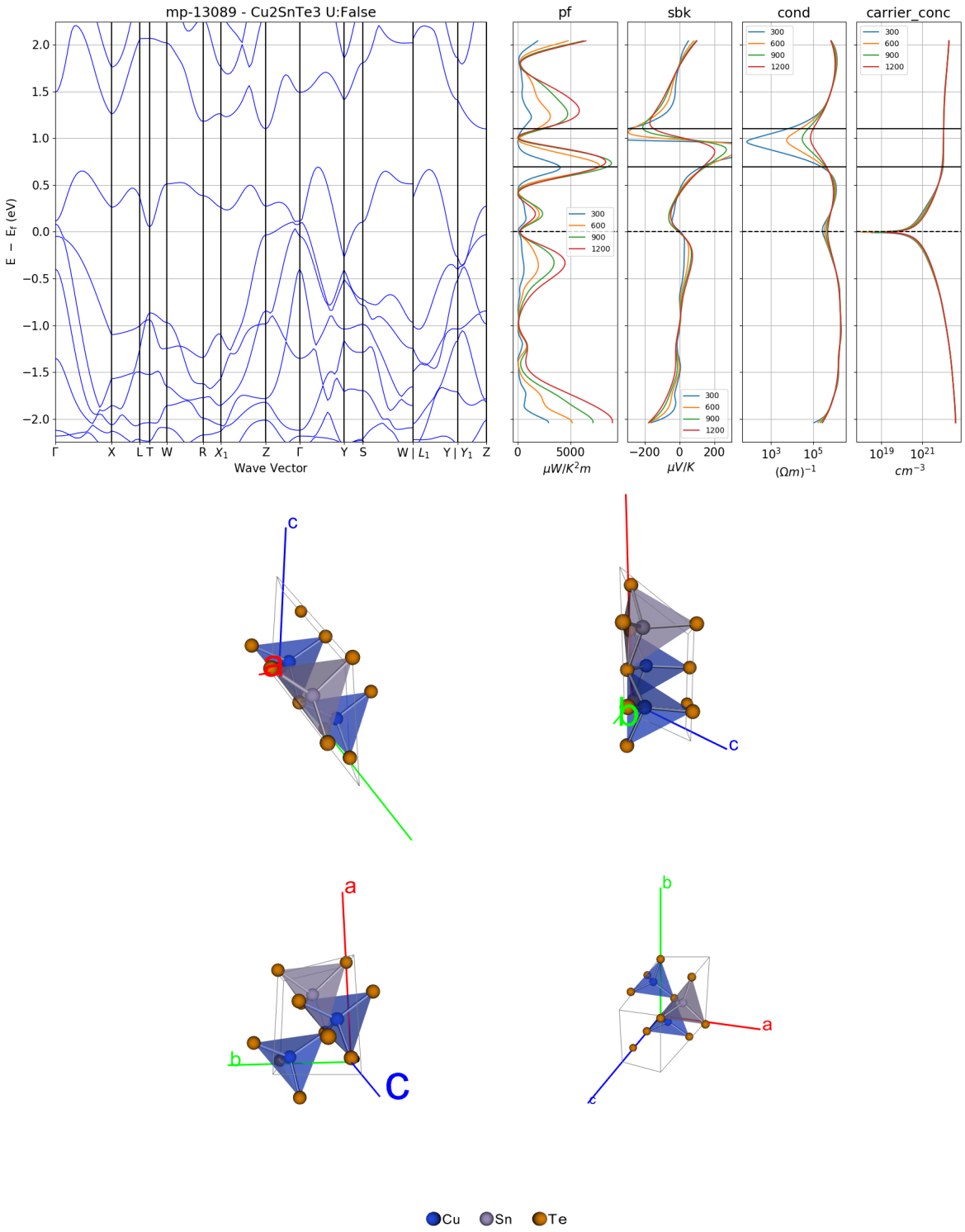

Figure 26: Structure (bottom), band structure, and calculated transport properties (top) of Cu_2SnTe_3 (mp-13089).

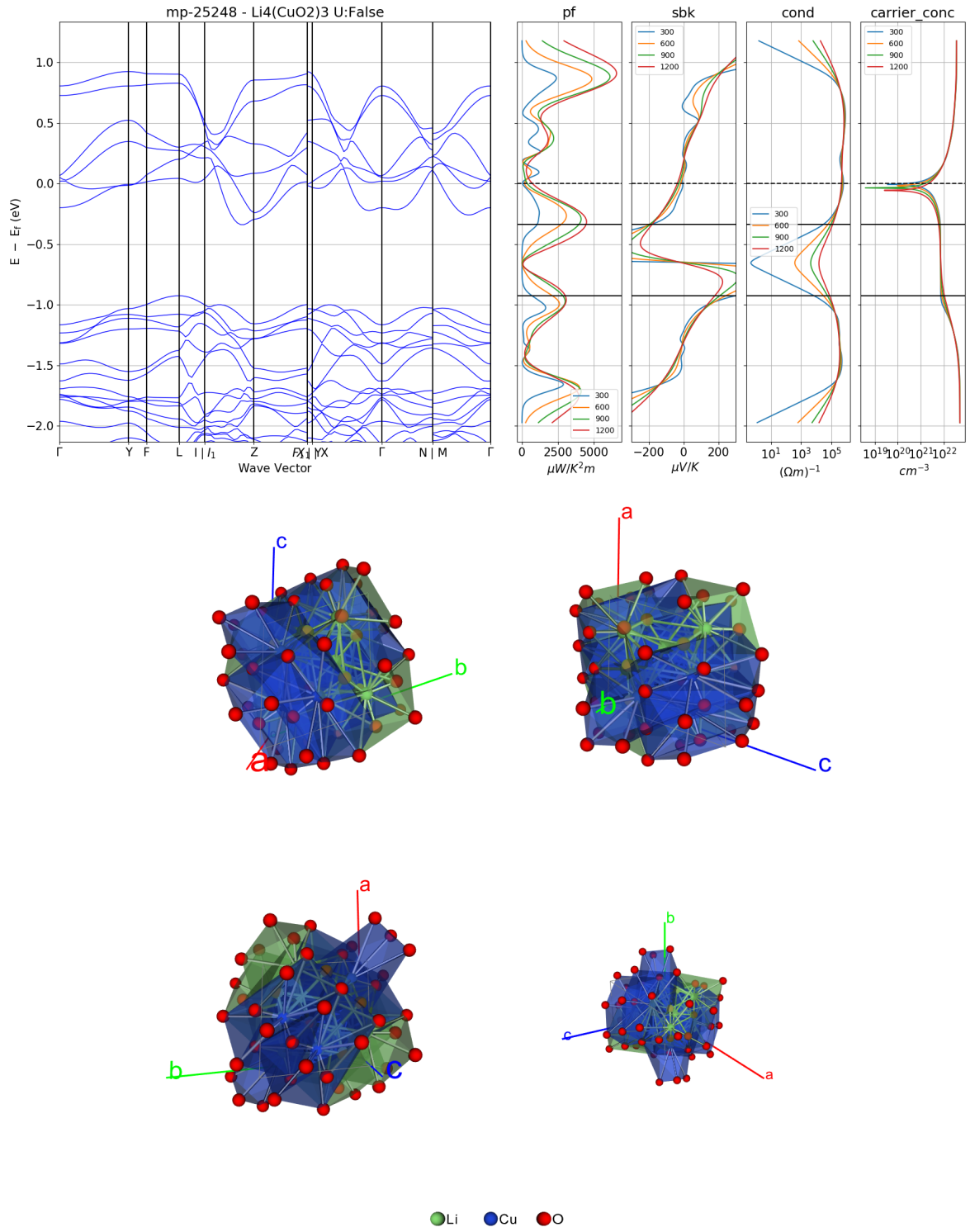

Figure 27: Structure (bottom), band structure, and calculated transport properties (top) of $Li_4(CuO_2)_3$ (mp-25248).

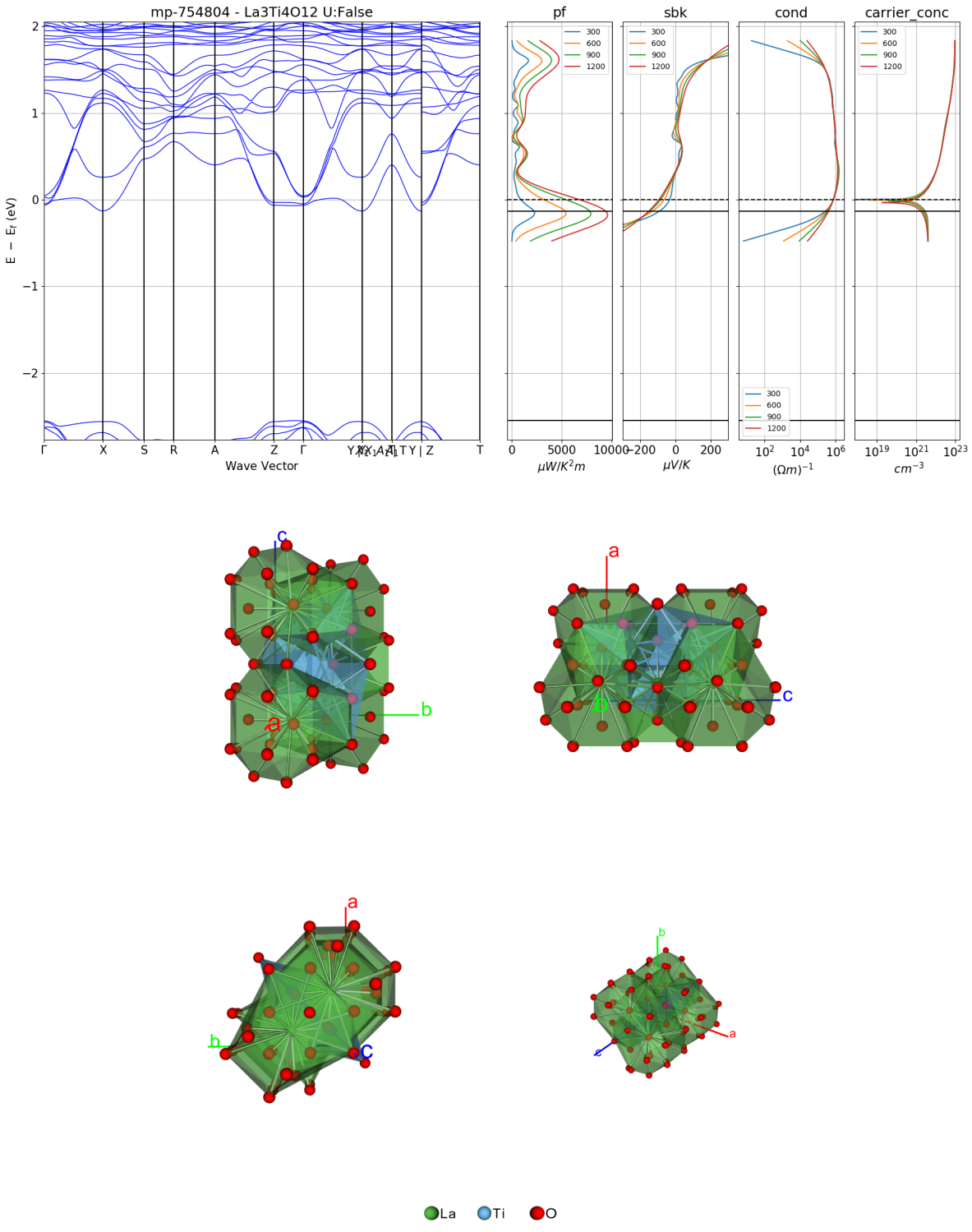

Figure 28: Structure (bottom), band structure, and calculated transport properties (top) of $La_3Ti_4O_{12}$ (mp-754804).

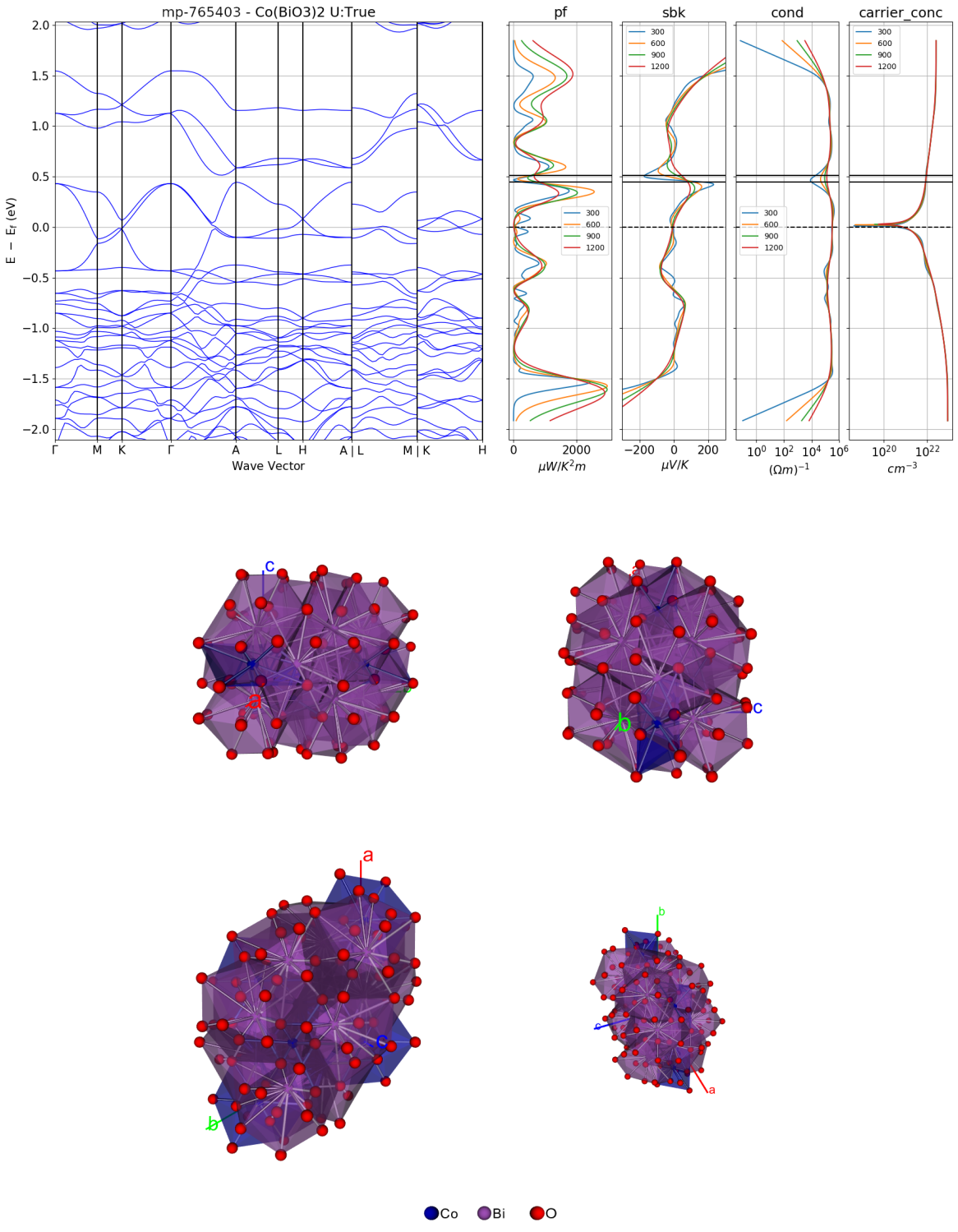

Figure 29: Structure (bottom), band structure, and calculated transport properties (top) of $Co(BiO_3)_2$ (mp-765403).

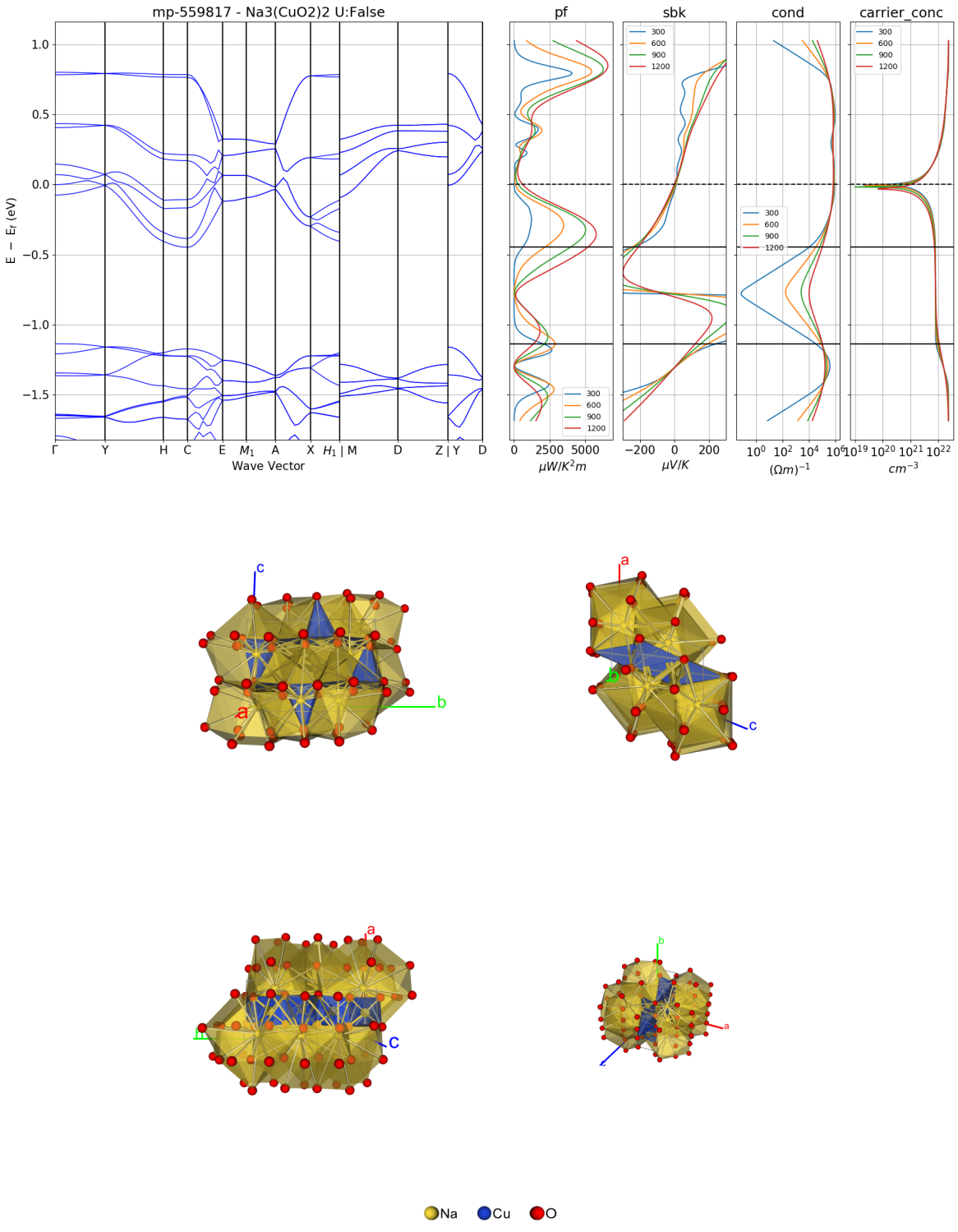

Figure 30: Structure (bottom), band structure, and calculated transport properties (top) of $Na_3(CuO_2)_2$ (mp-559817).

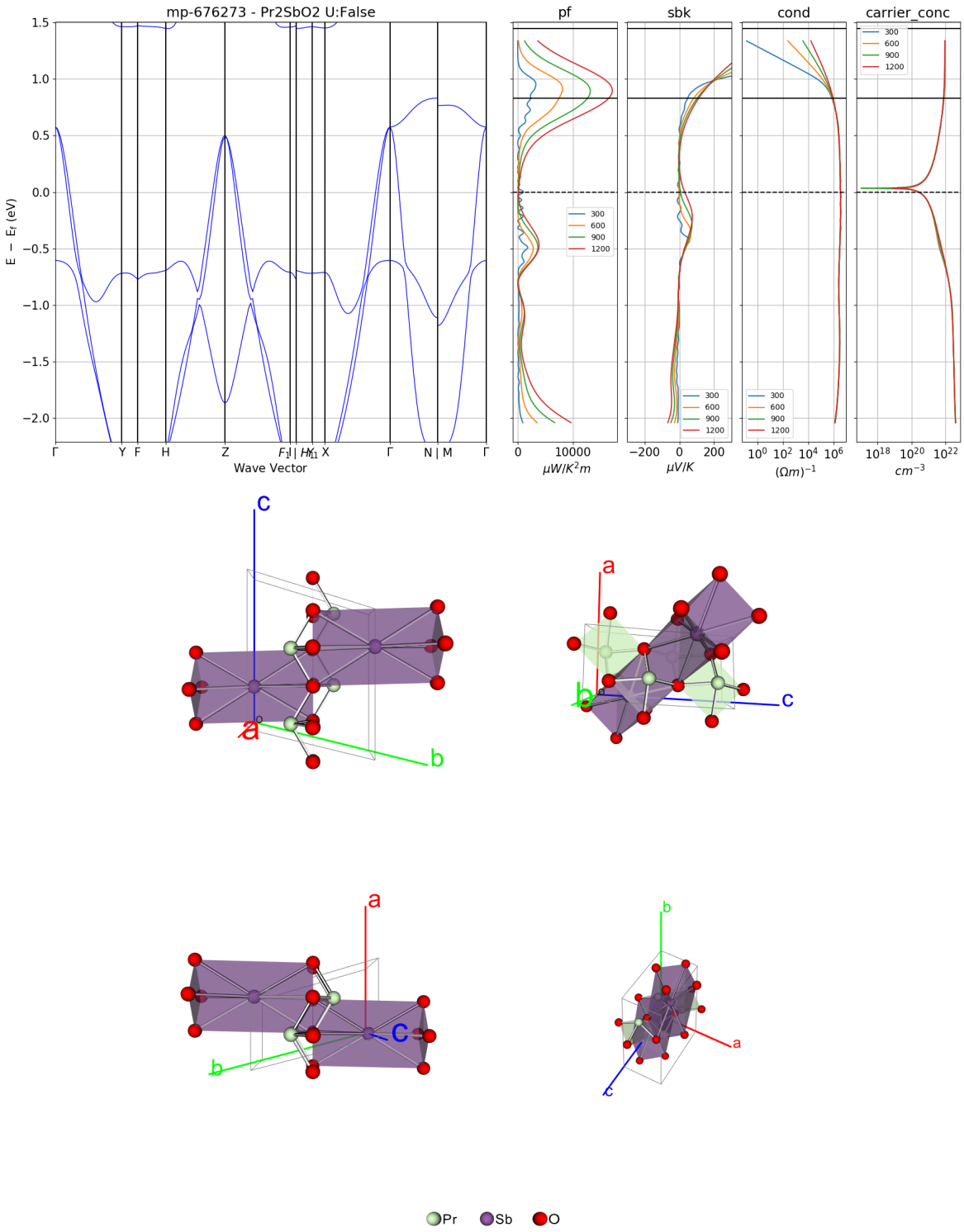

Figure 31: Structure (bottom), band structure, and calculated transport properties (top) of Pr_2SbO_2 (mp-676273).

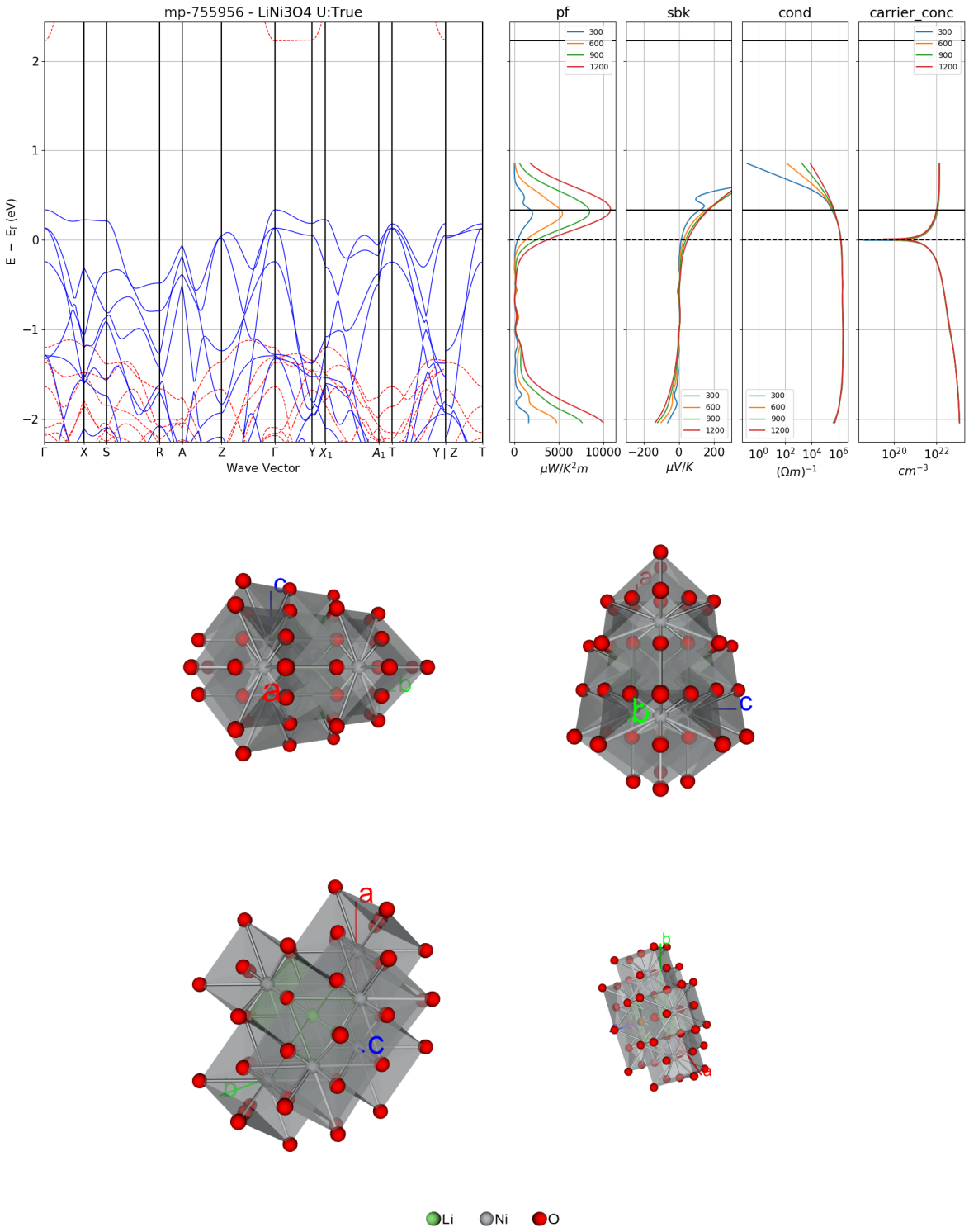

Figure 32: Structure (bottom), band structure, and calculated transport properties (top) of $LiNi_3O_4$ (mp-755956).

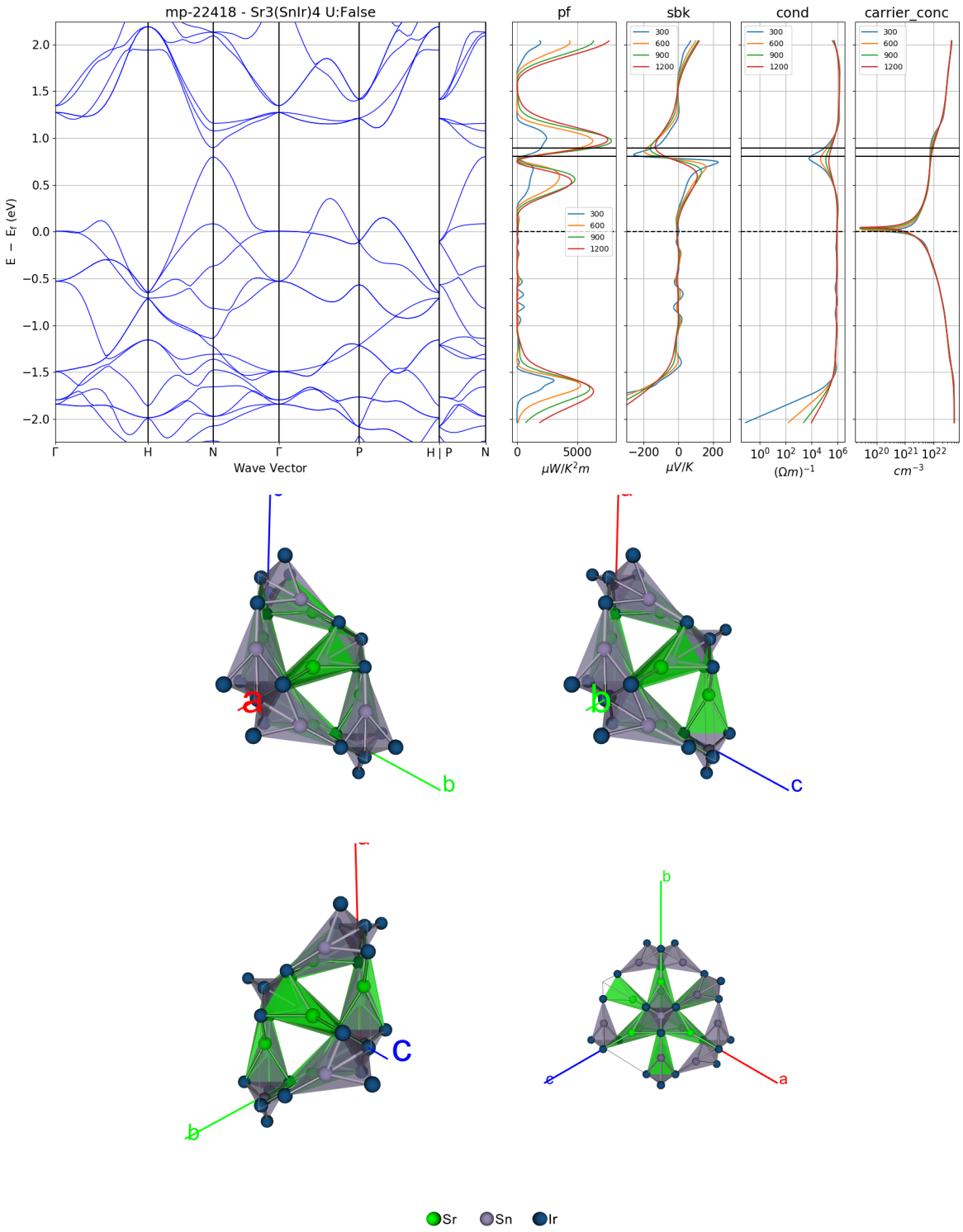

Figure 33: Structure (bottom), band structure, and calculated transport properties (top) of $Sr_3(SnIr)_4$ (mp-22418).

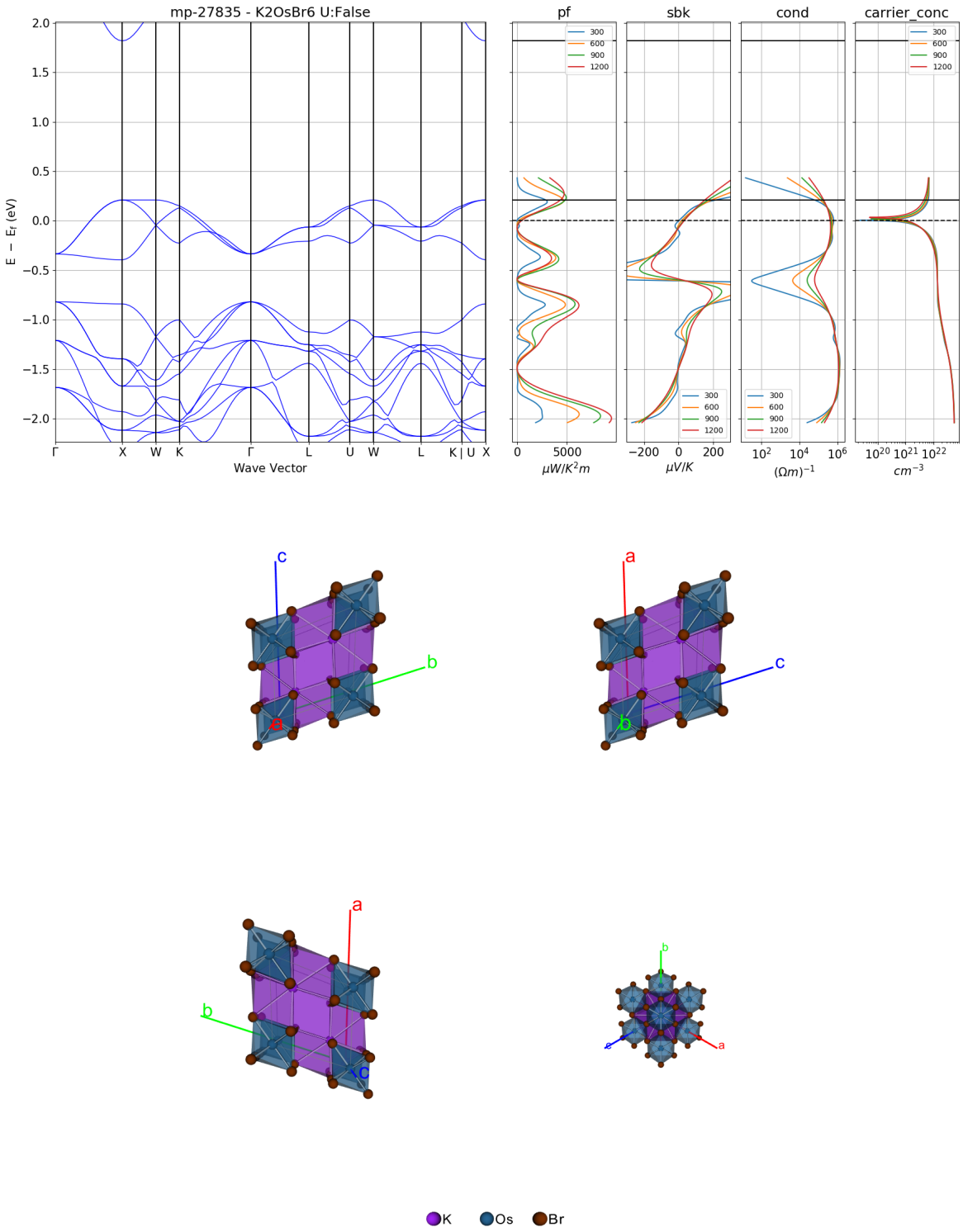

Figure 34: Structure (bottom), band structure, and calculated transport properties (top) of K_2OsBr_6 (mp-27835).

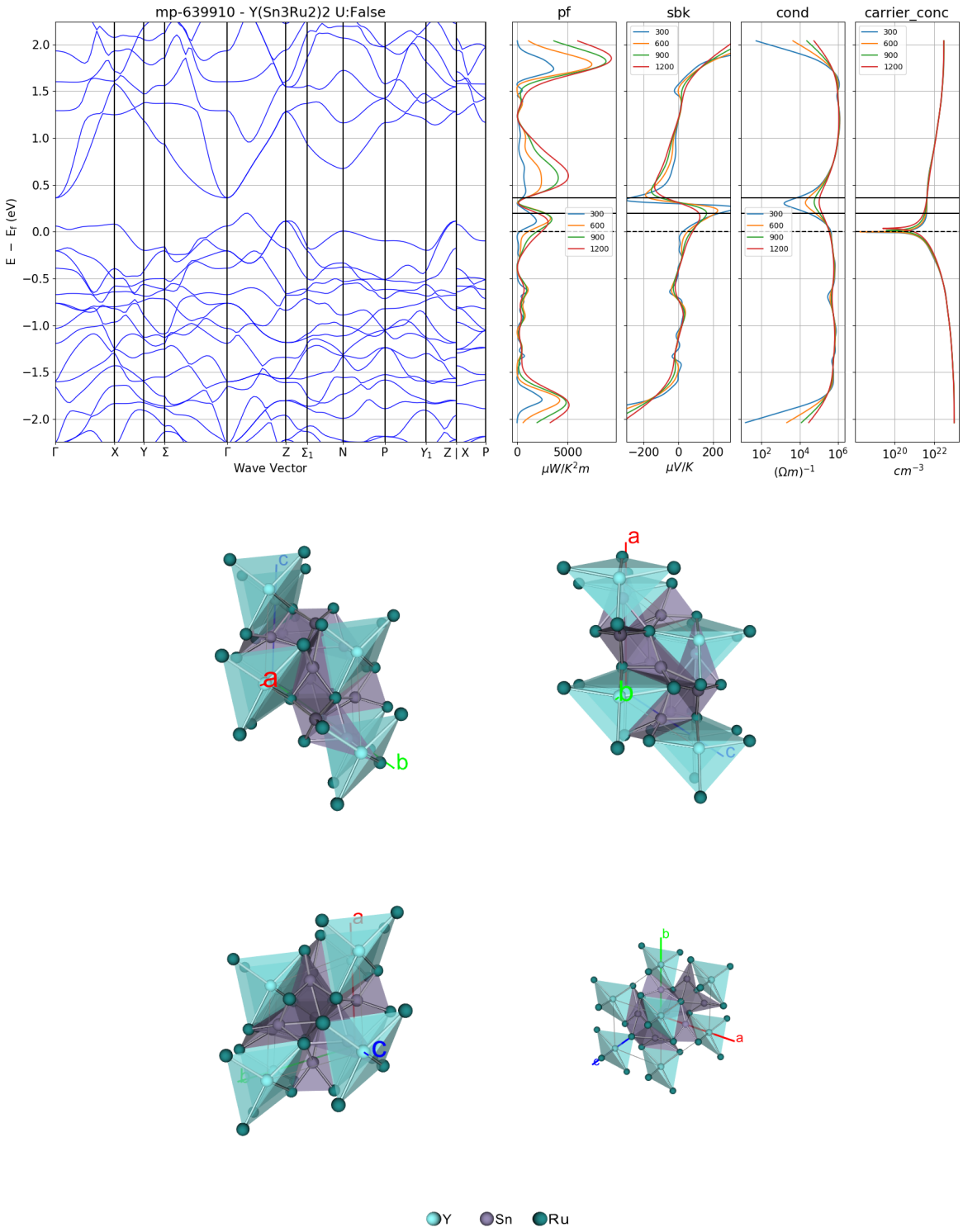

Figure 35: Structure (bottom), band structure, and calculated transport properties (top) of $Y(Sn_3Ru_2)_2$ (mp-639910).

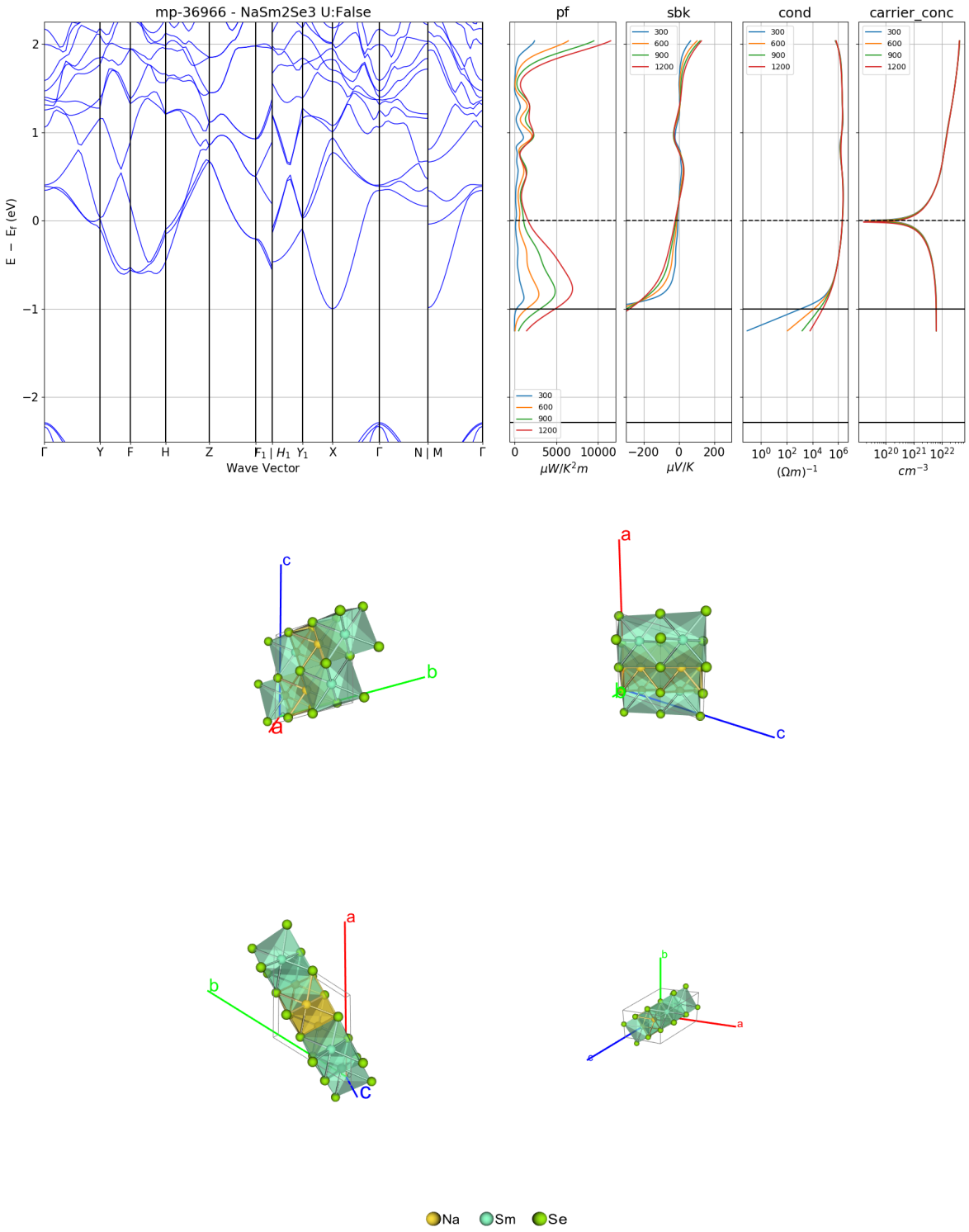

Figure 36: Structure (bottom), band structure, and calculated transport properties (top) of $NaSm_2Se_3$ (mp-36966).

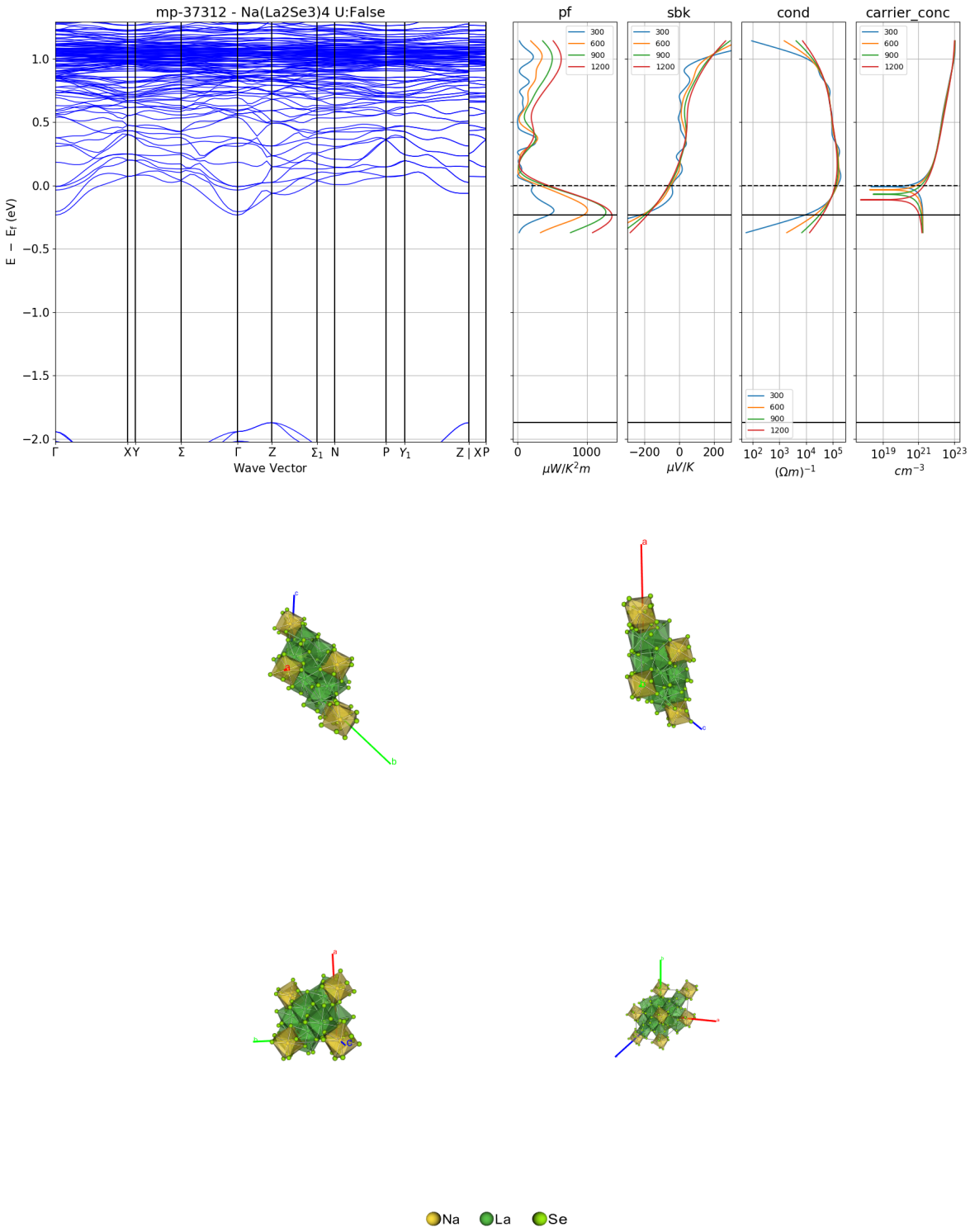

Figure 37: Structure (bottom), band structure, and calculated transport properties (top) of $Na(La_2Se_3)_4$ (mp-37312).

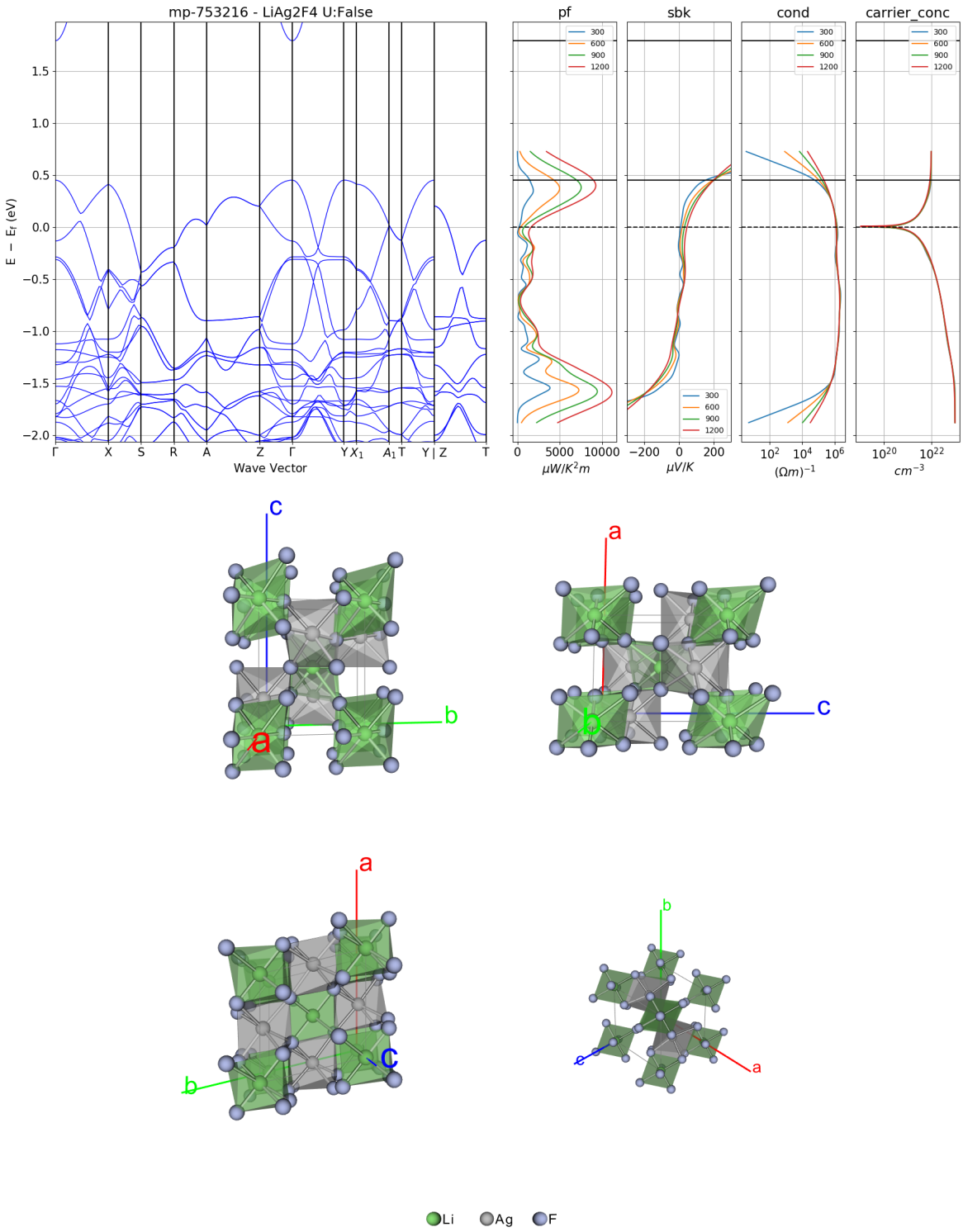

Figure 38: Structure (bottom), band structure, and calculated transport properties (top) of $LiAg_2F_4$ (mp-753216).

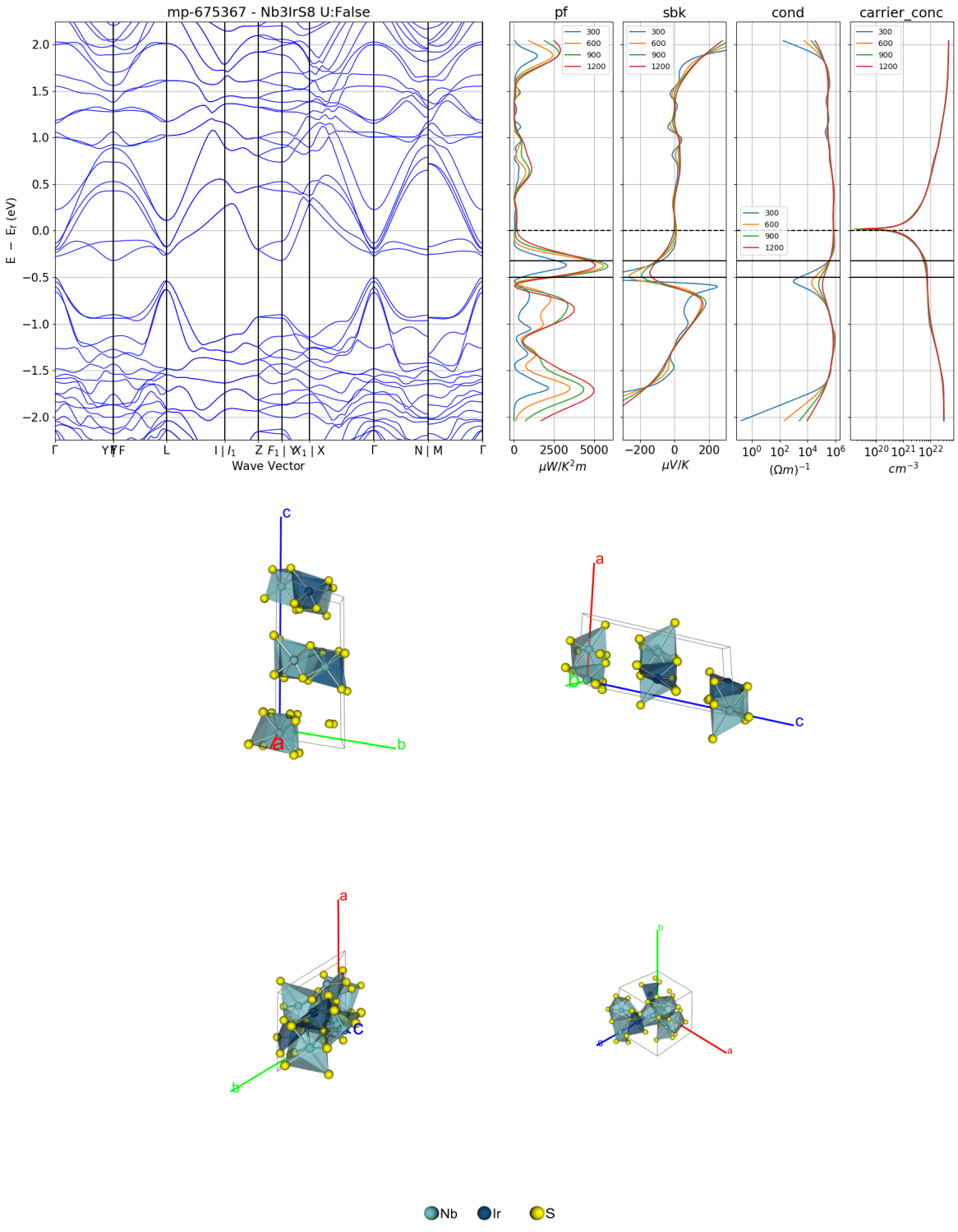

Figure 39: Structure (bottom), band structure, and calculated transport properties (top) of Nb_3IrS_8 (mp-675367).

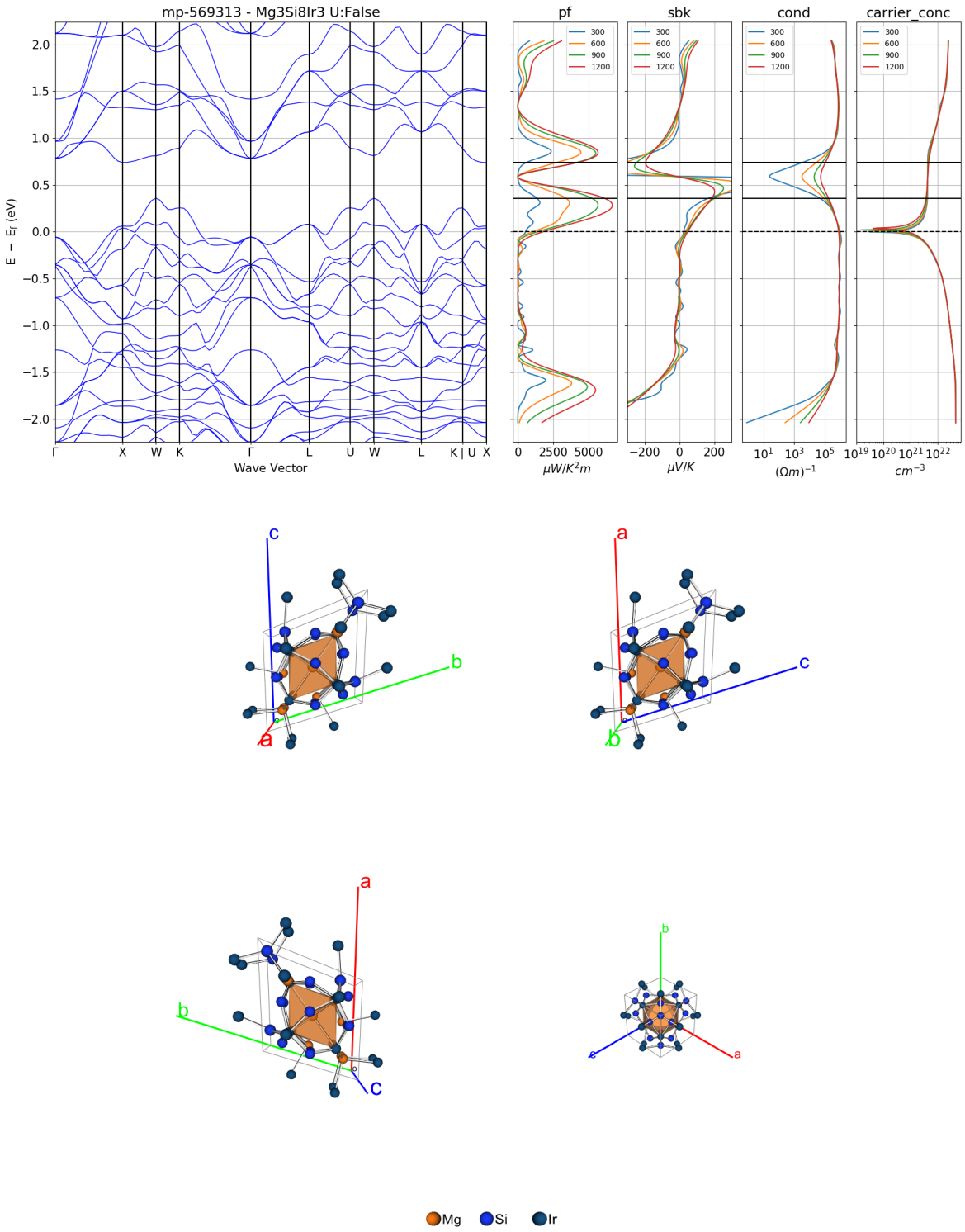

Figure 40: Structure (bottom), band structure, and calculated transport properties (top) of $Mg_3Si_8Ir_3$ (mp-569313).

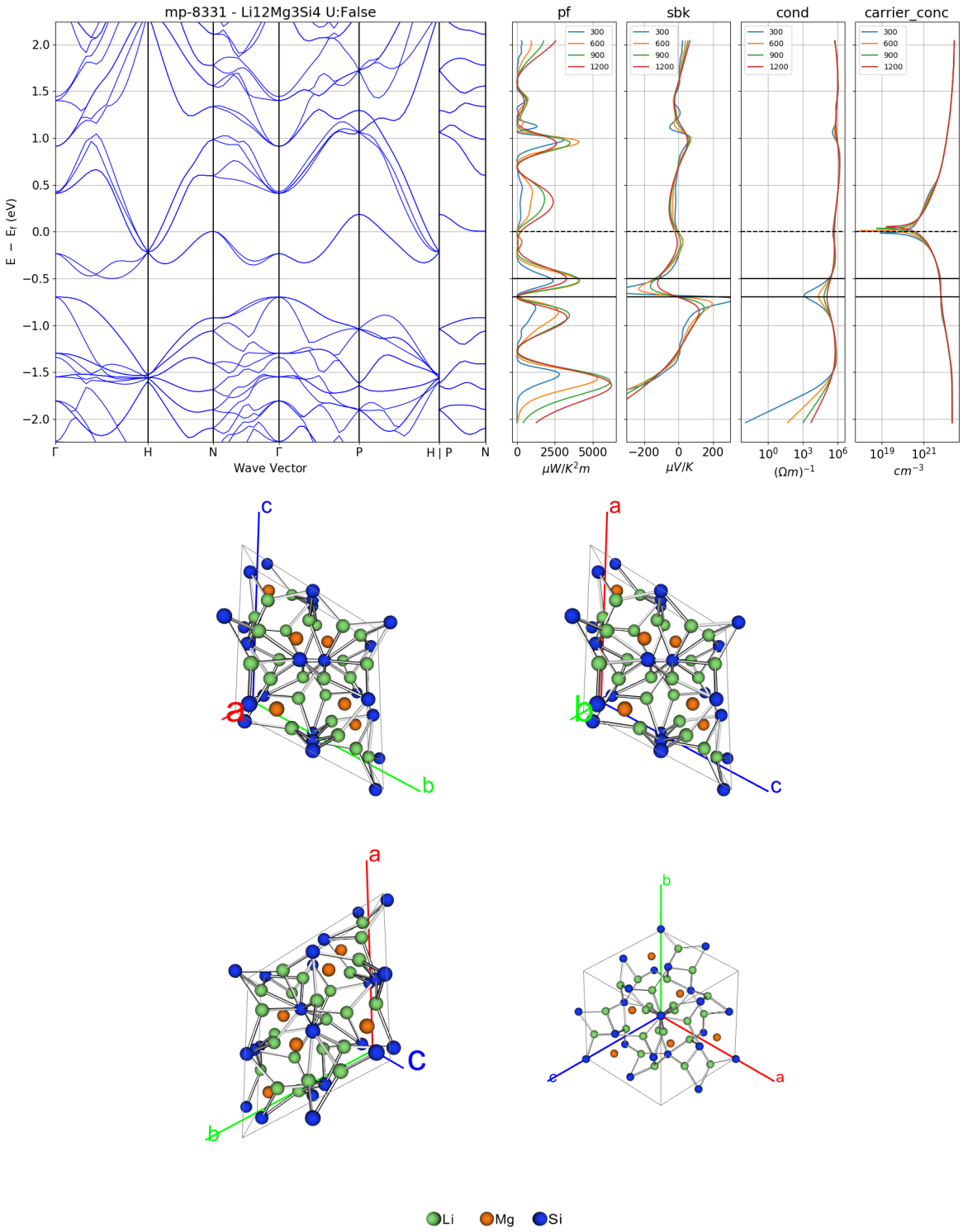

Figure 41: Structure (bottom), band structure, and calculated transport properties (top) of $Li_{12}Mg_3Si_4$ (mp-8331).

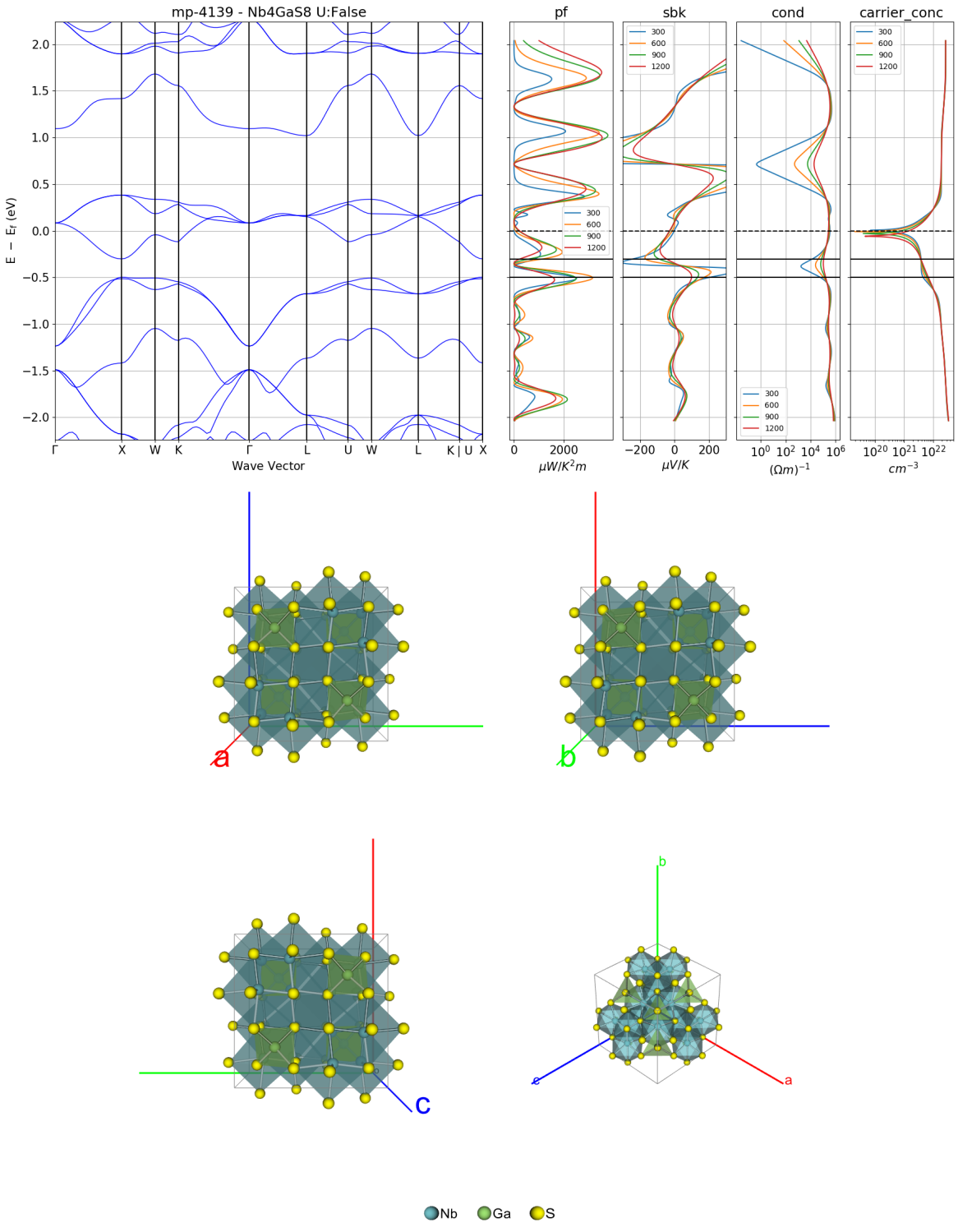

Figure 42: Structure (bottom), band structure, and calculated transport properties (top) of Nb_4GaS_8 (mp-4139).

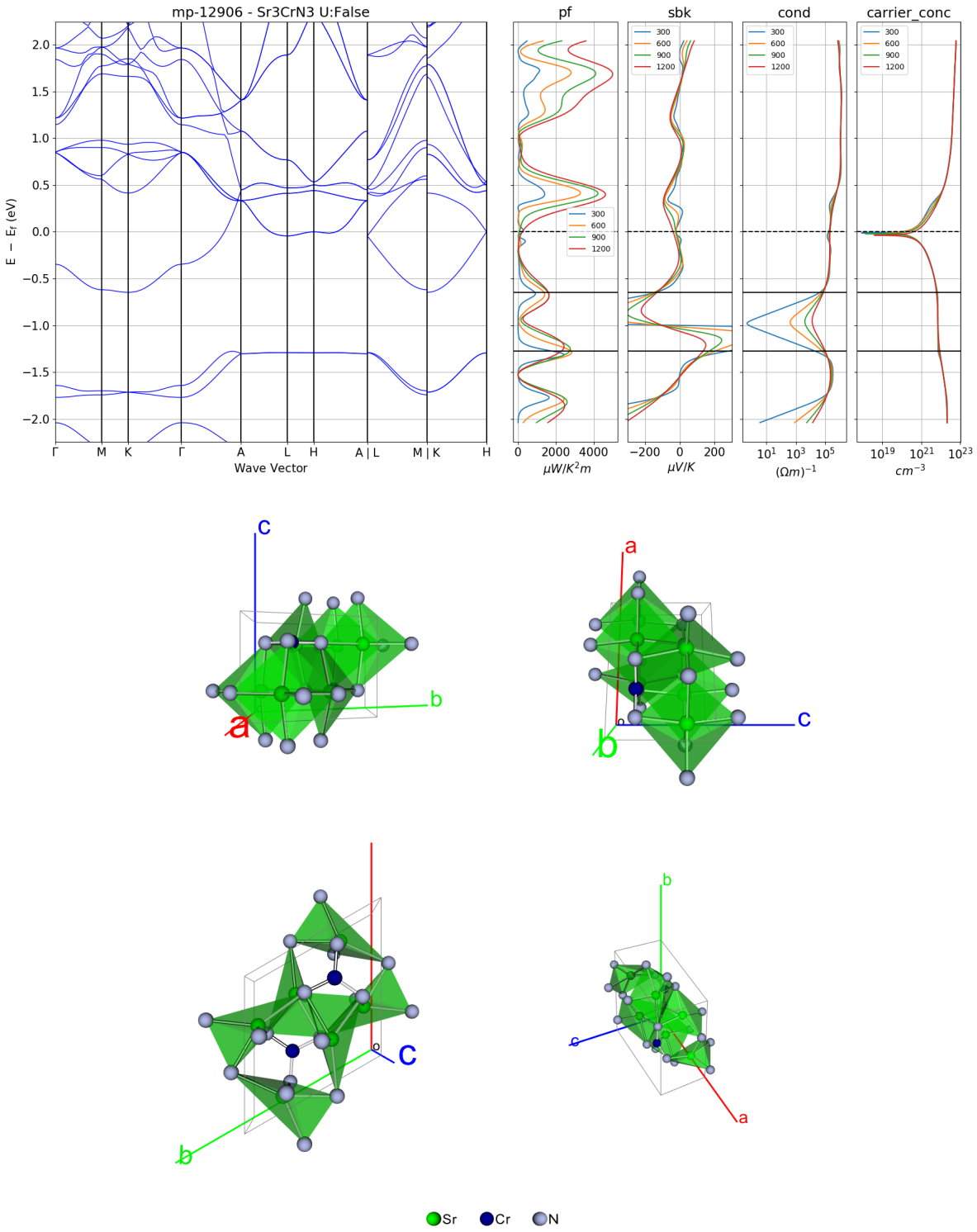

Figure 43: Structure (bottom), band structure, and calculated transport properties (top) of Sr_3CrN_3 (mp-12906).

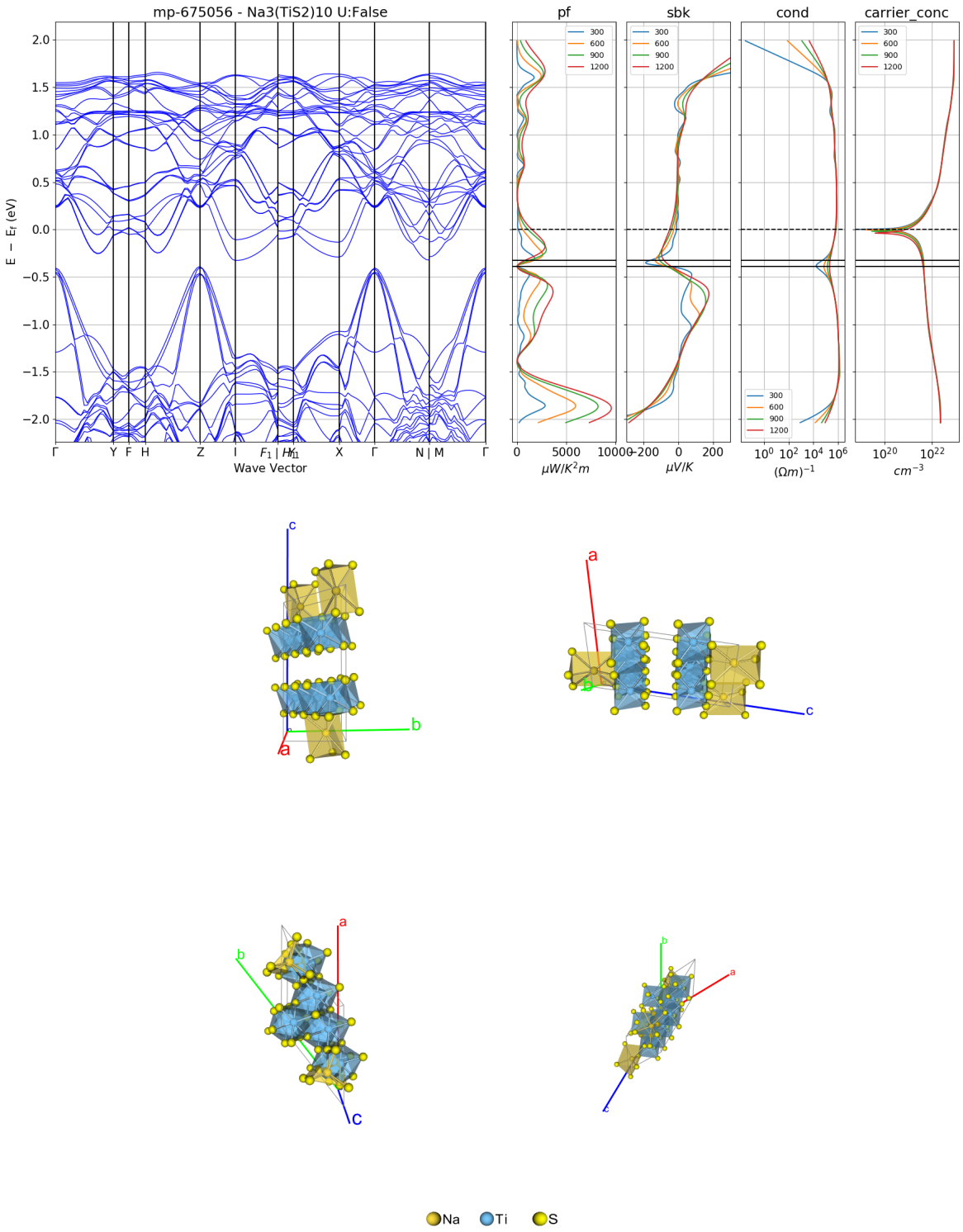

Figure 44: Structure (bottom), band structure, and calculated transport properties (top) of $Na_3(TiS_2)_{10}$ (mp-675056).

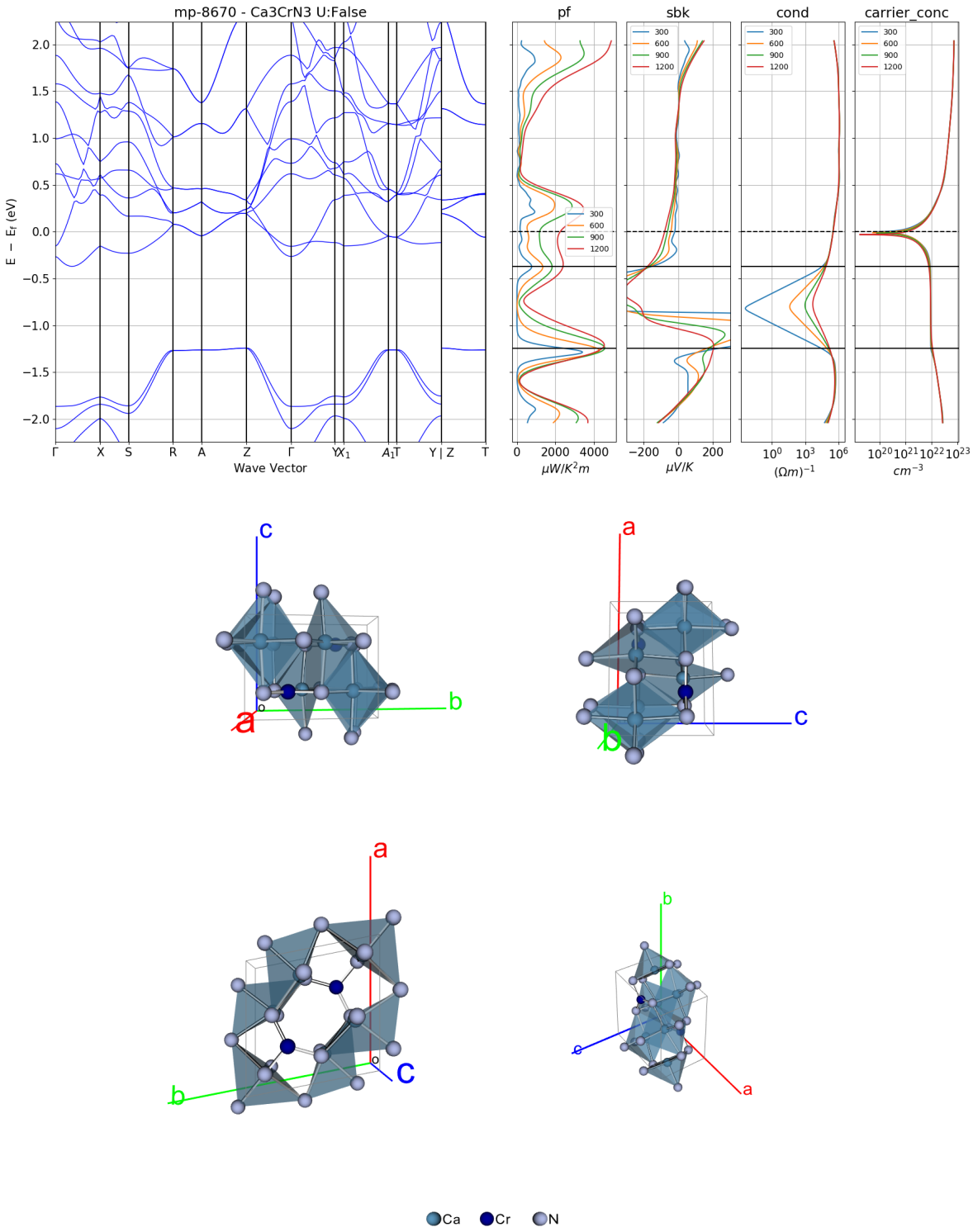

Figure 45: Structure (bottom), band structure, and calculated transport properties (top) of Ca_3CrN_3 (mp-8670).

References

[1] F. Ricci, W. Chen, U. Aydemir, G. J. Snyder, G.-M. Rignanese, A. Jain, and G. Hautier, "An ab initio electronic transport database for inorganic materials," *Scientific Data*, vol. 4, p. 170085, 2017.

[2] https://github.com/hackingmaterials/automatminer.

[3] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, "Benchmarking materials property prediction methods: The matbench test set and automatminer reference algorithm," 2020.

[4] L. Ward, A. Dunn, A. Faghaninia, N. E. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster, and A. Jain, "Matminer: An open source toolkit for materials data mining," *Computational Materials Science*, vol. 152, pp. 60 – 69, 2018.

[5] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," *npj Computational Materials*, vol. 2, p. 16028, 2016.

[6] A. M. Deml, R. O'Hayre, C. Wolverton, and V. Stevanovic, "Cheminform abstract: Predicting density functional theory total energies and enthalpies of formation of metal-nonmetal compounds by linear regression," *ChemInform*, vol. 47, no. 44, 2016.

[7] F. Faber, A. Lindmaa, O. A. Von Lilienfeld, and R. Armiento, "Crystal structure representations for machine learning models of formation energies," *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1094–1101, 2015.

[8] A. Salinas-Sanchez, J. Garcia-Muñoz, J. Rodriguez-Carvajal, R. Saez-Puche, and J. Martinez, "Structural characterization of R_2BaCuO_5 (r = y, lu, yb, tm, er, ho, dy, gd, eu and sm) oxides by x-ray and neutron diffraction," *Journal of Solid State Chemistry*, vol. 100, no. 2, pp. 201 – 211, 1992.

[9] https://epistasislab.github.io/tpot/citing/.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.