

SEARCH OF DENSE PLAIN NETS IN CRYSTAL STRUCTURES BY CLUSTER ANALYSIS

S.V. Rashchenko

*Sobolev Institute of Geology and Mineralogy, Siberian Branch, Russian Academy of Sciences,
Novosibirsk, Russia*

Novosibirsk State University, Novosibirsk, Russia

rashchenko@igm.nsc.ru

To the memory of S.V. Borisov (1930–2023)

The concept of dense plain atomic nets in crystal structures is commonly used in descriptive and comparative crystal chemistry and in polytype studies. To automatize the search of such nets, algorithms based on the Fourier analysis of atomic positions in unit cells have been developed since 1980s. However, these approaches have a number of limitations. As an alternative and addition to Fourier approaches, we propose a cluster analysis algorithm in the direct space and consider its implementation in the *nets* module of the *crystchemlib* open-source Python library.

Keywords: *plain atomic nets; cluster analysis; cation substructure; Python; crystchemlib.*

INTRODUCTION

Approaches allowing one to distinguish dense plain atomic nets (planar two-dimensional and "corrugated" pseudo- two-dimensional ones) in crystal structures were developed long ago and are well known to all crystallographers at least from the descriptions of densely packed structures. Far from being a purely descriptive abstraction, dense atomic nets of crystal structures is an essential physical entity in the processes of crystal growth and faceting, epitaxy, and phase transitions based on martensitic transformations. In descriptive crystal chemistry, distinguishing dense atomic nets allows establishing topological relationships between structures, *e.g.* to describe the diversity of polytypes. The role of dense plain nets in crystal chemistry was greatly clarified by the crystallographers Wells [1], O’Keeffe and Hyde [2], and by the Russian scientist Borisov [3].

Even though a qualified crystallographer often needs only to cast a glance at a crystal structure model to visually distinguish the presence of dense plain nets parallel to some crystallographic plane, efficient analysis in the general case requires an algorithm to find the most densely packed plain nets and to quantitatively compare their densities within one structure or between different structures. The most well-known realization of this algorithm has been developed by Borisov et al. [4, 5] since 1980s. In this approach, the structure is subjected to the Fourier transformation, and the atomic electron density is replaced by a delta-function, with an amplitude from 0 to 1, to account for the contribution of different atoms. This transformation results in a set of "conditional structure factors" F_{hkl}^0 adopting maximal values for the sets of (hkl) planes on which atoms are grouped with a smallest deviation. A definite advantage of this approach is that the "conditional structure factors" F_{hkl}^0 are closely related to the experimental reflection intensities, thus allowing one to identify dense plain nets of heavy atoms directly from single-crystal or powder XRD data, without solving the structure [6].

A disadvantage of this approach is that it reveals the fraction of atoms in a unit cell that are grouped on the (hkl) set of planes rather than the density of atomic nets *sensu stricto* (*i.e.* number of atoms per unit area). Namely, the quantity obtained in this approach is $\rho_a = F_{hkl}^0 / N$, where N is

number of atoms in the unit cell, which is not quite appropriately referred to by Borisov *et al.* as the "density of plain net filling". As an example, consider the NaCl structure for which both "conditional structure factors" F_{hkl}^0 and, respectively, "densities of plain net filling" ρ_a are identical for the (200), (220), and (222) planes due to the fact that all atoms of the unit cell lie exactly on these sets of planes, while the actual densities (atoms per unit area) of corresponding plane nets obviously differ strongly from each other and are equal to 0.13, 0.09, 0.07 Å⁻², respectively. Another example of this disadvantage can be illustrated by a hypothetical layered structure based on two equal-density plain nets parallel to (001). If they are separated by a distance of $0.5 \cdot c$, their presence is easily indicated by the "conditional structure factor" F_{002}^0 equal to the number of atoms in the unit cell. However, if the second net is shifted by $\sim 0.7 \cdot c$ with respect to the first net (*e.g.*, as a result of intercalation), the layered character of the structure in the [001] direction will be indicated by the "conditional structure factor" $F_{00.10}^0$ which can be easily ignored in the formal analysis in view of comparable F_{hkl}^0 values with larger interplanar distances.

The present work proposes an algorithm based on the cluster analysis in the direct space as an alternative and a supplement to the approach for the search of plain nets in the reciprocal space developed by Borisov *et al.*

ALGORITHM

As an example, we consider dense 3².4.3.4 type plain nets arranged parallel to (001) in the cationic substructure of the α -K₂Ca₃(CO₃)₄ high-pressure compound (*P*2₁2₁2₁, $a = 7.3905(2)$ Å, $b = 8.8156(3)$ Å, $c = 16.4817(5)$ Å) [7] (Fig. 1).

When searching for dense plain nets parallel to the (*hkl*) (hereinafter, *h*, *k*, and *l* are coprime integers) in the structure, the first step is the projection of atoms filling the unit cell onto the (*hkl*) normal (in our example, the [001] direction, Fig. 1c). The atoms on the obtained projection are grouped into one or more clusters if there are dense plain nets perpendicular to the chosen direction and are distributed relatively uniformly otherwise. As can be seen in Fig. 1c, the cation positions K1, K3, Ca4, and Ca5, constituting the 3².4.3.4 dense plain nets in the α -K₂Ca₃(CO₃)₄ structure, are projected on [001] as four clusters corresponding to four symmetrically equivalent plain nets along the *c* translation. On the contrary, the Ca2 position is "inter-net" one and is not part of the dense plain nets.

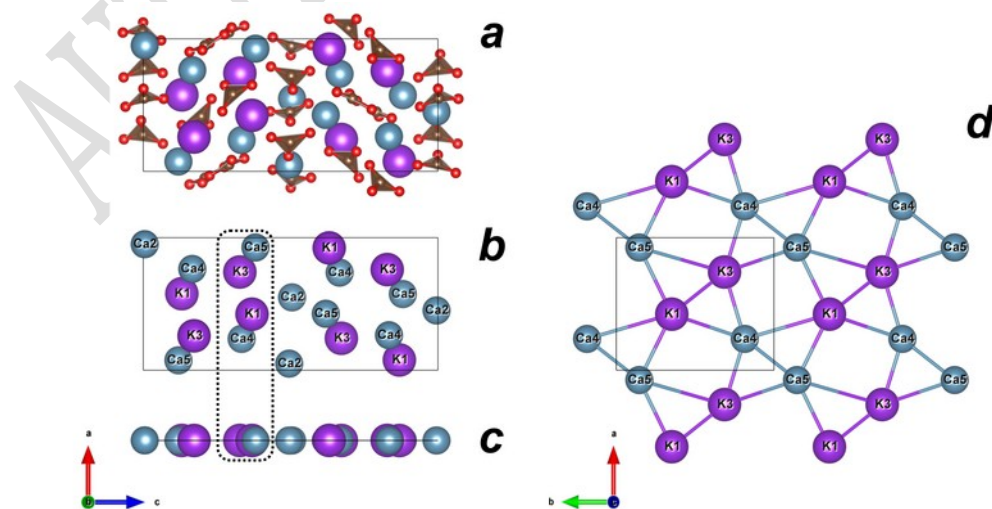


Fig. 1. α -K₂Ca₃(CO₃)₄ structure: projection along the *b* axis (a); cationic substructure (b); projection of the cationic substructure on the [001] direction (c); 3².4.3.4 plain cationic net marked in (b), (c) by a dotted rectangle (d)

Algebraically, projecting atoms filling the unit cell on the (hkl) normal corresponds to the finding of the φ_i coordinate of the i^{th} atom on the projection:

$$\varphi_i = \{x_i \cdot h + y_i \cdot k + z_i \cdot l\} \cdot d_{hkl},$$

where x_i, y_i, z_i are relative coordinates of the i^{th} atom in the unit cell; d_{hkl} is the corresponding interplanar distance; curly brackets represent the operation of taking the fractional part whereby all φ_i values fall within the interval $0 \leq \varphi_i < d_{hkl}$.

After projecting all atoms filling the unit cell on the (hkl) normal, one can find the $\Delta\varphi_{ij}$ distance between each pair of atoms on the resulting projection. Since the projection of the crystal structure on the (hkl) normal has a period of d_{hkl} , one obtains

$$\Delta\varphi_{ij} = \min(|\varphi_i - \varphi_j|, |\varphi_i - \varphi_j + d_{hkl}|, |\varphi_i - \varphi_j - d_{hkl}|).$$

Obviously, dense plain nets in the crystal structure will be formed by the groups of atoms with minimal $\Delta\varphi_{ij}$ values within the group. Therefore, dense plain nets can be searched using cluster analysis algorithms that can distinguish clusters, *i.e.* groups of objects separated from each other by systematically smaller distances than from the objects of other groups. Fig. 2 shows an example of hierarchical clustering by the "nearest neighbor" method applied for the $\alpha\text{-K}_2\text{Ca}_3(\text{CO}_3)_4$ cationic substructure in the projection on the [001] direction and visualized in the form of a *dendrogram*. The dendrogram shows how the atoms filling unit cell should be combined to obtain groups (clusters) characterized by the smallest φ_i variation corresponding to dense plain nets. The hierarchical nature of clustering is manifested in the fact that atoms with minimal $\Delta\varphi_{ij}$ (K3—Ca4, K1—Ca5, Ca2—Ca2) are paired at the lower (first) level, then the obtained pairs are combined at the next (second) level, and so on. The position of the connecting segment (link) along the vertical axis corresponds to the distance between the combined atoms / clusters along the (hkl) normal.

The dendrogram in Fig. 2 illustrates well the existence of four symmetrically equivalent $3^2.4.3.4$ plain cationic nets formed by K1, K3, Ca4, Ca5 positions perpendicularly to the [001] direction in the $\alpha\text{-K}_2\text{Ca}_3(\text{CO}_3)_4$ structure (Fig. 1).

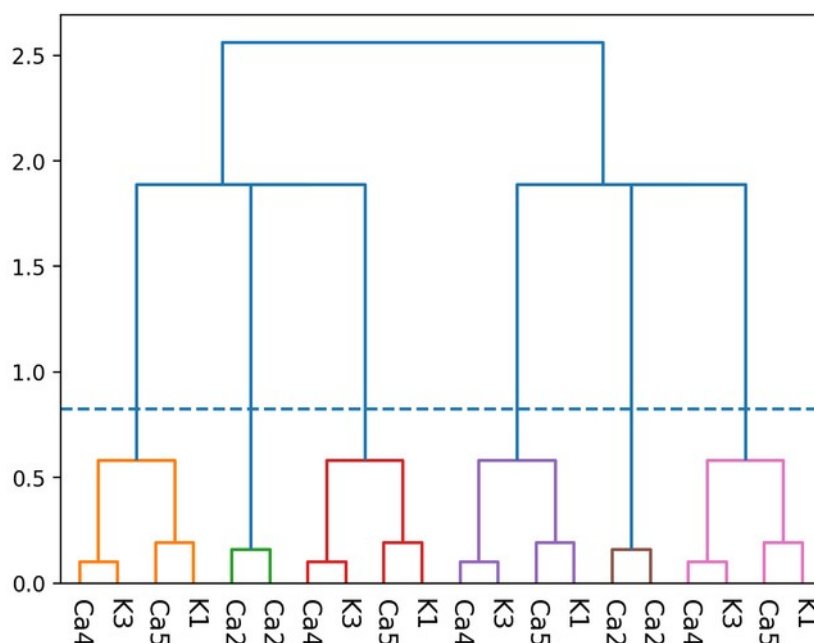


Fig. 2. Dendrogram of hierarchical clustering by the "nearest neighbor" method for of the $\alpha\text{-K}_2\text{Ca}_3(\text{CO}_3)_4$ cationic substructure in the projection on the [001] direction (vertical axis in Å)

For each of these nets, the distances between constituting clusters (*i.e.* positions of corresponding links on the vertical axis) does not exceed 0.6 Å, whereas characteristic distances between the nets are ~ 2 Å.

In contrast to the uniquely determined $\Delta\phi_{ij}$ interatomic distance, the distance between clusters can be defined differently in different hierarchical clustering methods. In the above example ("nearest neighbor" clustering), the distance between two clusters was the minimal $\Delta\phi_{ij}$ value for i^{th} and j^{th} atoms belonging to different clusters. The advantage of this approach is that on the dendrogram the sum of vertical coordinates of all links within a cluster has an important physical meaning: "thickness" w reflecting the corrugation of the corresponding plain net. The w value for the 3².4.3.4 plain cationic nets in the α -K₂Ca₃(CO₃)₄ structure is ≈ 0.9 Å.

Obviously, the number of clusters distinguished by hierarchical clustering depends on the maximum distance between the clusters at the moment when the procedure of their linkage is terminated. When searching for dense plain nets, such a limit distance can be determined as d_{hkl}/N , where N is the number of atoms in the unit cell. The physical meaning of this quantity is revealed by considering a hypothetical structure containing *no* dense plain nets perpendicular to the (hkl) normal. Obviously, atomic projections on the (hkl) normal in such structure are distributed most uniformly without forming clusters. In the ideal case of a perfectly uniform distribution, the distance between neighboring projections is equal to d_{hkl}/N . In the above example with the α -K₂Ca₃(CO₃)₄ cationic sublattice ($N=20$), the d_{001}/N value is ~ 0.8 Å (dashed line in Fig. 2), thus demonstrating the presence of four clusters corresponding to 3².4.3.4 plain cation nets alternating with the "inter-net" Ca₂ atoms.

In the proposed algorithm, dense plain nets perpendicular to some given (hkl) normal are searched. Obviously, the most dense plain net (nets) in the structure can be found by repeating this search for multiple (hkl) normals and then comparing the obtained numerical values. In the simplest case, the density of found clusters (plain nets) can be determined as a *formal density* $\sigma = M \cdot d_{hkl}/V$, where M is the number of atoms in the corresponding cluster; V is the unit cell volume. This value directly reflects the density of atoms per unit area of a plain net and is well suited to compare planar or weakly corrugated nets. However, in general case, the same formal density (σ) can correspond both to a planar net and to a strongly corrugated net 'rarefied' along the (hkl) normal. To ensure a preference to the first (more planar) net, a corrugation degree correction has to be introduced. If distances inside a cluster are limited by the d_{hkl}/N value, the maximum possible thickness of the net cannot exceed the w_{max} value of $(M-1) \cdot d_{hkl}/N$, the corrugation degree (varying from zero for the planar net to unity for the net with a thickness of w_{max}) can be determined as

$$w / w_{\text{max}} = w \cdot N / ((M-1) \cdot d_{hkl}).$$

From the corrugation degree we can easily obtain the "planarity degree" $(1 - w/w_{\text{max}})$, which in turn can be used as a weighting factor when comparing net densities:

$$\sigma_{\text{eff}} = \sigma \cdot (1 - w / w_{\text{max}}).$$

Thus, the proposed "effective density" σ_{eff} takes into account both the number of unit cell atoms belonging to the plain net and the corrugation of the latter.

ALGORITHM REALIZATION

The proposed strategy for the search of dense atomic nets in the direct space by cluster analysis was realized in the *nets* module of the *crytchemlib* open-source Python library [8]

available at <https://github.com/SergeyRa/crystchemlib>. The module is also available via library graphical interface (*Dense atomic nets* menu, Fig. 3).

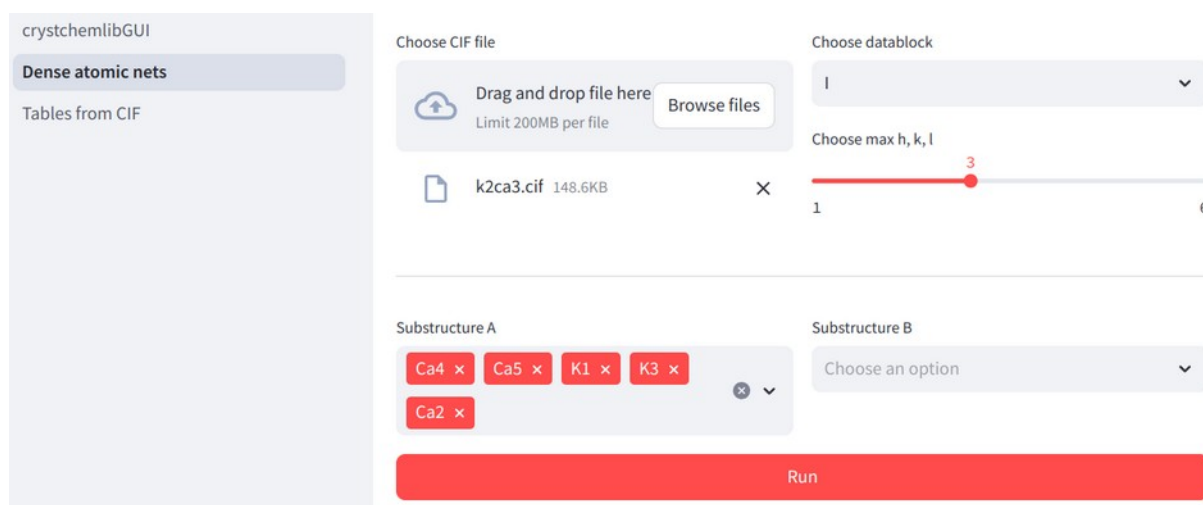


Fig. 3. Graphical interface of the *nets* module of the crystchemlib Python library

Using the graphical interface, one can load a CIF file for the analysis, select the required block of data (if the CIF file contains several structures), and set the maximum value of Miller indices for the search of dense atomic nets (in contrast to the Fourier algorithm, only directions with coprime h , k , l values are analyzed). Then the substructure for the search is defined by choosing the corresponding positions. The calculation starts with a *Run* button and can take up to few minutes for low-symmetry structures containing a large number of atoms in their unit cell (the estimated remaining time appears in the console where the GUI is started).

The calculation gives a table of analyzed (hkl) normals with the weighted-average values of the effective density of distinguished plane nets (according to the number of atoms in the nets) and a list of crystallographic directions that are symmetrically equivalent within a Laue class. The table is sorted by the decreasing weighted-average effective density of plane nets.

Detailed information on plane nets can be obtained by selecting the corresponding (hkl) normal from the drop-down list below the table. After that, a table with parameters of plane nets perpendicular to the chosen direction and the corresponding dendrogram are constructed. For each net, the following parameters are shown: atoms of the net, effective density σ_{eff} , thickness w , and the coordinate of the net along the (hkl) normal obtained by averaging the projections of the net's atoms on the (hkl) normal. A CIF file is generated for each chosen (hkl) normal. In this file, the distinguished plane nets are shown in separate data blocks. For convenience, the initial unit cell basis is transformed into a so-called "spread" basis [5] where the c^* direction coincides with the chosen (hkl) normal. As a result, the topologies of the nets and the character of their superimposition can be easily visualized using programs (e.g., VESTA [9]) able to read multiple data blocks from a CIF file (Fig. 4).

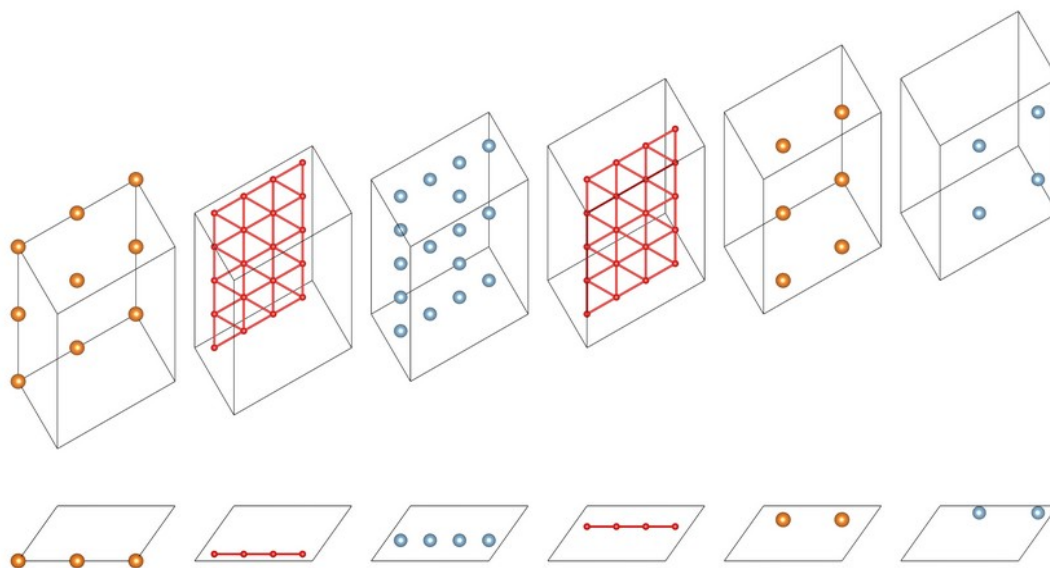


Fig. 4. VESTA visualization of dense plane nets in a spinel-type structure "spread" with respect to the [111] direction using the CIF file generated by the *nets* module the *crystchemlib* library: positions of densely packed nets (red), positions occupying octahedral (blue) and tetrahedral (yellow) voids (see the electronic version)

The *nets* graphical interface also allows analyzing relations between dense plane nets of cationic and anionic substructures. To this aim, the substructure elements should be placed into the corresponding "Substructure A" and "Substructure B" lists, after which the above analysis is supplemented by an additional step. At this stage, the effective density of type A and type B "subnets" will be additionally estimated for each found plane net; the effective density is also averaged for each "subnet". The obtained maximum averaged effective density will be observed for the crystallographic direction perpendicular not only to the most dense nets, but also the nets with the best separation of cationic and anionic subnets (the coefficient showing the degree of "subnet mixing" is also shown in the output table).

This strategy can be useful for describing crystal structures as alternating cationic and anionic nets. One such example is the NaCl structure, for which the standard search for dense atomic nets will give the [001] direction with a high (0.13 \AA^{-2}) density of *mixed* 4^4 -type atomic nets, whereas the search for cationic and anionic subnets will give the [111] direction of alternating purely cationic and purely anionic 3^6 -type nets that are used in the well-known close-packing description of this structure.

CONCLUSIONS

A cluster analysis based algorithm for the search of dense plane nets in crystal structures in the direct space was developed and implemented in the *crystchemlib* open-access Python library. This algorithm can be used for solving descriptive and comparative crystal chemistry problems and for the topological analysis of polytypes and crystal structures in general.

ACKNOWLEDGMENT

This work was funded by the Russian Science Foundation (project No. 23-77-10047, <https://rscf.ru/project/23-77-10047/>).

REFERENCES

1. A.F. Wells. Structural Inorganic Chemistry, 4th ed. Oxford, UK: Clarendon Press, **1975**.
2. M. O’Keeffe, B.G. Hyde. Plane nets in crystal chemistry. *Philos. Trans. R. Soc., A*, **1980**, 295(1417), 553–618. <https://doi.org/10.1098/rsta.1980.0150>
3. S.V. Borisov, S.A. Magarill, N.V. Pervukhina. Algoritmy i praktika kristallograficheskogo analiza atomnykh struktur (Algorithms and Practice of Crystallographic Analysis of Atomic Structures). Novosibirsk, Russia: Izd-vo SO RAN, **2012**. [In Russian]
4. N.A. Bliznyuk, E.N. Ipatova, S.V. Borisov, L.P. Solo’eva, S.V. Tsybulya. Programs for investigation of plane atomic networks in crystals. *J. Struct. Chem.*, **1988**, 28(4), 614–617. <https://doi.org/10.1007/bf00749605>
5. N.A. Bliznyuk, S.V. Borisov. Development of methods for geometric analysis of inorganic structures (review). *J. Struct. Chem.*, **1992**, 33(2), 284–304. <https://doi.org/10.1007/bf00781373>
6. S.A. Gromilov, E.A. Bykova, S.V. Borisov. Algorithms, software, and examples of pseudotranslational sub-lattice analysis for crystal structures. *Crystallogr. Rep.*, **2011**, 56(6), 947–952. <https://doi.org/10.1134/s1063774511060101>
7. S.V. Rashchenko, A.F. Shatskiy, M.A. Ignatov, A.V. Arefiev, K.D. Litasov. High-pressure synthesis and crystal structure of non-centrosymmetric $\text{K}_2\text{Ca}_3(\text{CO}_3)_4$. *CrystEngComm*, **2021**, 23(38), 6675–6681. <https://doi.org/10.1039/d1ce00882j>
8. S.V. Rashchenko. *crystchemlib*: A Python library and GUI for analysis of crystal structure datasets. *J. Appl. Crystallogr.*, **2025**, 58(1). <https://doi.org/10.1107/s1600576724011956>
9. K. Momma, F. Izumi. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.*, **2011**, 44(6), 1272–1276. <https://doi.org/10.1107/s0021889811038970>

This is a preprint of paper, accepted by the Journal of Structural Chemistry