Charge-density descriptors from a CHGCAR

Definitions, formulas and physical meaning of the 56 pydemi descriptors

Shubham Kumar Maurya, CMS Lab, IIT Kanpur • 30 October 2026

In one paragraph. Every descriptor below is a single number computed from one file, the VASP CHGCAR: the crystal structure in its header and the electron density ρ on a periodic grid. The descriptors fall into three groups. *Bonding* descriptors (19) ask how the charge is arranged around each atom and in the bonds. *Structural* descriptors (19) look at the density of the whole cell: how well it connects, its peaks and valleys, its directionality and its information content. *Heterogeneity* descriptors (18) ask how different the atoms of one structure are from each other. Every value is intensive (unchanged when the cell is replaced by a supercell) and unchanged by translating or rotating the crystal; the few exceptions to the letter of this rule are stated in the entries concerned.

How to read an entry. Each descriptor has a heading with its name as returned by `pydemi.featurize`, its units and its range. Below it: **Definition** what the number is, in words; **Formula** how it is computed; **Physics** what it tells about the material, with limiting cases; **Caveat** where present, what can mislead. A table of all 56 descriptors with page references is at the end (p. 16).

Contents

1 What a CHGCAR contains, and the common machinery **1**
1.1 The density in a CHGCAR . . . 1
1.2 Notation . . . 2

2 Bonding descriptors (19) **3**
2.1 Where the charge sits: radial moments and shell fractions . . . 3
2.2 Is the charge spherical? Gradient anisotropy . . . 4
2.3 Where charge is concentrated: the Laplacian . . . 5
2.4 Shape of the bonds: ellipticity . . . 6
2.5 Weak interactions (NCI) . . . 6
2.6 Electrostatics: the Hartree potential . . . 7
2.7 Density in the bonds . . . 8

3 Structural descriptors (19) **9**
3.1 Connectivity: percolation . . . 9
3.2 Peaks and valleys: the critical-point census . . . 10
3.3 The density floor . . . 11
3.4 Directionality: the charge-anisotropy tensor . . . 11
3.5 Information content . . . 12

4 Heterogeneity descriptors (18) **13**

5 All 56 descriptors at a glance **16**

1 What a CHGCAR contains, and the common machinery

1.1 The density in a CHGCAR

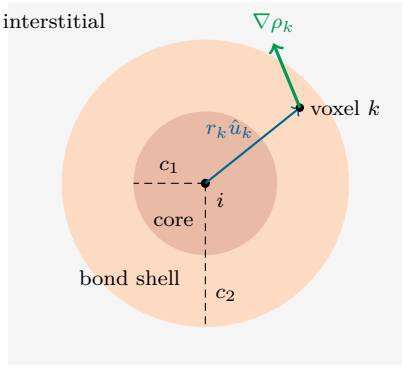
VASP uses the projector-augmented-wave (PAW) method [1]. The density written to the CHGCAR is the *valence pseudo-density*: it contains only the electrons the POTCAR treats as valence (so

semicore s, p or d states count when the POTCAR includes them, e.g. the `_pv` and `_sv` variants), and inside the augmentation sphere of each atom the true, nodal valence density is replaced by a smooth one. Three consequences run through this document:

- The deep core electrons are absent. Quantities that in an all-electron density are dominated by the core (the nuclear cusp, the atomic shell structure of the Laplacian) here describe the valence electrons only.
- Inside the augmentation spheres the density is smooth and can even be slightly negative. It need not peak at the nucleus: the pseudo-density of many s/p elements peaks on a shell or in the bonds.
- The numbers depend somewhat on the POTCAR chosen, because it fixes the number of valence electrons. Densities of one dataset should come from one consistent set of POTCARs.

1.2 Notation

Grid. The cell (volume V) is divided into N_v voxels k of volume $\Delta V = V/N_v$. ρ_k is the density at voxel k in e/Å^3 (the CHGCAR value divided by V). The total charge is $Q_{\mathrm{tot}} = \sum_k \rho_k \Delta V$, and $\langle\rho\rangle = Q_{\mathrm{tot}}/V$ is the mean density. A sum $\sum_k$ runs over all voxels; it approximates an integral, $\sum_k f_k \Delta V \approx \int f \, dV$.
Nearest nucleus. For each voxel, $i(k)$ is the nearest atom (periodic images included), r_k its distance and $\hat{u}_k$ the unit vector from that nucleus to the voxel.
Shells. Each voxel falls in one of three shells around its nearest nucleus (defaults $c_1 = 0.8$ Å, $c_2 = 1.5$ Å; optionally multiples of each atom's covalent radius):



Each voxel k is measured from its nearest nucleus.

core:	$r_k \le c_1$
bond shell:	$c_1 < r_k \le c_2$
interstitial:	$r_k > c_2$

The shell radii are absolute lengths, not atom-specific. Typical bond lengths are 1.6–3.0 Å, so the midpoints of most bonds lie 0.8–1.5 Å from both atoms, i.e. in the bond shell. For very large atoms (K, Rb, Cs, Ba, ...) part of an atom's own valence shell lies beyond 1.5 Å; the covalent-radius option scales the shells per atom when this matters.

Derivatives. The gradient $\nabla\rho$, the Laplacian $\nabla^2\rho$ and the Hessian $H_{ab} = \partial_a\partial_b\rho$ with eigenvalues $\lambda_1 \le \lambda_2 \le \lambda_3$ are computed spectrally (FFT) on the periodic, possibly non-orthogonal grid, with the full metric of the cell. For a plane-wave density this is exact up to round-off. Quantities that are defined in atomic units (the NCI thresholds and the information measures) are converted to bohr and e/bohr^3 first (1 e/bohr^3 = 6.748 e/Å^3).

Degenerate cases. A descriptor that is undefined for a special input (a perfectly uniform density, an empty shell, a structure without bonds) returns a documented constant, and a flag is set in the metadata. The only NaN is the within-element variance of a structure in which no element has two sites (Sec. 4).

Numerical stability. Candidate descriptors whose value changed by more than about 10% on the dataset when the derivatives were taken by finite differences instead of FFT, or when the grid was coarsened to 80%, were dropped (Sec. 5).

2 Bonding descriptors (19)

2.1 Where the charge sits: radial moments and shell fractions

The simplest picture of a solid is a set of atoms. These six descriptors ask how far the charge lies from its nearest nucleus, and how it is shared between the core, the bond shell and the space in between. They are moments of the *radial distribution* of the charge, $P(r) = \frac{1}{Q_{\text{tot}}} \sum_k \rho_k \, \Delta V \, \delta(r - r_k)$, the probability that an electron is at distance r from the nucleus it is closest to.

m1

Å ≥ 0

Definition The mean distance between an electron and its nearest nucleus, averaged over all electrons in the cell: every voxel is weighted by the charge it holds.

Formula

$$m_1 = \frac{\sum_k \rho_k \, r_k}{\sum_k \rho_k} = \int_0^\infty r \, P(r) \, dr.$$

Physics m_1 is a charge-weighted radius of the atoms in the crystal. For a single exponential (Slater 1s-type) density $\rho \propto e^{-2\zeta r}$ it equals $3/(2\zeta)$, so it grows as the valence charge becomes more diffuse (smaller orbital exponent ζ). In a solid the distance to the nearest nucleus cannot exceed the size of that atom's Voronoi cell, so m_1 also carries the packing: a density spread evenly over a sphere of radius R would give $m_1 = 3R/4$. Small values mean compact, tightly bound valence charge (first-row elements, compact d and f shells, short bonds); large values mean diffuse valence charge and a large volume per atom (alkali and alkaline-earth metals, open structures). Large m_1 usually goes with soft, polarizable atoms.

Caveat Only valence electrons are counted, so m_1 is a valence radius, not an all-electron one. Adding semicore states to the valence (another POTCAR) lowers it.

m2

Å^2 ≥ 0

Definition The mean *squared* distance between an electron and its nearest nucleus.

Formula

$$m_2 = \frac{\sum_k \rho_k \, r_k^2}{\sum_k \rho_k} = \int_0^\infty r^2 \, P(r) \, dr.$$

Physics Because r^2 grows faster than r, m_2 weights the outer charge more strongly than m_1. It is the radial "moment of inertia" of the charge about the nuclei. For the Slater density $m_2 = 3/\zeta^2$. The same moment $\langle r^2 \rangle$ sets the Langevin diamagnetic susceptibility of closed-shell atoms and ions, $\chi_{\text{dia}} = -\mu_0 n e^2 \langle r^2 \rangle / (6 m_e)$, per n atoms per volume [2]. So m_2 measures how much charge sits in the outer, easily deformed region of the atoms.

sigma_r2

Å^2 ≥ 0

Definition The variance of the electron–nucleus distance: how broad the radial distribution of the charge is about its mean.

Formula

$$\sigma_r^2 = m_2 - m_1^2 = \frac{\sum_k \rho_k \, (r_k - m_1)^2}{\sum_k \rho_k}.$$

Physics σ_r^2 separates the *width* of the radial distribution from its *position* (m_1). For a single exponential $\sigma_r^2 = 3/(4\zeta^2)$, so $\sigma_r^2 / m_1^2 = 1/3$ whatever ζ is. A ratio σ_r^2/m_1^2 well above $1/3$ therefore points to a distribution that is broader than a single shell: charge both close to the nuclei (compact d, f or semicore states) and far from them (diffuse s/p valence, interstitial charge). Narrow distributions belong to crystals whose valence charge occupies one well-defined shell.

`f_core` dimensionless $[0, 1]$

Definition The fraction of the electrons in the cell that lie within $c_1 = 0.8$ Å of a nucleus.

Formula

$$f_{\text{core}} = \frac{\sum_{k:\, r_k \le c_1} \rho_k}{\sum_k \rho_k}.$$

Physics The charge held tightly by the nuclei. It is high for elements whose valence (as the POTCAR defines it) is compact: small electronegative atoms (O, F, N), the $3d$ and $4f$ shells of transition metals and rare earths, and semicore p states. It is low for diffuse s/p valence charge. Since c_1 is fixed, f_{core} is not the chemist's core: it is the share of the *valence* charge that stays within 0.8 Å.

Caveat Depends on the POTCAR choice: semicore electrons treated as valence raise it.

`f_bond` dimensionless $[0, 1]$

Definition The fraction of the electrons in the shell $0.8 < r \le 1.5$ Å around the nearest nucleus.

Formula

$$f_{\text{bond}} = \frac{\sum_{k:\, c_1 < r_k \le c_2} \rho_k}{\sum_k \rho_k}.$$

Physics The shell holds the midpoints of most bonds (bond lengths 1.6–3.0 Å) and the outer part of typical valence shells, so f_{bond} is the share of the charge in the bonding region. It is large for covalent and moderately packed crystals, where the valence charge is drawn into the bonds. It is small when the charge is either kept on the atoms (large f_{core}) or spread into the interstitial (large f_{int}).

`f_int` dimensionless $[0, 1]$

Definition The fraction of the electrons farther than $c_2 = 1.5$ Å from every nucleus.

Formula

$$f_{\text{int}} = \frac{\sum_{k:\, r_k > c_2} \rho_k}{\sum_k \rho_k}, \qquad f_{\text{core}} + f_{\text{bond}} + f_{\text{int}} = 1.$$

Physics The charge in the space between the atoms: the nearly free valence electrons of metals made of large atoms, the diffuse valence shell of large anions, lone pairs pointing into voids. A large f_{int} points to delocalized, metallic charge; in close-packed crystals with short bonds few points are farther than 1.5 Å from any nucleus and f_{int} is small.

Caveat For very large atoms (K, Rb, Cs, Ba) part of the atom's own valence shell lies beyond 1.5 Å, so f_{int} then also measures the atom size. The covalent-radius option removes this.

2.2 Is the charge spherical? Gradient anisotropy

Around an isolated, spherical atom the density gradient points straight at the nucleus. Overlap with neighbours, bonds, lone pairs and aspherical d occupations all tilt it sideways.

`zeta` dimensionless $[0, 1]$

Definition The share of the density gradient that does *not* point along the line to the nearest nucleus, averaged over the cell with weight $|\nabla\rho|$.

Formula With θ_k the angle between $\nabla\rho_k$ and $\hat{u}_k$,

$$\zeta = 1 - \frac{\sum_k |\nabla\rho_k \cdot \hat{u}_k|}{\sum_k |\nabla\rho_k|} = 1 - \frac{\sum_k |\nabla\rho_k|\,|\cos\theta_k|}{\sum_k |\nabla\rho_k|}.$$

Physics $\zeta = 0$ for spherical atoms that do not overlap: the gradient is radial everywhere. If the gradient directions were random with respect to the nucleus, $\langle|\cos\theta|\rangle = 1/2$ and $\zeta = 1/2$. Two effects make ζ positive: (i) *overlap*: between two atoms the gradient is the sum of two radial fields pointing at different nuclei, so it is not radial with respect to either; (ii) *deformation*: bonds, lone

pairs and non-spherical orbital occupations. ζ is thus one number for how far the crystal's density departs from a superposition of isolated spherical atoms, the information a deformation density $\rho - \sum \rho_{\text{atom}}$ conveys, but without reference atoms. Because of the weight $|\nabla\rho|$ it is dominated by the steep outer parts of the atoms rather than by the flat interstitial. For most crystals it is a few per cent.

Limits Uniform density ($\nabla\rho = 0$): 0, flagged.

2.3 Where charge is concentrated: the Laplacian

The Laplacian compares the density at a point with its surroundings: averaged over a small sphere of radius h, $\langle\rho\rangle_{\text{sphere}} - \rho(\mathbf{r}) = \frac{h^2}{6}\nabla^2\rho + O(h^4)$. Where $\nabla^2\rho < 0$ the density is higher than around it, *locally concentrated*, as in atomic shells, covalent bonds and lone pairs. Where $\nabla^2\rho > 0$ it is depleted. In the theory of atoms in molecules the sign also tells which energy dominates: by the local virial theorem $\frac{1}{4}\nabla^2\rho = 2G + V$ (atomic units, $G > 0$ kinetic and $V < 0$ potential energy density), so a negative Laplacian marks regions where the potential energy dominates (shared, covalent interactions) and a positive one regions where the kinetic energy dominates (closed-shell, ionic or van der Waals interactions) [3, 4]. For a pseudo-density this reading is qualitative.

`lnf` dimensionless $[0, 1]$

Definition The fraction of the cell *volume* in which the charge is locally concentrated.

Formula

$$\text{lnf} = \frac{1}{N_v}\sum_k \mathbf{1}\left[\nabla^2\rho_k < 0\right].$$

Physics How much of the space is occupied by concentrated charge (bonds, lone pairs, valence shells) rather than depleted charge. A uniform electron gas has no curvature at all (flagged value 0).

Caveat Every voxel counts equally, including nearly empty ones, where the sign of a tiny Laplacian is decided by round-off. This makes lnf noisier than the charge-weighted version below.

`lnf_charge_weighted` dimensionless $[0, 1]$

Definition The fraction of the *electrons* that sit where the charge is locally concentrated.

Formula

$$\text{lnf}_\rho = \frac{\sum_{k:\,\nabla^2\rho_k<0}\rho_k}{\sum_k \rho_k}.$$

Physics The share of the charge in concentration regions (pseudo-cores, bonds, lone pairs) as opposed to the depleted regions around them. Low-density voxels carry little weight, so it is far less sensitive to the grid than lnf, and it is not interchangeable with it: a crystal can have a small concentrated volume holding most of its charge.

`lap_concentration` dimensionless always 1/2

Definition The share of the total absolute Laplacian that comes from concentration regions.

Formula

$$\frac{\sum_{k:\,\nabla^2\rho_k<0}|\nabla^2\rho_k|}{\sum_k|\nabla^2\rho_k|} = \frac{1}{2}.$$

Physics None: on a periodic cell the Laplacian integrates to zero, $\int\nabla^2\rho\,dV = 0$ (the divergence theorem with no boundary; in Fourier space the $\mathbf{G} = 0$ component of $\nabla^2\rho$ is $-|\mathbf{G}|^2\rho(\mathbf{G}) = 0$). The positive and negative parts therefore have equal weight and the ratio is $1/2$ for every crystal. It is kept so that the catalogue matches the specification, and it is the one descriptor excluded from the 55 informative ones.

`lap_concentration_valence` dimensionless $[0, 1]$

Definition The same ratio, restricted to the valence region outside the cores ($r_k > c_1$).

Formula

$$L_{\mathrm{val}} = \frac{\sum_{k:\, r_k > c_1,\ \nabla^2\rho_k < 0} |\nabla^2 \rho_k|}{\sum_{k:\, r_k > c_1} |\nabla^2 \rho_k|} = \frac{1}{2} - \frac{\int_{\mathrm{out}} \nabla^2 \rho \, dV}{2 \int_{\mathrm{out}} |\nabla^2 \rho| \, dV}.$$

Physics Outside the cores the identity above no longer forces 1/2, and by the divergence theorem

$$\int_{r>c_1} \nabla^2 \rho \, dV = -\oint_{\text{core surfaces}} \frac{\partial \rho}{\partial n} \, dS,$$

with $\partial\rho/\partial n$ the derivative along the outward normal of the core spheres. L_{val} thus measures the balance of concentration and depletion in the valence region, and it is tied to how the density behaves at the core surface. If the density still *falls* outward at 0.8 Å (a decaying valence tail), the outer region is a net depletion region and $L_{\mathrm{val}} < 1/2$; this is the common case. If it still *rises* outward (valence charge that peaks beyond 0.8 Å, charge piled up in bonds or lone pairs away from the atoms), L_{val} approaches or exceeds 1/2.

Limits Uniform density, or no voxel beyond c_1: 1/2, flagged.

2.4 Shape of the bonds: ellipticity

At a point on a bond the density is a maximum across the bond and a minimum along it: two Hessian eigenvalues are negative ($\lambda_1 \le \lambda_2 < 0$, curvatures across the bond) and one positive (λ_3, along it). If the two negative curvatures are equal the charge has a circular cross section, as in a σ bond; if they differ the cross section is elliptical, as in bonds with π character, where charge is spread in one plane [5].

`ellip_bond_bounded_avg` dimensionless $[0, 1)$

Definition The mean flattening of the charge cross section, over the bond-shell voxels where at least two curvatures are negative.

Formula With $B = \{k \in \text{bond shell} : \ \lambda_{2,k} < 0\}$ and the usual ellipticity $\varepsilon = \lambda_1/\lambda_2 - 1 \ge 0$,

$$\bar{b} = \frac{1}{|B|} \sum_{k \in B} \left(1 - \frac{\lambda_{2,k}}{\lambda_{1,k}}\right) = \frac{1}{|B|} \sum_{k \in B} \frac{\varepsilon_k}{1 + \varepsilon_k}.$$

Physics Each voxel contributes 0 when the charge around it is cylindrical ($\lambda_1 = \lambda_2$) and approaches 1 when it is flattened into a sheet ($\lambda_2 \to 0^-$). Unlike Bader's ellipticity, which is evaluated only at bond critical points, $\bar{b}$ averages over the whole bonding region, so it describes the typical cross-sectional shape of the charge ridges between the atoms: low for σ-type, cylindrical bonding, higher where the bonding charge is planar (π character, layered bonding).

Limits No such voxel: 0, flagged.

Caveat The bounded form is used because ε itself diverges as $\lambda_2 \to 0$: the mean of ε changed by a median 22% between FFT and finite-difference derivatives, while $\bar{b}$ changed by 0.2%.

2.5 Weak interactions (NCI)

The non-covalent-interaction (NCI) analysis of Johnson *et al.* [6] looks for regions of low density where the density also varies slowly. The measure of slow variation is the reduced density gradient, the dimensionless inhomogeneity that also enters GGA functionals:

$$s = \frac{|\nabla\rho|}{2(3\pi^2)^{1/3}\rho^{4/3}} = \frac{|\nabla\rho|}{2k_F\,\rho}, \qquad k_F = (3\pi^2\rho)^{1/3}.$$

s is the relative change of the density over a local Fermi wavelength: 0 in a uniform electron gas, small where the density is nearly flat, and large in exponentially decaying tails. Between two closed-shell fragments the density passes through a shallow trough, where s drops towards 0 at low ρ; that

is the NCI signature. pydemi's NCI region is

$$\mathrm{NCI} = \{k : \; s_k < 0.5, \quad \rho_k < 0.05 \text{ e/bohr}^3 \; (= 0.337 \text{ e/Å}^3)\}.$$

The sign of the middle Hessian eigenvalue separates the kinds of contact: $\lambda_2 < 0$ means the density is accumulated across the contact (an attractive, bonding-like interaction such as a hydrogen bond), $\lambda_2 > 0$ that it is depleted, as at the centre of a ring or cage (a non-bonded, sterically crowded contact).

`f_NCI` — dimensionless $[0, 1]$

Definition The fraction of the cell volume that belongs to the NCI region (low and slowly varying density).

Formula

$$f_{\mathrm{NCI}} = \frac{1}{N_v} \sum_k \mathbf{1}[s_k < 0.5 \text{ and } \rho_k < 0.05 \text{ a.u.}] \, .$$

Physics In molecules this volume marks the weak interactions. In crystals the same criterion is met above all by the low, nearly flat interstitial density of metals, which looks locally like a uniform electron gas ($s \approx 0$). In the wide empty gaps of molecular and layered crystals the density decays exponentially, s is large there, and only the thin troughs between molecules qualify. On the CMS Lab dataset f_{NCI} is indeed much larger for metals than for insulators. Read it as "volume of low, free-electron-like density" and together with f_{int} and the percolation levels.

Limits Uniform density: 0, flagged.

`NCI_attractive` — dimensionless $[0, 1]$

Definition The fraction of the NCI region in which $\lambda_2 < 0$.

Formula

$$\frac{\#\{k \in \mathrm{NCI} : \; \lambda_{2,k} < 0\}}{\#\{k \in \mathrm{NCI}\}}.$$

Physics The share of the weak-interaction volume in which the density is accumulated across the contact (attractive, bonding-like) rather than depleted (ring and cage centres, steric contacts).

Limits No NCI voxel: 0, flagged.

`sign_lambda2_rho_mean` — e/bohr^3 $(-0.05, 0.05)$

Definition The mean of the density, signed by λ_2, over the NCI region.

Formula

$$\overline{\mathrm{sign}(\lambda_2)\rho} = \frac{1}{\#\mathrm{NCI}} \sum_{k \in \mathrm{NCI}} \mathrm{sign}(\lambda_{2,k}) \, \rho_k \qquad (\rho \text{ in e/bohr}^3).$$

Physics It combines the *type* of the weak contacts (the sign) with their *strength*: the density at a weak contact grows with the strength of the interaction, lowest for dispersion contacts and higher for hydrogen bonds. Negative values mean the low-density region is dominated by attractive contacts, positive values by depleted, non-bonded ones.

Limits No NCI voxel: 0, flagged.

2.6 Electrostatics: the Hartree potential

The electrons create an electrostatic potential. Its potential energy for an electron, the Hartree potential, follows from ρ by one FFT (Poisson's equation in Fourier space),

$$V(\mathbf{G}) = \frac{4\pi k_e \, \rho(\mathbf{G})}{|\mathbf{G}|^2} \quad (\mathbf{G} \neq 0), \qquad V(\mathbf{G} = 0) = 0, \qquad k_e = \frac{e^2}{4\pi\varepsilon_0} = 14.40 \text{ eV Å}.$$

This is the repulsion an electron feels from the valence electrons only; the nuclei and cores are not included. It is large where the electrons are dense and small in the voids. The cell average is set to zero, so only differences matter.

`V_spread` eV ≥ 0

Definition The standard deviation, over the atoms of the cell, of the Hartree potential at the nuclei.

Formula With $V(\mathbf{R}_i)$ obtained by exact Fourier interpolation at nucleus i of the N_a atoms,

$$V_{\text{spread}} = \Big[\frac{1}{N_a}\sum_i \big(V(\mathbf{R}_i) - \bar{V}\big)^2\Big]^{1/2}, \qquad \bar{V} = \frac{1}{N_a}\sum_i V(\mathbf{R}_i).$$

Physics How different the electrostatic environments of the atoms are. It is exactly 0 when all atoms are symmetry-equivalent (most elemental crystals). It grows when the sites differ in how much electron charge surrounds them, which happens above all after charge transfer: an anion sits in a dense electron cloud and feels a strong repulsion, a cation that has given up its valence electrons much less. It is a rough, density-only signature of ionicity and of inequivalent sites.

`V_int_min` eV usually < 0

Definition The lowest Hartree potential in the interstitial region ($r_k > c_2$), relative to the cell average.

Formula

$$V_{\text{int,min}} = \min_{k:\, r_k > c_2} V_k, \qquad \textstyle\sum_k V_k = 0.$$

Physics The interstitial point where an electron feels the least repulsion from the other electrons, i.e. the most electron-empty part of the space between the atoms, compared with the average. Large negative values belong to open structures: van der Waals gaps of layered crystals, the space between chains and molecules, large cages. Values near zero mean a homogeneous density without deep voids, as in close-packed metals.

Limits No interstitial voxel: 0, flagged.

2.7 Density in the bonds

The neighbours of atom i are its first coordination shell: j is a neighbour if $|\mathbf{R}_{ij}| \leq 1.1\, d_i$, with d_i the distance from i to its nearest neighbour (the 10% allows for slightly distorted environments). The density at each bond midpoint $\frac{1}{2}(\mathbf{R}_i + \mathbf{R}_j)$ is obtained by exact Fourier interpolation.

`rho_mid_mean` e/Å^3

Definition The mean electron density at the midpoints of all nearest-neighbour bonds.

Formula

$$\bar{\rho}_{\text{mid}} = \frac{1}{N_b}\sum_{(i,j)} \rho\Big(\tfrac{1}{2}(\mathbf{R}_i + \mathbf{R}_j)\Big)\,.$$

Physics The midpoint density is a simple stand-in for the density at the bond critical point, which in the theory of atoms in molecules increases with bond order and bond strength [3]: high for shared, covalent bonds, low for ionic and metallic contacts. (For a homonuclear bond the two points coincide by symmetry; for a polar bond the critical point shifts towards the electropositive atom.) Only first-shell bonds are seen: in a molecular crystal these are the strong intramolecular bonds, so $\bar{\rho}_{\text{mid}}$ is high there even though the crystal is held together by weak forces.

Limits No bonds found: 0, flagged. With a PAW pseudo-density it can occasionally be negative.

`rho_mid_std` e/Å^3 ≥ 0

Definition The standard deviation of the same midpoint densities.

Formula

$$\sigma_{\mathrm{mid}} = \Big[\frac{1}{N_b}\sum_{(i,j)}\big(\rho_{\mathrm{mid},ij} - \bar{\rho}_{\mathrm{mid}}\big)^2\Big]^{1/2}.$$

Physics How uneven the bonds are. It is 0 when all first-shell bonds are equivalent (e.g. diamond, fcc metals) and large when the crystal mixes strong and weak bonds: different pairs of species, distorted polyhedra, coexisting covalent and ionic bonding.

3 Structural descriptors (19)

3.1 Connectivity: percolation

Draw the region where the density exceeds a level c, $\{\rho > c\}$, and lower c from the density maximum. At first the region consists of isolated islands around the atoms. At some level the islands join into a path that runs through the whole crystal along a lattice direction $\mathbf{a}_\alpha$: it connects a cell to its periodic image (a cluster that merely touches two opposite faces of the cell does not count; it must wind around the periodic cell).

`rho_perc_a, rho_perc_b, rho_perc_c` e/Å^3

Definition The highest density level at which the charge still forms a continuous path through the crystal along the lattice vector **a**, **b** or **c**.

Formula

$$\rho_{\mathrm{perc},\alpha} = \max\{c:\ \{\rho > c\} \text{ spans the crystal along } \mathbf{a}_\alpha\} = \max_{\text{paths along } \mathbf{a}_\alpha}\ \min_{\mathbf{r}\in\text{path}} \rho(\mathbf{r}).$$

Found by bisection over the voxel values; connectivity through shared faces, periodic boundaries handled by tracking how many times a cluster winds around the cell.

Physics The second form says it all: of all paths an electron could take through the crystal along $\mathbf{a}_\alpha$, take the one whose lowest point (bottleneck) is highest, and report that lowest density. The bottleneck is a saddle point of the density, a bond-type critical point (two negative curvatures, one positive). In a metal the valence charge forms a continuous sea and the bottleneck is high; in ionic and molecular solids every path has to cross a region of very low density, and ρ_{perc} is close to zero. In a layered crystal the path across the layers must cross the van der Waals gap, so ρ_{perc} in that direction drops to the gap density while the in-plane values stay high. This makes the three values, and their spread below, sensitive to the bonded dimensionality of the structure.

Caveat The labels a, b, c refer to the cell vectors as written in the CHGCAR. A different choice of the same lattice (another setting, a permutation of the vectors) relabels or changes the three values; their minimum, maximum and the anisotropy below are more robust.

`perc_anisotropy` dimensionless $[0, 3]$

Definition The spread of the three percolation levels relative to their mean.

Formula

$$A_{\mathrm{perc}} = \frac{\max_\alpha \rho_{\mathrm{perc},\alpha} - \min_\alpha \rho_{\mathrm{perc},\alpha}}{\frac{1}{3}\sum_\alpha \rho_{\mathrm{perc},\alpha}}.$$

Physics Whether the charge connects equally well in all directions. Ideal cases: isotropic 3-D bonding gives 0; a layered crystal with two equal in-plane levels and an empty gap direction gives $1/(2/3) = 1.5$; a chain crystal connected along one direction only gives 3. Real layered and chain compounds fall between these values and 0, well separated from 3-D bonded crystals.

Limits All three levels zero: 0, flagged.

3.2 Peaks and valleys: the critical-point census

The critical points of ρ (where $\nabla\rho = 0$) are classified by the signs of the Hessian eigenvalues: maxima (three negative), bond points (two negative), ring points (one negative) and minima or cage points (none negative) [3]. On a periodic cell their numbers obey the Morse relation $n_{\max} - n_{\text{bond}} + n_{\text{ring}} - n_{\min} = 0$, because the Euler characteristic of the 3-torus is zero [7]. pydemi counts the maxima and minima directly on the grid: a voxel is a maximum (minimum) if it is higher (lower) than all 26 neighbours; ties are broken by voxel index. The saddle counts are not descriptors (they changed by 17–20% on a coarser grid) but their Euler-sum check is reported in the metadata. Counts are divided by the cell volume.

`n_max` $1/\text{Å}^3$ ≥ 0

Definition The number of local maxima of the density per unit volume.

Formula

$$n_{\max} = \frac{\#\{k : \ \rho_k > \rho_{k'} \text{ for all 26 neighbours } k'\}}{V}.$$

Physics In an all-electron density every nucleus is a maximum, so $n_{\max}$ would equal the number density of atoms. In the pseudo-density the maxima can sit on a shell around the nucleus, in the bonds or in lone pairs, so $n_{\max}$ is the atom density plus the number of such additional peaks. Values well above the atom density point to non-spherical, structured valence charge.

`n_min` $1/\text{Å}^3$ ≥ 0

Definition The number of local minima of the density per unit volume.

Formula

$$n_{\min} = \frac{\#\{k : \ \rho_k < \rho_{k'} \text{ for all 26 neighbours } k'\}}{V}.$$

Physics The minima are the cage points: the centres of the empty spaces between the atoms, such as the interstitial holes of close-packed structures, cages of frameworks and the void centres in molecular crystals. $n_{\min}$ is therefore the number density of distinct voids.

`n_NNM` $1/\text{Å}^3$ ≥ 0

Definition The number of non-nuclear maxima per unit volume: local maxima farther than 0.8 Å from every nucleus.

Formula

$$n_{\text{NNM}} = \frac{\#\{k \text{ a maximum} : \ r_k > 0.8 \text{ Å}\}}{V}.$$

Physics A maximum away from the nuclei means that charge pools in the space between atoms. The textbook case is an electride, where electrons occupy cavities and act as anions (e.g. $[Ca_{24}Al_{28}O_{64}]^{4+}\cdot 4e^-$ and Ca_2N) [8, 9]. In a valence pseudo-density the descriptor also counts peaks in covalent bonds and lone pairs: the valence charge of Si, for instance, is well known to peak on the bonds. On the CMS Lab dataset non-nuclear maxima occur mainly in compounds of the heavier p-block elements (Si, Ge, Sn, Sb, Te, Bi, ...), so n_{NNM} is mostly a marker of directional covalent bonding and lone pairs, and only in rare cases of electrides.

`Q_NNM` dimensionless $[0, 1]$

Definition The fraction of the electrons that belong to the non-nuclear maxima.

Formula Every voxel is followed uphill, from neighbour to steepest-ascent neighbour (slope = density difference / distance), until a maximum is reached; the voxels that end at the same maximum form its basin, the grid analogue of a Bader basin [10]. Then

$$Q_{\text{NNM}} = \frac{\sum_{k \in \text{basins of the non-nuclear maxima}} \rho_k}{\sum_k \rho_k}.$$

Physics How much charge the non-nuclear maxima hold. In an electride it is the share of the electrons that act as anions; for bond and lone-pair maxima it is the charge assigned to the bonds rather than to the atoms. It is zero for most crystals.

3.3 The density floor

`rho_min` — e/Å^3

Definition The lowest density anywhere in the cell.

Formula

$$\rho_{\min} = \min_k \rho_k.$$

Physics In an all-electron density this would be the deepest void, the emptiest point of the crystal. In a CHGCAR it is usually slightly *negative* and lies near a nucleus, inside an augmentation sphere, where the PAW pseudization can undershoot. Then it reflects the POTCAR and the atoms present more than the bonding. A positive $\rho_{\min}$ is the density at the most open point of the interstitial.

`rho_min_ratio` — dimensionless

Definition The lowest density relative to the mean density of the cell.

Formula

$$\frac{\rho_{\min}}{\langle\rho\rangle}, \qquad \langle\rho\rangle = \frac{Q_{\text{tot}}}{V}.$$

Physics The same information on a scale that does not depend on how many electrons the cell has: 1 for a uniform density, close to 0 when there are deep, empty voids, negative when the PAW undershoot dominates.

Limits Zero mean density: 0, flagged.

`rho_int_mean` — e/Å^3

Definition The mean density over the interstitial region, the voxels farther than 1.5 Å from every nucleus.

Formula

$$\bar{\rho}_{\text{int}} = \frac{1}{N_{\text{int}}} \sum_{k:\, r_k > c_2} \rho_k.$$

Physics How densely the space between the atoms is filled. Where f_{int} gives the *fraction of the charge* in the interstitial, $\bar{\rho}_{\text{int}}$ gives the *density level* there. A high level means a well-filled, free-electron-like interstitial (metals); a low one, nearly empty voids (ionic, molecular and layered crystals). It can be compared with the density of a uniform electron gas, $\rho = 3/(4\pi r_s^3)$.

Limits No interstitial voxel: 0, flagged.

3.4 Directionality: the charge-anisotropy tensor

Averaging the outer product of the gradient with itself over the cell gives a symmetric 3×3 tensor with trace 1, the analogue of the structure tensor of image analysis:

$$T_{ab} = \frac{\sum_k \partial_a\rho_k\, \partial_b\rho_k}{\sum_k |\nabla\rho_k|^2}, \qquad \text{eigenvalues } t_1 \le t_2 \le t_3, \quad t_1 + t_2 + t_3 = 1.$$

T is weighted by $|\nabla\rho|^2$, so it is dominated by the steep parts of the density near the atoms. Being a second-rank tensor property of the crystal, it must be invariant under the crystal's point group (Neumann's principle). Hence: cubic crystals have $t_1 = t_2 = t_3 = 1/3$ exactly; tetragonal, hexagonal and trigonal crystals have two equal eigenvalues; only orthorhombic, monoclinic and triclinic crystals can have three different ones.

T_eigenvalues_t1, T_eigenvalues_t2, T_eigenvalues_t3 dimensionless $[0, 1]$

Definition The three eigenvalues of T, sorted in increasing order. Each is the share of the squared density gradient that points along one principal direction.

Formula $T\mathbf{e}_j = t_j\mathbf{e}_j$, $t_1 \le t_2 \le t_3$, $\sum_j t_j = 1$.

Physics If the density varies equally in all directions, all three are $1/3$. A *layered* density, which varies most strongly across the layers, puts most of the gradient along the stacking axis: one large eigenvalue (t_3) and two small ones. A *chain-like* density, uniform along the chains and varying across them, gives one small eigenvalue (t_1, the chain axis) and two large ones. The three values together tell which of these shapes the charge has.

Limits Uniform density: all $1/3$, flagged.

charge_FA dimensionless $[0, 1]$

Definition The fractional anisotropy of T: one number for how far the gradient distribution is from isotropic, as used for diffusion tensors [11].

Formula

$$\mathrm{FA} = \sqrt{\tfrac{3}{2}}\,\frac{\|T - \frac{1}{3}I\|_F}{\|T\|_F} = \sqrt{\tfrac{3}{2}}\,\frac{\sqrt{\sum_j (t_j - \frac{1}{3})^2}}{\sqrt{\sum_j t_j^2}}.$$

Physics 0 when T is isotropic (any cubic crystal), 1 when every gradient is parallel (a density that varies along one direction only). For most three-dimensional crystals it is close to zero; it is larger for strongly layered or chain-like densities and for low-symmetry cells.

Limits Uniform density: 0, flagged.

3.5 Information content

These treat the normalized density, the shape function $\tilde{\rho} = \rho/N_e$ with $N_e = \int \rho\, dV$ [12], as a probability distribution over the cell: $\tilde{\rho}(\mathbf{r})\, dV$ is the probability of finding a given electron in dV. Negative pseudo-density values are set to 0, and atomic units are used. The entropy and the disequilibrium are normalized by the cell volume, which makes them intensive and makes a uniform density the reference point.

Shannon entropy, disequilibrium and complexity depend only on the *distribution of density values* over the cell (a histogram of ρ), not on where the values are: they treat each voxel on its own. The Fisher information is the one measure that sees the spatial variation, through the gradient.

shannon_entropy dimensionless (nats) ≤ 0

Definition The Shannon entropy [13] of the shape function, measured from that of a uniform density.

Formula

$$S = -\int \tilde{\rho} \ln \tilde{\rho}\, dV - \ln V = -\int \tilde{\rho} \ln \frac{\tilde{\rho}}{1/V}\, dV.$$

Physics The second form is minus the Kullback–Leibler divergence of $\tilde{\rho}$ from the uniform distribution $1/V$: the information gained by knowing where the charge actually is, rather than assuming it is spread evenly. $S = 0$ for a uniform density (free-electron limit) and becomes more negative the more the charge concentrates in small regions: on compact atoms, in localized states, in crystals with large empty voids. e^S is the fraction of the cell volume that the charge effectively occupies.

`disequilibrium` dimensionless ≥ 1

Definition The volume-normalized integral of the squared shape function, a measure of how far the density is from uniform.

Formula

$$D = V \int \tilde{\rho}^2 \, dV = \frac{\langle \rho^2 \rangle}{\langle \rho \rangle^2} = 1 + \frac{\mathrm{Var}(\rho)}{\langle \rho \rangle^2},$$

with $\langle \cdot \rangle$ the average over the cell volume.

Physics $D - 1$ is the squared coefficient of variation of the density over the cell. $D = 1$ for a uniform density and grows as the charge piles up in some places and empties others. $1/D$ is the fraction of the cell an equivalent "flat-topped" density would fill (the second-order Rényi counterpart of e^S). Concentrated densities (compact atoms, empty voids) have large D.

`LMC_complexity` dimensionless ≥ 1

Definition The López-Ruiz–Mancini–Calbet statistical complexity [14] in its form for continuous distributions [15]: disequilibrium times the exponential of the entropy.

Formula

$$C = D\, e^S = \exp(H_1 - H_2), \qquad H_1 = -\int \tilde{\rho} \ln \tilde{\rho} \, dV, \quad H_2 = -\ln \int \tilde{\rho}^2 \, dV.$$

Physics H_1 (Shannon) and H_2 (Rényi of order 2) are two ways of counting the effective volume occupied by the charge; since $H_1 \geq H_2$, $C \geq 1$, with equality exactly when the density is constant wherever it is nonzero. The ratio of the two effective volumes does not change when the density is shifted or uniformly stretched or compressed, so C measures the *shape* of the density distribution independent of its overall scale: how strongly peaked it is compared with a flat plateau. It balances order (large D) against disorder (large e^S).

`fisher_information` $1/\mathrm{bohr}^2$ ≥ 0

Definition The Fisher information [16] of the shape function: an integral of the squared relative gradient of the density.

Formula

$$I = \int \frac{|\nabla \tilde{\rho}|^2}{\tilde{\rho}} \, dV = 4 \int |\nabla \sqrt{\tilde{\rho}}|^2 dV \qquad \text{(voxels with } \rho > 10^{-8} \text{ e/bohr}^3\text{)}.$$

Physics I measures how sharp the density is: large when it changes steeply over short distances (compact atoms, strong localization), small when it varies gently. It has a direct energetic meaning: the von Weizsäcker kinetic energy [17] $T_W = \frac{1}{8} \int |\nabla \rho|^2 / \rho \, dV$ (hartree) equals $N_e I/8$, so $I/8$ is the Weizsäcker kinetic energy per electron. T_W is the exact kinetic energy of a single-orbital density and a lower bound on the true (Kohn–Sham) kinetic energy. Unlike the three measures above, I depends on how the density values are arranged in space.

4 Heterogeneity descriptors (18)

The descriptors above describe the whole cell. These ask *how different the atoms are*. Each voxel is assigned to an atom; by default to its nearest nucleus, which gives Voronoi cells (smooth Becke partitions [18] give fractional weights). With $w_i(k) = 1$ if voxel k belongs to atom i and r_{ik}, $\hat{u}_{ik}$ measured from atom i, three per-atom quantities are computed. Each is the per-atom version of a bonding descriptor:

$$m_1^{(i)} = \frac{\sum_k w_i(k) \rho_k r_{ik}}{\sum_k w_i(k) \rho_k} \qquad \text{the radius of atom } i\text{'s charge cloud (cf. } \texttt{m1}\text{);}$$

$$f_{\mathrm{bond}}^{(i)} = \frac{\sum_{k:\, c_1 < r_{ik} \leq c_2} w_i(k) \rho_k}{\sum_k w_i(k) \rho_k} \qquad \text{the share of atom } i\text{'s charge in its bond shell (cf. } \texttt{f_bond}\text{);}$$

$$\zeta^{(i)} = 1 - \frac{\sum_k w_i(k) |\nabla \rho_k \cdot \hat{u}_{ik}|}{\sum_k w_i(k) |\nabla \rho_k|} \qquad \text{the asphericity of atom } i \text{ (cf. } \texttt{zeta}\text{).}$$

Physically, $m_1^{(i)}$ follows the size and charge state of an atom (a cation that has given away electrons has a compact cloud, an anion a diffuse one), $f_{\mathrm{bond}}^{(i)}$ how much of its charge it puts into bonds, and $\zeta^{(i)}$ how directional its bonding is. Symmetry-equivalent atoms have identical values, so every spread below comes from atoms that are *not* equivalent.

For each $X \in \{m_1, f_{\mathrm{bond}}, \zeta\}$ six statistics over the N_a atoms are reported. Let x_e be the fraction of the sites occupied by element e, $\bar{X}_e$ the mean over the sites of element e and $\bar{X}$ the mean over all sites. All variances and standard deviations are population ones (divided by the number of sites).

`m1_site_std, f_bond_site_std, zeta_site_std` — units of X — ≥ 0

Definition The standard deviation of the per-atom values over all atoms of the cell.

Formula $\mathrm{std}_i X^{(i)} = \left[\frac{1}{N_a}\sum_i (X^{(i)} - \bar{X})^2\right]^{1/2}$.

Physics The overall difference between the atoms. For m_1: how different the atoms are in size and charge state (large in strongly ionic compounds of a small cation with a large anion); for f_{bond}: whether some atoms are bonded much more strongly than others; for ζ: whether some atoms are spherical and others directionally bonded. Zero for crystals with a single inequivalent site.

`m1_site_range, f_bond_site_range, zeta_site_range` — units of X — ≥ 0

Definition The difference between the largest and the smallest per-atom value.

Formula $\max_i X^{(i)} - \min_i X^{(i)}$.

Physics The most extreme contrast between two atoms of the crystal. Unlike the standard deviation it does not depend on how many atoms of each kind there are, so a single unusual site (a dopant-like atom, a terminal atom, a site in an unusual coordination) shows up fully.

`m1_site_max, f_bond_site_max, zeta_site_max` — units of X

Definition The largest per-atom value.

Formula $\max_i X^{(i)}$.

Physics The extreme atom on the high side: the largest, most diffuse atom (m_1), the atom with the most charge in its bonds (f_{bond}), the most directionally bonded atom (ζ). Properties governed by the weakest or softest link of a structure can depend on this atom more than on the average.

`m1_site_min, f_bond_site_min, zeta_site_min` — units of X

Definition The smallest per-atom value.

Formula $\min_i X^{(i)}$.

Physics The extreme atom on the low side: the most compact atom, typically the most strongly charged cation (m_1); the atom with the least charge in its bond shell (f_{bond}); the most spherical atom (ζ).

`m1_within_element_var`, `f_bond_within_element_var`, `zeta_within_element_var`

(units of $X)^2$ ≥ 0 or NaN

Definition The part of the site-to-site variance that comes from atoms of the *same* element being different.

Formula $\sum_e x_e \operatorname{Var}_{i \in e} X^{(i)}$.

Physics Nonzero only when one element occupies inequivalent sites: different coordinations (tetrahedral and octahedral cations in a spinel), different oxidation states (mixed valence), bridging and terminal atoms, different layers of a stacking. It is the charge-density signature of crystallographically distinct sites of one element.

Limits NaN (flagged) when no element has two or more sites in the cell, because there is nothing to compare. Note that this depends on the cell: the primitive cell of rock-salt NaCl gives NaN, its conventional cell (four equivalent sites of each element) gives 0. Both mean "no inequivalent sites of one element".

`m1_between_element_var`, `f_bond_between_element_var`, `zeta_between_element_var`

(units of $X)^2$ ≥ 0

Definition The part of the site-to-site variance that comes from different elements behaving differently.

Formula $\sum_e x_e \, (\bar{X}_e - \bar{X})^2$.

Physics The site-fraction-weighted variance of the element means. For m_1 it measures the size contrast between the species, which grows with charge transfer (compact cations, diffuse anions); for f_{bond} and ζ, how differently the species bond. It plays the role that differences of tabulated atomic properties (radii, electronegativities) play in composition-based descriptors, but it is measured in the actual crystal.

Limits 0 (flagged) for a single-element crystal.

The last two split the variance exactly, as in a one-way analysis of variance:

$$\operatorname{Var}_i X^{(i)} = \underbrace{\sum_e x_e \operatorname{Var}_{i \in e} X^{(i)}}_{\text{within}} + \underbrace{\sum_e x_e \, (\bar{X}_e - \bar{X})^2}_{\text{between}} .$$

So the square of `X_site_std` is divided into a part due to inequivalent sites of one element and a part due to the chemistry of the different elements.

5 All 56 descriptors at a glance

descriptor	units	what it captures	page
Bonding (19)			
`m1`	Å	mean electron–nucleus distance	3
`m2`	Å^2	mean squared distance; outer charge	3
`sigma_r2`	Å^2	width of the radial distribution	3
`f_core`	–	charge within 0.8 Å of a nucleus	4
`f_bond`	–	charge in the bond shell	4
`f_int`	–	charge between the atoms	4
`zeta`	–	non-radial share of the gradient	4
`lnf`	–	volume of charge concentration	5
`lnf_charge_weighted`	–	charge in concentration regions	5
`lap_concentration`	–	always 1/2 (not informative)	5
`lap_concentration_valence`	–	concentration vs. depletion in the valence region	6
`ellip_bond_bounded_avg`	–	flattening of the bonding charge	6
`f_NCI`	–	volume of low, flat density	7
`NCI_attractive`	–	attractive share of weak contacts	7
`sign_lambda2_rho_mean`	e/bohr^3	signed strength of weak contacts	7
`V_spread`	eV	spread of the potential at the nuclei	8
`V_int_min`	eV	depth of the emptiest void	8
`rho_mid_mean`	e/Å^3	charge in the bonds	8
`rho_mid_std`	e/Å^3	unevenness of the bonds	9
Structural (19)			
`rho_perc_a`, `_b`, `_c`	e/Å^3	bottleneck density of the best path	9
`perc_anisotropy`	–	directionality of the connectivity	9
`n_max`	1/Å^3	density of maxima	10
`n_min`	1/Å^3	density of voids (minima)	10
`n_NNM`	1/Å^3	maxima away from the nuclei	10
`Q_NNM`	–	charge in those maxima's basins	10
`rho_min`	e/Å^3	lowest density	11
`rho_min_ratio`	–	lowest density / mean density	11
`rho_int_mean`	e/Å^3	density level between the atoms	11
`T_eigenvalues_t1`, `_t2`, `_t3`	–	principal shares of the squared gradient	12
`charge_FA`	–	anisotropy of the gradient, one number	12
`shannon_entropy`	–	spread of the charge vs. uniform	12
`disequilibrium`	–	1+ squared coefficient of variation	13
`LMC_complexity`	–	shape of the density distribution	13
`fisher_information`	1/bohr^2	sharpness; Weizsäcker kinetic energy	13
Heterogeneity (18): six statistics of $m_1^{(i)}$, $f_{\text{bond}}^{(i)}$, $\zeta^{(i)}$			
`*_site_std`	X	overall spread between atoms	14
`*_site_range`	X	most extreme difference	14
`*_site_max`	X	extreme atom, high side	14
`*_site_min`	X	extreme atom, low side	14
`*_within_element_var`	X^2	inequivalent sites of one element	15
`*_between_element_var`	X^2	differences between elements	15

Defaults: $c_1 = 0.8$ Å, $c_2 = 1.5$ Å; FFT derivatives; nearest-atom partition; `featurize` returns all 56. The 55 informative descriptors are all but `lap_concentration`.

Not included, and why

- *ELF and local energy densities* (G, V, H): a CHGCAR has no orbitals, so both would have to be estimated from the PAW pseudo-density with a density-only approximation of the kinetic energy. The results are not the true quantities.
- *Mean and spread of the unbounded ellipticity* ε: they diverge as $\lambda_2 \to 0$. On the dataset the mean changed by a median 22% (FD4 vs. FFT derivatives) and the spread by 67%; the bounded form above changed by 0.2%.
- *Saddle-point counts*: they changed by a median 17–20% when the grid was coarsened to 80%.

References

[1] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).

[2] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders, 1976), Ch. 31.

[3] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory* (Oxford University Press, 1990).

[4] R. F. W. Bader and H. Essén, The characterization of atomic interactions, J. Chem. Phys. **80**, 1943 (1984).

[5] R. F. W. Bader, T. S. Slee, D. Cremer and E. Kraka, Description of conjugation and hyperconjugation in terms of electron distributions, J. Am. Chem. Soc. **105**, 5061 (1983).

[6] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, Revealing noncovalent interactions, J. Am. Chem. Soc. **132**, 6498 (2010).

[7] A. Martín Pendás, A. Costales and V. Luaña, Ions in crystals: the topology of the electron density in ionic materials. I. Fundamentals, Phys. Rev. B **55**, 4275 (1997).

[8] S. Matsuishi *et al.*, High-density electron anions in a nanoporous single crystal: $[Ca_{24}Al_{28}O_{64}]^{4+}(4e^-)$, Science **301**, 626 (2003).

[9] K. Lee, S. W. Kim, Y. Toda, S. Matsuishi and H. Hosono, Dicalcium nitride as a two-dimensional electride with an anionic electron layer, Nature **494**, 336 (2013).

[10] G. Henkelman, A. Arnaldsson and H. Jónsson, A fast and robust algorithm for Bader decomposition of charge density, Comput. Mater. Sci. **36**, 354 (2006).

[11] P. J. Basser and C. Pierpaoli, Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI, J. Magn. Reson. B **111**, 209 (1996).

[12] R. G. Parr and L. J. Bartolotti, Some remarks on the density functional theory of few-electron systems, J. Phys. Chem. **87**, 2810 (1983).

[13] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[14] R. López-Ruiz, H. L. Mancini and X. Calbet, A statistical measure of complexity, Phys. Lett. A **209**, 321 (1995).

[15] R. G. Catalán, J. Garay and R. López-Ruiz, Features of the extension of a statistical measure of complexity to continuous systems, Phys. Rev. E **66**, 011102 (2002).

[16] R. A. Fisher, Theory of statistical estimation, Proc. Camb. Phil. Soc. **22**, 700 (1925).

[17] C. F. von Weizsäcker, Zur Theorie der Kernmassen, Z. Phys. **96**, 431 (1935).

[18] A. D. Becke, A multicenter numerical integration scheme for polyatomic molecules, J. Chem. Phys. **88**, 2547 (1988).