JOURNAL OF MATERIALS EDUCATION (Electronic)

Original Research

Teaching Conditional Comparison in Computational Materials Science through Stacking-Fault Energies

Adam Filip[1]

[1]*University of Alberta, Edmonton, Canada*

Correspondence to:
adamdilip214@gmail.com

RECEIVED:
08/05/2026
ACCEPTED:
27/09/2026
AVAILABLE ONLINE:
31/10/2026

ABSTRACT

Comparing material properties requires attention to the conditions that define each quantity, yet correct arithmetic can conceal an invalid comparison. This study develops a specialized course in computational materials science using paired stacking-fault energies for cubic borides, carbides, and nitrides. Eighteen complete fixed-cell and relaxed-cell comparisons are examined through signed differences, ordering reversals, diamond-relative selection rules, material-deletion checks, and elastic-energy curvature. The two energy rankings have a Spearman coefficient of 0.9690, although eight material pairs reverse order. Restricting the comparison to the twelve materials with positive energies under both conditions retains five reversals and reduces the coefficient to 0.9371. Three of seventeen non-reference materials change membership under a positive-energy rule referenced to diamond. Exact threshold intervals show that the maximum of four disagreements occupies a narrow multiplier range rather than a broad decision region. VN crosses zero, HfN has a positive condition difference, and the cubic PtN stiffness coefficients identify a descending tetragonal strain direction. These distinct cases support a thirty-hour course combining individual calculations, guided comparison, evidence requests, and technical argument. The instructional sequence, assessment criteria, worked answers, and transfer tasks are specified without claiming measured learning gains. The contribution connects mathematically checkable discrepancies with the disciplinary judgments required to interpret atomistic calculations and communicate their limits.

KEYWORDS

materials curriculum; stacking-fault energy; calculation conditions; elastic stability; paired comparison; technical judgment

© 2026 The Authors

This is an open access article published by InterPore under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0) (https://creativecommons.org/licenses/by-nc-nd/4.0/).

1. INTRODUCTION

A numerical ordering is often the most visible outcome of a computational materials exercise. Students calculate a property, place several compounds in a table, and identify an apparently preferable candidate. The educational difficulty begins when the comparison silently crosses the conditions under which those numbers acquire meaning. A defect energy calculated with a fixed simulation cell is not automatically interchangeable with an energy obtained after changing the cell shape. A negative value is not necessarily an exceptionally favorable instance of a positive property. The present article treats these distinctions as teachable disciplinary work, rather than as qualifications added after a ranking has already been accepted.

Current discussions of undergraduate materials education emphasize the relationship between foundational knowledge, emerging computational practices, and the capabilities expected of graduates [1]. Interdisciplinary course design similarly requires students to connect methods and explanations across disciplinary boundaries [2]. Those ambitions become concrete when a learner must decide whether two calculations answer the same question. Such a decision draws on crystallography, thermodynamics, numerical analysis, and scientific communication. It cannot be assessed adequately by asking only whether a software command executed or whether the resulting graph resembles an instructor's example. The conditions attached to a number must become part of the assessed answer.

Crystal-structure teaching provides a particularly relevant point of departure. A disciplinary-literacy review identifies extensive use of visual and symbolic representations, while observing that activities less often require learners to evaluate their information within a wider context [3]. The challenge addressed here lies at that evaluative stage. Students need to connect a formula and a crystallographic label to a modeled configuration, an energy reference, and a permitted set of atomic and cell relaxations. Recognizing a cubic structure is necessary background, but it does not establish that the modeled phase is stable or that its fault energy supports an engineering recommendation.

Materials-informatics tutorials demonstrate how competitions and focused exercises can engage participants in manipulating numerical material descriptions [4]. Workshop-based instruction also gives researchers practical access to algorithms and the decisions involved in applying them [5]. A specialized course can complement these activities by examining what remains unresolved after a calculation has produced a usable table. The emphasis is therefore on comparison, interpretation, and justification. Students are asked to identify the scope of a numerical result before attaching a claim to it. This emphasis permits substantial computational reasoning without requiring every learner to install or operate an electronic-structure package.

Online instruction in first-principles calculation has established the feasibility of teaching sophisticated computational methods beyond a conventional computer room [6]. However, changing the delivery medium does not by itself determine which intellectual activities students undertake. Reviews of virtual laboratories caution against evaluating technology chiefly through its novelty or a narrow content test [7]. The present course therefore specifies the decisions learners must make and the written evidence they must produce. A spreadsheet and a short program are alternative means of performing the same comparisons. Neither route substitutes for an explanation of why a particular comparison is valid.

The increased availability of computational material descriptions makes this educational distinction more consequential. Materials research faces persistent difficulties in combining numerical values whose definitions, quality, and calculation histories differ [8]. Modern infrastructure can preserve the sequence of operations that produced a result and expose the relationships among inputs and outputs [9]. It can also make calculations and associated scientific records accessible for inspection [10]. Students should encounter the intellectual purpose of these practices: recovering enough context to decide whether a result can answer the question they are asking. Record keeping is educationally useful when it changes the quality of a comparison, not merely when it produces a longer methods description.

Verification of electronic-structure implementations further illustrates why computational agreement requires a precise target. Large coordinated comparisons distinguish agreement among codes from wider claims about the physical accuracy and applicability of quantities [11]. Plane-wave convergence and pseudopotential choice can affect different properties unequally, especially when derivatives are involved [12]. These considerations provide a scientific reason to teach conditional interpretation before students attempt broad material screening. A consistent ranking within one numerical procedure is informative, but it does not remove the need to identify the modeled structure, relaxation convention, or physical situation to which that ranking belongs.

Educational research offers a rationale for beginning with a comparison that students cannot yet complete satisfactorily. Preparatory problem solving can support subsequent instruction when learners confront useful contrasts and the explanation addresses their attempted solutions [13, 14]. Carefully chosen scaffolding can direct attention toward the structure of a problem without supplying every solution step [15]. Here, the initial task deliberately withholds one clearly identified part of the calculation description. Students are told that the record is incomplete

and must request the missing information. The exercise rewards recognition of an unresolved comparison, rather than confidence in a premature choice.

The instructor's response to uncertainty is central to this sequence. Research on science teaching shows that uncertainty can be acknowledged while still being reduced too quickly for students to examine its significance [16]. A different teaching decision is to preserve a contradiction long enough for learners to identify what additional evidence would resolve it. Measurement-uncertainty instruction demonstrates the value of an explicitly organized sequence connecting numerical procedures with judgments about trustworthiness [17]. The course developed here applies a related educational concern to fault energies, while keeping printed-value rounding, calculation conditions, and physical model uncertainty distinct.

Feedback must also be attached to an identifiable act of reasoning. Educational feedback varies substantially in purpose and effect, so comments alone cannot be assumed to improve learning [18]. Students need opportunities to interpret comments and use them in later work [19]. A comparison exercise provides a concrete opportunity: a learner can contrast an initial recommendation with a condition-matched calculation and explain exactly why the conclusion changes. Internal feedback generated through comparison is especially relevant to this task [20]. The intellectual product is a reasoned correction of a claim, not simply a corrected cell in a spreadsheet.

Assessment should make this product visible. Reviews of rubric use indicate benefits that depend on implementation and do not justify treating a marking sheet as an automatically effective intervention [21]. More recent analysis emphasizes explicit explanation of criteria and a sequence in which students produce work, receive feedback or assess it, and then act on that comparison [22]. The present assessment therefore distinguishes numerical accuracy, treatment of calculation conditions, physical admissibility, and the quality of a bounded recommendation. A student can receive credit for identifying an unsupported inference even when a minor arithmetic error remains, while an accurate graph cannot compensate for a scientifically invalid conclusion.

The scientific vehicle is a set of stacking-fault energies because it produces several different interpretive demands within a compact exercise. The sign refers to an energy difference, the value depends on geometric constraints, and a low positive energy alone does not establish phase stability or mechanical performance. These features allow the course to move beyond a single repeated warning about uncertainty. Learners must distinguish an omitted result from zero, a change in ordering from a change in magnitude, and a numerical membership rule from a physically sufficient selection criterion. Each distinction can be checked against specific values rather than inferred from a general statement of good practice.

The distinction also matters when numerical work is divided among collaborators. One person may generate an energy, another normalize a table, and a third write the materials recommendation. A condition lost between those tasks can survive every arithmetic check. The course therefore asks students to carry the physical meaning of the quantity through successive representations. The energy definition, numerical operation, graphical position, and written conclusion must refer to the same object. This requirement makes communication part of technical correctness. It also gives an instructor a specific criterion for identifying where an otherwise polished analysis becomes insufficient, without reducing every weakness to a general problem of presentation.

The aim is to determine how condition-dependent differences in cubic-material calculations can support an assessable specialized course. The mathematical analysis examines complete paired energies, exact ordering changes, the intervals in which numerical labels differ, and the influence of individual materials. The educational design then converts those findings into a six-meeting sequence, worked tasks, and an individual technical memorandum. This is a curriculum-development study with completed numerical analyses. It contains no classroom intervention, learner measurements, or demonstrated improvement in achievement. Its educational claims concern the coherence and inspectability of the course design; its numerical claims concern the explicitly delimited material comparisons.

2. MATERIALS AND METHODS

2.1. Calculation Conditions and Paired Material Entries

Recent research investigated the stacking-fault energies of cubic borides, carbides, and nitrides using density functional theory calculations performed with the Vienna Ab Initio Simulation Package (VASP), the projector-augmented-wave method, and the Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional [23]. A plane-wave energy cutoff of 600 eV was employed to ensure computational accuracy. The investigation examined stacking-fault configurations at 0 K using atomic models consisting of five layers with four atoms per layer, while larger supercells were considered to assess interactions between neighboring faults. Atomic positions were relaxed under both fixed-cell and variable-cell conditions to evaluate the influence of structural relaxation on stacking-fault energies. The resulting energy values, presented in Table 1, provide a basis for comparing material behavior under different relaxation constraints. These findings also support subsequent comparative analyses of energy differences, material rankings, sensitivity to computational conditions, and their potential applications in materials science education.

Table 1: Paired SF2 energies and signed condition differences at 0 K.

Material	Fixed cell	Relaxed cell	Δ
Diamond	277.50	258.10	−19.40
cBN	183.80	176.50	−7.30
β-SiC	1.52	1.14	−0.38
PtN	318.00	293.70	−24.30
HfB	−26.20	−46.60	−20.40
ZrB	42.40	25.80	−16.60
HfC	1609.00	1604.00	−5.00
MoC	−929.90	−1211.00	−281.10
NbC	637.80	568.50	−69.30
ReC	−1021.00	−3206.00	−2185.00
TaC	772.60	633.40	−139.20
VC	895.50	892.60	−2.90
AlN	4813.00	4453.00	−360.00
HfN	234.50	297.90	+63.40
NbN	−130.00	−677.50	−547.50
TaN	−147.70	−970.80	−823.10
VN	91.00	−329.10	−420.10
ZrN	330.30	224.20	−106.10

The complete table contains twenty-seven material entries. Seven lack a relaxed-cell SF2 value: TiC, ZrC, GaN, InN, ScN, TiN, and YN. CrC and CrN contain alternative magnetic-branch values and are excluded from the primary paired comparison rather than resolved through an undocumented branch choice. The SF2 values for diamond and cBN are distinguished from their additional SF1 entries. Eighteen materials consequently supply one unambiguous value under each cell condition. The material is the comparison unit. The thirty-six numerical entries are paired condition values, not independent specimens, and chemical relationships among compounds remain relevant to any attempt at broader generalization.

This selection serves a finite mathematical question. It supplies enough distinct cases to test whether every ordering and every stated classification is preserved. The omitted relaxed-cell values also prevent treating complete-case selection as random. No unrecorded energy is interpolated, inferred from another compound, or replaced with zero. Retaining this boundary is particularly important because relatively large positive energies occur among the omitted entries. Conclusions about the eighteen eligible comparisons cannot be generalized automatically to all twenty-seven materials, all cubic ceramics, or all fault configurations.

2.2. Energy definitions and signed differences

For a faulted configuration and its corresponding reference, the energy per area is

$$\gamma = \frac{E_{\mathrm{fault}} - E_{\mathrm{ref}}}{A}. \tag{1}$$

The sign in Eq. (1) must be preserved. A negative numerator means that the chosen faulted configuration has lower energy than the chosen reference within the applicable convention. It does not mean that a positive fault-energy quantity has acquired an unusually advantageous magnitude. The conversion from electronic-structure units is $1\,\mathrm{eV\,\AA^{-2}} = 16021.76634\,\mathrm{mJ\,m^{-2}}$. Energies, energy densities, elastic constants, and energy per area are therefore treated as different quantities throughout the course.

Writing the fixed-cell and relaxed-cell values as γ_i^F and γ_i^R, the signed condition difference is

$$\Delta_i = \gamma_i^R - \gamma_i^F. \tag{2}$$

The analysis determines the number of positive and negative differences, their mean and median, and the median absolute difference. These summaries answer different questions. A signed average characterizes the aggregate direction of change, whereas an absolute difference characterizes displacement without cancellation. Because the individual energies span several orders of magnitude, a large change in one material can dominate a signed mean. That possibility is examined directly instead of being hidden through an automatic percentage transformation, which would be difficult to interpret near zero or across a sign change.

2.3. Ordering, normalized comparison, and exact intervals

An unordered pair reverses its relative position when

$$I_{ij} = \mathbf{1}\left[(\gamma_i^F - \gamma_j^F)(\gamma_i^R - \gamma_j^R) < 0\right]. \tag{3}$$

There are $18 \times 17/2 = 153$ such comparisons. They share materials and must not be treated as independent observations. Spearman and Kendall coefficients summarize ordering association, while Eq. (3) identifies individual reversals. Correlation alone does not establish agreement or the validity of a particular decision [24]. The analysis also restricts attention to materials with positive values under both conditions, thereby checking whether concordant negative-energy positions inflate the impression of stability in the region where a positive-energy selection rule operates.

Diamond supplies a condition-matched comparator, with $d_F = 277.5$ and $d_R = 258.1\,\mathrm{mJ\,m^{-2}}$. Define

$$x_i = \gamma_i^F / d_F, \qquad y_i = \gamma_i^R / d_R, \tag{4}$$

and the numerical membership rule

$$L_i^c(\alpha) = \mathbf{1}[0 < \gamma_i^c \le \alpha d_c], \quad c \in \{F, R\}. \tag{5}$$

Diamond is excluded from the membership denominator, leaving seventeen materials. The primary comparison sets $\alpha = 1$. The multiplier is then varied between 0.5 and 1.5. Eq. (5) is an instructional numerical rule, not a sufficient criterion for twinning, hardness, ductility, or suitability in service. The strict lower boundary is necessary because sign and small positive magnitude carry different interpretations.

When both values are positive, the two labels differ on $[\min(x_i, y_i), \max(x_i, y_i))$. Intersecting that interval with the examined multiplier range yields exact disagreement regions. The half-open boundary follows from the inclusive upper limit in Eq. (5). A material positive under only one condition requires a separate sign-aware treatment. Exact event positions are used rather than relying on a uniformly spaced multiplier grid, which can miss short intervals. The normalized coordinates also permit the locations of entry and exit to be examined geometrically without substituting a reference value from the other cell condition.

2.4. Influence, numerical perturbations, and physical admissibility

Each material is removed once, and ordering quantities are recalculated for the remaining seventeen entries. Each deletion leaves 136 unordered comparisons. This deterministic operation identifies influence within the available comparisons; it does not create confidence limits or a new validation sample. The sum of absolute differences is also decomposed by material to distinguish influence on the mean displacement from influence on ordering. These are different properties of an unusual value. A large change can dominate a magnitude total while preserving every ordering involving that material.

Printed-value sensitivity is examined by allowing each energy and the diamond reference to vary independently by $\pm 0.5\,\mathrm{mJ\,m^{-2}}$. This conservative bound exceeds the final-place rounding of the decimal entries. The test asks whether a displayed-value perturbation can change the primary numerical label. It provides no calibrated estimate of electronic-structure error. Reliable prediction uncertainty requires evidence about calibration and the behavior of an error description, not simply an arbitrary interval around a result [25]. The energies needed to distinguish reference-state effects, final fault areas, and distinct local minima are unavailable, which limits interpretation of the exceptional HfN difference.

Physical admissibility is tested separately for cubic PtN using $C_{11} = 174$, $C_{12} = 202$, and $C_{44} = 6$ GPa. The relevant small-strain conditions include $C_{11} - C_{12} > 0$, $C_{11} + 2C_{12} > 0$, and $C_{44} > 0$. Elastic-analysis tools distinguish these stability checks from calculation of aggregate moduli [26]. For normal strain $(\delta, -\delta, 0)$ and engineering shear η, the corresponding quadratic energy densities are

$$\Delta u_{\mathrm{tet}} = (C_{11} - C_{12})\delta^2, \qquad \Delta u_{\mathrm{shear}} = \tfrac{1}{2} C_{44} \eta^2. \tag{6}$$

These expressions are analytic consequences of the stiffness entries. Their curves are not new atomistic trajectories or mechanical-test observations. Keeping the two strain definitions explicit also prevents a factor-of-two error in the shear contribution.

3. RESULTS

3.1. Local ordering changes within strong aggregate association

The eighteen pairs have Spearman's coefficient 0.9690 and Kendall's coefficient 0.8954. Nevertheless, eight unordered pairs reverse order, representing 5.23% of the 153 comparisons. The affected pairs are diamond–HfN, diamond–ZrN, β-SiC–VN, PtN–HfN, PtN–ZrN, HfB–VN, ZrB–VN, and HfN–ZrN. Figure 1 localizes those reversals without turning the entire comparison into a single coefficient. Materials that do not participate in a reversal remain part of the calculation and help determine the aggregate association. Their presence does not make the affected local choices disappear.

The comparison between HfN and ZrN is particularly direct. Under the fixed-cell condition, their energies are 234.5 and 330.3 $\mathrm{mJ\,m^{-2}}$, respectively, a difference of $-95.8\,\mathrm{mJ\,m^{-2}}$. Under the relaxed-cell condition, the order changes: the values are 297.9 and 224.2 $\mathrm{mJ\,m^{-2}}$, giving a difference of $73.7\,\mathrm{mJ\,m^{-2}}$. The result establishes an ordering reversal within the two calculation conventions. It does not establish which convention is preferable for every application. A defensible answer requires the application to specify the geometric constraints and the physical interpretation of the comparison.

Twelve materials have positive energies under both conditions. Their restricted comparison retains five reversals among sixty-six pairs, giving Kendall's coefficient 0.8485 and Spearman's coefficient 0.9371. This restriction removes the sign-crossing VN entry and the materials negative under both conditions, but does not eliminate the ordering problem. The stronger coefficient for the full collection partly reflects agreement among positions outside the common positive-energy region. For teaching, the restricted comparison prevents the convenient but incorrect explanation that every reversal is merely a consequence of one material crossing zero. The HfN–ZrN reversal remains even when both values are positive.

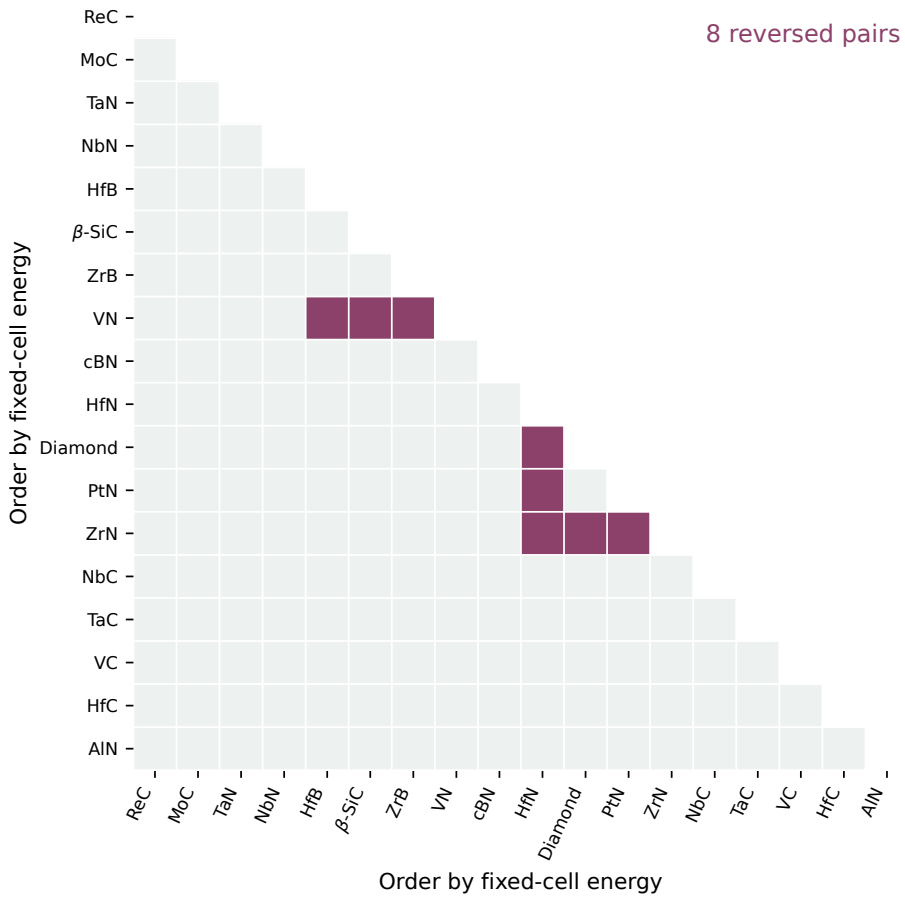


Figure 1: Location of the eight ordering reversals among the eighteen material entries. Each colored square represents one reversed unordered pair; the unused upper triangle is blank. Rows and columns follow ascending fixed-cell energy. Pale squares retain their ordering. This display identifies local changes that a single rank coefficient cannot locate.

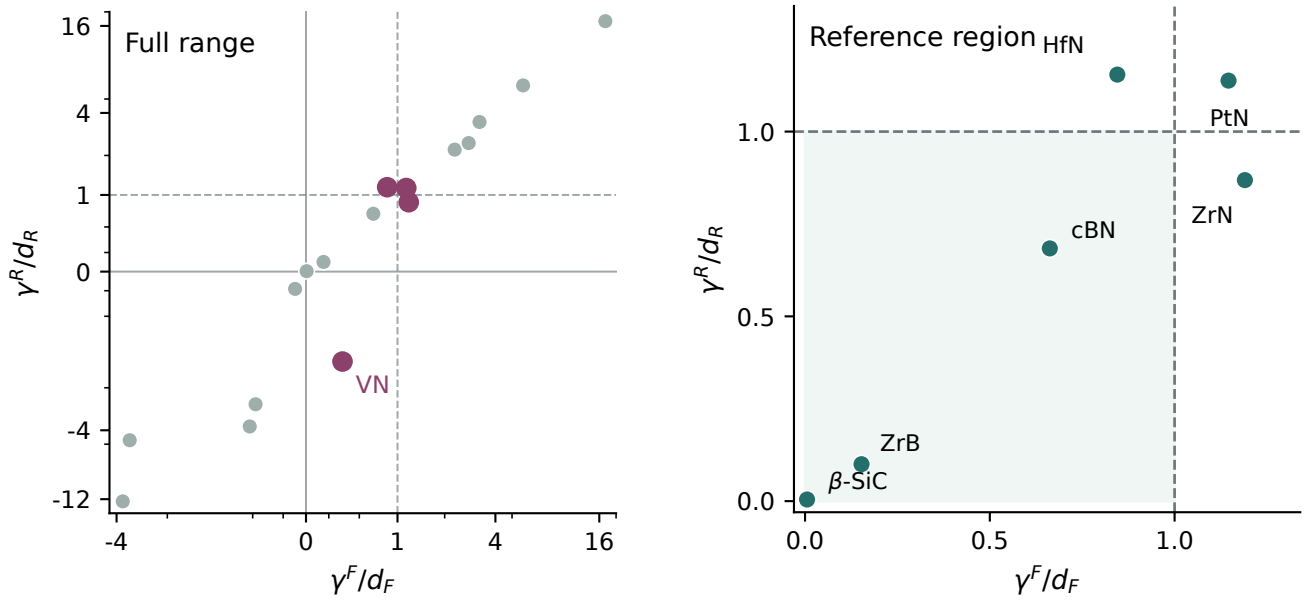
3.2. Changes in positive-energy membership

At the primary multiplier, the fixed-cell rule includes cBN, β-SiC, ZrB, HfN, and VN. The relaxed-cell rule includes cBN, β-SiC, ZrB, and ZrN. Their intersection contains three materials and their union contains six. The difference between set sizes is only one, yet three individual labels change. HfN exits through the upper boundary, ZrN enters through that boundary, and VN exits because its relaxed-cell energy is negative. These are three different relations to the declared rule, not interchangeable examples of a material becoming better or worse.

Figure 2 separates the full numerical range from the region near the reference. In the normalized plane, the upper membership boundaries are unity on the two coordinate axes, while the positivity conditions remain explicit. HfN has coordinates approximately $(0.8450, 1.1542)$ and ZrN approximately $(1.1903, 0.8687)$. They lie on opposite sides of the reference boundary in opposite directions. VN has a negative relaxed coordinate and therefore belongs to a different interpretive case. A comparison based only on whether a point lies below a diagonal would fail to express these distinct sign and reference conditions.

A uniform positive scaling would not produce these normalized differences. If every relaxed-cell energy and its diamond comparator were multiplied by the same positive factor relative to the fixed-cell values, the normalized coordinates would coincide and both labels would agree for every multiplier. The observed sign change and the different locations of HfN and ZrN exclude that simple description. This argument identifies what the normalization removes and what it preserves. It removes a common multiplicative scale while retaining material-dependent responses to the cell condition. Students can establish this result algebraically before discussing any microscopic explanation, keeping the mathematical conclusion separate from a claim about the origin of the differences.

The condition-matched normalization is consequential. The diamond value itself changes by 19.4 mJ m^{-2}, so borrowing the fixed-cell reference for the relaxed-cell values would mix two operations: a change in the target energy and a change in the comparator convention. The present calculation keeps them matched. Normalization therefore clarifies the instructional rule, although it does not remove uncertainty in the underlying energies or make the chosen comparator physically universal. Students are expected to explain that distinction before presenting a numerical membership list as a materials recommendation.



(a) Full numerical range.

(b) Linear view near the reference.

Figure 2: Condition-matched normalized SF2 energies for seventeen non-reference materials. Dashed boundaries at unity implement the diamond-relative upper limit. Panel (a) uses inverse-hyperbolic-sine axes to retain both signs and the full range; panel (b) uses linear axes and shades the region positive and below unity on both coordinates. VN lies below zero in the relaxed coordinate. HfN and ZrN occupy opposite disagreement regions.

3.3. The disagreement intervals have different widths and causes

Table 2 and Figure 3 give the material-specific disagreement intervals. cBN changes label over a short range near a multiplier of two-thirds. HfN and ZrN differ over substantially wider ranges that straddle unity. PtN contributes only between approximately 1.137931 and 1.145946. VN differs across the entire examined range because only its fixed-cell value satisfies the positivity requirement. Thus, an aggregate count conceals both the materials responsible and the mathematical reason for each change. Exact intervals also show why a coarsely spaced sweep cannot support a detailed statement about threshold behavior.

The maximum disagreement count of four occurs only on the PtN interval, whose width is approximately 0.008015. Outside that narrow region, the count falls to three, two, or one depending on the overlap of the HfN, ZrN, and cBN intervals. At unity there are three disagreements. At multipliers of 0.5, 0.75, 1.25, and 1.5, VN is the sole disagreement. It would therefore be misleading to describe four disagreements as typical across the examined range. Equally, selecting only one of those four illustrative multipliers would obscure the HfN and ZrN transitions that matter near the primary comparison.

The conservative printed-value perturbation does not change any primary non-reference label within either condition. This result is useful because it separates the observed membership changes from the final-place display of the energies. It is also limited: no conclusion follows about pseudopotential error, temperature dependence, or alternative magnetic states. The exercise illustrates a general analytical requirement without claiming a general error bound. A sensitivity calculation must identify what is being perturbed, by how much, and why that perturbation is relevant to the particular question.

Table 2: Multiplier intervals in which fixed-cell and relaxed-cell membership differ.

Material	Disagreement interval	Reason
cBN	$[0.662342, 0.683843)$	Different positive-energy entry thresholds.
HfN	$[0.845045, 1.154204)$	The fixed-cell normalized value enters first.
ZrN	$[0.868656, 1.190270)$	The relaxed-cell normalized value enters first.
PtN	$[1.137931, 1.145946)$	A narrow difference between entry thresholds; physical instability is assessed separately.
VN	$[0.500000, 1.500000]$	Positivity fails under the relaxed-cell condition throughout the examined range.

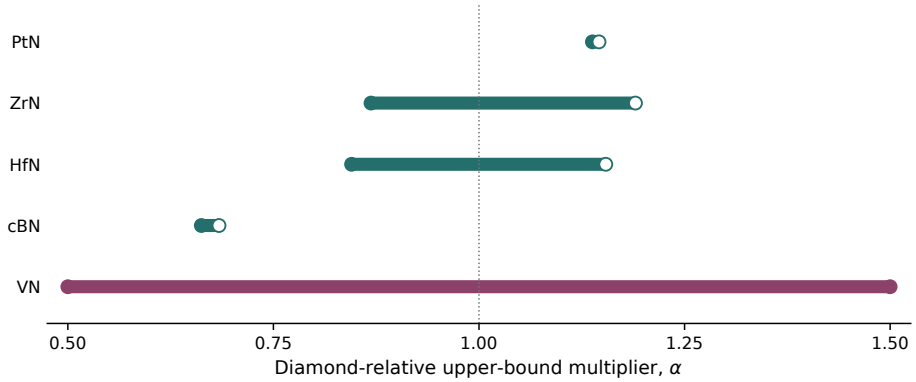


Figure 3: Material-specific intervals of disagreement over the diamond multiplier. Filled endpoints are included and open endpoints excluded. VN differs throughout the examined range because its relaxed-cell value is negative. The much shorter PtN interval creates the narrow maximum of four disagreements when it overlaps those of HfN, VN, and ZrN. The dotted line marks the primary multiplier of unity.

3.4. Influence on ordering differs from influence on magnitude

Removing HfN, VN, or ZrN separately leaves five ordering reversals among the remaining 136 pairs. Each of those deletions gives Spearman's coefficient 0.9779 and Kendall's coefficient 0.9265. Removing diamond or PtN leaves six reversals. Removing β-SiC, HfB, or ZrB leaves seven. Deleting any of the remaining ten entries leaves eight. Figure 4 displays this pattern. No single-material deletion eliminates the reversal finding, so the result cannot be attributed wholly to one exceptional compound. These counts describe sensitivity of the finite comparison, not the probability of a reversal in a wider materials population.

Seventeen condition differences are negative, and HfN supplies the only positive difference, $63.4\,\mathrm{mJ\,m^{-2}}$. The median signed difference is $-46.80\,\mathrm{mJ\,m^{-2}}$, the median absolute difference is $66.35\,\mathrm{mJ\,m^{-2}}$, and the mean signed difference is $-275.79\,\mathrm{mJ\,m^{-2}}$. The absolute differences sum to $5091.08\,\mathrm{mJ\,m^{-2}}$. ReC contributes 42.92% of that sum. ReC, TaN, NbN, and VN together contribute 78.09%, as shown in Figure 5. Such concentration explains why one signed average is a poor substitute for inspection of individual entries.

ReC illustrates the distinction particularly clearly. It has the largest absolute difference, $2185\,\mathrm{mJ\,m^{-2}}$, but occupies the lowest-energy rank under both conditions and participates in no reversal. Removing it changes the mean signed difference to approximately $-163.49\,\mathrm{mJ\,m^{-2}}$, while all eight ordering reversals remain. The median signed difference becomes $-24.30\,\mathrm{mJ\,m^{-2}}$. An instructor can therefore ask two separate questions about the same material: whether it controls a magnitude summary and whether it changes a local ordering. A single label such as outlier would not explain both effects adequately.

3.5. Sign interpretation and elastic stability

VN changes from 91.0 to $-329.1\,\mathrm{mJ\,m^{-2}}$, giving a signed difference of $-420.1\,\mathrm{mJ\,m^{-2}}$. The number of negative values increases from five to six. The newly negative entry demands attention to the reference configuration and permitted relaxation. The calculation alone does not establish a transformation rate, a kinetic pathway, or a

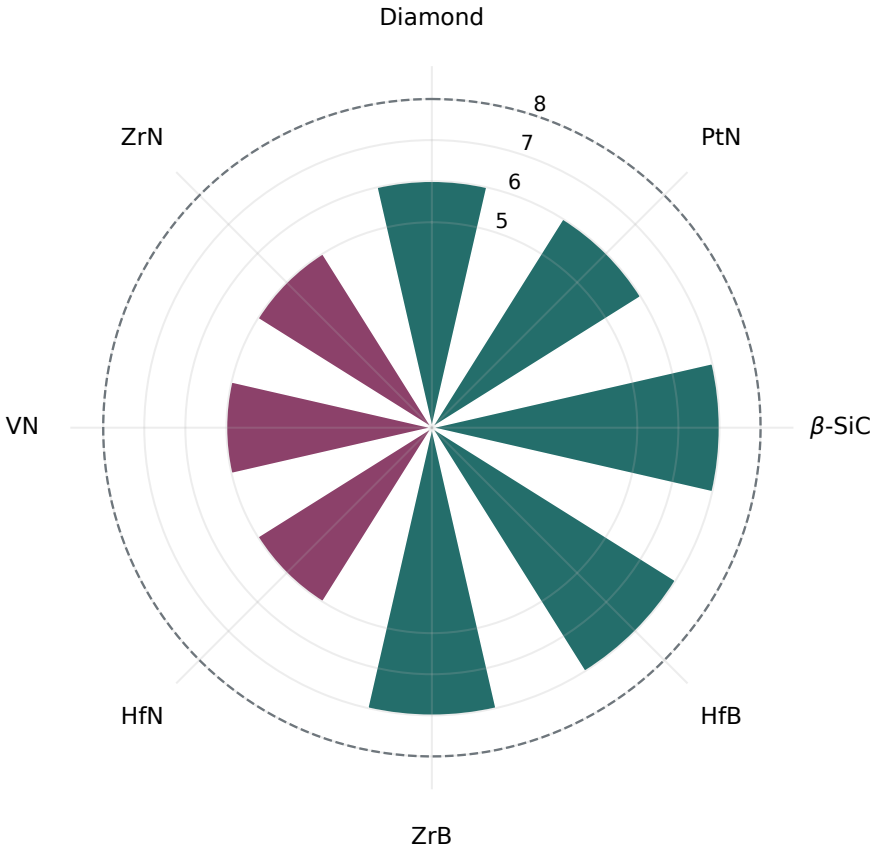


Figure 4: Ordering reversals remaining after removal of each material that participates in a reversal. Radial length gives the count among the remaining 136 pairs; the dashed circle marks the eight reversals with all entries retained. Removing HfN, VN, or ZrN leaves five. Each of the ten other materials, omitted from the radial display for clarity, leaves eight reversals when removed.

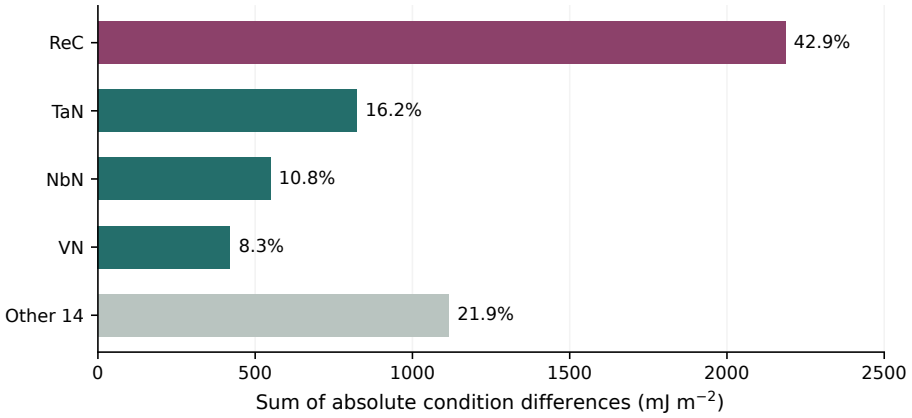


Figure 5: Contribution to the total absolute condition difference of $5091.08\,\mathrm{mJ\,m^{-2}}$. ReC supplies 42.92%; ReC, TaN, NbN, and VN together supply 78.09%. The final bar aggregates the fourteen remaining entries. Contribution to a magnitude total is distinct from participation in an ordering reversal: ReC participates in none.

material-processing recommendation. Calling the relaxed value merely a lower member of the positive-energy ordering would erase the change in the meaning of the comparison. The course consequently treats sign interpretation as a separate task from ranking and percentage change.

HfN requires a different response. Its relaxed-cell energy exceeds the fixed-cell value, although relaxing the same total-energy problem over a larger admissible set cannot increase the exact minimum. The quantity here is an energy difference divided by a fault area. Changes in the reference energy, area convention, local minimum, or numerical

treatment can therefore matter. The component energies, final areas, and relaxed structures are insufficiently detailed to diagnose the cause from the paired values alone. The positive difference is retained. Its educational value lies in requiring a precise evidence request rather than an unsupported correction that forces the entry to follow the dominant trend.

For cubic PtN, $C_{11} - C_{12} = -28$ GPa, while $C_{11} + 2C_{12} = 578$ GPa and $C_{44} = 6$ GPa. Passing the latter two conditions does not compensate for failing the first. Figure 6 converts that failed condition into a visible energy argument: the tetragonal contribution decreases away from zero strain, whereas the isolated shear contribution is positive. At a strain magnitude of 0.01, the respective energy-density changes are −2.8 and 0.3 MPa. The plotted curves express a local quadratic elastic approximation, with different normal and engineering-shear coordinates explicitly identified. They are not measured deformation curves.

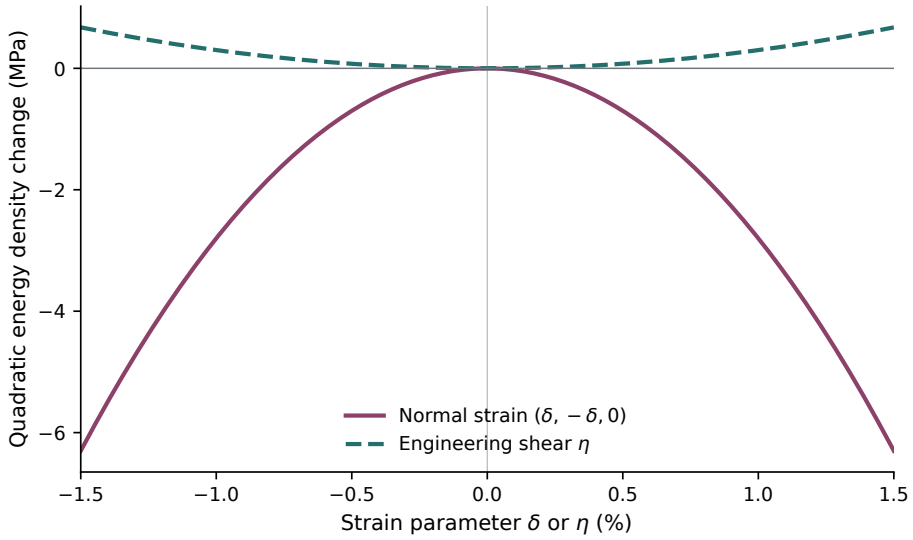


Figure 6: Analytical small-strain energy-density contributions for cubic PtN using its zero-pressure stiffness coefficients. Normal strain $(\delta, -\delta, 0)$ gives $-28\delta^2$ GPa, while engineering shear gives $3\eta^2$ GPa. The negative normal-strain curvature identifies an unstable elastic direction despite positive shear curvature. The two abscissa parameters denote different strain paths. These are quadratic elastic calculations, not deformation measurements or new DFT trajectories.

PtN remains in the arithmetic comparisons because those comparisons ask how the numerical entries behave. Its instability prevents an unrestricted physical recommendation for the cubic phase. This distinction avoids two opposite errors: deleting a difficult value merely because it complicates a plot, and accepting that value as physically sufficient merely because it can be ranked. Modern elastic-analysis software provides many derived quantities, but the availability of a modulus or anisotropy display does not replace a stability check [27, 28]. The scientific conclusion is conditional inclusion for numerical examination, followed by a separate judgment about the modeled phase.

4. COURSE DESIGN AND ASSESSMENT

4.1. Position within a materials programme

The course is intended for senior undergraduates or beginning postgraduates who have encountered crystal planes, elementary defects, energy minimization, and basic matrix elasticity. It comprises twelve contact hours and eighteen hours of independent work. The thirty-hour allocation is a design specification, not a measured estimate of student effort. It permits insertion into an existing computational-materials elective or a short specialist course. The teaching objective is to connect calculations with defensible comparison statements. Students are not expected to learn an entire electronic-structure package while simultaneously learning the physical meaning of every quantity used in the exercise.

A spreadsheet route and a programming route are available because installation difficulties should not determine access to the disciplinary task. Cloud-based instructional tools can reduce technical entry barriers for beginners, although they introduce their own delivery choices [29]. The course can instead use local software and printed

input tables when connectivity is unreliable. In either route, a learner must expose intermediate calculations and state the rule being applied. An executable result without an inspectable explanation does not meet the outcome. Conversely, a correct explanation supported by a transparent spreadsheet is not treated as inferior merely because it lacks a particular programming language.

The course has five assessable performances: recovering the conditions of a quantity, preserving distinctions among missing values and alternative branches, checking paired comparisons with the correct counting unit, separating numerical membership from physical admissibility, and communicating a conclusion with explicit limits. These performances are connected in the teaching sequence rather than taught as independent vocabulary items. Table 3 specifies the six meetings and their required products. The course content moves from an incomplete comparison toward a technical argument whose assumptions another reader can inspect. No teaching observation is represented as having occurred merely because the sequence is specified in detail.

Figure 7 allocates planned time among calculation, comparison, physical interpretation, and feedback within the six meetings.

Table 3: Instructional sequence and required evidence.

Meeting	Disciplinary task	Required product	Independent hours
1	Compare incomplete fixed-cell entries; identify the missing condition.	Preliminary ordering with prioritized requests for calculation details.	2
2	Recover energy references, units, fault configurations, and branch choices.	Unit conversion and a complete account of inclusion and exclusion.	3
3	Calculate paired differences and reversed orders.	Verified numerical table and eight manually inspectable pair reversals.	4
4	Interpret VN, investigate HfN, and test cubic PtN stability.	Three distinct evidence-based explanations and requests.	3
5	Apply matched diamond references and exact multiplier intervals.	Condition-specific membership and criticism of the numerical rule.	3
6	Analyze withheld comparisons and defend an individual conclusion.	Technical memorandum with decisive calculations and a stated inference boundary.	3

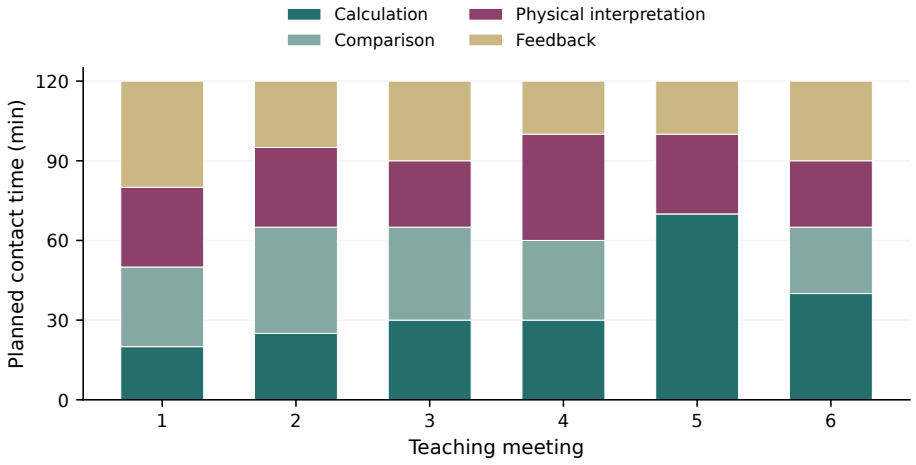


Figure 7: Allocation of the 720 planned contact minutes across the six meetings. The totals are 215 minutes of calculation, 160 of comparison, 180 of physical interpretation, and 165 of feedback. Every meeting occupies 120 minutes. The bars express instructional choices and do not represent classroom observations.

4.2. Initial comparison and recovery of calculation conditions

The opening handout contains fixed-cell values for cBN, β-SiC, HfB, ZrB, NbC, TaC, VC, and HfC. It identifies the quantity and unit correctly, while stating openly that part of the calculation description is unavailable at this stage. Students individually rank the entries and write what further information is needed before a recommendation could be justified. The preliminary answer is retained for later comparison but is not graded for guessing the withheld condition. The instructor then assembles the different information requests and uses them to distinguish uncertainty about arithmetic from uncertainty about what the arithmetic is comparing.

Small-group discussion follows the individual attempt so that every participant has an initial position to examine. Productive collaboration requires more than placing learners around the same table; attention must be directed toward what makes a difficult problem difficult [30]. Students compare the assumptions behind their preliminary choices and identify which missing item would most change their confidence. The instructor supplies the cell-shape distinction only after these requests have become explicit. This sequencing follows the instructional value of linking explanation to prior attempts, while avoiding the idea that prolonged unsupported failure is intrinsically beneficial [31].

The second meeting connects energy references with dimensional reasoning. Learners derive the conversion in Eq. (1), identify SF1 and SF2 as different configurations, and account for every excluded entry. They must explain why a second number in a table cell is not automatically a replicate. The two magnetic-branch entries are treated as unresolved choices in the present exercise, while the seven absent relaxed-cell entries remain absent. The required product is a concise record of inclusion decisions and their consequences. The exercise does not reward completing every table cell at the expense of changing the meaning of its contents.

The instructor should separate help with notation from help with interpretation. A learner unfamiliar with scientific notation may need a worked unit conversion before undertaking the full comparison, whereas a learner confident with arithmetic may need the stronger challenge of identifying an invalid inference. Research connecting cognitive load with regulation of effort supports attending to these different demands [32]. The course therefore permits brief preparation activities for mathematical prerequisites. Such preparation is an access measure, not evidence that prerequisite instruction has already produced a learning gain in the intended population.

4.3. Paired calculations and disagreement analysis

The third meeting requires students to calculate signed differences and enumerate ordering reversals. The checkable output contains eighteen material rows and a separate list of the eight reversed pairs. Students must also distinguish the paired values from the 153 dependent comparisons. A script can generate the pair list, but the learner must verify at least one reversal manually. The HfN–ZrN comparison is suitable because the energies remain positive while their order changes. This prevents a correct answer from being explained entirely through sign handling and gives the instructor a way to test whether the relation itself has been understood.

The difference between correlation and local invariance becomes an explicit written task. Students are given the full coefficient and asked what it permits them to conclude about one specified pair. They then inspect the reversal matrix and compare the full collection with the twelve common-positive entries. A successful answer identifies the different objects summarized by a rank coefficient and an individual comparison. It does not need a significance test. The pedagogical target is a logically sufficient conclusion from the available arithmetic, including the recognition that a large association coefficient cannot answer a stronger question merely because it appears statistically impressive.

The fifth meeting, after the physical-admissibility work, develops the diamond-relative rule. Students first state both boundaries in Eq. (5), then calculate membership under each condition. They identify entry and exit materials before varying the multiplier. The exact interval for PtN is useful because it is narrower than a commonly chosen numerical step. Students must explain which events a coarse sweep would miss, not simply generate a smoother-looking curve. The final comparison distinguishes a sign-driven disagreement from disagreements caused by the location of an upper cutoff. The threshold is an object of criticism as well as an arithmetic input.

4.4. Physical reasoning and requests for additional evidence

The fourth meeting examines VN, HfN, and PtN as different diagnostic cases. For VN, the expected response concerns the energy reference and the consequence of crossing zero. For HfN, the learner must retain the positive difference and identify which missing component quantities would be needed to investigate it. For PtN, the learner calculates the negative tetragonal coefficient and explains why a positive shear coefficient is insufficient for stability. A general assertion that the values are uncertain receives less credit than a specific account of what is known, what is not determined, and which calculation would distinguish the competing explanations.

This treatment also positions automated elastic tools appropriately. Packages that calculate stiffness coefficients can implement several strain-matrix choices and temperature treatments [33]. Tools for visualizing elastic anisotropy can make directional behavior accessible, but their displays remain consequences of supplied tensor entries [34]. Students should therefore learn to ask whether the tensor is appropriate and stable before interpreting an attractive directional surface. The course does not require use of these packages. It uses their scientific capabilities to clarify the distinction between producing a derived quantity and establishing the physical premises under which that quantity is meaningful.

The evidence request is a substantive technical product. For HfN, a useful request identifies the faulted and reference energies under each constraint, the area used in normalization, the final geometry, and convergence behavior. For VN, it asks how the permitted relaxation changes the reference relationship and whether the final structure preserves the intended interpretation of a fault. For PtN, it asks whether a different phase or structure is appropriate before advancing a materials recommendation. These requests do not invent missing values. They show that recognizing the limit of a calculation can lead to a specific next scientific action.

4.5. Assessment products and individual transfer

The assessment comprises a calculation notebook worth thirty-five points, a comparator analysis worth twenty-five, and a technical memorandum worth forty. Table 4 identifies the evidence attached to those weights. The distribution is a course-design choice and has not been psychometrically validated. Clear criteria are accompanied by examples of sufficient and insufficient reasoning. Students compare their own work against the criteria before the final submission, consistent with the distinction between formative self-assessment and simply assigning oneself a mark [35]. The aim is an inspectable judgment about work quality rather than a claim that the rubric itself guarantees competence.

Peer review is used to expose the assumptions of a comparison, while the final memorandum remains an individual product. Evidence on peer assessment supports its educational potential but does not remove the need to design the activity and interpret its context [36]. Each reviewer identifies the calculation condition, comparator, numerical claim, and physical inference in a partner's draft. Comments must point to an equation or statement that can be checked. The author then explains which comment changed the analysis and which did not. This sequence creates an observable use of feedback rather than counting the number of comments received as evidence of learning.

ZrN and the PtN stiffness check can be withheld until the individual transfer task, provided their worked answers have not already been distributed. ZrN must be compared with diamond under the same condition: its fixed-cell energy exceeds the comparator by $52.8\,\mathrm{mJ\,m^{-2}}$, while its relaxed-cell energy is lower by $33.9\,\mathrm{mJ\,m^{-2}}$. For PtN, the decisive calculation is the negative difference between its two normal stiffness coefficients. If all answers are available beforehand, the instructor must select a different pair or rule and describe it honestly. Recalling a disclosed answer is not equivalent to interpreting an unfamiliar comparison.

The final memorandum states the modeled material, quantity, constraint, comparator, decisive calculation, and limit of the conclusion. It includes one justified request for further evidence and one explanation of why an apparently attractive alternative conclusion is unsupported. Assessors distinguish an arithmetic error from a failure to match conditions. They also distinguish the latter from a physically invalid recommendation based on otherwise correct arithmetic. This separation prevents graphical polish from overwhelming scientific judgment and permits feedback to address the part of the reasoning that actually needs improvement.

Table 4: Assessment weights and evidence for full credit.

Product and criterion	Points	Evidence
Notebook: calculation identity	12	Correct quantity, condition, configuration, reference, and citation.
Notebook: dimensions	10	Correct energy-per-area conversion, units, and sign.
Notebook: paired entries	13	Eighteen complete pairs; seven absent values and two branch exclusions explained; counting units distinguished.
Comparator analysis: arithmetic	15	Correct differences, reversal pairs, and condition-matched membership.
Comparator analysis: sensitivity	10	Exact intervals and a clear distinction between rounding and physical uncertainty.
Memorandum: bounded claim	15	A conclusion that names its modeled object, conditions, and permitted inference.
Memorandum: physical reasoning	15	Appropriate treatment of sign, unresolved HfN evidence, and the PtN stability condition.
Memorandum: inspectable argument	10	Sufficient intermediate calculations and explanation for another reader to check the decision.

5. DISCUSSION

5.1. What the course adds to computational materials teaching

The numerical findings support a specific curricular argument. Agreement across a collection, stability of a local ordering, membership under a chosen rule, and physical admissibility are different achievements. The eighteen comparisons make those differences visible without changing the central scientific topic during the course. A learner cannot dismiss every issue as an unfamiliar technique encountered in a separate exercise. The same energies must be revisited with a more precise question. This continuity lets the instructor evaluate the quality of the learner's reasoning as the information and the required inference change, rather than merely increasing the computational complexity of successive tasks.

The approach complements workshop instruction in algorithms and first-principles methods by concentrating on the interpretation that follows their use. A well-executed calculation can still be compared under an inappropriate constraint or communicated with an excessive physical claim. Modern materials repositories make it easier to obtain numerical results, but they also expose users to heterogeneous definitions and calculation histories [37]. Teaching a condition-matched comparison therefore addresses a practical responsibility of computational work. It is not an argument against automation or broad numerical access. It specifies a form of judgment that becomes more necessary as access to quantities increases.

The distinction between numerical verification and physical inference is also educationally valuable. Students can reproduce the arithmetic in this article while remaining unable to diagnose the origin of HfN's positive condition difference. That is an appropriate outcome when essential component quantities are missing. It would be inappropriate to reward a confident microscopic explanation unsupported by those quantities. A course should make clear that an accurately bounded conclusion is scientifically stronger than an elaborate explanation whose premises cannot be checked. This standard connects scientific communication directly to the extent of the available evidence, rather than to the amount of technical terminology in an answer.

5.2. Why local decisions require more than a coefficient

The contrast between the full rank coefficient and the eight reversals is not a failure of correlation analysis. The coefficient summarizes the ordering relation it is designed to summarize. The error would be asking it to certify invariance of every local choice. The common-positive restriction makes that distinction harder to overlook, because five reversals remain after the negative-energy cases are removed. Similarly, the deletion analysis shows persistence after any single material is excluded. These results support a teaching exercise in matching an analytical quantity to a question, rather than an exercise in finding a single statistic that can be described as universally decisive.

Magnitude and ordering provide a further distinction. ReC dominates the absolute-change total while participating in no ordering reversal. HfN, VN, and ZrN influence the reversal count more directly. This difference gives students a concrete reason to examine multiple representations without treating them as interchangeable decorations. The reversal matrix answers where comparisons change; the normalized plane answers how values relate to a stated comparator; the interval display answers when labels differ; and the concentration display answers which differences dominate an aggregate magnitude. Each figure has a separate analytical purpose and is assessed through a corresponding written explanation.

The threshold analysis also limits the temptation to select a convenient numerical rule after inspecting the results. The multiplier and positivity requirement are declared before the primary membership comparison. Varying the multiplier then examines the dependence of that rule on its own definition. The narrow four-disagreement interval is especially useful: it prevents a visually prominent maximum from becoming a broad scientific claim. The numerical rule remains instructional throughout. Its role is to support reasoning about conditional comparison, not to establish that diamond-relative fault energies alone determine the processing or performance of another compound.

5.3. Feedback, judgment, and disciplinary communication

The course requires feedback to operate through a change in the learner's work. Effective feedback depends on the design of the surrounding activities and the conditions in which students can act on it [38]. The preliminary comparison, intermediate analysis, and final memorandum are consequently connected products. Their relationship lets the instructor ask whether a learner has changed a condition, corrected an equation, narrowed a claim, or identified an unresolved premise. Such changes are more informative than the existence of an annotated script. They also make the student's decision process available for discussion without assuming that every initial error has the same educational significance.

Teachers need capabilities for eliciting, interpreting, and supporting these acts of comparison [39]. Students in turn need to understand the purpose of feedback, seek relevant information, and judge which comments affect the quality of their work [40]. The course assigns these responsibilities explicitly: the teacher selects contrasting cases and clarifies the criteria; the student exposes an initial judgment, tests it against the later evidence, and explains the resulting decision. Research on the interplay between teacher and student feedback literacy supports attention to both sides of that relationship [41]. It does not establish that the present allocation of responsibilities has already succeeded in a classroom.

Evaluative judgment extends beyond recognizing a correct numerical answer. It includes deciding whether a claim is adequate for its purpose and whether the supporting evidence is sufficient [42]. The final memorandum operationalizes this expectation through a bounded scientific argument. A recommendation must name its comparison conditions, preserve the sign and units, and distinguish a numerical threshold from a physical requirement. An evidence request must specify a missing quantity that could change the conclusion. These requirements make disciplinary communication an assessable component of computational work instead of a generic writing task added after the calculations are complete.

The marking process should preserve the distinction between types of error. Two assessors can first score a small common set of memoranda, compare their reasons, and clarify the criteria before marking the remaining work. They should record disagreements about whether a conclusion exceeds its evidence, rather than resolving every difference through an unexplained average. A learner who correctly rejects a physical inference but makes a minor subtraction error should remain distinguishable from one who calculates accurately and endorses an unstable phase. This procedure specifies how the assessment could be implemented. It does not supply an agreement coefficient or establish that the present criteria already produce reliable judgments across instructors.

The notebook is similarly a scientific communication product. Guidance on computational notebooks emphasizes the connection between readable explanation, executable steps, and inspectable numerical outputs [43]. In this course, the corresponding requirement applies equally to code and spreadsheets. A reviewer must be able to locate the inputs to the decisive comparison and understand why the stated rule was used. An instructor should not reward a

hidden calculation merely because the final number matches an answer key. Nor should every intermediate software action be documented at the expense of explaining the physical comparison. The record should be sufficient for the specific claim it supports.

5.4. Educational evaluation and limits of inference

No learning-effect estimate can be drawn from the material calculations. The course has not supplied participant observations, test scores, or implementation records, and its workload and marking weights remain untested design choices. A future evaluation should distinguish the ability to repeat the worked examples from the ability to interpret a new property under unfamiliar constraints. An equally timed comparison course could teach the same numerical content with direct instruction preceding all problem solving. A delayed individual task could then assess condition matching, treatment of missing values, physical admissibility, and evidence requests. Its sampling and analysis plan would have to precede educational data collection.

A transfer task should also separate missing knowledge from inaccessible presentation. The same unfamiliar comparison can be supplied with readable units, an explicit energy definition, and a clear statement of which condition is absent. Learners would then have to identify the necessary comparison and justify an evidence request, rather than infer an instructor's preferred interpretation from ambiguous wording. Using parallel tasks with different materials could reduce dependence on memorized answers, provided their scientific demands are checked for equivalence. Such preparation is necessary before attributing differences in student responses to the teaching sequence. It avoids treating a poorly specified question as a meaningful test of sophisticated materials reasoning.

Satisfaction should be examined separately from achievement. Active engagement can alter learners' perceptions of learning in ways that do not track measured performance directly [44]. Broader evidence on active learning also emphasizes educational equity, making access and participation relevant design considerations rather than incidental administrative matters [45]. The spreadsheet alternative, individual preliminary attempt, and explicit discussion of criteria help define how such considerations could be addressed here. They do not demonstrate equitable outcomes. A classroom evaluation would need to examine prior programming experience, materials knowledge, language demands, and opportunities to participate, rather than assuming that the course serves all learners equally.

Physical generalization is limited as well. The paired values concern particular cubic configurations, a specific fault arrangement, and zero-temperature calculations. The seven absent relaxed-cell values and two unresolved magnetic-branch entries narrow the eligible comparison. The eighteen materials are not a random sample of all compounds, and related chemistries share a computational procedure. Temperature dependence, alternative phases, detailed magnetic order, and kinetic accessibility require additional evidence. These limitations are part of the scientific content of the course. Removing them would make the narrative smoother at the cost of teaching an inference that the available calculations cannot support.

6. CONCLUSION

A specialized computational-materials course can make the conditions of a numerical comparison an explicit learning and assessment target. The paired stacking-fault energies show why this target requires more than correct arithmetic: strong overall rank association coexists with local ordering reversals, sign changes, and condition-dependent membership under a declared rule. Exact intervals identify a narrow maximum-disagreement region, while individual-material deletions show that the reversal finding persists beyond any single exceptional entry. The PtN stiffness calculation adds a separate physical check by identifying a descending tetragonal strain direction.

The six-meeting design connects these results to unit analysis, recovery of calculation conditions, paired reasoning, evidence requests, and an individual technical memorandum. Its assessment distinguishes numerical correctness from the validity of the physical inference and the quality of its communication. The completed contribution is the mathematical examination and the specification of a coherent course. Classroom effectiveness, transfer to other material properties, and the reliability of the assessment require subsequent investigation. Those questions can

be tested without changing the central educational standard: a material comparison is adequate only when its quantity, conditions, comparator, and permitted conclusion are stated together.

Conflict of Interests

The author declares that they have no conflicts of interest.

Funding Information

This research received no external funding.

Data Availability

The data supporting the findings of this study are available from the author.

References

1. J. Choy, K. Peltz, Z. Sayers, N. A. Spaldin and M. A. Wells. Reimagining undergraduate teaching in materials science and engineering. *Nature Reviews Materials* **9** (2024), 95–99. https://doi.org/10.1038/s41578-023-00621-6.
2. M. S. Khan and M. A. Wells. Integrating interdisciplinary education in materials science and engineering. *Nature Reviews Materials* **8** (2023), 491–493. https://doi.org/10.1038/s41578-023-00576-8.
3. V. Caro, B. A. Carter, J. Millunchick and S. Reeves. Teaching crystal structures in undergraduate courses: a systematic review from a disciplinary literacy perspective. *Chemistry Education Research and Practice* **24** (2023), 394–406. https://doi.org/10.1039/D2RP00153E.
4. S. Sun, K. Brown and A. G. Kusne. Teaching machine learning to materials scientists: Lessons from hosting tutorials and competitions. *Matter* **5** (2022), 1620–1622. https://doi.org/10.1016/j.matt.2022.04.019.
5. A. Mannodi-Kanakkithodi, A. McDannald, S. Sun, S. Desai, K. A. Brown and A. G. Kusne. A framework for materials informatics education through workshops. *MRS Bulletin* **48** (2023), 560–569. https://doi.org/10.1557/s43577-023-00531-6.
6. M. Geshi. Computational Materials Design (CMD) workshop: A hands-on workshop of first-principles calculations online. *Journal of Materials Education* **44** (2022), 57–66. https://researchmap.jp/geshi/published_papers/40843394.
7. S. M. Reeves and K. J. Crippen. Virtual Laboratories in Undergraduate Science and Engineering Courses: a Systematic Review, 2009–2019. *Journal of Science Education and Technology* **30** (2021), 16–30. https://doi.org/10.1007/s10956-020-09866-0.
8. L. Himanen, A. Geurts, A. S. Foster and P. Rinke. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Advanced Science* **6** (2019), 1900808. https://doi.org/10.1002/advs.201900808.
9. S. P. Huber, S. Zoupanos, M. Uhrin et al.. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Scientific Data* **7** (2020), 300. https://doi.org/10.1038/s41597-020-00638-4.
10. L. Talirz, S. Kumbhar, E. Passaro et al.. Materials Cloud, a platform for open computational science. *Scientific Data* **7** (2020), 299. https://doi.org/10.1038/s41597-020-00637-5.
11. E. Bosoni, L. Beal, M. Bercx et al.. How to verify the precision of density-functional-theory implementations via reproducible and universal workflows. *Nature Reviews Physics* **6** (2024), 45–58. https://doi.org/10.1038/s42254-023-00655-3.
12. G. Prandini, A. Marrazzo, I. E. Castelli, N. Mounet and N. Marzari. Precision and efficiency in solid-state pseudopotential calculations. *npj Computational Materials* **4** (2018), 72. https://doi.org/10.1038/s41524-018-0127-2.
13. T. Sinha and M. Kapur. When Problem Solving Followed by Instruction Works: Evidence for Productive Failure. *Review of Educational Research* **91** (2021), 761–798. https://doi.org/10.3102/00346543211019105.

14. K. Loibl, I. Roll and N. Rummel. Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educational Psychology Review* **29** (2017), 693–715. https://doi.org/10.1007/s10648-016-9379-x.
15. T. Sinha, M. Kapur, R. West, M. Catasta, M. Hauswirth and D. Trninic. Differential benefits of explicit failure-driven and success-driven scaffolding in problem-solving prior to instruction. *Journal of Educational Psychology* **113** (2021), 530–555. https://doi.org/10.1037/edu0000483.
16. E. Starrett, M. Jordan, Y.-C. Chen, J. Park and C. Meza-Torres. Desirable uncertainty in science teaching: Exploring teachers' perceptions and practice of using student scientific uncertainty as a pedagogical resource. *Teaching and Teacher Education* **140** (2024), 104456. https://doi.org/10.1016/j.tate.2023.104456.
17. R. Bärenthaler-Pachner and C. Nagel. Development and evaluation of a GUM-based curriculum on measurement uncertainty for upper secondary. *Progress in Science Education* **7** (2024), 56–77. https://doi.org/10.25321/prise.2024.1428.
18. B. Wisniewski, K. Zierer and J. Hattie. The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology* **10** (2020), 3087. https://doi.org/10.3389/fpsyg.2019.03087.
19. D. Carless and D. Boud. The development of student feedback literacy: enabling uptake of feedback. *Assessment and Evaluation in Higher Education* **43** (2018), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354.
20. D. Nicol. The power of internal feedback: exploiting natural comparison processes. *Assessment and Evaluation in Higher Education* **46** (2021), 756–778. https://doi.org/10.1080/02602938.2020.1823314.
21. E. Panadero, A. Jonsson, L. Pinedo and B. Fernández-Castilla. Effects of Rubrics on Academic Performance, Self-Regulated Learning, and Self-Efficacy: a Meta-analytic Review. *Educational Psychology Review* **35** (2023), 113. https://doi.org/10.1007/s10648-023-09823-4.
22. A. Jonsson, E. Panadero, L. Pinedo and B. Fernández-Castilla. Using rubrics for formative purposes: identifying factors that may affect the success of rubric implementations. *Assessment in Education: Principles, Policy and Practice* **32** (2025), 192–211. https://doi.org/10.1080/0969594X.2025.2486947.
23. Y. Zhang, Z.-R. Liu, D.-W. Yuan, Q. Shao, J.-H. Chen, C.-L. Wu and Z.-L. Zhang. Elastic properties and stacking fault energies of borides, carbides and nitrides from first-principles calculations. *Acta Metallurgica Sinica (English Letters)* **32** (2019), 1099–1110. https://doi.org/10.1007/s40195-019-00873-8.
24. P. Schober, C. Boer and L. A. Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia* **126** (2018), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864.
25. P. Pernot. The long road to calibrated prediction uncertainty in computational chemistry. *Journal of Chemical Physics* **156** (2022), 114109. https://doi.org/10.1063/5.0084302.
26. S. Singh, L. Lang, V. Dovale-Farelo, U. Herath, P. Tavadze, F.-X. Coudert and A. H. Romero. MechElastic: A Python library for analysis of mechanical and elastic properties of bulk and 2D materials. *Computer Physics Communications* **267** (2021), 108068. https://doi.org/10.1016/j.cpc.2021.108068.
27. S. Yalameha, Z. Nourbakhsh and D. Vashaee. ElATools: A tool for analyzing anisotropic elastic properties of the 2D and 3D materials. *Computer Physics Communications* **271** (2022), 108195. https://doi.org/10.1016/j.cpc.2021.108195.
28. S. Yalameha, Z. Nourbakhsh and D. Vashaee. Erratum: ElATools: A tool for analyzing anisotropic elastic properties of the 2D and 3D materials [Comput. Phys. Commun. 271 (2022) 108195]. *Computer Physics Communications* **288** (2023), 108728. https://doi.org/10.1016/j.cpc.2023.108728.
29. B. Kim and G. Henke. Easy-to-Use Cloud Computing for Teaching Data Science. *Journal of Statistics and Data Science Education* **29** (2021), S103–S111. https://doi.org/10.1080/10691898.2020.1860726.
30. J. Kerrigan, K. Weber and C. Chinn. Effective collaboration in the productive failure process. *Journal of Mathematical Behavior* **63** (2021), 100892. https://doi.org/10.1016/j.jmathb.2021.100892.
31. M. Kapur, J. Hattie, I. Grossman and T. Sinha. Fail, flip, fix, and feed – Rethinking flipped learning: A review of meta-analyses and a subsequent meta-analysis. *Frontiers in Education* **7** (2022), 956416. https://doi.org/10.3389/feduc.2022.956416.
32. A. B. H. de Bruin, J. Roelle, S. K. Carpenter, M. Baars and EFG-MRE. Synthesizing Cognitive Load and Self-regulation Theory: a Theoretical Framework and Research Agenda. *Educational Psychology Review* **32** (2020), 903–915. https://doi.org/10.1007/s10648-020-09576-4.

33. Z.-L. Liu, C. E. Ekuma, W.-Q. Li, J.-Q. Yang and X.-J. Li. ElasTool: An automated toolkit for elastic constants calculation. *Computer Physics Communications* **270** (2022), 108180. https://doi.org/10.1016/j.cpc.2021.108180.
34. Z. Ran, C. Zou, Z. Wei and H. Wang. VELAS: An open-source toolbox for visualization and analysis of elastic anisotropy. *Computer Physics Communications* **283** (2023), 108540. https://doi.org/10.1016/j.cpc.2022.108540.
35. H. L. Andrade. A Critical Review of Research on Student Self-Assessment. *Frontiers in Education* **4** (2019), 87. https://doi.org/10.3389/feduc.2019.00087.
36. K. S. Double, J. A. McGrane and T. N. Hopfenbeck. The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educational Psychology Review* **32** (2020), 481–509. https://doi.org/10.1007/s10648-019-09510-3.
37. C. Draxl and M. Scheffler. The NOMAD laboratory: from data sharing to artificial intelligence. *Journal of Physics: Materials* **2** (2019), 036001. https://doi.org/10.1088/2515-7639/ab13bb.
38. M. Henderson, M. Phillips, T. Ryan, D. Boud, P. Dawson, E. Molloy and P. Mahoney. Conditions that enable effective feedback. *Higher Education Research and Development* **38** (2019), 1401–1416. https://doi.org/10.1080/07294360.2019.1657807.
39. D. Boud and P. Dawson. What feedback literate teachers do: an empirically-derived competency framework. *Assessment and Evaluation in Higher Education* **48** (2023), 158–171. https://doi.org/10.1080/02602938.2021.1910928.
40. E. Molloy, D. Boud and M. Henderson. Developing a learning-centred framework for feedback literacy. *Assessment and Evaluation in Higher Education* **45** (2020), 527–540. https://doi.org/10.1080/02602938.2019.1667955.
41. D. Carless and N. Winstone. Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education* **28** (2023), 150–163. https://doi.org/10.1080/13562517.2020.1782372.
42. J. Tai, R. Ajjawi, D. Boud, P. Dawson and E. Panadero. Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education* **76** (2018), 467–481. https://doi.org/10.1007/s10734-017-0220-3.
43. A. Rule, A. Birmingham, C. Zuniga et al.. Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks. *PLoS Computational Biology* **15** (2019), e1007007. https://doi.org/10.1371/journal.pcbi.1007007.
44. L. Deslauriers, L. S. McCarty, K. Miller, K. Callaghan and G. Kestin. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences* **116** (2019), 19251–19257. https://doi.org/10.1073/pnas.1821936116.
45. E. J. Theobald, M. J. Hill, E. Tran et al.. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences* **117** (2020), 6476–6483. https://doi.org/10.1073/pnas.1916903117.