Toward library-free mineral identification from EDS spectra

Bilal Aslan[a,b,*], Geoff Nitschke[a], Jonathan Knapp[b], Chi Ly[c]

[a] *Department of Computer Science, University of Cape Town, Cape Town, South Africa*
[b] *Hitachi High-Tech America, Inc., Chicago, Illinois, USA*
[c] *Bruker Nano Analytics, Houston, Texas, USA*

Abstract

Identifying minerals from energy-dispersive X-ray spectroscopy (EDS) spectra is one of the hardest everyday problems in automated mineralogy. Mixed-phase pixels, fine intergrowths, varying solid-solution chemistry, light elements like oxygen and carbon, surface coatings, and minerals that aren't in the reference library all conspire to throw off conventional spectral matching, and every new ore body tends to mean building a new library from scratch. We are developing a single, comprehensive pipeline that identifies minerals directly from raw EDS spectra without that manual library work. Instead of memorizing a fixed catalogue, it learns from physics-based simulated spectra and reasons through the messy realities mineralogists deal with daily: it weighs competing explanations for light elements and coatings, handles mixtures and solid solutions, folds in geological context, and knows when to flag an unfamiliar phase rather than guess. The goal is dependable, library-free mineral identification that travels across ore bodies and slots into existing SEM-EDS workflows.

Keywords: automated mineralogy, SEM-EDS, mineral identification, X-ray spectrum simulation, machine learning, process mineralogy

*Corresponding author.

Email addresses: aslbil001@myuct.ac.za (Bilal Aslan), gnitschke@cs.uct.ac.za (Geoff Nitschke), jonathan.knapp@hitachi-hightech.com (Jonathan Knapp), Chi.Ly@bruker.com (Chi Ly)

1. Introduction

Automated mineralogy underpins modern process mineralogy: modal mineralogy, liberation and deportment, and the geometallurgical models built on them (Gottlieb et al., 2000; Fandrich et al., 2007; Becker et al., 2016; Hrstka et al., 2018; Schulz et al., 2020). At every measurement point the instrument records an energy-dispersive X-ray (EDS) spectrum, and the software names the mineral by comparing that spectrum with a list of reference spectra or classification rules.

That list is the weak link. An experienced mineralogist builds and edits it for each ore body, and it has to be checked against other techniques wherever minerals vary in composition (Kearney et al., 2025). The work is skilled, slow and repeated for every new deposit, and the result depends on who did it. Differences in how it is set up are one source of the spread seen when laboratories measure the same material (Lastra and Paktunc, 2016). A mineral missing from the list cannot be reported, however clearly its spectrum shows it. Machine-learning classifiers have been trained on labelled measured data (Hao et al., 2019), but they inherit the same dependence on measured, labelled examples of every mineral.

This paper describes work toward identification that does not depend on a measured library. By *library-free* we mean that no measured reference spectra are needed. The system knows minerals through their chemistry: a catalogue of 7,525 mineral entries (species and named varieties) with formulas. It also knows them through physics, which predicts what each mineral's spectrum should look like on an SEM. We outline the approach, report where it stands on four sets of measured spectra, and state plainly what it cannot yet do.

2. Identification from EDS: Challenges

An EDS spectrum records which elements are present and in what proportions; it does not record crystal structure. Polymorphs such as pyrite and marcasite, calcite and aragonite, or the SiO_2 minerals therefore cannot be told apart by EDS alone. This is common, not exceptional. Of the 6,198 minerals our system can report, 3,289 (53%) have the same measurable elements in near-identical proportions as at least one other mineral. A system that always prints a single species name is over-claiming for half of mineralogy.

Light elements add a second difficulty. Hydrogen and lithium give no signal at all, and beryllium and boron are marginal at best. Carbon and

oxygen give low-energy peaks that are strongly absorbed in the sample and the detector window, so their measured intensities are hard to trust, and carbon is also the usual conductive coating. Third, automated mineralogy measures many points quickly, so each pixel typically carries a few thousand counts, far fewer than a careful point analysis (Newbury and Ritchie, 2013). Fourth, the volume that emits X-rays is typically a micrometre or more across, so in fine intergrowths a single pixel records two or more phases at once and its spectrum matches no single mineral. Finally, real samples bring solid-solution chemistry that drifts across a deposit, minerals nobody expected, and instruments that differ in detector window, resolution and geometry (Goldstein et al., 2018).

3. Methods

The central idea is to replace the measured library with physics. We built our own Monte Carlo X-ray simulator, which runs on a single desktop graphics card (GPU). Like DTSA-II and PENEPMA (Ritchie, 2009; Llovet and Salvat, 2017), it follows electrons through a material of known composition and predicts the spectrum that material emits. From the chemistry catalogue it produces a reference spectrum for every candidate mineral without a single measurement (Figure 2). The same simulator produces the examples from which the system's learned components are trained. They handle four jobs: detecting elements, finding similar spectra, recognising a spectrum unlike anything seen in training, and deciding whether an answer should be given at all. The first three read spectra directly. They are neural networks, that is, pattern-recognition models that learn from examples, and they learn from simulated spectra only. The decision to answer or decline is made by a separate statistical model calibrated on measured cases.

Identification then follows the reasoning of an experienced analyst (Figure 1):

1. *Which elements are really present?* Detector artefacts, and the peak of one element mimicking another, are rejected.
2. *Which minerals could explain them?* Oxygen, carbon and hydrogen are never allowed to rule a candidate out, because their peaks are too unreliable to prove absence.
3. *How well does each candidate's simulated spectrum match the measured one?*

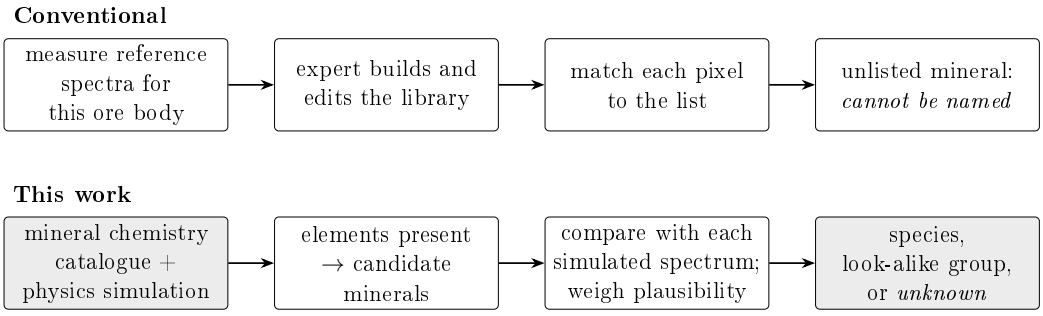

Figure 1: Conventional classification (Schulz et al., 2020; Kearney et al., 2025) depends on a measured reference library built for each ore body. The approach described here draws its reference knowledge from mineral chemistry and physics, and grades how specific its answer can be.

4. *How plausible is each candidate?* How common a mineral is, and optional geological context (deposit type, lithology, location), are weighed with probability rules. They break ties the spectrum cannot break, but never override clear spectral evidence.
5. *How specific can the answer be?* The answer has three levels: an exact species where the spectrum supports one; a look-alike group, such as "pyrite–marcasite", with its members listed in order of commonness, where EDS cannot separate them; or "unknown" when the evidence is weak or the spectrum is unlike anything the system knows (Figure 3).

Three further features address the difficulties of Section 2. The detector description stored in the spectrum file (window, resolution and take-off angle) is read and used in element detection, detection limits and quantification. Oxygen, carbon and hydrogen contents are inferred from the charge balance of the identified mineral rather than read from their weak peaks. Whether a carbon signal is better explained by a coating or by the mineral itself is tested and reported. On mapped samples, pixels that look like mixtures of phases identified elsewhere in the same sample are flagged. Solid solutions are treated as continuous series, and an estimated position along common series, such as the anorthite content of plagioclase, is reported. Spectra are read directly from Bruker `.spx` files, AMICS projects and the standard EMSA/MSA format, so existing measurements can be re-examined without re-acquisition.

Table 1: Identification on measured spectra. "Group": correct look-alike group ranked first (a decline counts as wrong). "Species": correct species ranked first. No geological context was supplied for any set.

Data set	Spectra	Median counts	Group	Species	Declined
(a) Certified mineral standards (Astimex) and pure-element mounts; Bruker EDS, 5–20 kV; 44 materials	112	625 k	75%	67%	0%
(b) Characterised mineral samples; Hitachi SEM with Bruker detector*	34	14.5 k	79%	65%	0%
(c) Held-out mineral reference spectra; Bruker EDS, 25 kV; 9 materials	45	19 M	42%	33%	0%
(d) AMICS particle-mapping pixels; agreement with the commercial classification	2,523	2.2 k	47%	30%	11%

*Recorded as 20 kV; for about half of these files the true beam energy is uncertain. Sets (a), (b) and (d) contributed to development; set (c) never did.

4. Results

Table 1 summarises four sets of measured spectra, all scored with the system's current settings in September 2026. The main measure is whether the correct look-alike group was ranked first; a declined answer counts as wrong. The group is the fair measure because no EDS method can do better than the group for polymorphs. We also report whether the exact species was ranked first, which shows how often the commonness ordering picks the right member.

Sets (a) and (b) were used to guide development, so their figures (75% and 79%) are optimistic; set (d) also helped calibrate the decision to decline. Set (c), 45 spectra of 9 mineral reference materials, was never used during development and was acquired at 25 kV, a beam energy the system was not tuned for. There the correct group came first for 42% of spectra, rising to 55% on the 31 spectra whose reference name is a single species rather than a family name such as "mica". The errors were concentrated: calcite was identified as fluorite, and magnetite–hematite as a rare iron nitride. In both, the deciding information sits in weak light-element peaks.

Set (d) comes from an AMICS particle map at a median of 2,222 counts per pixel, the regime process mineralogists actually work in. There the sys-

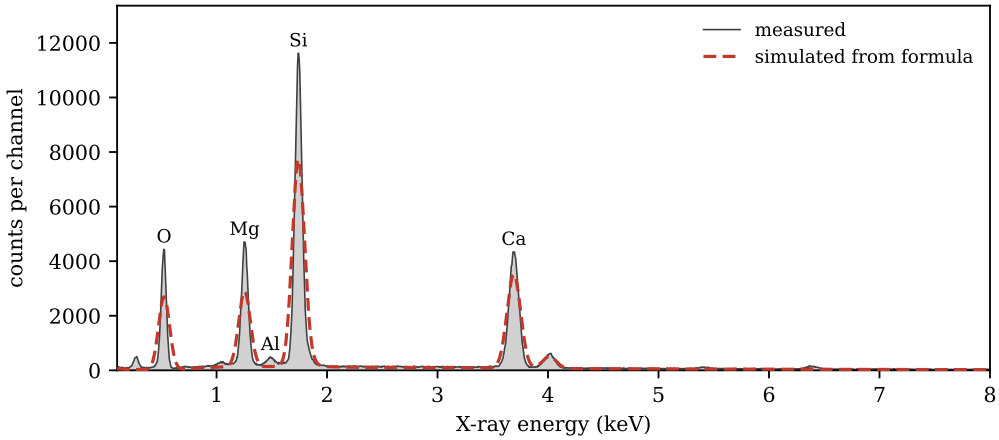

Figure 2: Measured spectrum of a certified diopside standard, $CaMgSi_2O_6$ (20 kV, Bruker detector; grey, filled), and the reference the system compares it with, simulated from the mineral's formula alone (red, dashed). No measured reference was used. Peak positions and relative heights agree. The simulated peaks are wider because references are prepared at a single typical detector resolution, broader than this detector's.

tem and the commercial classification agreed on the mineral group for 47% of pixels, and for 52% of those it answered. Those labels were not independently checked, so the disagreements are a mixture of our errors and uncertain labels. At this count level the system declined 11% of pixels rather than guess.

Figure 2 shows what the matching step works with: a measured diopside standard against its reference spectrum simulated from the formula alone. Figure 3 shows two answers as a user sees them. For solid solutions, the estimated anorthite content of a certified plagioclase was An61 against a certified An59. For a certified olivine, however, the estimate was Fo99.9 against a certified Fo92, so series positions are not yet validated.

5. Limitations and outlook

The results are an honest starting point rather than a finished product. The held-out and mapping figures, not the development-set figures, are the fair guide to what a new user should expect. Low-count mapping pixels are the frontier: they are what automated mineralogy produces, and they are where the system is weakest today. Light-element-driven distinctions, such

Certified pyrite standard, 20 kV	**Certified diopside standard, 20 kV**
Elements confirmed: Fe, S	Elements confirmed: Ca, Mg, Si, O
Answer level: **look-alike group**	Answer level: **exact species**
Answer: **pyrite–marcasite** (FeS_2). EDS cannot separate these two polymorphs; pyrite is listed first as the more common. Next distinct possibility: a group led by pyrrhotite and troilite (FeS).	Answer: **diopside**. Look-alikes with similar chemistry (monticellite, åkermanite, tremolite) were considered and set aside. Next distinct possibility: a hornblende group, far behind.

Figure 3: Two answers as a user sees them, taken from live output on certified standards. Left: the system reports a look-alike group because EDS physically cannot choose between its members. Right: the spectrum supports an exact name.

as carbonate against fluoride or oxide against nitride, remain the commonest error. Look-alike groups will always need another technique, such as XRD or Raman, to resolve; the system's contribution there is to say so rather than guess. Estimates of position along solid-solution series need validation before they are relied on. Finally, each spectrum is identified on its own. Apart from the mixture flag, neither the backscattered-electron image nor the answers for neighbouring pixels inform the identification yet, although an analyst would use both.

In the meantime, the system can already run alongside existing automated-mineralogy software as a second opinion, where disagreement with the library flags pixels worth an analyst's attention.

6. Conclusions

Mineral identification from EDS spectra can draw its reference knowledge from mineral chemistry and physics rather than from a measured library. The system described here does so, reports look-alike groups where EDS cannot separate species, and declines when the evidence is too weak. On certified standards it ranks the correct mineral group first for three spectra in four. On spectra it was never tuned for it does so for 42%, and on low-count mapping pixels it agrees with the commercial classification for 47%. Two developments target that gap: a new generation of simulated training spectra covering the full range of beam energies used in practice, mixed-phase pixels and carbon-coated surfaces, and validation on process samples (plant feed, concentrates and tailings) with partner laboratories.

References

Becker, M., Wightman, E.M., Evans, C.L. (Eds.), 2016. Process Mineralogy. Number 6 in JKMRC Monograph Series in Mining and Mineral Processing, Julius Kruttschnitt Mineral Research Centre, Brisbane.

Fandrich, R., Gu, Y., Burrows, D., Moeller, K., 2007. Modern SEM-based mineral liberation analysis. International Journal of Mineral Processing 84, 310–320. doi:10.1016/j.minpro.2006.07.018.

Goldstein, J.I., Newbury, D.E., Michael, J.R., Ritchie, N.W.M., Scott, J.H.J., Joy, D.C., 2018. Scanning Electron Microscopy and X-Ray Microanalysis. 4th ed., Springer, New York. doi:10.1007/978-1-4939-6676-9.

Gottlieb, P., Wilkie, G., Sutherland, D., Ho-Tun, E., Suthers, S., Perera, K., Jenkins, B., Spencer, S., Butcher, A., Rayner, J., 2000. Using quantitative electron microscopy for process mineralogy applications. JOM 52, 24–25. doi:10.1007/s11837-000-0126-9.

Hao, H., Guo, R., Gu, Q., Hu, X., 2019. Machine learning application to automatically classify heavy minerals in river sand by using SEM/EDS data. Minerals Engineering 143, 105899. doi:10.1016/j.mineng.2019.105899.

Hrstka, T., Gottlieb, P., Skála, R., Breiter, K., Motl, D., 2018. Automated mineralogy and petrology – applications of TESCAN Integrated Mineral Analyzer (TIMA). Journal of Geosciences 63, 47–63. doi:10.3190/jgeosci.250.

Kearney, L.I., Christy, A.G., Belousova, E.A., Hines, B.R., Kontonikas-Charos, A., de Bruyn, M., Cathey, H.E., Lisitsin, V., 2025. Automated, not autonomous: integrating automated mineralogy with complementary techniques to refine and validate phase libraries in complex mineral systems. Minerals 15, 1118. doi:10.3390/min15111118.

Lastra, R., Paktunc, D., 2016. An estimation of the variability in automated quantitative mineralogy measurements through inter-laboratory testing. Minerals Engineering 95, 138–145. doi:10.1016/j.mineng.2016.06.025.

Llovet, X., Salvat, F., 2017. PENEPMA: a Monte Carlo program for the simulation of X-ray emission in electron probe microanalysis. Microscopy and Microanalysis 23, 634–646. doi:10.1017/S1431927617000526.

Newbury, D.E., Ritchie, N.W.M., 2013. Is scanning electron microscopy/energy dispersive X-ray spectrometry (SEM/EDS) quantitative? Scanning 35, 141–168. doi:10.1002/sca.21041.

Ritchie, N.W.M., 2009. Spectrum simulation in DTSA-II. Microscopy and Microanalysis 15, 454–468. doi:10.1017/S1431927609990407.

Schulz, B., Sandmann, D., Gilbricht, S., 2020. SEM-based automated mineralogy and its application in geo- and material sciences. Minerals 10, 1004. doi:10.3390/min10111004.