SEEING THROUGH SPECTRAL OVERLAP: RECOVERING MIXTURE COMPOSITION AS A MEASURE ON A LEARNED INFRARED SPECTRAL MANIFOLD

Anonymous authors
Paper under double-blind review

ABSTRACT

Infrared mixture analysis seeks to recover component spectra and their abundances from a single observation, but spectral overlap allows different compositions to produce similar signals, making reconstruction alone insufficient for identification. We introduce SpecMUSE, which represents mixture composition as a finitely supported probability measure on a learned spectral manifold. Its support locations specify component spectra, while its masses encode their abundances. Masked reconstruction and coordinate alignment establish the manifold from pure spectra. A set estimator jointly predicts component coordinates, presence probabilities, and abundances from the observed mixture. A differentiable realization field and discrete projection connect coordinates to individual component spectra, while a straight-through surrogate carries spectral-loss gradients to the continuous selection queries. This formulation couples component recovery with spectral reconstruction while accommodating unknown component counts at inference. Experiments on mixtures constructed from simulated QM9S spectra and measured NIST spectra demonstrate improved recovery across mixture complexities and generalization to unseen mixtures. Anonymous code is available at `https://anonymous.4open.science/r/SpecMUSE-9F68`.

1 INTRODUCTION

Infrared (IR) spectroscopy is a foundational technique for chemical characterization because molecular vibrations produce absorption bands that encode functional groups, bonding environments, and molecular structure (Colthup et al., 2012; Griffiths, 1983). Fourier-transform infrared (FTIR) spectroscopy enables rapid spectral acquisition for chemical identification and reaction monitoring, with applications in environmental sensing, pharmaceutical analysis, and industrial process control (Ashikari et al., 2025). Many practical samples, however, contain multiple compounds whose absorption bands overlap. Interpreting these observations requires moving beyond single-molecule recognition to identify the contributing components and estimate their abundances, a central challenge in automated IR analysis (Melle et al., 2026; Tan et al., 2025). Appendix C provides background on infrared absorption, gas-phase spectra, and library identification.

Under an additive absorbance model, mixture recovery must resolve several coupled unknowns (Swinehart, 1962). Similar component spectra can compensate for one another through changes in their weights, and weak components can be missed while dominant bands remain well reconstructed. A small spectral residual therefore need not identify the correct composition. The number of components is also unknown, and their order has no physical meaning.

Chemometric methods approach this problem through nonnegative fitting over reference spectra and multivariate curve resolution (Zhang et al., 2014; de Juan & Tauler, 2021). Neural models learn component detection or structure prediction from mixtures (Tan et al., 2025; Alberts et al., 2025), while generative unmixing models component participation and count (Chang & Chen, 2025). In parallel, large pure-spectrum datasets and masked spectral modeling provide representations of molecular band patterns (Alberts et al., 2024b; Wang et al., 2025). The challenge is to carry this single-component prior into mixture recovery, where overlapping spectral signatures require component identities and abundances to be inferred jointly from a single observation.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We introduce SpecMUSE (Spectral Manifold-based Unordered Set Estimation) to estimate mixture composition from an observed spectrum, as illustrated in Figure 1. We represent the unordered composition as a finitely supported probability measure on a learned spectral manifold, with support locations specifying component spectra and masses encoding their abundances. This representation separates spectral identity from mixture-specific weight and is invariant to component ordering. Pure-spectrum learning combines masked reconstruction, coordinate alignment, and spectral realization, yielding component coordinates and a shared field that maps these coordinates to individual spectral patterns. A set estimator then predicts component coordinates, presence probabilities, and abundance scores from the mixture. Component supervision constrains each predicted spectrum, while additive reconstruction evaluates how the recovered components explain the observation together. Hard selection chooses reference coordinates and realizes one spectrum per component; a soft gradient surrogate carries spectral-loss feedback to the continuous queries used for selection. At inference, presence predictions determine the active component count.

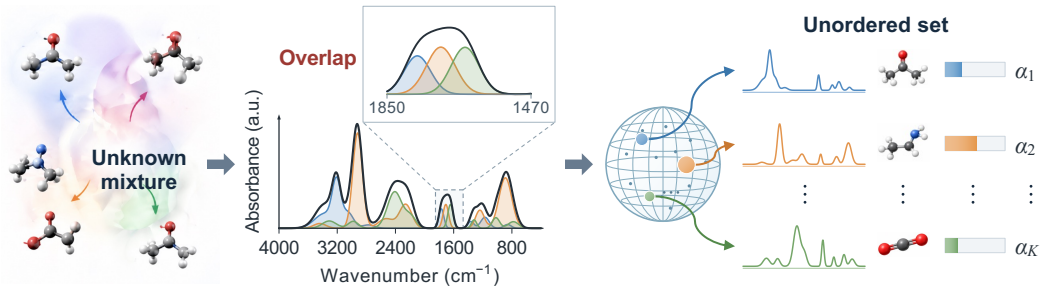

Figure 1: SpecMUSE recovers an unordered composition from overlapping IR bands. On the learned spectral manifold, measure supports specify component spectra and masses α_k encode abundances. Molecular structures illustrate recovered component identities.

In summary, our contributions are as follows:

- We formulate infrared mixture recovery as support and mass estimation on a learned spectral manifold, coupling an unordered component set to its additive spectrum.
- We develop SpecMUSE with a pure-spectrum prior and unordered component supervision, complemented by joint mixture reconstruction. Hard selection realizes individual components, while a soft surrogate trains continuous queries from spectral feedback.
- Experiments on mixtures constructed from simulated and experimentally measured spectra show consistent improvements over strong baselines and effective generalization to unseen mixtures.

2 Related Work

Spectral mixture analysis. Reference fitting combines spectral matching with nonnegative estimation (Zhang et al., 2014; Zhao et al., 2022), while multivariate curve resolution jointly estimates spectra and concentrations (de Juan & Tauler, 2021). BIPU jointly models abundance, presence, and latent endmember variability through Bayesian inference (Wu et al., 2024). Online unmixing updates spectra sequentially (Kouakou et al., 2024); IR libraries support reaction monitoring and liquid-mixture identification (Ashikari et al., 2025; Melle et al., 2026). Regression estimates mixture compositions from FTIR spectra for process optimization (Angulo et al., 2022).

Neural models estimate gas concentrations (Prischepa et al., 2023; Tian et al., 2023) or detect Raman components (Fan et al., 2019; Zhao et al., 2023; Wang et al., 2024). RamanFormer quantifies Raman components with self-attention (Koyun et al., 2024). DeepRaman and SpecRecFormer match Raman mixtures to references (Fan et al., 2023; Yu et al., 2025), while DeepMIR performs reference-conditioned IR identification (Tan et al., 2025). Transformers generate structures from binary IR mixtures (Alberts et al., 2025); EB-gMCR gates a learned component pool (Chang & Chen, 2025). For multiphase X-ray diffraction, XDecomposer combines query-based decomposition and

reconstruction (Gao et al., 2026). PhaseDifformer uses conditional diffusion to recursively remove phases from powder patterns, recovering each by subtraction (Ito et al., 2026).

Spectral representations and learned priors. Simulated corpora support molecular spectrum modeling (Zou et al., 2023; Alberts et al., 2024b). Transformers predict structures from IR spectra and molecular formulas (Alberts et al., 2024a). SMEN uses contrastive spectrum–molecule alignment for library ranking and structure generation (Kanakala et al., 2024). ACT combines learnable processing with attention within and across IR spectra for calibration transfer (Huang et al., 2025). MolSpectra combines masked spectral reconstruction and contrastive alignment for 3D molecular pretraining (Wang et al., 2025). RamanMAE reconstructs masked Raman patches to learn transferable representations (Paidi & Maheshwari, 2025). DSCF learns component filtering for preprocessing, quantification, and interpretation (Xue et al., 2025). For inverse recovery, generative decoders constrain signals to their output range (Bora et al., 2017), while diffusion priors support Bayesian inference and audio source separation (Janati et al., 2025).

Measure-valued inverse problems and set prediction. Measure recovery jointly estimates support locations and masses (Bredies & Pikkarainen, 2013; Candès & Fernandez-Granda, 2014). Sparse super-resolution analyzes separation and non-degeneracy conditions (Duval & Peyré, 2015; Denoyelle et al., 2017). Conditional-gradient and particle methods optimize continuous supports and weights (Boyd et al., 2017; Chizat, 2022). Measure optimization fits individual observations; set prediction learns shared inference. Deep Sets and Set Transformer encode sets (Zaheer et al., 2017; Lee et al., 2019). Probabilistic set prediction models elements, cardinality, and permutation (Rezatofighi et al., 2022). Query decoders and permutation-invariant matching support unordered outputs (Carion et al., 2020; Yu et al., 2017). Vector quantization and straight-through estimators connect discrete choices to gradients (van den Oord et al., 2017; Bengio et al., 2013; Liu et al., 2023). SpecMUSE combines set matching, hard anchor realization, and additive reconstruction with surrogate gradients for continuous queries.

3 PRELIMINARIES

Let $\Omega = \{\omega_\ell\}_{\ell=1}^{L}$ be the shared wavenumber grid and $\mathbf{s} \in \mathcal{X} \subseteq [0,1]^L$ a peak-normalized pure spectrum. For dilute, non-reactive gas mixtures, the Beer–Lambert law motivates additive absorbance (Swinehart, 1962; Mayerhöfer et al., 2020):

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{s}_k + \boldsymbol{\epsilon}, \qquad \sum_{k=1}^{K} \alpha_k = 1. \tag{1}$$

Here $\alpha_k > 0$ denotes component abundance, its normalized weight in the spectral sum; $\boldsymbol{\epsilon}$ captures deviations from this additive absorbance model.

Let $\mathcal{M}_\psi \subseteq \mathbb{S}^{d_z - 1}$ be the learned component coordinate space, with spherical distance $d_{\mathcal{M}}(\mathbf{a}, \mathbf{b}) = \arccos(\mathbf{a}^\top \mathbf{b})$. The realization field R_ψ maps each coordinate to a predicted pure spectrum. Composition and its additive realization are

$$\mu = \sum_{k=1}^{K} \alpha_k \delta_{\mathbf{z}_k}, \qquad \mathcal{A}_\psi(\mu) = \sum_{k=1}^{K} \alpha_k R_\psi(\mathbf{z}_k). \tag{2}$$

Each support $\mathbf{z}_k \in \mathcal{M}_\psi$ specifies a component spectrum, and its associated mass encodes that component's abundance; the measure is invariant to component ordering. Given $\mathbf{y}$, we estimate $\widehat{\mu}_{\mathbf{y}}$ with at most $K_{\max}$ supports. Learning R_ψ from pure spectra constrains individual components before mixing.

4 METHODOLOGY

Figure 2 summarizes the two stages. Stage 1 learns component coordinates and a shared realization field from pure spectra. Stage 2 trains a set estimator whose slots each realize an individual spectral signature. Component supervision identifies the supports, while mixture reconstruction evaluates whether their combined abundances explain the overlapping bands.

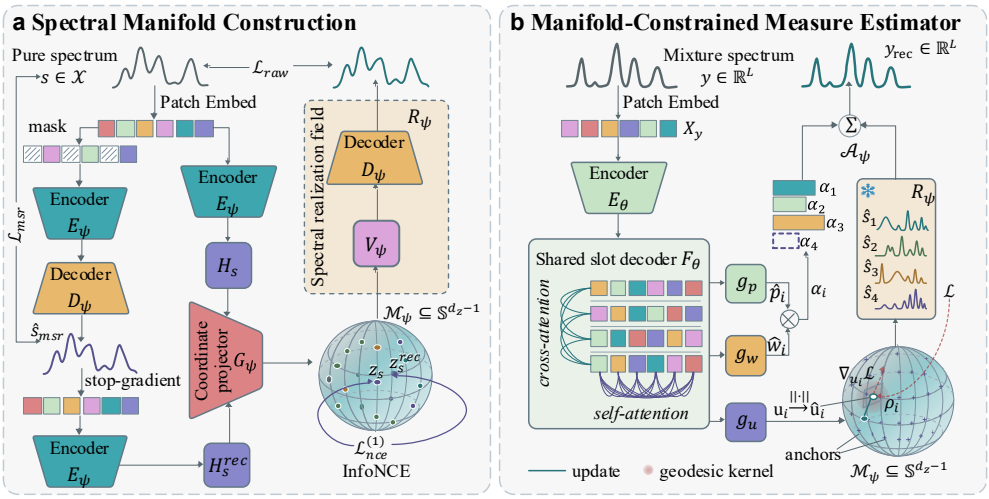

Figure 2: SpecMUSE framework. **(a)** Pure-spectrum pretraining learns component coordinates and the realization field R_ψ. **(b)** Mixture slots predict coordinates, presence, and abundances. Hard projection selects one realized spectrum per slot; a soft kernel supplies query gradients. The additive operator $\mathcal{A}_\psi$ reconstructs the mixture.

4.1 Spectral Manifold Construction

Masked reconstruction. Following spectral patch modeling (Wang et al., 2025), we embed N overlapping patches of width q as $\mathbf{X}_s \in \mathbb{R}^{N \times d_e}$. We mask wavenumber intervals and exclude every intersecting patch, preventing leakage through overlapping patches.

An encoder E_ψ processes the visible tokens $\mathbf{X}_s^{\text{vis}}$, a restoration module U_ψ reinserts mask tokens, and a decoder D_ψ predicts nonnegative patches, following the asymmetric masked autoencoder design (He et al., 2022). Uniform overlap-add $\mathcal{U}_{\text{ola}}$ averages patch predictions at each wavenumber:

$$\hat{\mathbf{s}}^{\text{msr}} = \mathcal{U}_{\text{ola}}\big(D_\psi\big(U_\psi\big(E_\psi(\mathbf{X}_s^{\text{vis}})\big)\big)\big)\,. \tag{3}$$

The reconstruction loss $\mathcal{L}_{\text{msr}} = d_{\text{sp}}(\hat{\mathbf{s}}^{\text{msr}}, \mathbf{s})$ combines Huber absorbance error, spectral angle, and a first-order Savitzky–Golay derivative term (Huber, 1964; Kruse et al., 1993; Savitzky & Golay, 1964). These terms compare amplitudes, band patterns, and local spectral changes. Masked reconstruction encourages the encoder to use context across spectral regions. Appendix A.1 defines the patching and loss operators.

Coordinate alignment. Reconstruction alone does not specify how spectra should be organized in coordinate space. We therefore encode the complete spectrum and its reconstructed view with shared patch embedding, encoder, and projector $G_\psi = g_\psi \circ \text{Pool}_\psi$. Normalizing $G_\psi(E_\psi(\mathbf{X}_s))$ to unit length gives $\mathbf{z}_s$; applying the same mapping to the detached reconstructed view gives $\mathbf{z}_s^{\text{rec}}$. InfoNCE aligns these coordinates using in-batch negatives (Oord et al., 2018):

$$\mathcal{L}_{\text{nce}}^{(1)} = -\frac{1}{|\mathcal{B}|}\sum_{s\in\mathcal{B}} \log \frac{\exp((\mathbf{z}_s^{\text{rec}})^\top \mathbf{z}_s/\tau_c)}{\sum_{s'\in\mathcal{B}} \exp((\mathbf{z}_s^{\text{rec}})^\top \mathbf{z}_{s'}/\tau_c)}. \tag{4}$$

Here τ_c is the contrastive temperature. Paired views tie each reconstructed spectrum to its source, while batch negatives separate pure-spectrum examples for later matching and realization. Detaching the reconstruction lets both views train the encoder and projector without passing contrastive gradients through the reconstruction path.

Spectral realization. Masked reconstruction and spectral realization share D_ψ but use different inputs. The restoration module supplies visible patch features and mask tokens; the lifting block supplies tokens derived from one component coordinate. Sharing the decoder links the spectral structure

learned from patch context to the component representation. Following the spatial broadcast principle (Watters et al., 2019; Locatello et al., 2020), each patch receives the same projected component coordinate. Learned positional embeddings distinguish the tokens along the wavenumber grid:

$$\mathbf{H}_z^0 = \mathbf{P}_{\text{lift}} + \mathbf{1}_N(\mathbf{W}_z\mathbf{z} + \mathbf{b}_z)^\top. \tag{5}$$

Here $\mathbf{P}_{\text{lift}}$ contains the learned positional embeddings for the spectral patches. A lifting block V_ψ prepares these tokens for the shared decoder D_ψ. Overlap-add combines the predicted patches as $\mathbf{s}_z^{\text{raw}} = \mathcal{U}_{\text{ola}}\big(D_\psi(V_\psi(\mathbf{H}_z^0))\big)$. Peak normalization yields $R_\psi(\mathbf{z}) = \mathbf{s}_z^{\text{raw}}/\|\mathbf{s}_z^{\text{raw}}\|_\infty$ for nonzero outputs. We supervise realization with $\mathcal{L}_{\text{raw}} = d_{\text{sp}}(R_\psi(\mathbf{z}_s), \mathbf{s})$. Stage 1 minimizes the combined objective $\mathcal{L}_{\text{S1}} = \mathcal{L}_{\text{msr}} + \lambda_{\text{align}}\mathcal{L}_{\text{nce}}^{(1)} + \lambda_{\text{raw}}\mathcal{L}_{\text{raw}}$. Stage 2 uses field evaluations at pure-spectrum anchors as its admissible component realizations.

4.2 Manifold-Constrained Measure Estimation

Mixture encoding and slots. A mixture encoder E_θ maps spectral patches $\mathbf{X}_y$ to contextual tokens $\mathbf{C}_y = E_\theta(\mathbf{X}_y)$. A slot decoder F_θ combines cross-attention to $\mathbf{C}_y$ with self-attention among $K_{\max}$ learned queries $\mathbf{Q}$, producing component vectors $\mathbf{h}_i = [F_\theta(\mathbf{Q}, \mathbf{C}_y)]_i$ (Vaswani et al., 2017; Carion et al., 2020). Cross-attention gathers evidence across separated wavenumber regions for each component, while self-attention lets queries exchange information before prediction.

Shared heads produce a query $\mathbf{u}_i = g_u(\mathbf{h}_i)$, a sigmoid presence probability $\widehat{p}_i$, and a positive softplus abundance score $\widehat{w}_i$. We normalize the coordinate queries as

$$\widehat{\mathbf{u}}_i = \frac{\mathbf{u}_i}{\max(\|\mathbf{u}_i\|_2, \varepsilon_u)}. \tag{6}$$

The query has unit norm when $\|\mathbf{u}_i\|_2 \geq \varepsilon_u$; below this threshold, it lies inside the unit ball.

Anchor realization and surrogate gradients. The pretrained encoder maps the reference pool to M anchors, stacked as rows of $\mathbf{Z} \in \mathbb{R}^{M\times d_z}$; their realized spectra form $\mathbf{S}_\psi \in \mathbb{R}^{M\times L}$. Section 5.1 specifies the reference pools. To select individual components while training coordinate queries, we combine a soft neighborhood kernel with a hard selector:

$$\boldsymbol{\rho}_i = \text{softmax}(\mathbf{Z}\widehat{\mathbf{u}}_i/\tau_z), \qquad \boldsymbol{\rho}_i^{\text{h}} = \text{onehot}(\arg\max_m \rho_{im}). \tag{7}$$

For unit-normalized queries, inner products decrease with geodesic distance; τ_z controls the concentration around nearby anchors. The straight-through selector (Bengio et al., 2013; Liu et al., 2023) combines soft and hard weights during training:

$$\widetilde{\boldsymbol{\rho}}_i = \boldsymbol{\rho}_i + \text{sg}(\boldsymbol{\rho}_i^{\text{h}} - \boldsymbol{\rho}_i). \tag{8}$$

Here sg stops gradients. The selector returns $\widehat{\mathbf{z}}_i = \mathbf{Z}^\top\widetilde{\boldsymbol{\rho}}_i$ and $\widehat{\mathbf{s}}_i = \mathbf{S}_\psi^\top\widetilde{\boldsymbol{\rho}}_i$. The continuous query $\widehat{\mathbf{u}}_i$ guides selection, while the forward support $\widehat{\mathbf{z}}_i$ is a reference coordinate with spectrum $R_\psi(\widehat{\mathbf{z}}_i)$. Backpropagation carries spectral feedback through the smooth weighted map $\mathbf{S}_\psi^\top\boldsymbol{\rho}_i$. Let $\mathbf{J}_i = \text{Diag}(\boldsymbol{\rho}_i) - \boldsymbol{\rho}_i\boldsymbol{\rho}_i^\top$. For $\|\mathbf{u}_i\|_2 > \varepsilon_u$, the spectral-loss path has surrogate gradient

$$\widetilde{\nabla}_{\mathbf{u}_i}\mathcal{L}_{\text{spectral}} = \frac{\mathbf{I} - \widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i^\top}{\tau_z\|\mathbf{u}_i\|_2}\mathbf{Z}^\top\mathbf{J}_i\mathbf{S}_\psi\nabla_{\widehat{\mathbf{s}}_i}\mathcal{L}_{\text{spectral}}. \tag{9}$$

In this branch, the leading projector keeps the update tangent to the coordinate sphere. The remaining factors translate spectral discrepancy into changes in anchor weights and then into a coordinate direction. Appendix A.2 gives the Jacobian for the clipped branch.

Composition and reconstruction. During training, presence probabilities gate the positive abundance scores, linking each component's presence to its weight in the mixture. Normalizing across all slots yields the predicted composition:

$$\widetilde{\alpha}_i = \frac{\widehat{p}_i\widehat{w}_i}{\sum_j \widehat{p}_j\widehat{w}_j}, \qquad \widetilde{\mu}_{\mathbf{y}} = \sum_i \widetilde{\alpha}_i\delta_{\widehat{\mathbf{z}}_i}. \tag{10}$$

The masses are nonnegative and sum to one. Their shared denominator couples the slots, so changing one unnormalized weight also changes the relative masses of the others. The additive operator gives $\mathbf{y}_{\text{rec}} = \mathcal{A}_\psi(\widetilde{\mu}_{\mathbf{y}}) = \sum_i \widetilde{\alpha}_i\widehat{\mathbf{s}}_i$. Each spectrum retains its band pattern, scaled by its mass. Comparing the reconstruction with the observation evaluates component spectra and abundances jointly.

4.3 Training and Inference

Set supervision. We train the mixture estimator θ using the learned spectral representation and annotated components $\{(\mathbf{s}_k, \mathbf{z}_k, \alpha_k)\}_{k=1}^{K}$. Permutation-invariant matching aligns components to unordered slots (Yu et al., 2017):

$$\pi^\star = \arg\min_{\pi \in \Pi_{K,K_{\max}}} \frac{1}{K} \sum_{k=1}^{K} d_{\mathrm{sp}}(\widehat{\mathbf{s}}_{\pi(k)}, \mathbf{s}_k). \tag{11}$$

Here $\Pi_{K,K_{\max}}$ contains injective assignments, and the minimum mean cost defines $\mathcal{L}_{\mathrm{comp}}$. Matched slots receive coordinate and abundance targets. Presence targets are positive for matched slots and negative for unmatched slots. The slot loss is $\mathcal{L}_{\mathrm{slot}} = \lambda_z \mathcal{L}_{\mathrm{nce}}^{(2)} + \lambda_p \mathcal{L}_{\mathrm{prs}} + \lambda_\alpha \mathcal{L}_\alpha$, combining coordinate InfoNCE, presence binary cross-entropy, and abundance MAE. The joint training objective is

$$\mathcal{L}_{\mathrm{S2}} = \lambda_{\mathrm{comp}} \mathcal{L}_{\mathrm{comp}} + \lambda_{\mathrm{slot}} \mathcal{L}_{\mathrm{slot}} + \lambda_{\mathrm{mix}} d_{\mathrm{sp}}(\mathbf{y}_{\mathrm{rec}}, \mathbf{y}). \tag{12}$$

Along the mixture-loss path, each slot's spectral gradient scales with its abundance, giving weak components little influence on the residual. The component loss averages spectral discrepancies across matched slots without weighting by mixture mass, directly supervising weak signatures. Presence supervision penalizes spurious slots even when their small masses make them negligible in reconstruction. Appendix A.2 details the losses and assignment.

Inference. At inference, we retain the active set $\mathcal{I}_{\mathbf{y}} = \{i : \widehat{p}_i \geq \tau_p\}$. If this set is empty, we add the slot with the highest presence probability. Abundances are then normalized over the active set as $\bar{\alpha}_i = \widehat{w}_i / \sum_{j \in \mathcal{I}_{\mathbf{y}}} \widehat{w}_j$. Each active slot takes the N_r anchors with the largest query inner products as candidates $\mathcal{N}(\widehat{\mathbf{u}}_i)$. Holding abundances and all other slots fixed, refinement chooses the candidate that best reconstructs the observed mixture:

$$\widehat{\mathbf{z}}_i \leftarrow \arg\min_{\mathbf{z} \in \mathcal{N}(\widehat{\mathbf{u}}_i)} d_{\mathrm{sp}} \left(\sum_{j \in \mathcal{I}_{\mathbf{y}} \setminus \{i\}} \bar{\alpha}_j R_\psi(\widehat{\mathbf{z}}_j) + \bar{\alpha}_i R_\psi(\mathbf{z}), \mathbf{y} \right). \tag{13}$$

One sweep visits slots in decreasing abundance order and accepts strict improvements. Duplicate anchors are then merged and their masses added. The resulting supports and masses define $\widehat{\mu}_{\mathbf{y}}$. Appendix A.3 also defines the ranked alternatives used for identification.

5 Experiments

5.1 Experimental Settings

Data and reference pools. Stage 1 learns spectral representations from the IR subset of the Multimodal Spectroscopic Dataset, comprising approximately 790k unique molecules (Alberts et al., 2024b). Stage 2 evaluates mixture recovery on simulated QM9Spectra (QM9S) spectra (Zou et al., 2023) and 5,228 experimentally measured gas-phase spectra from the NIST/EPA collection (Stein, 2008). For dilute, non-reactive gas mixtures, Beer–Lambert additivity allows mixture absorbance to be constructed from pure-component spectra (Swinehart, 1962; Mayerhöfer et al., 2020). Spectra are resampled to $L = 1800$ points over 400–$4000\,\mathrm{cm}^{-1}$ and peak-normalized. QM9S molecules are split $8{:}1{:}1$ by structure, with a separate unlabeled reference pool for each partition. NIST compounds are divided into five folds; each run fine-tunes the QM9S-trained estimator on four folds and evaluates on the fifth, using a reference pool spanning all five folds. Appendices B.1 and B.4 detail mixture sampling, model hyperparameters, and training configurations.

Baselines. We compare against NNLS (Melle et al., 2026), a fixed-library adaptation of EB-gMCR (Chang & Chen, 2025), a Transformer spectral set predictor (Vaswani et al., 2017), and a DeepCID-based CNN retrieval adaptation (Fan et al., 2019). The Transformer baseline takes the mixture spectrum as its sole input and predicts an unordered set of component spectra, presence probabilities, and abundances. Molecular identities are obtained by matching the predicted spectra to the reference collection. The CNN ranks reference spectra and applies NNLS to selected candidates. These methods span additive fitting, regularized coefficient estimation, direct spectrum prediction, and embedding-based retrieval, exposing different approaches to resolving spectral overlap. Appendix B.2 details the baseline adaptations.

Metrics. Set F1 is the harmonic mean of precision and recall computed from exact molecular identities after duplicate consolidation. Spectral information similarity (SIS) quantifies agreement between predicted and reference spectra (McGill et al., 2021). Abund. denotes the mean absolute error of component abundances after bipartite matching that maximizes SIS. This error is compared within each mixture size K, for which the mean true abundance is $1/K$. Top-1 and Top-10 measure whether the correct molecular identity ranks first or appears among the ten highest-ranked candidates for each matched prediction, with unmatched components counted as failures. Appendix B.3 provides the detailed metric definitions.

5.2 Main Results

Simulated Dataset. SpecMUSE achieves the best results across all five metrics and mixture sizes in Table 1. At $K = 4$, its F1 is 0.712, compared with 0.462 for NNLS, the strongest baseline. Higher SIS and lower abundance error accompany this identification gain, showing that the recovered identities also correspond to more accurate component spectra and weights.

Table 1: Comparison with baselines on QM9S for different mixture sizes K. Abund. denotes abundance MAE; Top-1 and Top-10 are reported as percentages. Bold and underlined values indicate the best and second-best results for each K, respectively.

K	Method	F1↑	SIS↑	Abund.↓	Top-1↑	Top-10↑
2	NNLS	<u>0.719</u>	<u>0.811</u>	<u>0.083</u>	<u>71.9</u>	<u>72.3</u>
	EB-gMCR	0.717	0.809	0.087	71.7	72.2
	Transformer	0.592	0.690	0.097	62.3	63.6
	DeepCID	0.604	0.722	0.130	60.4	61.3
	SpecMUSE	**0.975**	**0.977**	**0.023**	**98.2**	**98.7**
3	NNLS	<u>0.575</u>	<u>0.714</u>	<u>0.094</u>	<u>57.5</u>	<u>58.4</u>
	EB-gMCR	0.569	0.709	0.100	56.9	58.2
	Transformer	0.413	0.535	0.103	40.7	43.9
	DeepCID	0.429	0.597	0.125	42.9	44.4
	SpecMUSE	**0.854**	**0.917**	**0.034**	**87.3**	**90.6**
4	NNLS	<u>0.462</u>	<u>0.640</u>	0.087	<u>46.2</u>	<u>47.8</u>
	EB-gMCR	0.454	0.635	0.091	45.4	47.3
	Transformer	0.311	0.407	<u>0.083</u>	32.7	34.1
	DeepCID	0.333	0.531	0.106	33.3	35.0
	SpecMUSE	**0.712**	**0.820**	**0.027**	**71.1**	**78.4**

As K increases, identification becomes harder for every method. SpecMUSE retains its lead, but its widening gap between Top-1 and Top-10 indicates that correct identities increasingly remain among the alternatives without ranking first. Overlapping bands allow several candidate spectra to explain similar features. A candidate that matches prominent peaks in isolation may fit less well alongside the other selected components. The pure-spectrum prior links bands within each component, while additive reconstruction evaluates their compatibility within the recovered composition. Appendix B.6 examines the geometry of the learned spectral coordinates and their correspondence with chemical annotations.

Measured Dataset. On NIST, SpecMUSE leads in F1, SIS, Top-1, and Top-10 for every mixture size (Table 2). The strongest identification baseline shifts from NNLS for binary mixtures to DeepCID for larger mixtures. Lower identification scores and a wider Top-10–Top-1 gap than on QM9S indicate greater ambiguity among leading candidates. SpecMUSE also achieves the lowest abundance error at $K = 3$ and 4, whereas DeepCID is better at $K = 2$. This difference separates abundance accuracy from molecular identification: because abundance error follows spectral matching, an incorrect identity can receive a weight close to the matched component's abundance.

Case study. Figure 3 shows a three-component mixture dominated by component 2. SpecMUSE recovers all three identities, including the two weak components. Panels (a–c) retain their characteristic bands, while the mixture reconstruction and residual in panel (d) show how the recovered

Table 2: Comparison with baselines on NIST across mixture sizes K (mean $\pm$ standard deviation over five folds). Notation and highlighting follow Table 1.

K	Method	F1↑	SIS↑	Abund.↓	Top-1↑	Top-10↑
2	NNLS	0.589 ± 0.007	0.797 ± 0.004	0.131 ± 0.001	58.7 ± 0.7	64.3 ± 0.6
	EB-gMCR	0.555 ± 0.005	0.784 ± 0.004	0.144 ± 0.001	55.3 ± 0.5	62.3 ± 0.6
	Transformer	0.452 ± 0.011	0.709 ± 0.016	0.161 ± 0.007	51.2 ± 0.1	58.4 ± 0.2
	DeepCID	0.576 ± 0.017	0.772 ± 0.008	$\mathbf{0.091 \pm 0.004}$	58.0 ± 1.3	61.8 ± 0.9
	SpecMUSE	$\mathbf{0.678 \pm 0.011}$	$\mathbf{0.862 \pm 0.008}$	0.108 ± 0.004	$\mathbf{73.2 \pm 1.3}$	$\mathbf{85.5 \pm 1.2}$
3	NNLS	0.406 ± 0.003	0.695 ± 0.002	0.108 ± 0.001	40.6 ± 0.3	44.7 ± 0.3
	EB-gMCR	0.383 ± 0.002	0.680 ± 0.002	0.122 ± 0.001	38.2 ± 0.2	43.2 ± 0.2
	Transformer	0.373 ± 0.008	0.522 ± 0.015	0.122 ± 0.004	32.3 ± 0.1	37.9 ± 0.5
	DeepCID	0.421 ± 0.011	0.675 ± 0.006	0.107 ± 0.002	42.3 ± 0.7	46.6 ± 0.6
	SpecMUSE	$\mathbf{0.493 \pm 0.008}$	$\mathbf{0.753 \pm 0.007}$	$\mathbf{0.090 \pm 0.001}$	$\mathbf{52.1 \pm 1.0}$	$\mathbf{69.0 \pm 1.3}$
4	NNLS	0.279 ± 0.003	0.627 ± 0.001	0.083 ± 0.002	27.9 ± 0.3	31.0 ± 0.4
	EB-gMCR	0.261 ± 0.003	0.611 ± 0.001	0.095 ± 0.005	26.1 ± 0.3	29.7 ± 0.4
	Transformer	0.243 ± 0.001	0.538 ± 0.012	0.064 ± 0.003	24.3 ± 0.1	27.9 ± 0.2
	DeepCID	0.333 ± 0.006	0.621 ± 0.004	0.107 ± 0.001	33.4 ± 0.3	37.8 ± 0.2
	SpecMUSE	$\mathbf{0.407 \pm 0.006}$	$\mathbf{0.680 \pm 0.003}$	$\mathbf{0.056 \pm 0.001}$	$\mathbf{40.0 \pm 0.5}$	$\mathbf{58.1 \pm 0.8}$

spectra combine. The abundances in panel (e) preserve the imbalance. Recovering the weak components retains composition information that contributes little to the dominant peaks. Appendix D provides additional cases across both datasets and mixture sizes.

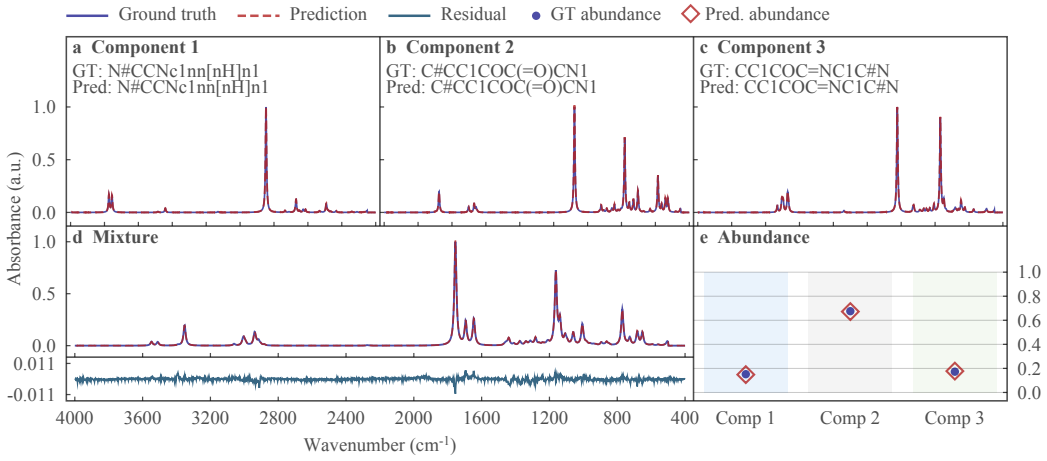

Figure 3: Recovery of an imbalanced three-component mixture: **(a–c)** component spectra, **(d)** reconstructed mixture and residual, and **(e)** component abundances.

5.3 Sensitivity Analysis

Table 3 examines the minimum component abundance $\alpha_{\min}$ on QM9S. Raising this floor strengthens weak spectral signatures, with the largest identification gains in crowded mixtures. It also caps the largest abundance at $1 - (K-1)\alpha_{\min}$, so the decrease in MAE reflects both easier component recovery and a narrower range of possible weights. Gains diminish as the mixtures become more balanced. We choose $\alpha_{\min} = 0.15$, near the onset of this plateau, to retain abundance imbalance while making weak components more distinguishable.

Table 4 examines the refinement candidate count N_r. A larger neighborhood offers alternatives for bands that admit several component assignments. Comparing each alternative through the joint residual evaluates its fit alongside selected spectra. Refinement holds slot abundances unchanged;

Table 3: Sensitivity to minimum component abundance on QM9S ($N_r = 50$).

	Abundance MAE↓ / Top-1↑ (%)		
$\alpha_{\min}$	$K = 2$	$K = 3$	$K = 4$
0.05	0.030 / 95.3	0.049 / 78.7	0.056 / 60.2
0.10	**0.023 / 98.2**	0.039 / 84.7	0.039 / 66.5
0.15	**0.023 / 98.2**	**0.034 / 87.3**	0.027 / 71.1
0.20	0.024 / 98.1	0.035 / 87.2	**0.023 / 71.7**

Table 4: Sensitivity to the refinement candidate count on QM9S ($\alpha_{\min} = 0.15$).

	Abundance MAE↓ / Top-1↑ (%)		
N_r	$K = 2$	$K = 3$	$K = 4$
10	0.028 / 96.9	0.039 / 81.6	0.035 / 63.2
20	0.026 / 97.6	0.036 / 84.6	0.031 / 67.2
40	0.025 / **98.2**	0.035 / 86.8	0.028 / 70.3
50	**0.023 / 98.2**	**0.034 / 87.3**	**0.027 / 71.1**

abundance MAE changes through support selection, duplicate consolidation, and spectral matching. Identification saturates earlier for binary mixtures, while larger mixtures continue to benefit from additional candidates. We choose $N_r = 50$ to retain the gains observed for larger mixtures.

5.4 Ablation Study

Component representations. Table 5 compares ablations on QM9S at $K = 4$. Removing Stage 1 alignment causes the largest degradation in SIS and abundance accuracy. Spectral reconstruction can fit a source without placing its reconstructed view at a consistent coordinate, weakening the targets used to learn component selection. The *w/o* R_ψ variant removes the spectral realization branch and its Stage 1 loss $\mathcal{L}_{\text{raw}}$, using stored reference spectra for component realization. Its lower accuracy supports the contribution of learned realization to recovering mixture composition.

Gradient propagation. Removing the soft surrogate preserves individual component outputs but blocks spectral gradients to the queries. Coordinate targets still supervise selection, while discrepancies in the selected spectra and their weighted sum can no longer adjust the queries. The resulting decline in identification shows the value of this additional feedback: nearby spectral alternatives inform how a query should move when its current component choice fails to explain the mixture.

Table 5: QM9S component ablations.

Variant	F1↑	SIS↑	Abund.↓	Top-1↑
Full	**0.712**	**0.820**	**0.027**	**71.1**
w/o R_ψ	0.398	0.698	0.128	42.7
w/o $\boldsymbol{\rho}_i$	0.250	0.663	0.129	25.0
w/o $\mathcal{L}_{\text{mix}}$	0.668	0.791	0.029	66.7
w/o $\mathcal{L}_{\text{slot}}$	0.306	0.579	0.131	34.6
w/o $\mathcal{L}_{\text{nce}}^{(1)}$	0.252	0.437	0.149	30.1

Supervision objectives. Removing $\mathcal{L}_{\text{slot}}$ discards coordinate, presence, and abundance targets. The resulting loss of identification accuracy reflects the ambiguity left by spectral supervision when different component sets explain similar bands. Removing $\mathcal{L}_{\text{mix}}$ has a smaller effect, reducing F1 from 0.712 to 0.668 while component and slot supervision remain. Direct targets guide which components are present and their abundances; the mixture loss adds a joint constraint on how those predictions combine into the observed spectrum.

6 Conclusion

SpecMUSE recovers mixture composition by estimating support locations and masses on a learned spectral manifold. A pure-spectrum prior provides component representations, and unordered set prediction estimates component coordinates, presence, and abundances. Additive reconstruction links these component estimates to the observed mixture. Experiments on QM9S and NIST show improved identification and spectral recovery for mixtures containing two to four components, with generalization to unseen mixtures. The recovered spectra and abundances describe component signatures and their relative weights. This description could support automated experimental analysis through comparisons across samples and help associate changes during reaction monitoring with individual components. **Limitations and future work.** Evaluation covers a small fraction of chemical space, using simulated spectra of small molecules (H, C, N, O, F) and fewer measured gas-phase spectra. Fixed output slots with presence prediction may limit flexibility when component counts vary more widely. Advances in representation learning, including JEPA-style prediction (Chen et al., 2026), could improve feature stability under baseline variation, peak shifts, and broadening. Progress in generative modeling (Geng et al., 2025; Janati et al., 2025) could broaden spectral coverage and support physically constrained recovery with uncertainty estimates.

AI USE STATEMENT

We used generative AI tools to assist with experimental code implementation. We also used these tools to improve the language and readability of the manuscript, generate selected illustrative icons, and refine figure color schemes and layouts. All AI-assisted text, code, and visual elements were reviewed by the authors. The authors take full responsibility for the final content of this work, including all text, claims, experimental code, figures, and reported results.

ETHICS STATEMENT

This study analyzes molecular spectra from existing scientific datasets and does not involve human participants or personal data. The reported abundances describe normalized spectral mixtures and should be interpreted within the stated experimental setting. Use and redistribution of the underlying spectral collections are subject to their providers' access and licensing terms.

REPRODUCIBILITY STATEMENT

Section 4 specifies the two-stage model and inference procedure. Appendix A details the spectral losses, assignment, slot supervision, and refinement. Appendix B records the data partitions, available anchor pools, mixture generation, baseline adaptations, metrics, training configuration, and computational costs. Appendix B.5 analyzes structural overlap and spectral variation.

REFERENCES

Marvin Alberts, Teodoro Laino, and Alain C Vaucher. Leveraging infrared spectroscopy for automated structure elucidation. *Communications Chemistry*, 7:268, 2024a.

Marvin Alberts, Oliver Schilter, Federico Zipoli, Nina Hartrampf, and Teodoro Laino. Unraveling molecular structure: A multimodal spectroscopic dataset for chemistry. In *Advances in neural information processing systems*, volume 37, pp. 125780–125808, 2024b.

Marvin Alberts, Filippo Ficarra, and Teodoro Laino. Language model enabled structure prediction from infrared spectra of mixtures. In *AI for Accelerated Materials Design-NeurIPS 2025*, 2025.

Andrea Angulo, Lankun Yang, Eray S Aydil, and Miguel A Modestino. Machine learning enhanced spectroscopic analysis: towards autonomous chemical mixture characterization for rapid process optimization. *Digital Discovery*, 1(1):35–44, 2022.

Yosuke Ashikari, Takashi Tamaki, Kyosuke Tomite, Yuya Yonekura, and Aiichiro Nagaki. Real-time inline-ir-analysis via linear-combination strategy and machine learning for automated reaction optimization. *Communications Chemistry*, 8:287, 2025.

Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International conference on machine learning*, volume 70, pp. 537–546. PMLR, 2017.

Nicholas Boyd, Geoffrey Schiebinger, and Benjamin Recht. The alternating descent conditional gradient method for sparse inverse problems. *SIAM Journal on Optimization*, 27(2):616–639, 2017.

Kristian Bredies and Hanna Katriina Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pp. 213–229. Springer, 2020.

Yu-Tang Chang and Shih-Fang Chen. Eb-gmcr: Energy-based generative modeling for signal unmixing and multivariate curve resolution. *arXiv preprint arXiv:2507.23600*, 2025.

Delong Chen, Mustafa Shukor, Théo Moutakanni, Willy Chung, Lei Yu, Tejaswi Kasarla, Allen Bolourchi, Yann LeCun, and Pascale Fung. VL-JEPA: Joint Embedding Predictive Architecture for Vision-Language. In *International Conference on Learning Representations*, 2026.

Lénaïc Chizat. Sparse optimization on measures with over-parameterized gradient descent. *Mathematical Programming*, 194:487–532, 2022.

Pamela M. Chu, Franklin R. Guenther, George C. Rhoderick, and Walter J. Lafferty. The nist quantitative infrared database. *Journal of Research of the National Institute of Standards and Technology*, 104(1):59–81, 1999.

Norman B. Colthup, Lawrence H. Daly, and Stephen E. Wiberley. *Introduction to Infrared and Raman Spectroscopy*. Academic Press, 3 edition, 2012.

David F Crouse. On implementing 2d rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, 52(4):1679–1696, 2016.

Anna de Juan and Romà Tauler. Multivariate curve resolution: 50 years addressing the mixture analysis problem–a review. *Analytica Chimica Acta*, 1145:59–78, 2021.

Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

Xiaqiong Fan, Wen Ming, Huitao Zeng, Zhimin Zhang, and Hongmei Lu. Deep learning-based component identification for the raman spectra of mixtures. *Analyst*, 144(5):1789–1798, 2019.

Xiaqiong Fan, Yue Wang, Chuanxiu Yu, Yuanxia Lv, Hailiang Zhang, Qiong Yang, Ming Wen, Hongmei Lu, and Zhimin Zhang. A universal and accurate method for easily identifying components in raman spectroscopy based on deep learning. *Analytical Chemistry*, 95(11):4863–4870, 2023.

Hanyu Gao, Bin Cao, Yunyue Su, Tong-Yi Zhang, and Qiang Liu. Xdecomposer: Learning prior-free set decomposition for multiphase x-ray diffraction. *arXiv preprint arXiv:2605.05866*, 2026.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, volume 38, pp. 75460–75482, 2025.

Peter R Griffiths. Fourier transform infrared spectrometry. *Science*, 222(4621):297–302, 1983.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16000–16009, June 2022.

Shiluo Huang, Yining Jin, Wei Jin, and Ying Mu. Analytical-chemistry-informed transformer for infrared spectra modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17440–17448, 2025.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

Yusei Ito, Naoya Chiba, Tatsunori Taniai, Ryo Igarashi, Yuta Suzuki, Kotaro Saito, Yoshitaka Ushiku, and Kanta Ono. Enabling single-observation decomposition of multi-phase x-ray diffraction patterns via generative deep learning. *npj Computational Materials*, 12:238, 2026.

Yazid Janati, Badr Moufad, Mehdi Abou El Qassime, Alain Oliviero Durmus, Eric Moulines, and Jimmy Olsson. A mixture-based framework for guiding diffusion models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 26830–26876. PMLR, 2025.

Ganesh Chandan Kanakala, Bhuvanesh Sridharan, and U Deva Priyakumar. Spectra to structure: Contrastive learning framework for library ranking and generating molecular structures for infrared spectra. *Digital Discovery*, 3(12):2417–2423, 2024.

Hugues Kouakou, José Henrique de Morais Goulart, Raffaele Vitale, Thomas Oberlin, David Rousseau, Cyril Ruckebusch, and Nicolas Dobigeon. On-the-fly spectral unmixing based on kalman filtering. *Chemometrics and Intelligent Laboratory Systems*, 255:105252, 2024.

Onur Can Koyun, Reyhan Kevser Keser, Safa Onur Sahin, Damla Bulut, Mustafa Yorulmaz, Veysel Yucesoy, and Behcet Ugur Toreyin. Ramanformer: A transformer-based quantification approach for raman mixture components. *ACS omega*, 9(22):23241–23251, 2024.

Fred A Kruse, Adam B Lefkoff, Joseph W Boardman, Kathleen B Heidebrecht, AT Shapiro, PJ Barloon, and Alexander FH Goetz. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3): 145–163, 1993.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pp. 3744–3753. PMLR, 2019.

Peter J Linstrom and William G Mallard. The nist chemistry webbook: A chemical data resource on the internet. *Journal of Chemical & Engineering Data*, 46(5):1059–1063, 2001.

Liyuan Liu, Chengyu Dong, Xiaodong Liu, Bin Yu, and Jianfeng Gao. Bridging discrete and backpropagation: Straight-through and beyond. *Advances in Neural Information Processing Systems*, 36:12291–12311, 2023.

Francesco Locatello, Dirk Weissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, and Thomas Kipf. Object-centric learning with slot attention. *Advances in neural information processing systems*, 33:11525–11538, 2020.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Thomas G Mayerhöfer, Susanne Pahlow, and Jürgen Popp. The bouguer-beer-lambert law: Shining light on the obscure. *ChemPhysChem*, 21(18):2029–2046, 2020.

Charles McGill, Michael Forsuelo, Yanfei Guan, and William H Green. Predicting infrared spectra with message passing neural networks. *Journal of chemical information and modeling*, 61(6): 2594–2609, 2021.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction, 2020. URL `https://arxiv.org/abs/1802.03426`.

Yannah JU Melle, Thanh Nguyen, Jeffrey Lopez, and Daniel Schwalbe-Koda. Automatic identification of compounds in molecular mixtures from liquid-phase infrared spectra. *Chemical science*, 17(27):13465–13478, 2026.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Santosh Kumar Paidi and Parul Maheshwari. Ramanmae: Masked autoencoders enable efficient molecular imaging by learning biologically meaningful spectral representations. *Analytical Chemistry*, 97(48):26812–26820, 2025.

Vladimir V Prischepa, VE Skiba, Denis A Vrazhnov, and Yu V Kistenev. Gas mixtures ir absorption spectra decomposition using a deep neural network. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 301:108521, 2023.

Hamid Rezatofighi, Tianyu Zhu, Roman Kaskman, Farbod T. Motlagh, Javen Qinfeng Shi, Anton Milan, Daniel Cremers, Laura Leal-Taixe, and Ian Reid. Learn to predict sets using feed-forward neural networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 44(12):9011–9025, 2022.

Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

George Socrates. *Infrared and Raman characteristic group frequencies: tables and charts*. John Wiley & Sons, 3 edition, 2004.

Stephen E. Stein. Nist 35. NIST/EPA Gas-Phase Infrared Database - JCAMP Format, October 2008.

Barbara H. Stuart. *Infrared Spectroscopy: Fundamentals and Applications*. John Wiley & Sons, 2004. doi: 10.1002/0470011149.

Donald F Swinehart. The beer-lambert law. *Journal of chemical education*, 39(7):333, 1962.

Lin Tan, Yue Wang, Hailiang Zhang, Jinyu Sun, Qiong Yang, Xiao Yang, Zhimin Zhang, and Hongmei Lu. Deepmir: A hybrid convolutional neural network-transformer framework for accurate identification of target components from mid-infrared spectra of mixtures. *Analytical Chemistry*, 97(50):27706–27715, 2025.

Linbo Tian, Jinbao Xia, Alexandre A Kolomenskii, Hans A Schuessler, Feng Zhu, Yanfeng Li, Jingliang He, Qian Dong, and Sasa Zhang. Gas phase multicomponent detection and analysis combining broadband dual-frequency comb absorption spectroscopy and deep learning. *Communications Engineering*, 2(1):54, 2023.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6306–6315, 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, volume 30, 2017.

Liang Wang, Shaozhen Liu, Yu Rong, Deli Zhao, Qiang Liu, Shu Wu, and Liang Wang. MolSpectra: Pre-training 3d molecular representation with multi-modal energy spectra. In *International Conference on Learning Representations*, 2025.

Zilong Wang, Yunfeng Li, Jinglei Zhai, Siwei Yang, Biao Sun, and Pei Liang. Deep learning-based raman spectroscopy qualitative analysis algorithm: A convolutional neural network and transformer approach. *Talanta*, 275:126138, 2024.

Nicholas Watters, Loic Matthey, Christopher P Burgess, and Alexander Lerchner. Spatial broadcast decoder: A simple architecture for learning disentangled representations in vaes. *arXiv preprint arXiv:1901.07017*, 2019.

Edgar Bright Wilson, John Courtney Decius, and Paul C. Cross. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. McGraw-Hill, New York, 1955.

Dong Wu, Mingmin Chi, Xuan Zang, and Bo Peng. Solving spectrum unmixing as a multi-task bayesian inverse problem with latent factors for endmember variability. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 15878–15887, 2024.

Bingsen Xue, Xinyuan Bi, Zheyi Dong, Yunzhe Xu, Minghui Liang, Xin Fang, Yizhe Yuan, Ruoxi Wang, Shuyu Liu, Rushi Jiao, et al. Deep spectral component filtering as a foundation model for spectral analysis demonstrated in metabolic profiling. *Nature Machine Intelligence*, 7(5): 743–757, 2025.

Dong Yu, Morten Kolbæk, Zheng-Hua Tan, and Jesper Jensen. Permutation invariant training of deep models for speaker-independent multi-talker speech separation. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 241–245. IEEE, 2017.

Xinna Yu, Tianyuan Liu, Lili Kong, Tianshuo Lan, Qifang Sun, Fanhua Qu, Meichun Liu, Jie Chen, and Meizhen Huang. Specrecformer: Deep learning-driven adaptive component identification of pah mixtures based on single-component raman spectra. *Analytical Chemistry*, 97(18):9876–9883, 2025.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander Smola. Deep sets. In *Advances in neural information processing systems*, volume 30, 2017.

Zhi-Min Zhang, Xiao-Qing Chen, Hong-Mei Lu, Yi-Zeng Liang, Wei Fan, Deng Xu, Jack Zhou, Fei Ye, and Zheng-Yu Yang. Mixture analysis using reverse searching and non-negative least squares. *Chemometrics and Intelligent Laboratory Systems*, 137:10–20, 2014.

Xin Zhao, Caizheng Liu, Ziyan Zhao, Qibing Zhu, and Min Huang. Performance improvement of handheld raman spectrometer for mixture components identification using fuzzy membership and sparse non-negative least squares. *Applied Spectroscopy*, 76(5):548–558, 2022.

Ziyan Zhao, Zhenfang Liu, Mingqiang Ji, Xin Zhao, Qibing Zhu, and Min Huang. Conincedeep: A novel deep learning method for component identification of mixture based on raman spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 234:104757, 2023.

Zihan Zou, Yujin Zhang, Lijun Liang, Mingzhi Wei, Jiancai Leng, Jun Jiang, Yi Luo, and Wei Hu. A deep learning model for predicting selected organic molecular spectra. *Nature Computational Science*, 3:957–964, 2023.

A METHODOLOGY DETAILS

This appendix specifies the operators and training objectives used in Section 4.

A.1 MANIFOLD CONSTRUCTION DETAILS

Spectral patches and masked reconstruction. With patch width q, stride $r_p < q$, and $L - q$ divisible by r_p, patch n starts at $a_n = 1 + (n-1)r_p$ and the sequence contains $N = 1 + (L-q)/r_p$ patches. For $n = 1, \dots, N$, extraction and embedding give

$$\mathbf{p}_{s,n} = (s_{a_n}, \dots, s_{a_n+q-1})^\top, \qquad \mathbf{x}_{s,n} = \mathbf{W}_p \mathbf{p}_{s,n} + \mathbf{e}_n^p. \tag{14}$$

Here $\mathbf{W}_p \in \mathbb{R}^{d_e \times q}$ and $\mathbf{e}_n^p \in \mathbb{R}^{d_e}$ are the patch projection and positional embedding. Stacking these tokens gives $\mathbf{X}_s \in \mathbb{R}^{N \times d_e}$.

Masking operates on wavenumber intervals because adjacent patches share spectral values. At each pretraining step, we sample $\lfloor r_m N \rfloor$ seed patch indices uniformly without replacement. Their index set $\mathcal{A}_s^{\mathrm{msk}}$ defines the masked positions $\mathcal{W}_s^{\mathrm{msk}} = \bigcup_{n \in \mathcal{A}_s^{\mathrm{msk}}} \{a_n, \dots, a_n + q - 1\}$. A patch is visible when none of its positions belongs to this set:

$$\mathcal{I}_{\mathrm{vis}} = \{n \mid \{a_n, \dots, a_n + q - 1\} \cap \mathcal{W}_s^{\mathrm{msk}} = \varnothing\}. \tag{15}$$

The remaining indices form $\mathcal{I}_{\mathrm{msk}} = \{1, \dots, N\} \setminus \mathcal{I}_{\mathrm{vis}}$. Partially overlapping patches are therefore excluded, and r_m specifies the seed ratio rather than the final proportion of hidden tokens.

The encoder produces $\mathbf{H}_s^{\mathrm{vis}} = E_\psi(\mathbf{X}_s^{\mathrm{vis}})$. If $j(n)$ locates visible patch n within this shorter sequence, restoration assigns $\mathbf{h}_{s,n}^{\mathrm{rst}} = \mathbf{W}_{ed} \mathbf{h}_{s,j(n)}^{\mathrm{vis}} + \mathbf{e}_n^d$. Hidden positions instead receive $\mathbf{h}_{s,n}^{\mathrm{rst}} = \mathbf{m}_\psi + \mathbf{e}_n^d$, where $\mathbf{m}_\psi$ is the learned mask token and $\mathbf{e}_n^d$ identifies the patch location. The ordered restored sequence implements U_ψ. This asymmetric encoder and decoder follow masked autoencoders (He et al., 2022); spectral patching follows MolSpectra (Wang et al., 2025).

The shared decoder produces nonnegative patch predictions $\widehat{\mathbf{P}}_s = D_\psi(\mathbf{H}_s^{\mathrm{rst}}) \in \mathbb{R}_+^{N \times q}$. Let $\mathcal{N}_\ell = \{n \mid a_n \le \ell \le a_n + q - 1\}$ index the patches covering position ℓ. Their predictions are averaged as

$$\widehat{s}_\ell^{\mathrm{msr}} = \frac{1}{|\mathcal{N}_\ell|} \sum_{n \in \mathcal{N}_\ell} \widehat{P}_{s,n,\ell - a_n + 1}. \tag{16}$$

Here $\widehat{P}_{s,n,t}$ is the t-th prediction within patch n, and the equation applies at every $\ell = 1, \dots, L$. This operation defines $\mathcal{U}_{\mathrm{ola}}$ for both masked reconstruction and spectral realization.

Spectral discrepancy. The discrepancy d_{sp} combines amplitude, derivative, and angular terms. For a scalar residual e, the Huber penalty with threshold $\delta_{\mathrm{H}} > 0$ is (Huber, 1964)

$$h_{\delta_{\mathrm{H}}}(e) = \frac{1}{2} \min\{e^2, \delta_{\mathrm{H}}^2\} + \delta_{\mathrm{H}} \max\{|e| - \delta_{\mathrm{H}}, 0\}. \tag{17}$$

This penalty is quadratic up to the threshold and linear beyond it. For spectra $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^L$, let $\boldsymbol{\xi}^{(1)} = \mathbf{D}_{\mathrm{SG}}^{(1)}(\mathbf{a} - \mathbf{b})$, where $\mathbf{D}_{\mathrm{SG}}^{(1)}$ is the first-order Savitzky–Golay derivative operator on Ω (Savitzky & Golay, 1964). The amplitude and derivative discrepancies are

$$d_{\mathrm{amp}}(\mathbf{a}, \mathbf{b}) = \frac{1}{L} \sum_{\ell=1}^{L} h_{\delta_{\mathrm{H}}}(a_\ell - b_\ell), \qquad d_{\mathrm{der}}(\mathbf{a}, \mathbf{b}) = \frac{1}{L} \sum_{\ell=1}^{L} h_{\delta_{\mathrm{H}}}(\xi_\ell^{(1)}). \tag{18}$$

The derivative term responds to local differences in peak position and width. The angular term compares relative band patterns (Kruse et al., 1993), with $\varepsilon_n > 0$ and $0 < \varepsilon_c < 1$ stabilizing the normalization and inverse cosine:

$$d_{\mathrm{sam}}(\mathbf{a}, \mathbf{b}) = \arccos\left[\mathrm{clip}_{[-1+\varepsilon_c,\, 1-\varepsilon_c]}\left(\frac{\mathbf{a}^\top \mathbf{b}}{\max\{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2, \varepsilon_n\}}\right)\right]. \tag{19}$$

Nonnegative coefficients η_a, η_s, and η_d combine these terms:

$$d_{\mathrm{sp}}(\mathbf{a}, \mathbf{b}) = \eta_a d_{\mathrm{amp}}(\mathbf{a}, \mathbf{b}) + \eta_s d_{\mathrm{sam}}(\mathbf{a}, \mathbf{b}) + \eta_d d_{\mathrm{der}}(\mathbf{a}, \mathbf{b}). \tag{20}$$

Coordinate alignment and spectral realization. The complete spectrum and detached reconstruction $\mathrm{sg}(\widehat{\mathbf{s}}^{\mathrm{msr}})$ share the patch embedding, encoder, and projector. Their tokens give $\mathbf{H}_s = E_\psi(\mathbf{X}_s)$ and $\mathbf{H}_s^{\mathrm{rec}} = E_\psi(\widehat{\mathbf{X}}_s)$. Attention pooling followed by a two-layer MLP defines $G_\psi = g_\psi \circ \mathrm{Pool}_\psi$. Normalizing its nonzero outputs gives

$$\mathbf{z}_s = \mathrm{Norm}_2(G_\psi(\mathbf{H}_s)), \qquad \mathbf{z}_s^{\mathrm{rec}} = \mathrm{Norm}_2(G_\psi(\mathbf{H}_s^{\mathrm{rec}})). \tag{21}$$

Here $\mathrm{Norm}_2(\mathbf{v}) = \mathbf{v}/\|\mathbf{v}\|_2$ for $\mathbf{v} \neq \mathbf{0}$. Equation 4 applies InfoNCE to these paired coordinates using in-batch negatives at temperature $\tau_c > 0$ (Oord et al., 2018). Both views update the shared encoder and projector; detachment blocks alignment gradients through reconstruction.

The realization field R_ψ uses the coordinate lifting in Equation 5 and the shared decoder D_ψ. For a pure-spectrum batch $\mathcal{B}_p$, masked reconstruction and realization minimize

$$\mathcal{L}_{\mathrm{msr}} = \frac{1}{|\mathcal{B}_p|} \sum_{\mathbf{s} \in \mathcal{B}_p} d_{\mathrm{sp}}(\widehat{\mathbf{s}}^{\mathrm{msr}}, \mathbf{s}), \qquad \mathcal{L}_{\mathrm{raw}} = \frac{1}{|\mathcal{B}_p|} \sum_{\mathbf{s} \in \mathcal{B}_p} d_{\mathrm{sp}}(R_\psi(\mathbf{z}_s), \mathbf{s}). \tag{22}$$

Masked reconstruction is supervised over the complete grid. The realization loss updates the field and backpropagates through the original-view coordinate $\mathbf{z}_s$ to the encoder.

A.2 Estimator and Stage-2 Objectives

Slot prediction and reference realization. The mixture uses the same patch width and stride as the pure spectra. The $K_{\max}$ query slots (Vaswani et al., 2017; Carion et al., 2020; Lee et al., 2019) use shared MLP heads to predict coordinate queries, presence probabilities, and abundance scores. For slot vector $\mathbf{h}_i$, the latter two outputs are

$$\widehat{p}_i = \sigma(g_p(\mathbf{h}_i)), \qquad \widehat{w}_i = \mathrm{softplus}(g_w(\mathbf{h}_i)) + \varepsilon_w. \tag{23}$$

The constant $\varepsilon_w > 0$ keeps abundance scores nonzero. Coordinate queries $\mathbf{u}_i = g_u(\mathbf{h}_i)$ use the clipped normalization in Equation 6, with $\varepsilon_u > 0$. Targets are assigned after prediction through permutation-invariant matching.

The original-view coordinates of reference spectra and their field evaluations form the matrices

$$\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_M]^\top, \qquad \mathbf{S}_\psi = [R_\psi(\mathbf{z}_1), \ldots, R_\psi(\mathbf{z}_M)]^\top. \tag{24}$$

The reference pools follow the QM9S partitions and NIST folds in Appendix B.1. The positive kernel temperature τ_z is selected on validation data.

Hard realization and surrogate gradients. Using the straight-through weights in Equation 8 (Bengio et al., 2013; Liu et al., 2023), each slot returns

$$\widehat{\mathbf{z}}_i = \mathbf{Z}^\top \widetilde{\boldsymbol{\rho}}_i, \qquad \widehat{\mathbf{s}}_i = \mathbf{S}_\psi^\top \widetilde{\boldsymbol{\rho}}_i. \tag{25}$$

The forward pass selects a reference coordinate $\widehat{\mathbf{z}}_i$ and its field evaluation $R_\psi(\widehat{\mathbf{z}}_i)$. Backpropagation differentiates the soft spectral map $\mathbf{S}_\psi^\top \boldsymbol{\rho}_i$ with respect to the continuous query.

Write $r_i = \|\mathbf{u}_i\|_2$ and let $\mathbf{B}_i$ denote the Jacobian of the normalization in Equation 6. Away from $r_i = \varepsilon_u$, its two branches are

$$\mathbf{B}_i = \frac{\mathbf{I} - \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i^\top}{r_i} \ (r_i > \varepsilon_u), \qquad \mathbf{B}_i = \frac{\mathbf{I}}{\varepsilon_u} \ (r_i < \varepsilon_u). \tag{26}$$

For $r_i > \varepsilon_u$, the query has unit norm and the Jacobian projects onto its tangent space. For $r_i < \varepsilon_u$, normalization is linear rescaling and the query lies inside the unit ball. With $\mathbf{J}_i = \mathrm{Diag}(\boldsymbol{\rho}_i) - \boldsymbol{\rho}_i \boldsymbol{\rho}_i^\top$, both branches yield

$$\widetilde{\nabla}_{\mathbf{u}_i} \mathcal{L}_{\mathrm{spectral}} = \tau_z^{-1} \mathbf{B}_i^\top \mathbf{Z}^\top \mathbf{J}_i \mathbf{S}_\psi \nabla_{\widehat{\mathbf{s}}_i} \mathcal{L}_{\mathrm{spectral}}. \tag{27}$$

Equation 9 is the unit-norm branch of this expression. At the clipping boundary, the normalization is nondifferentiable; the two formulas above describe its smooth regions.

Set matching and slot supervision. For annotated components $\{(\mathbf{s}_k, \mathbf{z}_k, \alpha_k)\}_{k=1}^{K}$, let $c_{ik} = d_{\mathrm{sp}}(\widehat{\mathbf{s}}_i, \mathbf{s}_k)$ and let $\Pi_{K,K_{\max}}$ contain injective assignments from components to slots. Spectral matching and the corresponding component loss are (Yu et al., 2017)

$$\pi^\star = \arg\min_{\pi \in \Pi_{K,K_{\max}}} \frac{1}{K} \sum_{k=1}^{K} c_{\pi(k),k}, \qquad \mathcal{L}_{\mathrm{comp}} = \frac{1}{K} \sum_{k=1}^{K} c_{\pi^\star(k),k}. \tag{28}$$

We solve the rectangular linear assignment with the Jonker–Volgenant algorithm (Crouse, 2016). At each step, the assignment is recomputed and held constant when differentiating the losses.

Let $\mathcal{B}_m$ be a mixture mini-batch and $\mathcal{C}_{\mathcal{B}} = \{(b,k) \mid b \in \mathcal{B}_m,\ 1 \le k \le K_b\}$ index its annotated components. Each matched query $\mathbf{v}_{bk} = \widehat{\mathbf{u}}_{b,\pi_b^\star(k)}$ has the original-view pure-spectrum coordinate $\mathbf{z}_{bk}$ as its target. The negative set $\mathcal{N}_{bk} \subseteq \mathcal{C}_{\mathcal{B}}$ contains components whose corpus identity differs from that of (b,k), excluding repeated occurrences of the positive component. With $\kappa(\mathbf{a},\mathbf{b}) = \exp(\mathbf{a}^\top \mathbf{b}/\tau_s)$ and $\tau_s > 0$, the coordinate loss is

$$\mathcal{L}_{\mathrm{nce}}^{(2)} = -\frac{1}{|\mathcal{C}_{\mathcal{B}}|} \sum_{(b,k)\in\mathcal{C}_{\mathcal{B}}} \log \frac{\kappa(\mathbf{v}_{bk}, \mathbf{z}_{bk})}{\kappa(\mathbf{v}_{bk}, \mathbf{z}_{bk}) + \sum_{(b',k')\in\mathcal{N}_{bk}} \kappa(\mathbf{v}_{bk}, \mathbf{z}_{b'k'})}. \tag{29}$$

The target t_i is one for slots matched by $\pi^\star$ and zero for unmatched slots. Binary cross-entropy averages these targets over all $K_{\max}$ outputs:

$$\mathcal{L}_{\mathrm{prs}} = -\frac{1}{K_{\max}} \sum_{i=1}^{K_{\max}} \left[t_i \log \widehat{p}_i + (1 - t_i) \log(1 - \widehat{p}_i) \right]. \tag{30}$$

The implementation evaluates binary cross-entropy from logits for numerical stability.

Matched slots receive abundance targets. We normalize their positive scores within the matched set and minimize the mean absolute error:

$$\widehat{\alpha}_{\pi^\star(k)} = \frac{\widehat{w}_{\pi^\star(k)}}{\sum_{\ell=1}^{K} \widehat{w}_{\pi^\star(\ell)}}, \qquad \mathcal{L}_{\alpha} = \frac{1}{K} \sum_{k=1}^{K} |\widehat{\alpha}_{\pi^\star(k)} - \alpha_k|. \tag{31}$$

The spectral assignment supplies the correspondence for coordinate, presence, and abundance supervision, which together form

$$\mathcal{L}_{\mathrm{slot}} = \lambda_z \mathcal{L}_{\mathrm{nce}}^{(2)} + \lambda_p \mathcal{L}_{\mathrm{prs}} + \lambda_\alpha \mathcal{L}_{\alpha}. \tag{32}$$

Joint objective and ablation variants. Mixture reconstruction uses all slots with masses proportional to $\widehat{p}_i \widehat{w}_i$ (Equation 10), whereas abundance supervision normalizes scores within the matched set. The consistency loss $\mathcal{L}_{\mathrm{mix}} = d_{\mathrm{sp}}(\mathbf{y}_{\mathrm{rec}}, \mathbf{y})$ enters the joint objective in Equation 12.

The *w/o* R_ψ variant removes the spectral realization branch and its associated Stage 1 loss $\mathcal{L}_{\mathrm{raw}}$, using stored reference spectra for component realization. Removing $\mathcal{L}_{\mathrm{slot}}$ drops the coordinate, presence, and abundance terms while retaining component spectral supervision and mixture reconstruction. Removing $\mathcal{L}_{\mathrm{mix}}$ retains component and slot supervision.

The Stage 2 objective is averaged over each mixture mini-batch when optimizing θ. Both stages use AdamW (Loshchilov & Hutter, 2019) with linear warmup, cosine decay, gradient-norm clipping, and mixed precision. Appendix B.4 lists the training configurations.

A.3 Inference Details

Activation and candidate refinement. A presence threshold selects the active components, whose abundance scores are normalized to sum to one:

$$\mathcal{I}_{\mathbf{y}} = \{i \mid \widehat{p}_i \ge \tau_p\}, \qquad \bar{\alpha}_i = \frac{\widehat{w}_i}{\sum_{j\in\mathcal{I}_{\mathbf{y}}} \widehat{w}_j}. \tag{33}$$

Slot indices range from 1 to $K_{\max}$, and masses are assigned for $i \in \mathcal{I}_{\mathbf{y}}$. If no slot reaches the threshold, we retain the highest-presence slot before normalization.

For each active slot, $\mathcal{N}(\widehat{\mathbf{u}}_i)$ contains the N_r anchors with the largest query inner products. Holding abundances and other supports unchanged, define the background and candidate mixture as

$$\mathbf{b}_i = \sum_{j \in \mathcal{I}_\mathbf{y} \setminus \{i\}} \bar{\alpha}_j R_\psi(\widehat{\mathbf{z}}_j), \qquad \mathbf{y}_i(\mathbf{z}) = \mathbf{b}_i + \bar{\alpha}_i R_\psi(\mathbf{z}). \tag{34}$$

The candidate that best explains the observed mixture is

$$\mathbf{z}_i^\star = \arg\min_{\mathbf{z} \in \mathcal{N}(\widehat{\mathbf{u}}_i)} d_{\mathrm{sp}}(\mathbf{y}_i(\mathbf{z}), \mathbf{y}). \tag{35}$$

One sweep visits slots in decreasing abundance order. We replace $\widehat{\mathbf{z}}_i$ with $\mathbf{z}_i^\star$ only when it strictly lowers the discrepancy. Accepted updates enter the background for later slots. For identification, we retain the N_o candidates with the lowest discrepancies for each slot, together with their molecular identities, for per-slot Top-1 and Top-10 evaluation.

Recovered composition and refinement cost. After the sweep, $\widehat{\mathbf{z}}_i$ denotes the support actually retained by slot i. Slots selecting the same anchor are merged by adding their abundances. The recovered supports and masses are

$$\widehat{\mathcal{Z}}_\mathbf{y} = \{\widehat{\mathbf{z}}_i \mid i \in \mathcal{I}_\mathbf{y}\}, \qquad \widehat{\alpha}(\mathbf{z}) = \sum_{i \in \mathcal{I}_\mathbf{y}} \bar{\alpha}_i \mathbb{I}[\widehat{\mathbf{z}}_i = \mathbf{z}]. \tag{36}$$

They define the final composition measure

$$\widehat{\mu}_\mathbf{y} = \sum_{\mathbf{z} \in \widehat{\mathcal{Z}}_\mathbf{y}} \widehat{\alpha}(\mathbf{z}) \delta_\mathbf{z}. \tag{37}$$

Consolidation preserves total mass and the reconstructed spectrum while combining duplicate supports into individual components.

The realized candidate spectra are available as rows of $\mathbf{S}_\psi$. For each candidate, refinement retrieves one row, adds its weighted spectrum to $\mathbf{b}_i$, and evaluates d_{sp}. A sweep examines $N_r|\mathcal{I}_\mathbf{y}|$ candidates, each requiring a lookup and operations over the length-L grid.

B Experimental Details

B.1 Datasets and Online Mixture Generation

Spectral sources and preprocessing. Stage 1 uses approximately 790k unique molecules from the infrared subset of the Multimodal Spectroscopic Dataset (Alberts et al., 2024b). Mixture recovery uses simulated spectra from QM9Spectra (QM9S) (Zou et al., 2023) and 5,228 measured gas-phase spectra from the NIST/EPA database (Stein, 2008). All spectra are resampled to $L = 1800$ points on Ω spanning 400–4000 cm^{-1}. Each nonzero resampled spectrum is peak-normalized as $\mathbf{s} = \widetilde{\mathbf{s}}/\|\widetilde{\mathbf{s}}\|_\infty$. NIST spectra are zero-filled beyond their measured intervals before normalization.

Data partitions and reference pools. QM9S is divided 8:1:1 into training, validation, and test partitions by molecular structure. Each partition supplies its own unlabeled pure-spectrum anchors, and its mixtures are sampled from those components. NIST compounds are divided into five folds. Each run fine-tunes the QM9S-trained estimator on mixtures constructed from four folds and evaluates on mixtures from the held-out fold; the NIST reference pool spans all five folds.

The pretraining corpus supplies examples for learning the encoder and realization field. Downstream reference pools supply the candidate pure spectra, which are encoded by that learned pathway. QM9S evaluation uses the held-out partition's reference spectra and NIST evaluation uses the shared five-fold pool. SpecMUSE and the reference-based baselines access these same chemical candidates: SpecMUSE evaluates the learned field at their coordinates, whereas fitting and retrieval baselines use their stored spectra. Molecular identities provide training annotations and evaluation labels; the reference spectra provide the candidate information. Appendix B.5 reports whole-collection structural intersections, which are distinct from the partition-specific reference protocol.

Measured pure spectra are much less numerous than the simulated collections. Their gas-phase absorbance nevertheless provides a basis for constructing mixtures: under dilute, non-reactive conditions, the Beer–Lambert law adds the absorbance of independently absorbing species (Swinehart, 1962; Mayerhöfer et al., 2020). The NIST mixtures therefore combine measured pure-component spectra through this additive model.

Online mixture generation. During data loading, we draw $K \in \{2, 3, 4\}$ with probabilities $(0.5, 0.3, 0.2)$ and select K distinct components from the relevant partition. A draw $\boldsymbol{\beta} \sim \mathrm{Dirichlet}(\mathbf{1}_K)$ determines their abundances and additive mixture:

$$\alpha_k = \alpha_{\min} + (1 - K\alpha_{\min})\beta_k, \qquad \mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{s}_k. \tag{38}$$

We use $\alpha_{\min} = 0.15$, giving $\sum_k \alpha_k = 1$ and $\alpha_k \geq \alpha_{\min}$. Component combinations and abundances vary across epochs. Mixtures are generated without added noise or baseline perturbations.

The estimator uses $K_{\max} = 4$ output slots, with active components selected through their predicted presence. The sampled K supplies annotations and groups the reported results by mixture size. Each α_k is a coefficient of a peak-normalized spectrum and therefore represents an abundance. Recovering a chemical concentration would additionally require the component's absorption strength and measurement calibration.

B.2 Baseline Implementations

Common protocol. Baselines use the data partitions and resampled mixtures described above. QM9S methods access the reference spectra associated with the corresponding partition, and NIST methods use the reference pool across its five folds. Activity thresholds are chosen on validation data, and outputs contain at most $K_{\max}$ components. Repeated identities are consolidated and retained abundances are normalized before evaluation. Candidate ranking follows each method's prediction: coefficients for NNLS and EB-gMCR, spectral discrepancy for the Transformer, cosine similarity for DeepCID, and refinement discrepancy for SpecMUSE.

NNLS. NNLS estimates a nonnegative combination of the reference spectra (Melle et al., 2026). Let $\mathbf{S}_{\mathrm{adm}} \in \mathbb{R}_+^{L \times M}$ contain the peak-normalized reference spectra as columns. For each mixture, we solve

$$\widehat{\mathbf{c}} = \arg\min_{\mathbf{c} \geq \mathbf{0}} \frac{1}{2} \|\mathbf{y} - \mathbf{S}_{\mathrm{adm}}\mathbf{c}\|_2^2. \tag{39}$$

Coefficients rank the molecular identities. A coefficient threshold selects active candidates, retaining the $K_{\max}$ largest if necessary. Their reference spectra supply the recovered component spectra for evaluation, and normalization of the retained coefficients gives their abundances.

EB-gMCR. We adapt EB-gMCR (Chang & Chen, 2025) to a single observed infrared spectrum using the reference library as its component matrix. The adaptation optimizes one nonnegative coefficient per library spectrum under the reconstruction objective and energy-based regularization. Coefficients determine the identity ranking; a validation-selected activity threshold and the $K_{\max}$ cap select the reported components. Their reference spectra are used for spectral evaluation, and their normalized coefficients provide the abundances.

Transformer. The Transformer baseline performs spectrum-to-set prediction from the mixture spectrum (Vaswani et al., 2017; Carion et al., 2020). Patches of width 20 are embedded with learned positional tokens. An encoder and a decoder with $K_{\max}$ learned queries predict a nonnegative spectrum, presence probability, and abundance score for each slot. Permutation-invariant spectral matching supplies the training correspondence.

Training supervises component spectra, binary presence, abundances, and additive mixture reconstruction. At inference, predicted presence selects the active components and their abundance scores are normalized. Comparison of their spectra with the reference collection using d_{sp} gives molecular identity rankings by increasing discrepancy.

DeepCID. The DeepCID baseline is a CNN retrieval adaptation of the component-detection approach originally developed for Raman mixtures (Fan et al., 2019). A shared one-dimensional convolutional encoder maps the mixture and reference spectra to normalized embeddings. A multi-positive contrastive objective aligns each mixture embedding with all its annotated components. Cosine similarity to the reference embeddings then ranks candidate identities.

NNLS estimates abundances within a validation-selected subset of the highest-ranked candidates. A coefficient threshold and the $K_{\max}$ cap determine the output set. Evaluation uses the selected reference spectra and the retained NNLS coefficients normalized to unit sum.

B.3 Evaluation Metrics

For J evaluated mixtures, let $\mathcal{G}_j$ and $\widehat{\mathcal{G}}_j$ be the annotated and predicted identity sets of mixture j. Predicted duplicates are consolidated before scoring. We write $K_j = |\mathcal{G}_j|$ and $K_\Sigma = \sum_{j=1}^{J} K_j$. Set F1 averages over mixtures; SIS, abundance error, and identification accuracy average over annotated components. Results are reported separately for each mixture size. NIST scores are summarized as means and standard deviations across the five held-out folds.

Set-level F1. Precision measures the fraction of predicted identities that belong to the annotated composition, while recall measures the fraction of annotated identities recovered:

$$P_j = \frac{|\widehat{\mathcal{G}}_j \cap \mathcal{G}_j|}{\max\{1, |\widehat{\mathcal{G}}_j|\}}, \qquad R_j = \frac{|\widehat{\mathcal{G}}_j \cap \mathcal{G}_j|}{|\mathcal{G}_j|}. \tag{40}$$

The per-mixture F1 and its mean across the evaluation set are

$$F_{1,j} = \frac{2P_jR_j}{P_j + R_j}, \qquad F_1 = \frac{1}{J}\sum_{j=1}^{J} F_{1,j}. \tag{41}$$

We set $F_{1,j} = 0$ when $P_j + R_j = 0$. Identity sets make the metric invariant to component ordering.

Spectral information similarity. SIS compares recovered and annotated spectral shapes (McGill et al., 2021). For nonnegative spectra $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^L$, smoothing by $\varepsilon > 0$ prevents logarithmic singularities, and normalization gives the intensity distributions

$$p_\ell = \frac{a_\ell + \varepsilon}{\sum_{r=1}^{L}(a_r + \varepsilon)}, \qquad q_\ell = \frac{b_\ell + \varepsilon}{\sum_{r=1}^{L}(b_r + \varepsilon)}. \tag{42}$$

Their symmetric spectral information divergence and corresponding similarity are

$$\mathrm{SID} = \sum_{\ell=1}^{L}(p_\ell - q_\ell)\log\frac{p_\ell}{q_\ell}, \qquad \mathrm{SIS}(\mathbf{a}, \mathbf{b}) = \frac{1}{1 + \mathrm{SID}}. \tag{43}$$

For each mixture, a bipartite assignment maximizes total SIS between annotated and predicted spectra. Let σ_{jk} be the similarity assigned to annotated component k, with zero assigned to an unmatched component. The reported score averages over all annotated components:

$$\mathrm{SIS} = \frac{1}{K_\Sigma}\sum_{j=1}^{J}\sum_{k=1}^{K_j}\sigma_{jk}. \tag{44}$$

This spectral assignment is used for evaluation; training uses the d_{sp} assignment in Equation 28.

Abundance MAE (Abund.). Abundance error uses the same SIS-maximizing assignment. Let α_{jk} be the annotated abundance and $\bar{\alpha}_{jk}$ the mass of its matched prediction, assigning $\bar{\alpha}_{jk} = 0$ to an unmatched component. We report

$$\mathrm{Abund.} = \frac{1}{K_\Sigma}\sum_{j=1}^{J}\sum_{k=1}^{K_j}|\bar{\alpha}_{jk} - \alpha_{jk}|. \tag{45}$$

The score measures relative spectral weights under spectral correspondence. As the mean annotated abundance is $1/K$, comparisons are made within each mixture size.

Top-1 and Top-10 molecular identification. For annotated identity g_{jk}, let $\mathcal{H}_{jk}^{(r)}$ contain the first r identities returned for its matched prediction. We set $h_{jk}^{(r)} = \mathbb{I}[g_{jk} \in \mathcal{H}_{jk}^{(r)}]$, with zero assigned to unmatched components. Accuracy in percent is

$$\text{Top-}r = \frac{100}{K_\Sigma} \sum_{j=1}^{J} \sum_{k=1}^{K_j} h_{jk}^{(r)}. \tag{46}$$

We report $r = 1$ and $r = 10$ from the candidate ranking associated with each spectrally matched prediction. This ranking is evaluated separately from the final recovered identity set used for F1; its first-ranked candidate need not be the identity retained in the recovered composition.

B.4 Implementation Details

We implement SpecMUSE in Python 3.10 and PyTorch 2.11.0+cu128 on Ubuntu 22.04.5 LTS. The random seed is 42 for Python, NumPy, and PyTorch. Table 6 lists the architecture and training hyperparameters for both stages.

Table 6: Hyperparameter configurations for the two stages of SpecMUSE.

Parameter	Stage 1	Stage 2
Spectrum length L	1800	1800
Patch / stride	20 / 10	20 / 10
Embedding dimension	768	768
Encoder layers / heads	6 / 12	6 / 12
Decoder layers / heads	4 / 8	4 / 8
Encoder / decoder FFN	2304 / 1536	2304 / 1536
Mask seed ratio r_m	40%	–
Maximum slots $K_{\max}$	–	4
Optimizer	AdamW	AdamW
Learning rate	5×10^{-4}	4×10^{-4}
Weight decay	10^{-5}	10^{-5}
Batch size / rank	256	128
Epochs	500	1000
Warmup ratio	0.10	0.05
Gradient clipping	1.0	1.2
Presence threshold τ_p	–	0.5
Refinement candidates N_r	–	50
Retained alternatives N_o	–	10

Computational resources and cost. Training uses Distributed Data Parallel on eight NVIDIA A100 GPUs with 40 GB of memory each, hosted on a server with two Intel Xeon Gold 5218 CPUs at 2.30 GHz. Stage 1 and Stage 2 take approximately 3 and 74 hours, respectively, totaling 77 hours. Stage 1 contains 47.53 million parameters, all trainable. Stage 2 contains 105.92 million parameters, of which 58.40 million are trainable. The per-sample computational costs are 19.81 and 593.55 GFLOPs, with latencies of 19.84 and 93.97 ms for Stage 1 and Stage 2, respectively.

B.5 Dataset Characteristics

Table 7 summarizes the chemical coverage of the Multimodal, QM9S, and NIST collections. The spectral analysis uses the peak-normalized $L = 1800$ grid described in Appendix B.1. The NIST/EPA release documents measured ranges of 450–$3966\,\mathrm{cm}^{-1}$ for its EPA spectra and 550–$3846\,\mathrm{cm}^{-1}$ for its NIST spectra (Stein, 2008); the shared model grid extends these intervals through zero filling.

Chemical coverage. Multimodal spans larger molecules than QM9S and NIST, with median atom counts of 40, 18, and 23, respectively. QM9S is concentrated around small molecules, while NIST has a broader size distribution (Table 7).

Table 7: Summary statistics and atom-count distributions for the infrared spectral datasets.

Metric	Multimodal	QM9S	NIST
Total entries	794,386	129,817	5,228
Unique canonical structures	789,255	123,720	5,072
Unique elements	10	5	14
Minimum recorded atom count	8	3	0
Atom-count Q1	31	16	18
Median atom count	40	18	23
Atom-count Q3	51	20	29
Atom-count P95	62	23	45
Maximum atom count	101	29	92
Dominant atom-count bin	41–80	11–20	21–40

Note. Statistics cover all entries. The recorded NIST minimum of 0 denotes unavailable structure metadata. Q1, Q3, and P95 denote percentiles 25, 75, and 95.

QM9S contains H, C, N, O, and F. Multimodal extends this set with P, S, Cl, Br, and I, while NIST additionally includes B, Si, Se, and Fe. NIST has the widest elemental coverage despite containing the fewest records.

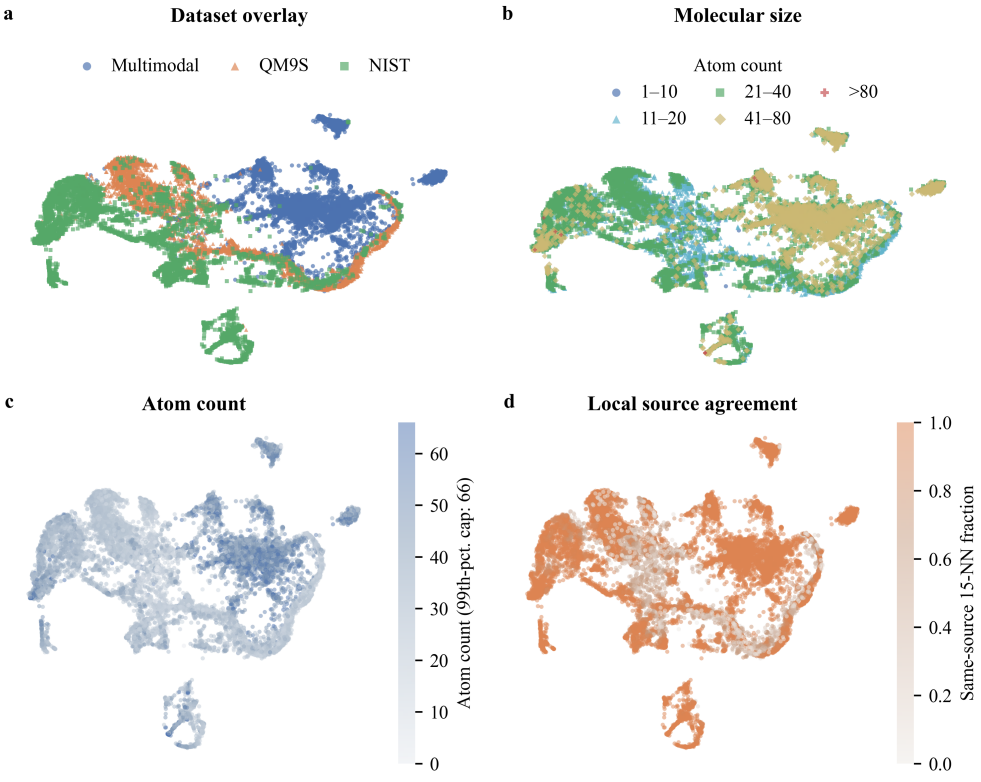

Figure 4: Joint embedding of normalized input spectra, colored by **(a)** dataset, **(b)** molecular-size category, **(c)** atom count with a capped color scale, and **(d)** the local fraction of neighbors from the same dataset.

Spectral distributions. Figure 4(a) shows a central region shared by the datasets and branches enriched in individual sources. Panels (b) and (c) show variation in molecular size, with similar atom counts appearing across different source-enriched regions. Panel (d) reveals more dataset-specific neighborhoods along these branches and more mixed neighborhoods near the center.

Neighborhood-preservation analysis supports the local organization shown by the embedding. Source classification in the principal-component representation also distinguishes the datasets effec-

tively, indicating that source-related differences remain in the spectral distributions after resampling and peak normalization.

B.6 Geometry and Chemical Organization of the Learned Manifold

We examine the local geometry of the learned spectral coordinates and their association with chemical annotations. After controlling for heavy-atom count, coordinate neighbors show greater functional-group Jaccard and Morgan-fingerprint Tanimoto similarity than random neighbors. This correspondence indicates that nearby spectral coordinates tend to represent chemically related molecules beyond similarities in molecular size.

Figures 5 and 6 show regional enrichment of carbonyl and alcohol O–H motifs, with overlap among categories. Local angular spread and functional-group similarity also vary across the coordinate space.

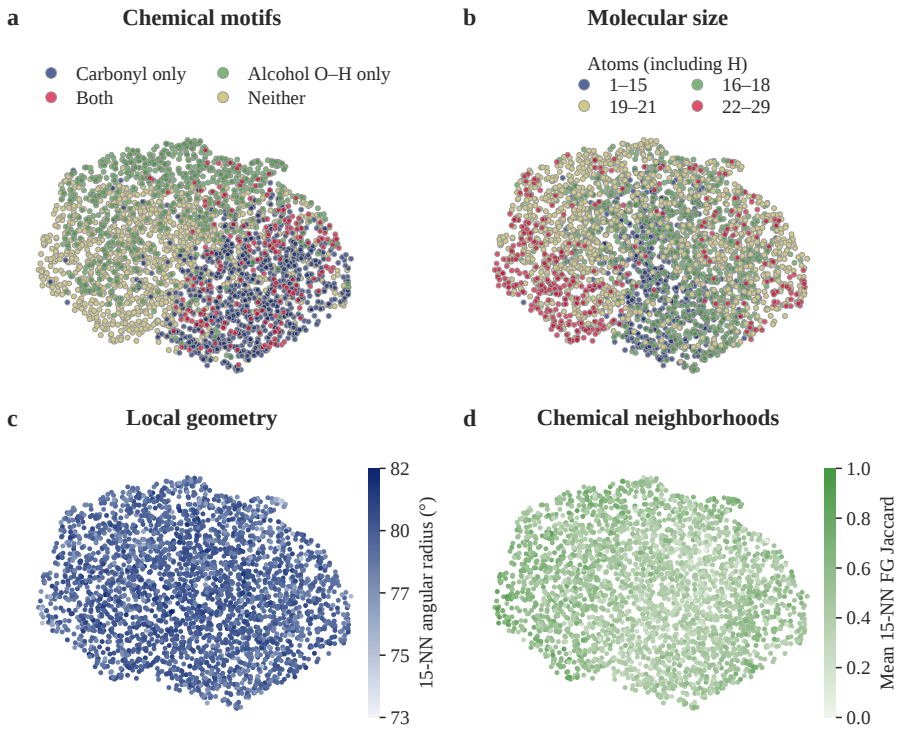

Figure 5: Planar UMAP view of learned spectral coordinates (McInnes et al., 2020). **(a)** Carbonyl and alcohol O–H motif membership. **(b)** Total atom count, including hydrogen. **(c)** Local angular neighborhood radius. **(d)** Mean functional-group Jaccard similarity within each neighborhood. Neighborhood statistics are computed before projection, excluding self-neighbors.

C Infrared Spectroscopy Background

This section connects molecular band patterns and gas-phase measurements to the additive absorbance model used for mixture construction.

C.1 Infrared Absorption and Molecular Vibrations

Infrared absorption records how molecular transitions attenuate radiation at different wavenumbers (Wilson et al., 1955; Colthup et al., 2012; Stuart, 2004). A molecule with N_a atoms has $3N_a - 6$

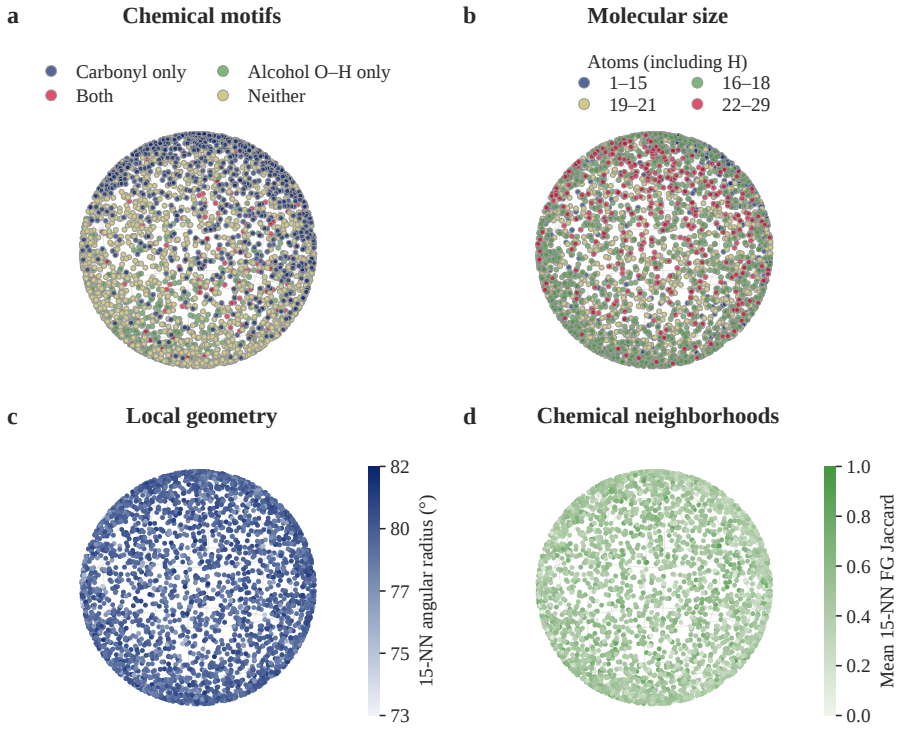

Figure 6: Spherical view of learned spectral coordinates. Panels retain the annotations, color scales, and original-space neighborhood statistics of Figure 5.

vibrational normal modes if nonlinear and $3N_{\mathrm{a}} - 5$ if linear. In the harmonic approximation, the mass-weighted Hessian determines their displacement vectors and frequencies:

$$\mathbf{H}_{\mathrm{mw}}\mathbf{e}_q = \lambda_q \mathbf{e}_q, \qquad \widetilde{\nu}_q = \frac{\sqrt{\lambda_q}}{2\pi c}. \tag{47}$$

Here $\lambda_q > 0$ is an eigenvalue after rigid translations and rotations have been excluded. If λ_q is expressed in s^{-2} and the speed of light c in $\mathrm{cm\,s}^{-1}$, the wavenumber $\widetilde{\nu}_q$ is in cm^{-1}.

An IR-active vibration changes the molecular electric dipole. In the harmonic electric-dipole approximation, the activity condition and integrated band intensity follow from the dipole derivative along normal coordinate Q_q:

$$\left.\frac{\partial \boldsymbol{\mu}}{\partial Q_q}\right|_{\mathrm{eq}} \neq \mathbf{0}, \qquad I_q \propto \left\| \left.\frac{\partial \boldsymbol{\mu}}{\partial Q_q}\right|_{\mathrm{eq}} \right\|_2^2. \tag{48}$$

Atomic masses and force constants influence band positions, while charge redistribution during vibration determines their intensities (Wilson et al., 1955; Colthup et al., 2012).

For a harmonic fundamental, absorption changes the vibrational quantum number by $\Delta v_q = +1$, and the downward transition has $\Delta v_q = -1$. Anharmonicity produces overtones, combination bands, and resonance interactions. Coupling between modes can shift their positions and redistribute intensity (Stuart, 2004; Griffiths, 1983). A pure spectrum therefore contains relationships among bands across the wavenumber range, which provide information for recovering a component whose local peaks overlap with others.

C.2 Spectral Regions and Band Interpretation

The mid-infrared range extends approximately from 4000 to $400\,\mathrm{cm}^{-1}$, with spectra commonly plotted as absorbance or transmittance against wavenumber (Stuart, 2004). The group-frequency region, approximately 4000–$1500\,\mathrm{cm}^{-1}$, contains characteristic stretching bands associated with O–H, N–H, C–H, C≡C, C≡N, C=O, and C=C groups (Colthup et al., 2012; Socrates, 2004). Their positions depend on the bonds and their molecular environment, including substitution, conjugation, hydrogen bonding, symmetry, and coupling.

The fingerprint region, approximately 1500–$400\,\mathrm{cm}^{-1}$, contains dense patterns of bending, skeletal, torsional, and coupled vibrations. These patterns reflect coordinated motion across the molecule. Compounds with similar functional groups may therefore share individual group-frequency bands while differing across the fingerprint region.

A band is characterized by its center, integrated intensity, width, and line shape. A schematic spectrum combines these contributions as

$$S(\widetilde{\nu}) = \sum_q I_q \phi_q(\widetilde{\nu}; \widetilde{\nu}_q, \Gamma_q). \tag{49}$$

Here $\widetilde{\nu}_q$ and Γ_q are the band center and width. Each line shape ϕ_q has unit area over wavenumber, so I_q denotes integrated intensity.

C.3 Gas-Phase Rovibrational Spectra

Gas-phase infrared spectra combine vibrational and rotational structure (Colthup et al., 2012; Griffiths, 1983). The schematic energy expression $E(v, J)/(hc) = G(v) + F_v(J)$ contains vibrational and rotational terms $G(v)$ and $F_v(J)$. Both are in cm^{-1} when hc uses the corresponding energy–length units. Here h is Planck's constant, and additional rotational quantum numbers are omitted.

Allowed transitions with $\Delta J = -1$, 0, and $+1$ form P, Q, and R branches, respectively, subject to molecular symmetry and the transition dipole. At high resolution, individual rotational lines may be resolved; at lower resolution, they form band envelopes. Linear, symmetric-top, and asymmetric-top molecules have different contours because their rotational constants and selection rules differ.

Thermal motion contributes approximately Gaussian Doppler broadening, while collisions contribute approximately Lorentzian broadening. A Voigt profile combines these effects. Temperature changes rotational-state populations and the intensity distribution within a band; pressure changes collision rates and broadening (Griffiths, 1983). Gas-phase spectra retain rotational, thermal, collisional, and instrumental effects while excluding solvent and crystal-lattice contributions.

C.4 Fourier-Transform Infrared Spectrometry

An FTIR spectrometer records an interferogram as a function of optical path difference and recovers the spectrum by Fourier transformation (Griffiths, 1983). In a Michelson interferometer, radiation travels along two paths that are recombined as one path length changes. Each wavenumber contributes a corresponding modulation to the interferogram.

The maximum optical path difference determines nominal resolution. Apodization windows the finite interferogram, trading spectral sidelobes against effective line width. Zero filling of the interferogram increases the displayed sampling density without increasing intrinsic resolving power. These choices affect peak widths and the resolution of adjacent bands.

For single-beam intensity, the instrumental response can be modeled as

$$I_{\mathrm{obs}}(\widetilde{\nu}) = [I_{\mathrm{true}} * h_{\mathrm{inst}}](\widetilde{\nu}) + b(\widetilde{\nu}) + \eta(\widetilde{\nu}). \tag{50}$$

Here h_{inst} is the instrumental line shape, b is the background, η is measurement noise, and $*$ denotes convolution (Griffiths, 1983; Chu et al., 1999).

The sample-to-background intensity ratio compensates for the source, optical path, and detector responses; absorbance is the negative logarithm of this ratio. This nonlinear transformation means intensity convolution does not generally become the same convolution in absorbance. Differences in atmospheric or optical-path conditions can leave residual water and carbon dioxide absorption.

C.5 Transmittance, Absorbance, and the Beer–Lambert Law

The incident intensity $I_0(\widetilde{\nu})$ and transmitted intensity $I(\widetilde{\nu})$ define transmittance and absorbance:

$$T(\widetilde{\nu}) = \frac{I(\widetilde{\nu})}{I_0(\widetilde{\nu})}, \qquad A(\widetilde{\nu}) = -\log_{10} T(\widetilde{\nu}). \tag{51}$$

The Beer–Lambert law relates absorbance to molar absorptivity $\varepsilon(\widetilde{\nu})$, concentration c, and path length ℓ (Mayerhöfer et al., 2020; Chu et al., 1999). For independent absorbers in a homogeneous gas mixture, absorbance adds at sufficiently narrow spectral resolution when scattering and intermolecular perturbations are negligible. The single-species and mixture expressions are

$$A(\widetilde{\nu}) = \varepsilon(\widetilde{\nu}) c \ell, \qquad A_{\mathrm{mix}}(\widetilde{\nu}) = \ell \sum_{k=1}^{K} c_k \varepsilon_k(\widetilde{\nu}). \tag{52}$$

Each species contributes at every wavenumber according to its concentration and absorptivity. Overlapping bands therefore combine into peaks that can contain signals from several components. This motivates the additive forward model in Section 3. In the normalized mixtures used here, masses describe the abundances defined in Appendix B.1.

Integrated absorbance over $\mathcal{B}$ is $\mathcal{A}_{\mathcal{B}} = \int_{\mathcal{B}} A(\widetilde{\nu})\,\mathrm{d}\widetilde{\nu}$. These band areas summarize absorption across resolved or partially resolved features and support quantitative gas-phase spectroscopy (Chu et al., 1999).

C.6 Reference Spectra and Library Identification

Library identification compares an unknown spectrum with spectra of chemically identified compounds (Linstrom & Mallard, 2001; Stein, 2008). Reference entries store chemical identities and the phase, spectral range, resolution, and conditions of acquisition.

Comparison requires a common spectral representation. Procedures align wavenumber ranges, sampling intervals, resolution, baseline, and intensity conventions, then rank candidates through peak positions, relative intensities, band shapes, derivatives, or full-spectrum distances. Agreement across both group-frequency and fingerprint regions informs the identity assignment.

The NIST Chemistry WebBook provides compound-resolved infrared spectra and chemical and physical-property data (Linstrom & Mallard, 2001). The NIST/EPA Gas-Phase Infrared Database used here contains 5,228 spectra in JCAMP-DX format from the EPA Vapor-Phase IR Library and NIST measurements (Stein, 2008). The separate NIST Quantitative Infrared Database provides absorption-coefficient spectra from primary gas standards at multiple resolutions and with multiple apodization functions (Chu et al., 1999).

D Additional Case Studies

Figures 7–12 show mixtures with two, three, and four components on QM9S and NIST. Component panels compare annotated and predicted spectra alongside molecular structures. Mixture panels compare the observed and reconstructed spectra and display their residuals. Abundance panels compare the annotated and predicted component abundances.

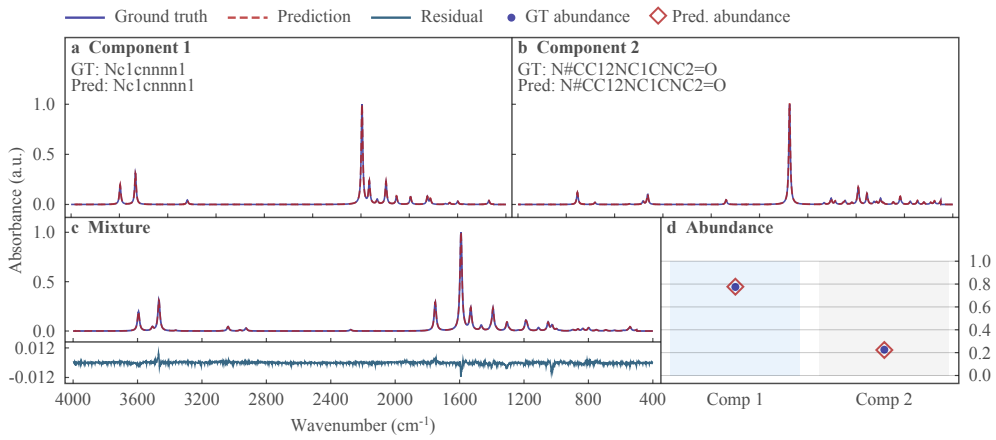

Figure 7: Case study of a binary QM9S mixture ($K = 2$).

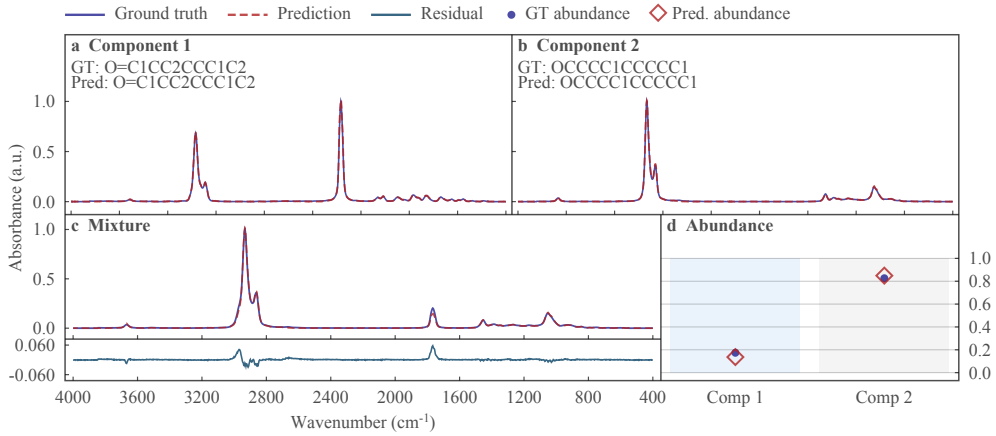

Figure 8: Case study of a binary NIST mixture ($K = 2$).

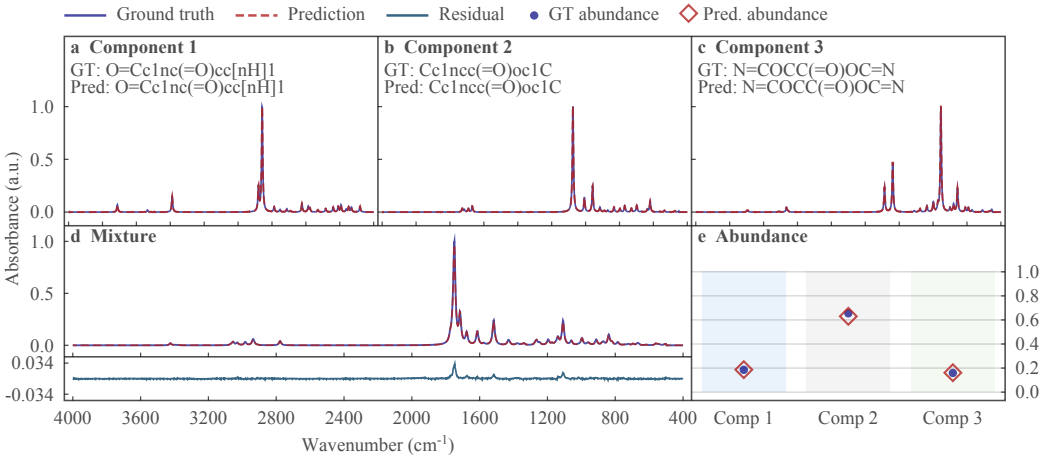

Figure 9: Case study of a ternary QM9S mixture ($K = 3$).

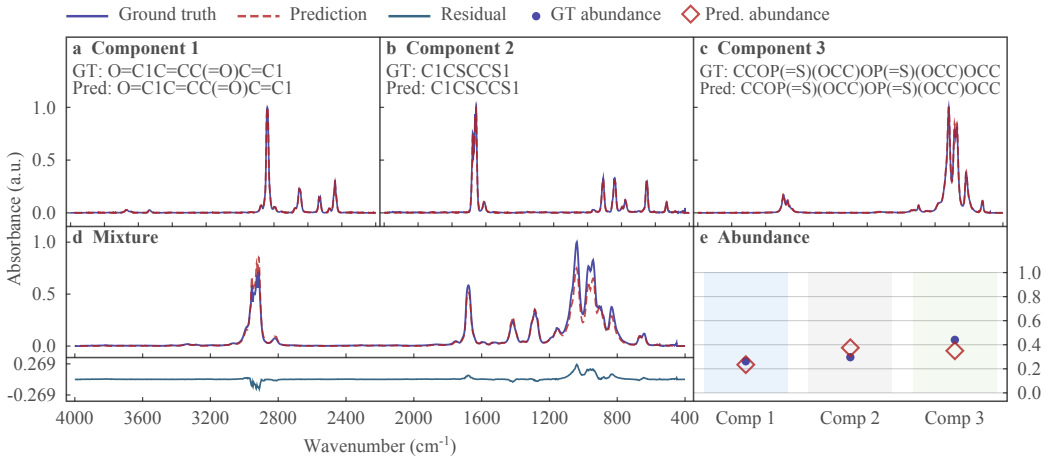

Figure 10: Case study of a ternary NIST mixture ($K = 3$).

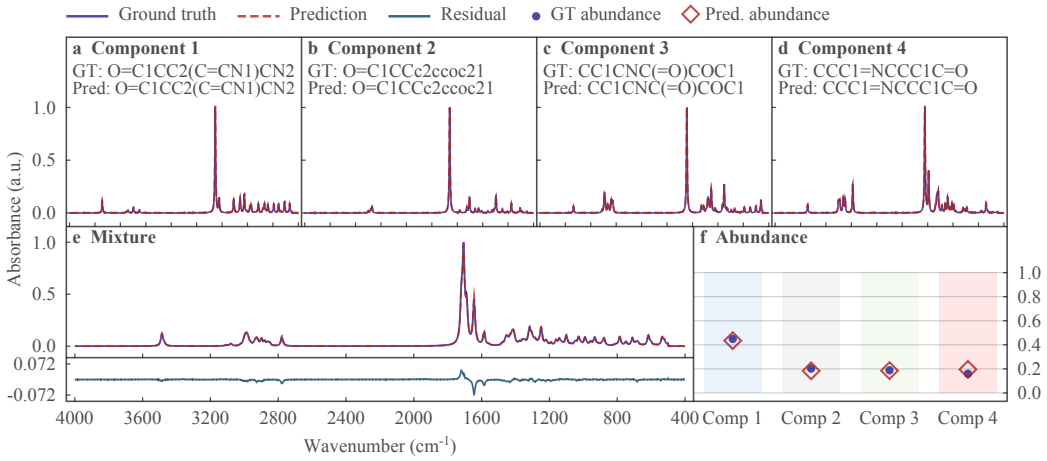

Figure 11: Case study of a four-component QM9S mixture ($K = 4$).

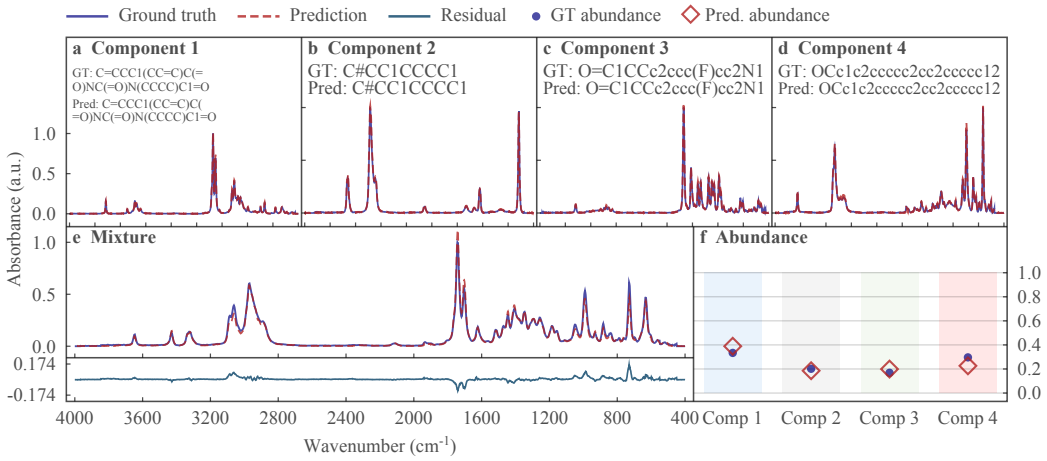

Figure 12: Case study of a four-component NIST mixture ($K = 4$).