Multiscale Residual Tokenization for Continuous Spectral Coordinates: Semantics and Physical Stability

Susanna Di Vita
EPFL
Securecell AG
susanna.divita@epfl.ch

Philippe Schwaller
EPFL
philippe.schwaller@epfl.ch

Abstract

Spectral coordinates are continuous, but many foundation-model pipelines require discrete vocabularies. We introduce **multiscale residual tokenization (MRT)**, which represents each coordinate by level-specific coarse-to-fine residual digits. On frozen NMR Sim→Exp retrieval, MRT improves ^{13}C Top-1 from 0.1363 to 0.2006, compared with 0.1071 for fixed-width quantization, and improves ^{1}H Top-1 from 0.0141 to 0.0281. We use controlled coordinate jitter to identify the perturbation scale at which this representation pays off. MRT reduces degradation relative to adaptive quantization, but the absolute comparison depends on perturbation scale: at the smallest tested jitter, post-perturbation kNN is tied on ^{1}H and IR and lower on ^{13}C; by jitter comparable to the finest residual scale, MRT overtakes adaptive on both NMR coordinates and marginally on IR. Using normalized jitter $\rho = \sigma/\Delta_{\text{fine}}$, at $\rho = 1.50$ on ^{1}H, $\rho = 1.00$ on ^{13}C, and $\rho = 1.00$ on IR, MRT is ahead by 0.0243, 0.0202, and 0.0032 kNN Tanimoto, respectively. Paired bootstrap intervals on the jitter degradation difference exclude zero at every tested NMR and IR jitter level. A packed-token ablation removes most of this advantage, supporting the role of explicit level-wise geometry. These results identify a tradeoff between clean semantic geometry and stability in spectral tokenization.

1 Introduction

Discrete tokenization fixes the geometry a downstream model sees [10, 13, 16, 21]. For spectral coordinates, an arbitrarily small displacement $x \mapsto x + \varepsilon$ can cross a flat quantization boundary and replace an entire categorical symbol even though the underlying physical measurement changes negligibly. High nominal resolution can be poorly aligned with the *scale* of physical variation [9, 11]. **Multiscale residual tokenization (MRT)** factorizes a coordinate into nested coarse-to-fine residual digits and retains those digits as level-specific features. A small displacement may change a fine component while preserving coarse components; progressively larger perturbations reach progressively coarser levels. Residual and hierarchical quantization are established in learned discrete representations and compression [21, 26]; MRT uses the hierarchy to expose the *physical scale at which a spectral coordinate changes*. Molecular spectroscopy provides a natural testbed for this question because spectral peak positions are continuous physical coordinates with experimental uncertainty, while molecular structure supplies an independent notion of semantic geometry. We study two complementary modalities: NMR chemical shifts and experimental IR band positions. Machine-learning approaches to molecular spectroscopy have used NMR sequence, set, grid, and continuous representations [1, 4, 12, 18, 22, 23, 25], while recent IR work has learned functional-group and molecular-structure representations directly from experimental infrared spectra [15, 17, 27]. These approaches primarily optimize predictive or semantic performance rather than the geometry induced by physically small coordinate changes. We isolate the tokenizer itself and avoid a large

40th Conference on Neural Information Processing Systems (NeurIPS 2026). Workshop: Representations for the Physical Sciences Workshop.

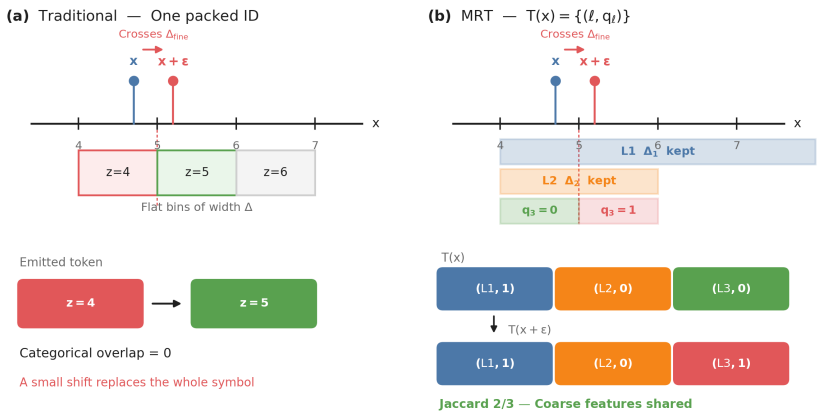

Figure 1: **Flat tokenization versus multiscale residual tokenization. (a)** A flat tokenizer assigns one categorical identity to each coordinate bin. A small displacement that crosses a quantization boundary replaces the entire token, so the original and perturbed representations have zero categorical overlap. **(b)** MRT represents the same coordinate using level-prefixed residual features, $T(x) = \{(\ell, q_\ell(x))\}_{\ell=1}^{L}$. When a perturbation crosses only the finest-scale boundary, the fine component changes while coarser components remain shared. In the three-level example, two of three level-specific features are preserved. MRT therefore makes representation change reflect the physical scale of the coordinate perturbation.

learned spectral encoder, whose architecture, objective, and capacity would confound a tokenization-level stability comparison. Across NMR and IR, the same tradeoff appears: MRT is weaker on clean molecular-neighborhood geometry, but becomes more competitive as coordinate uncertainty grows and overtakes adaptive quantization around the finest-scale noise regime. We use controlled perturbations to identify that regime and NMR Sim→Exp retrieval to test whether the same inductive bias helps under a natural domain shift.

Our contributions are: **(1) Multiscale spectral tokenization.** We introduce a level-prefixed residual representation for continuous spectral coordinates. **(2) Scale-dependent stability.** We show that level-ℓ flips are confined to boundary neighborhoods whose measure scales with $|\varepsilon|/\Delta_\ell$. **(3) Cross-modality validation.** The predicted clean–stability reversal and multiscale contraction appear on both NMR and experimental IR. **(4) Experimental transfer.** The same hierarchy improves NMR Sim→Exp retrieval, including a large ^{13}C Top-1 gain.

2 MRT, scale-dependent stability, and evaluation

Spectral coordinates and semantic geometry. A spectral coordinate is a measured peak position: a chemical shift in ppm for NMR or a band wavenumber in cm^{-1} for IR. Each spectrum enters the tokenizer as a variable-length list of these coordinates, with each position discretized into features and pooled to represent the molecule. By *semantic geometry* we mean whether nearby spectral representations correspond to structurally similar molecules. We assess this using the mean Morgan-fingerprint Tanimoto similarity of each molecule's five nearest spectral neighbors: Morgan fingerprints encode local molecular substructures, and Tanimoto measures their shared versus combined fingerprint bits [19]. Murcko-scaffold enrichment measures how much more often these neighbors share the same ring-and-linker molecular framework than expected from the dataset baseline [3].

Data and baselines. For NMR we use NMRGym peak lists for approximately 216k molecules with ^{1}H/^{13}C chemical shifts [7]. For IR we use 42,890 structure-resolved experimental spectra from the public IRexp/IRSpectra-Bench resource [24], after deduplication by InChIKey, with integer-valued band positions from 400 to 3998 cm^{-1}. Because the source positions are reported as discrete band locations on an approximately $1\,\mathrm{cm}^{-1}$ grid [24], $\sigma = 1\,\mathrm{cm}^{-1}$ jitter should be interpreted as roughly one source-grid step of coordinate uncertainty rather than sub-grid noise. Both evaluations use 4,000 held-out molecules, 5-NN Morgan-fingerprint Tanimoto, and Murcko-scaffold enrichment [3, 19].

We compare (i) fixed-width uniform quantization [11], (ii) quantile-based adaptive quantization [5], (iii) MRT, and (iv) continuous Fourier coordinates [20]. For IR, discrete baselines are closely rate-matched; NMR uses the frozen baseline configurations specified below. For NMR, the main tables retain frozen adaptive ($k = 512$) and fixed-width ($\Delta = 0.02/0.25$ ppm for ^{1}H/^{13}C) configurations. Since matching nominal bit rate does not control spatial resolution, Appendix H separately repeats the coordinate-jitter evaluation using adaptive and fixed single-codebook baselines with MRT's product vocabulary ($V = 450/520$, corresponding to $8.81/9.02$ bits/peak), together with two-phase coarse-grid ensembles at Δ_1. These controls separate fine-codebook rate from coarse-resolution stability; the packed-token ablation separately tests level-wise exposure for fixed residual digits.

Multiscale residual tokenization. Given resolutions $\Delta_1 > \Delta_2 > \cdots > \Delta_L$ and range $[x_{\min}, x_{\max}]$, MRT extracts digits by recursively quantizing the leftover residual:

$$q_\ell = \mathrm{clip}(\lfloor r_{\ell-1}/\Delta_\ell \rfloor, 0, V_\ell - 1), \qquad r_\ell = r_{\ell-1} - q_\ell \Delta_\ell, \tag{1}$$

with $r_0 = x - x_{\min}$, $V_1 = \lceil (x_{\max} - x_{\min})/\Delta_1 \rceil + 1$, and $V_\ell = \mathrm{round}(\Delta_{\ell-1}/\Delta_\ell)$ for $\ell > 1$. Algorithm 1 and a worked ^{1}H example appear in Appendix B. The exposed representation is the level-prefixed bag

$$T(x) = \{(\ell, q_\ell(x))\}_{\ell=1}^{L},$$

not the packed integer $z = \big((q_1 V_2 + q_2)V_3 + q_3\big)$. Thus $(q_1, q_2, q_3) \to (q_1, q_2, q_3')$ preserves the first two level-specific features rather than replacing the whole symbol. A reconstructed coordinate is $\hat{x} = x_{\min} + \sum_\ell q_\ell \Delta_\ell$.

We fix all scales *a priori*. NMR uses $(1.0, 0.1, 0.0333)$ ppm for ^{1}H and $(10, 1, 0.5)$ ppm for ^{13}C, spanning previously used chemical-shift resolution regimes [5, 14, 15]. IR uses $(100, 20, 4)\,\mathrm{cm}^{-1}$, with the finest level fixed relative to its $1\,\mathrm{cm}^{-1}$ measurement grid. For NMR, MRT and adaptive are nominally rate-matched to within 0.04 bits/peak: MRT/adaptive/fixed require $8.81/8.77/9.45$ bits/peak for ^{1}H and $9.02/8.99/9.97$ for ^{13}C. Unique quantiles yield effective adaptive alphabets of $k_{\mathrm{eff}} = 437$ and 510, respectively. Empirical token entropy is lower for MRT than adaptive on both nuclei (7.66 vs. 8.65 bits/peak for ^{1}H; 8.17 vs. 8.97 for ^{13}C), so the clean-geometry gap is not explained by MRT using a higher-rate code. For IR, adaptive uses $k = 925$ and gives 9.82 bits/peak versus 9.85 for MRT. No scale or tokenizer hyperparameter is selected using perturbation or transfer performance. MRT emits $L = 3$ level-specific features per coordinate; a literal token-sequence implementation therefore triples sequence length relative to a one-token-per-coordinate baseline, even when information rate is similar.

Scale-dependent stability. Let ∂_ℓ be the discontinuity set of level ℓ and define

$$d_{\mathrm{MRT}}(x, x+\varepsilon) = \sum_{\ell=1}^{L} w_\ell \mathbf{1}[q_\ell(x) \neq q_\ell(x+\varepsilon)],$$

where $w_\ell \geq 0$ weights the contribution of level ℓ.

Proposition 1 (Scale-confined flips)**.** *A level-ℓ digit can change only if $x \in N_{|\varepsilon|}(\partial_\ell)$. Under a uniform-within-cell residual phase,*

$$\Pr[q_\ell(x) \neq q_\ell(x+\varepsilon)] \approx \min\left(1, \frac{|\varepsilon|}{\Delta_\ell}\right), \qquad \mathbb{E}[d_{\mathrm{MRT}}] \approx \sum_{\ell=1}^{L} w_\ell \min\left(1, \frac{|\varepsilon|}{\Delta_\ell}\right). \tag{2}$$

Fine digits destabilize first, while coarser level-specific features can remain shared. The full proof appears in Appendix A. We normalize Gaussian coordinate jitter by $\rho = \sigma/\Delta_{\mathrm{fine}}$.

Evaluation. Semantic degradation is $\Delta\mathrm{kNN} = \mathrm{kNN}_{\mathrm{clean}} - \mathrm{kNN}_{\mathrm{pert}}$ and semantic retention is $\mathrm{SR} = \mathrm{kNN}_{\mathrm{pert}}/\mathrm{kNN}_{\mathrm{clean}}$. We additionally measure digit retention, bag Jaccard, reconstructed-peak similarity, global coordinate shifts, and peak dropout. Paired bootstrap 95% CIs are computed on identical held-out molecules for $\Delta\mathrm{kNN}_{\mathrm{adaptive}} - \Delta\mathrm{kNN}_{\mathrm{MRT}}$; positive values mean that MRT degrades less. Appendix C also reports per-method intervals on absolute post-jitter kNN.

3 Results and Discussion

The main comparison has two parts. First, frozen Sim→Exp retrieval tests a real domain shift and gives the clearest task-level gain. Second, controlled jitter identifies the perturbation regime in which

Table 1: **Clean semantics, absolute post-jitter performance, and transfer.** Post-jitter values are 5-NN Morgan-fingerprint Tanimoto. "Gain" is MRT minus adaptive at the stated ρ; bold marks the best absolute post-jitter score among Adaptive, MRT, and Fixed. Sim→Exp uses 1,849 paired NMR spectra.

A. Clean semantic geometry

Modality	Adaptive	MRT	Fixed
^{1}H	0.2545	0.1983	0.2492
^{13}C	0.2735	0.2228	0.2690
IR	0.1348	0.1296	**0.1419**

B. Absolute post-jitter kNN and crossover

Modality	ρ	Adaptive	MRT	Fixed	MRT–Adapt.
^{1}H	0.30	0.1867	0.1881	**0.1947**	+0.0014
	1.50	0.1432	**0.1675**	0.1432	**+0.0243**
^{13}C	0.20	**0.2415**	0.2138	0.2350	−0.0277
	1.00	0.1744	**0.1946**	0.1605	**+0.0202**
IR	0.25	0.1288	0.1286	**0.1371**	−0.0002
	1.00	0.1215	**0.1247**	0.1242	**+0.0032**

C. NMR Sim→Exp retrieval

Nucleus	Adaptive Top-1	MRT Top-1	Fixed Top-1	Adaptive MRR	MRT MRR
^{1}H	0.0141	**0.0281**	0.0206	0.0351	**0.0620**
^{13}C	0.1363	**0.2006**	0.1071	0.1985	**0.2656**

MRT's multiscale bias pays off. Relative degradation is useful for diagnosing stability, but we report absolute post-perturbation kNN as the primary comparison.

Clean geometry. On clean data, MRT trails adaptive by 0.0562 on ^{1}H and 0.0507 on ^{13}C; on IR the gap is only 0.0052, while fixed-width quantization is best at 0.1419. These clean scores show the cost of exposing coarse levels; MRT is not a uniformly better representation.

Crossover scale. The relative reductions in $\Delta\mathrm{kNN}$ are large at small jitter, but absolute post-perturbation performance gives the more useful comparison. At the smallest tested jitter, ^{1}H is effectively tied (0.1867 adaptive versus 0.1881 MRT), IR is tied (0.1288 versus 0.1286), and adaptive remains ahead on ^{13}C (0.2415 versus 0.2138); fixed-width leads both ^{1}H and IR at this jitter (0.1947 and 0.1371). As perturbation approaches the finest residual scale, the ordering changes. At $\rho = 1.50$ on ^{1}H, MRT reaches 0.1675 versus 0.1432 for adaptive; at $\rho = 1.00$ on ^{13}C, 0.1946 versus 0.1744; and at $\rho = 1.00$ on IR, MRT is marginally ahead at 0.1247 versus 0.1215. The gains are $+0.0243$, $+0.0202$, and $+0.0032$, respectively. This is the regime in which the multiscale bias improves absolute semantic geometry. On NMR, the corresponding post-jitter intervals (Appendix C) do not overlap at these ρ values.

The pattern persists at larger perturbations. Relative degradation reductions contract from 85.0% to 66.0% on ^{1}H, from 71.9% to 68.5% on ^{13}C, and from 82.7% to 57.4% on IR over the tested ranges. For NMR, the corresponding relative-reduction 95% CIs are $[82.6, 87.5]\%$, $[70.6, 74.2]\%$, and $[64.3, 67.7]\%$ for ^{1}H, and $[66.3, 76.9]\%$, $[69.5, 73.7]\%$, and $[66.7, 70.2]\%$ for ^{13}C. For IR, we report paired CIs on the absolute Δ difference rather than percentage CIs (Appendix D). The contraction matches Proposition 1: fine levels fail first, while coarser levels can remain shared after the finest scale is crossed.

Packed-token ablation. Packing (q_1, q_2, q_3) into one categorical ID removes the level-wise overlap. On IR, packed MRT is numerically equivalent to the matched $4\,\mathrm{cm}^{-1}$ fixed quantizer and becomes worse than adaptive for $\rho \geq 1.00$, whereas level-prefixed MRT yields reductions of $82.7\%/64.8\%/63.4\%/57.4\%$. On NMR, the headline reductions increase from $49.0\%/47.0\%$ with packed IDs to $85.0\%/71.9\%$ with explicit levels (Appendix E). The packed ablation holds the residual digits fixed, so the gain is from exposing those digits rather than from hierarchical coding itself. That is not a claim that hierarchy outperforms coarser resolution: two-phase Δ_1 grids are more jitter-stable than MRT (Appendix H). Independent level histograms can alias residual digits across distant coordinate regions (Appendix E).

Sim→Exp transfer. On 1,849 frozen paired spectra [6], MRT improves ^{13}C Top-1 from 0.1363 to 0.2006, compared with 0.1071 for fixed and 0.0719 for continuous coordinates. MRR rises from 0.1985 to 0.2656. On ^{1}H, Top-1 improves from 0.0141 to 0.0281 and MRR from 0.0351 to 0.0620, although absolute retrieval remains weak. The high cosine similarity between paired MRT representations (0.743 for ^{1}H and 0.860 for ^{13}C) is consistent with simulated and experimental coordinates differing at fine levels while retaining coarser residual features.

Global shifts and dropout separate related effects. Global coordinate shifts leave semantic kNN nearly unchanged for every representation, but MRT retains more level-specific symbols (e.g., 0.853 versus 0.357 on a 0.01 ppm ^{1}H shift). Dropout is weaker and less consistent: at $p = 0.40$, MRT reduces degradation by 36.8%, 20.8%, and 17.3% on ^{1}H, ^{13}C, and IR, respectively. The largest and most consistent effect remains coordinate perturbation.

4 Conclusion

MRT exposes an inductive bias: spectral coordinates should preserve coarse representation identity when only their finest physical scale changes. Across NMR and experimental IR, this yields a reproducible crossover from weaker clean geometry to stronger post-perturbation geometry, while frozen NMR Sim→Exp retrieval shows that the same bias is useful under a real domain shift. The packed-token ablation isolates the bags-versus-packed gap: preserving level identity, rather than hierarchical coding alone. Coarse Δ_1 grids can be more jitter-stable than MRT, so we do not claim that hierarchy, rather than resolution, drives robustness (Appendix H); residual-digit aliasing is reported in Appendix E. Our study is limited to extracted peak/band coordinates and deliberately avoids large learned encoders; rankings may change for full spectral traces or after large-scale pretraining. Within that scope, MRT provides a concrete design principle for scientific tokenization: encode the scale at which the measurement changes, not just its resolution. The full codebase will be released upon acceptance.

References

[1] Marvin Alberts, Federico Zipoli, and Alain C. Vaucher. Learning the language of NMR: Structure elucidation from NMR spectra using transformer models. In *NeurIPS Workshop on AI for Accelerated Materials Design (AI4Mat)*, 2023.

[2] C. F. Barnes. *Residual Quantizers*. PhD thesis, Brigham Young University, 1989.

[3] Guy W. Bemis and Mark A. Murcko. The properties of known drugs. 1. molecular frameworks. *Journal of Medicinal Chemistry*, 39(15):2887–2893, 1996. doi: 10.1021/jm9602928.

[4] Roman Bushuiev, Anton Bushuiev, Raman Samusevich, Corinna Brungs, Josef Sivic, and Tomáš Pluskal. Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS. *Nature Biotechnology*, 44:630–640, 2026. doi: 10.1038/s41587-025-02663-3.

[5] Thomas De Meyer, Davy Sinnaeve, Bjorn Van Gasse, Elena Tsiporkova, Ernst R. Rietzschel, Marc L. De Buyzere, Thierry C. Gillebert, Sofie Bekaert, José C. Martins, and Wim Van Criekinge. Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform. *Chemometrics and Intelligent Laboratory Systems*, 85(1): 144–154, 2007. doi: 10.1016/j.chemolab.2006.08.014.

[6] Susanna Di Vita. Sim → exp-MMNMR: A benchmark for simulation-to-experiment generalization in multimodal NMR with chemistry-aware metrics. In *AI for Accelerated Materials Design – NeurIPS 2025*, 2025.

[7] Zheng Fang, Chen Yang, Hai-tao Yu, Haoming Luo, Haitao He, Jiaqing Xie, Zhuo Yang, Yuqiang Li, and Jun Xia. NMRGym: A comprehensive benchmark for nuclear magnetic resonance based molecular structure elucidation. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2026.

[8] R. L. Frost, C. F. Barnes, and F. Xu. Design and performance of residual quantizers. In *Proceedings of the Data Compression Conference*, pages 129–138, 1991.

[9] Allen Gersho and Robert M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston, MA, 1992. doi: 10.1007/978-1-4615-3626-0.

[10] Robert M. Gray. Vector quantization. *IEEE ASSP Magazine*, 1(2):4–29, 1984. doi: 10.1109/MASSP.1984.1162229.

[11] Robert M. Gray and David L. Neuhoff. Quantization. *IEEE Transactions on Information Theory*, 44(6):2325–2383, 1998. doi: 10.1109/18.720541.

[12] Chao Han, Xiaolin Pan, and Yingkai Zhang. End-to-end molecular structure elucidation from multimodal NMR spectra images using vision transformers. *Chemical Science*, 17(26): 13072–13082, 2026. doi: 10.1039/D6SC02352E.

[13] Hervé Jégou, Matthijs Douze, and Cordelia Schmid. Product quantization for nearest neighbor search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):117–128, 2011. doi: 10.1109/TPAMI.2010.57.

[14] Stefan Kuhn, Björn Egert, Steffen Neumann, and Christoph Steinbeck. Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction. *BMC Bioinformatics*, 9:400, 2008. doi: 10.1186/1471-2105-9-400.

[15] Gwanho Lee, Hyekyoung Shim, Juhyun Cho, and Sang-Il Choi. Machine-learning approach to identify organic functional groups from FT-IR and NMR spectral data. *ACS Omega*, 10(12): 12717–12723, 2025. doi: 10.1021/acsomega.5c01903.

[16] Fabian Mentzer, David Minnen, Eirik Agustsson, and Michael Tschannen. Finite scalar quantization: VQ-VAE made simple. In *International Conference on Learning Representations*, 2024.

[17] Gyoung S. Na and Yecheol Rho. Explainable machine learning for characterizing unknown molecular structures in infrared spectra. *Analytical Chemistry*, 97(38):20869–20878, 2025. doi: 10.1021/acs.analchem.5c03126.

[18] Hiu-Lok Ngan, Jialing Zhang, Kenneth Kin-Leung Kwan, Jacinth Wing-Sum Cheu, Li Zhong, Yike Guo, Xian Yang, Carmen Chak-Lui Wong, Hong Yan, and Zongwei Cai. Resolution-adaptive binning enhances machine learning modeling by interbatch and multiplatform Orbitrap-based shotgun mass spectrometry data integration. *Analytical Chemistry*, 97(48):26877–26885, 2025. doi: 10.1021/acs.analchem.5c05874.

[19] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t.

[20] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[21] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[22] Fanjie Xu, Jinyuan Hu, Jingxiang Zou, Junjie Wang, Boying Huang, Zhifeng Gao, Xiaohong Ji, Weinan E, Zhong-Qun Tian, Fujie Tang, and Jun Cheng. Synergistic cross-modal learning for experimental NMR-based structure elucidation. *arXiv preprint arXiv:2602.08752*, 2026.

[23] Xi Xue, Hanyu Sun, Jingying Sun, Luc Patiny, Xiangying Liu, Kai Chen, Jingjie Yan, Liangning Li, Xue Liu, Shu Xu, Dongming Zhang, Yafeng Deng, Yingda Zang, Yaling Gong, Jie Ma, and Xiaojian Wang. NMRMind: A transformer-based model enabling the elucidation from multidimensional NMR to structures. *Analytical Chemistry*, 97(41):22603–22614, 2025. doi: 10.1021/acs.analchem.5c03783.

[24] Ilkham Yabbarov, Rudra Sondhi, and Rodrigo A. Vargas-Hernández. IRSpectra-Bench and IRexp: Candidate recall, not verification, limits LLM elucidation from real experimental IR and NMR. 2026.

[25] Liujia Yang, Zhuo Yang, Jiaqing Xie, Yubin Wang, Ben Gao, Tianfan Fu, Xingjian Wei, Jiaxing Sun, Jiang Wu, Conghui He, Yuqiang Li, and Qinying Gu. NMRTrans: Structure elucidation from experimental NMR spectra via set transformers. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2026. doi: 10.1145/3770855.3818935.

[26] Neil Zeghidour, Alejandro Luebs, Ahmed Omran, Jan Skoglund, and Marco Tagliasacchi. SoundStream: An end-to-end neural audio codec. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:495–507, 2022. doi: 10.1109/TASLP.2021.3129994.

[27] Colin Zhang and Yang Ha. Toward complete molecular structure prediction from infrared spectroscopy using deep learning. *Journal of Chemical Information and Modeling*, 66(1): 100–109, 2026. doi: 10.1021/acs.jcim.5c01834.

A Proof of Proposition 1

Let q_ℓ denote the level-ℓ residual quantizer with resolution Δ_ℓ, and let ∂_ℓ be its set of discontinuities [9, 11]. MRT exposes the hierarchy as

$$T(x) = \{(\ell, q_\ell(x))\}_{\ell=1}^{L}.$$

For any set A, write $N_\eta(A) = \{x : \mathrm{dist}(x, A) \le \eta\}$. A level-$\ell$ digit can change only if the interval between x and $x + \varepsilon$ crosses ∂_ℓ. Hence

$$q_\ell(x) \neq q_\ell(x+\varepsilon) \quad \Longrightarrow \quad x \in N_{|\varepsilon|}(\partial_\ell).$$

For a uniform residual quantizer, level-ℓ boundaries have spacing of order Δ_ℓ [9–11]. If residual phase is uniform within a cell,

$$\Pr[q_\ell(x) \neq q_\ell(x+\varepsilon)] \approx \min\left(1, \frac{|\varepsilon|}{\Delta_\ell}\right),$$

up to clipping and boundary effects. More generally, under bounded coordinate density this probability scales as

$$O\left(\min\left\{1, \frac{|\varepsilon|}{\Delta_\ell}\right\}\right).$$

If $|\varepsilon| < \Delta_k/2$ and x lies outside the boundary neighborhoods of all $\ell < k$,

$$q_1(x) = q_1(x+\varepsilon), \ldots, q_{k-1}(x) = q_{k-1}(x+\varepsilon),$$

so perturbations may change fine level-specific features while preserving coarser ones. This construction is related to classical multistage and residual quantization [2, 8, 26].

Finally, for

$$d_{\mathrm{MRT}}(x, x+\varepsilon) = \sum_{\ell=1}^{L} w_\ell \mathbf{1}[q_\ell(x) \neq q_\ell(x+\varepsilon)],$$

linearity of expectation gives

$$\mathbb{E}[d_{\mathrm{MRT}}] = \sum_{\ell=1}^{L} w_\ell \Pr[q_\ell(x) \neq q_\ell(x+\varepsilon)] \approx \sum_{\ell=1}^{L} w_\ell \min\left(1, \frac{|\varepsilon|}{\Delta_\ell}\right).$$

Larger perturbations reach coarser representation components. □

Comparison with flat and packed representations. A flat quantizer emits one categorical identity, so crossing one boundary replaces the entire symbol. Packing the residual tuple $(q_1, \ldots, q_L)$ into one categorical ID has the same problem: a fine-level change destroys whole-token equality even if all coarse digits survive. Level-prefixed MRT instead exposes those invariant components explicitly. The distinction is not the absence of boundaries, but whether the representation preserves *which physical scales did not change*.

B Residual digit extraction

As a numerical example, take a ^{1}H shift $x = 7.23\,\mathrm{ppm}$ with $(x_{\min}, x_{\max}) = (0, 14)$ and $(\Delta_1, \Delta_2, \Delta_3) = (1.0, 0.1, 0.0333)$, so $(V_1, V_2, V_3) = (15, 10, 3)$. Then $q_1 = \lfloor 7.23/1.0 \rfloor = 7$ and $r_1 = 0.23$; $q_2 = \lfloor 0.23/0.1 \rfloor = 2$ and $r_2 = 0.03$; $q_3 = \lfloor 0.03/0.0333 \rfloor = 0$. Hence $T(7.23) = \{(1, 7), (2, 2), (3, 0)\}$ and $\hat{x} = 7.20\,\mathrm{ppm}$. The same residual digits at a different coarse bin (e.g., $x = 3.23$) yield $\{(1, 3), (2, 2), (3, 0)\}$: the level-2 feature is reused across the axis. The reconstructed coordinate is $\hat{x} = x_{\min} + \sum_\ell q_\ell \Delta_\ell$.

Algorithm 1 Residual digit extraction

Require: coordinate x; widths $\Delta_1 > \cdots > \Delta_L$; range $[x_{\min}, x_{\max}]$
Ensure: digits $q_1, \ldots, q_L$ and bag $T(x)$

$r_0 \leftarrow x - x_{\min}$
$V_1 \leftarrow \lceil (x_{\max} - x_{\min})/\Delta_1 \rceil + 1$
for $\ell = 1$ to L **do**
  if $\ell > 1$ **then**
    $V_\ell \leftarrow \mathrm{round}(\Delta_{\ell-1}/\Delta_\ell)$
  end if
  $q_\ell \leftarrow \mathrm{clip}\big(\lfloor r_{\ell-1}/\Delta_\ell \rfloor,\, 0,\, V_\ell - 1\big)$
  $r_\ell \leftarrow r_{\ell-1} - q_\ell \Delta_\ell$
end for
return $T(x) = \{(\ell, q_\ell)\}_{\ell=1}^{L}$

Table 2: **NMR tokenizer rate.** Nominal rate is computed from alphabet size; empirical rate is token entropy $H(z)$ on the zero-training corpus.

Nucleus	Tokenizer	Nominal	Empirical H	Alphabet
^{1}H	MRT	8.814	7.660	$[15, 10, 3]$
	Adaptive	8.771	8.654	$k_{\text{eff}} = 437$
	Fixed	9.451	8.343	700 bins
^{13}C	MRT	9.022	8.169	$[26, 10, 2]$
	Adaptive	8.994	8.971	$k_{\text{eff}} = 510$
	Fixed	9.966	9.129	1000 bins

C NMR clean-data and perturbation diagnostics

Rate accounting. Nominal rate is $\log_2 V$ for flat tokenizers and $\sum_\ell \log_2 V_\ell$ for MRT; empirical rate is token entropy $H(z)$. MRT and adaptive differ by only 0.043 bits/peak for ^{1}H and 0.028 bits/peak for ^{13}C in nominal rate, while MRT has lower empirical entropy on both nuclei.

Table 3 shows that MRT is never the best clean representation. Because MRT and adaptive are nominally rate-matched and MRT has lower empirical entropy (Table 2), this gap is not attributable to MRT using more bits. This rate comparison does not control spatial resolution, which can independently explain differences in clean geometry and stability (Appendix H). The cost is especially pronounced on NMR, where exposing coarse levels adds shared structure that is useful under perturbation but less discriminative on trusted coordinates.

Coordinate-jitter analysis. The clean-data results above establish the semantic cost of MRT on trusted coordinates. We next report the corresponding NMR perturbation results, including absolute post-jitter kNN and paired bootstrap confidence intervals.

The relative reductions in ΔkNN are 85.0% $[82.6, 87.5]$, 72.4% $[70.6, 74.2]$, and 66.0% $[64.3, 67.7]$ for ^{1}H at $\rho = 0.30/1.50/3.00$, and 71.9% $[66.3, 76.9]$, 71.6% $[69.5, 73.7]$, and 68.5% $[66.7, 70.2]$ for ^{13}C at $\rho = 0.20/1.00/2.00$.

Global shifts. Semantic retention remains approximately one for every representation. Symbolic retention instead exposes the hierarchy. For ^{1}H shifts of 0.01, 0.05, and 0.10 ppm, MRT retains $0.853/0.489/0.632$ of level features versus $0.357/0.027/0.021$ for adaptive. For ^{13}C shifts $0.10/0.50/1.00$ ppm, MRT retains $0.901/0.483/0.631$ versus $0.620/0.066/0.024$. Non-monotonic retention at exact bin multiples reflects residual realignment rather than semantic change.

Dropout. Paired $\Delta_{\text{adaptive}} - \Delta_{\text{MRT}}$ estimates for ^{1}H are 0.0001 $[-0.0016, 0.0018]$, 0.0032 $[0.0010, 0.0054]$, and 0.0227 $[0.0198, 0.0256]$ at $p = 0.10/0.20/0.40$. For ^{13}C they are -0.0009 $[-0.0025, 0.0007]$, -0.0010 $[-0.0032, 0.0012]$, and 0.0122 $[0.0094, 0.0151]$. Small-dropout differences are therefore indistinguishable from zero in three of the four $p \leq 0.20$ comparisons, whereas both nuclei favor MRT at $p = 0.40$. At $p = 0.40$, these correspond to relative degradation reductions

Table 3: **Complete clean semantic comparison.** Intervals are paired-bootstrap 95% CIs for held-out 5-NN Morgan-fingerprint Tanimoto.

Modality	Representation	kNN Tanimoto	Scaffold enrich.
^{1}H	Adaptive	$0.2545\ [0.2497, 0.2591]$	$10.41\times$
	MRT	$0.1983\ [0.1949, 0.2019]$	$5.62\times$
	Fixed	$0.2492\ [0.2444, 0.2536]$	$10.03\times$
	Continuous	$0.2188\ [0.2148, 0.2228]$	$7.53\times$
^{13}C	Adaptive	$0.2735\ [0.2688, 0.2785]$	$12.33\times$
	MRT	$0.2228\ [0.2190, 0.2266]$	$8.03\times$
	Fixed	$0.2690\ [0.2642, 0.2738]$	$12.42\times$
	Continuous	$0.2290\ [0.2246, 0.2333]$	$8.96\times$
IR	Adaptive	$0.1348\ [0.1329, 0.1369]$	$6.31\times$
	MRT	$0.1296\ [0.1280, 0.1313]$	$4.89\times$
	Fixed	$0.1419\ [0.1397, 0.1440]$	$7.69\times$
	Continuous	$0.1255\ [0.1240, 0.1270]$	$4.27\times$

Table 4: **NMR coordinate-jitter analysis.** "Post" is clean kNN minus ΔkNN; brackets in the Post column are per-method bootstrap 95% CIs on the 4,000 held-out means, not paired MRT–adaptive intervals on the absolute gap. The final column instead reports paired bootstrap 95% CIs on $\Delta\mathrm{kNN}_{\mathrm{adaptive}} - \Delta\mathrm{kNN}_{\mathrm{MRT}}$; positive values mean that MRT degrades less.

Nucleus	Rep.	ρ	ΔkNN	Post kNN	SR	Paired $\Delta_{\mathrm{adaptive}} - \Delta_{\mathrm{MRT}}$ [95% CI]
^{1}H	Adaptive	0.30	0.0678	$0.1867\ [0.1836, 0.1898]$	0.734	–
	MRT	0.30	0.0102	$0.1881\ [0.1850, 0.1914]$	0.949	$0.0576\ [0.0546, 0.0608]$
	Fixed	0.30	0.0545	**0.1947** $[0.1913, 0.1980]$	0.781	–
	Adaptive	1.50	0.1113	$0.1432\ [0.1415, 0.1450]$	0.563	–
	MRT	1.50	0.0308	**0.1675** $[0.1653, 0.1698]$	0.845	$0.0806\ [0.0772, 0.0840]$
	Fixed	1.50	0.1060	$0.1432\ [0.1415, 0.1448]$	0.575	–
	Adaptive	3.00	0.1206	$0.1339\ [0.1326, 0.1353]$	0.526	–
	MRT	3.00	0.0410	**0.1573** $[0.1555, 0.1593]$	0.793	$0.0796\ [0.0761, 0.0830]$
	Fixed	3.00	0.1129	$0.1363\ [0.1348, 0.1377]$	0.547	–
^{13}C	Adaptive	0.20	0.0320	**0.2415** $[0.2372, 0.2461]$	0.883	–
	MRT	0.20	0.0090	$0.2138\ [0.2101, 0.2176]$	0.960	$0.0230\ [0.0206, 0.0254]$
	Fixed	0.20	0.0340	$0.2350\ [0.2306, 0.2393]$	0.873	–
	Adaptive	1.00	0.0991	$0.1744\ [0.1718, 0.1772]$	0.638	–
	MRT	1.00	0.0282	**0.1946** $[0.1917, 0.1977]$	0.874	$0.0709\ [0.0676, 0.0743]$
	Fixed	1.00	0.1085	$0.1605\ [0.1583, 0.1628]$	0.597	–
	Adaptive	2.00	0.1191	$0.1544\ [0.1523, 0.1565]$	0.564	–
	MRT	2.00	0.0375	**0.1853** $[0.1827, 0.1881]$	0.832	$0.0816\ [0.0783, 0.0847]$
	Fixed	2.00	0.1236	$0.1454\ [0.1435, 0.1472]$	0.540	–

of 36.8% on ^{1}H and 20.8% on ^{13}C. The dropout effect is weaker and less consistent than coordinate jitter.

D Experimental IR replication

The IR experiment contains 42,890 unique InChIKeys with integer band positions from 400 to $3998\,\mathrm{cm}^{-1}$, with mean 11.4 and median 9 bands per molecule. MRT uses $(100, 20, 4)\,\mathrm{cm}^{-1}$ with vocabularies $(37, 5, 5)$ and 9.85 bits/peak; adaptive uses $k = 925$ (9.82 bits/peak), fixed uses $\delta = 4\,\mathrm{cm}^{-1}$ (9.81 bits/peak), and continuous coordinates use 16 Fourier features. Evaluation uses the same 4,000-molecule protocol and molecular semantic metrics as NMR.

MRT digit retention remains $0.907/0.833/0.725/0.613$ across the same jitter sequence, demonstrating that coarse components remain shared after finer levels begin to change.

For global shifts of $2/4/8\,\mathrm{cm}^{-1}$, MRT level retention is $0.797/0.591/0.511$ while kNN SR remains $0.997/0.995/0.992$. For dropout, paired $\Delta_{\mathrm{adaptive}} - \Delta_{\mathrm{MRT}}$ is $0.0011\ [-0.0003, 0.0026]$ at $p = 0.10$,

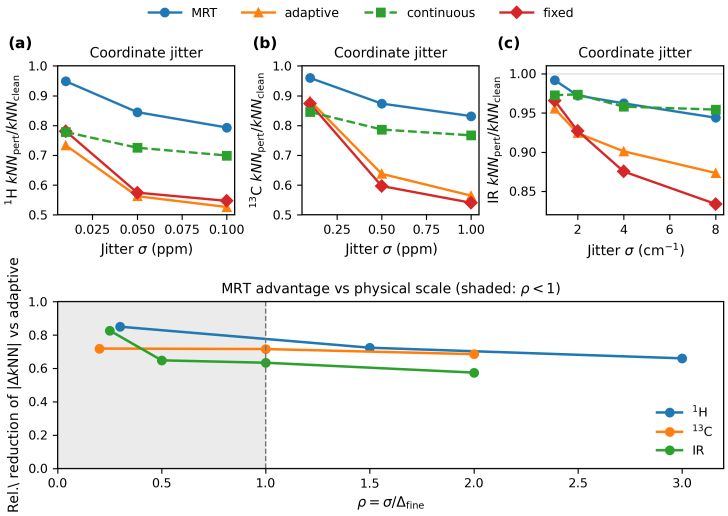

Figure 2: **Absolute and relative stability across perturbation scale.** The upper panels show absolute post-jitter kNN for ^{1}H, ^{13}C, and IR; the lower panel shows relative degradation reduction versus adaptive quantization. MRT becomes competitive around the finest-scale noise regime and retains more semantic geometry as perturbations increase.

Table 5: **IR coordinate-jitter replication.** IR source positions are integer-valued at $1\,\mathrm{cm}^{-1}$. The final column reports paired bootstrap 95% CIs on $\Delta\mathrm{kNN}_{\mathrm{adaptive}} - \Delta\mathrm{kNN}_{\mathrm{MRT}}$; IR percentage CIs are not reported. Post-kNN values are point estimates; we do not treat the $+0.0032$ absolute gap as established.

Rep.	ρ	ΔkNN	Post kNN	SR	Paired $\Delta_{\mathrm{adaptive}} - \Delta_{\mathrm{MRT}}$ [95% CI]
Adaptive	0.25	0.00600	0.12880	0.955	–
MRT	0.25	0.00104	0.12856	0.992	0.0050 [0.0030, 0.0069]
Fixed	0.25	0.00482	0.13708	–	–
Adaptive	0.50	0.01020	0.12460	0.924	–
MRT	0.50	0.00359	0.12601	0.972	0.0066 [0.0044, 0.0090]
Fixed	0.50	0.01030	0.13160	–	–
Adaptive	1.00	0.01333	0.12147	0.901	–
MRT	1.00	0.00489	0.12471	0.962	0.0084 [0.0062, 0.0106]
Fixed	1.00	0.01768	0.12422	–	–
Adaptive	2.00	0.01707	0.11773	0.873	–
MRT	2.00	0.00726	0.12234	0.944	0.0098 [0.0076, 0.0120]
Fixed	2.00	0.02356	0.11834	–	–

0.0011 $[-0.0006, 0.0028]$ at $p = 0.20$, and 0.0023 $[0.0003, 0.0044]$ at $p = 0.40$. Only the highest-dropout interval excludes zero; at $p = 0.40$ this is a 17.3% relative degradation reduction. IR reproduces the reduced-degradation jitter pattern and the weaker missing-peak effect; it does not establish an absolute post-jitter advantage.

E Packed-token ablation

The level-prefixed representation is necessary for the packed-versus-bags gap. If a residual tuple is instead packed into one ID,

$$(q_1, q_2, q_3) \mapsto z,$$

then (q_1, q_2, q_3') is a completely different symbol even though its coarse digits are unchanged.

IR provides a controlled demonstration because the packed MRT code is effectively isomorphic to its finest matched $4\,\mathrm{cm}^{-1}$ flat quantizer: both obtain clean kNN 0.1419, and their perturbation curves are numerically indistinguishable. Packed MRT gives only a non-significant 19.8% improvement at $\rho = 0.25$, no improvement at $\rho = 0.50$, and is worse than adaptive for $\rho \geq 1.00$. Exposing the levels changes these results to significant reductions of $82.7\%/64.8\%/63.4\%/57.4\%$.

NMR exhibits the same representation effect. At the headline jitter levels ($\rho = 0.30$ for ^{1}H and $\rho = 0.20$ for ^{13}C), packed MRT reduced degradation by 49.0% and 47.0%, respectively; the level-prefixed representation raises these to 85.0% and 71.9%. This ablation separates *hierarchical coding* from *hierarchical geometry*: the latter requires the downstream representation to preserve level identity. Independent level histograms do not bind residual digits to a parent bin, so distinct quantized peak lists such as $\{7.23, 3.41\}$ and $\{7.41, 3.23\}$ can share a bag. On the same 4,000 holdout, a residual re-pairing leaves the bag unchanged for 98.8% of ^{1}H and 99.9% of ^{13}C molecules, and the finest residual alphabet is saturated for $90.6\%/99.4\%$ of them; however, unpaired bags do not merge additional holdout molecules beyond packed IDs (3,960 distinct bags and packed codes on ^{1}H; 3,976 on ^{13}C). Thus residual unpairing is a combinatorial alias and a contributor to weak fine-scale localization, not a collapse of molecule identity.

F Sim→Exp details

The external experiment uses 1,849 paired simulated and experimental NMR peak lists [6], with all tokenizers frozen before evaluation.

Table 6: **NMR Sim→Exp retrieval.**

Nucleus	Representation	Top-1	Top-5	MRR	Cosine	Jaccard
^{1}H	Adaptive	0.0141	0.0411	0.0351	0.092	0.054
	MRT	**0.0281**	**0.0811**	**0.0620**	**0.743**	**0.596**
	Fixed	0.0206	0.0395	0.0384	0.093	0.055
	Continuous	0.0124	0.0476	0.0388	0.410	0.000
^{13}C	Adaptive	0.1363	0.2515	0.1985	0.201	0.122
	MRT	**0.2006**	**0.3250**	**0.2656**	**0.860**	**0.726**
	Fixed	0.1071	0.1898	0.1533	0.180	0.105
	Continuous	0.0719	0.1882	0.1342	0.568	~ 0

For ^{1}H, MRT improves Top-1 by $+0.0141$ (95% CI $[0.0054, 0.0227]$) and MRR by $+0.0269$ ($[0.0170, 0.0365]$); discordances are 50 MRT-only versus 24 adaptive-only Top-1 successes. For ^{13}C, Top-1 improves by $+0.0644$ ($[0.0449, 0.0849]$) and MRR by $+0.0670$ ($[0.0478, 0.0858]$), with 244 versus 125 exclusive successes.

Experimental ^{1}H peak lists contain integral-like integer artifacts. Without sanitization, MRT Top-1 is 0.0065 versus 0.0027 for adaptive. Even after sanitization, ^{1}H Top-1 is only 0.0281, so the transfer gain is real relative to adaptive but weak in absolute terms and depends on this preprocessing step.

G Leakage and fitting hygiene

- Adaptive bin edges are fit using training coordinates only.
- MRT widths are fixed before perturbation, IR robustness, or Sim→Exp evaluation.
- No tokenizer hyperparameter is selected using jitter, dropout, global shifts, or transfer performance.
- NMR and IR semantic evaluations use fixed held-out molecule sets.
- Paired MRT–adaptive intervals apply to the frozen-baseline degradation difference $\Delta\mathrm{kNN}_{\mathrm{adaptive}} - \Delta\mathrm{kNN}_{\mathrm{MRT}}$. Absolute post-jitter kNN intervals in Appendix C are per-method bootstrap CIs, not paired absolute-gap intervals.
- The rate-matched and coarse-grid NMR controls (Appendix H) are unpaired point estimates.
- Sim→Exp uses frozen tokenizers; ^{1}H sanitization is a documented preprocessing step rather than a tuned tokenizer component.

H Rate-matched NMR baselines

Bit-rate matching does not imply resolution matching. MRT uses 8.81 bits/peak on ^{1}H ($V = 450$) and 9.02 on ^{13}C ($V = 520$). The main-text adaptive ($k = 512$) and fixed ($\Delta = 0.02/0.25$ ppm)

baselines are frozen and do not control the spatial scale of MRT's coarse digits. We therefore repeat the 4,000-molecule jitter protocol of Appendix C with (i) single-codebook adaptive/fixed codes at MRT's product vocabulary and (ii) coarse grids at Δ_1 and Δ_2. These are unpaired point estimates; scaffold enrichment is omitted because the packed NMRGym cache has no SMILES.

Fine rate-matched codes. Adaptive@V uses at most V quantiles ($k_{\text{eff}} = 400/518$, $8.64/9.02$ bits). Fixed@V is a uniform grid with V cells ($\Delta = 0.0311/0.481$ ppm; $8.81/9.02$ bits). Both spend the bit budget on one fine codebook.

Coarse-grid controls. A single Δ_1 grid cannot spend MRT's bit budget (3.91 bits at 1.0 ppm; 4.70 bits at 10 ppm), so we use two phase-shifted Δ_1 grids: 1.0 ppm×2 on ^{1}H (7.81 bits) and 10 ppm×2 on ^{13}C (9.40 bits). Single Δ_2 grids (0.1 ppm / 1.0 ppm; $7.14/7.97$ bits) are additional resolution controls.

Table 7: **Fine single-codebook baselines at MRT's nominal bit rate.** Post is 5-NN fingerprint Tanimoto. Adaptive@450 has 400 unique quantiles. Point estimates; no paired-bootstrap intervals.

Nucleus	Rep.	bits	Clean	ρ	ΔkNN	Post	SR
^{1}H	Adaptive@450	8.64	0.2533	0.30	0.0637	0.1896	0.749
	Fixed@450	8.81	0.2445	0.30	0.0377	**0.2068**	0.846
	MRT	8.81	0.1983	0.30	0.0102	0.1881	0.949
	Adaptive@450	8.64	0.2533	1.50	0.1087	0.1446	0.571
	Fixed@450	8.81	0.2445	1.50	0.0946	0.1499	0.613
	MRT	8.81	0.1983	1.50	0.0308	**0.1675**	0.845
	Adaptive@450	8.64	0.2533	3.00	0.1186	0.1347	0.532
	Fixed@450	8.81	0.2445	3.00	0.1062	0.1383	0.566
	MRT	8.81	0.1983	3.00	0.0410	**0.1573**	0.793
^{13}C	Adaptive@520	9.02	0.2743	0.20	0.0346	0.2397	0.874
	Fixed@520	9.02	0.2655	0.20	0.0155	**0.2500**	0.942
	MRT	9.02	0.2228	0.20	0.0090	0.2138	0.960
	Adaptive@520	9.02	0.2743	1.00	0.1011	0.1732	0.631
	Fixed@520	9.02	0.2655	1.00	0.0821	0.1835	0.691
	MRT	9.02	0.2228	1.00	0.0282	**0.1946**	0.874
	Adaptive@520	9.02	0.2743	2.00	0.1180	0.1563	0.570
	Fixed@520	9.02	0.2655	2.00	0.1061	0.1594	0.600
	MRT	9.02	0.2228	2.00	0.0375	**0.1853**	0.832

Against these fine codes, Table 1B's crossover is unchanged: at $\rho = 1.50$ on ^{1}H, MRT reaches 0.1675 versus $0.1446/0.1499$ for adaptive@450/fixed@450; at $\rho = 1.00$ on ^{13}C, 0.1946 versus $0.1732/0.1835$. At the smallest ρ, MRT is tied with adaptive@450 on ^{1}H and behind both fine codes on ^{13}C.

Table 8: **Coarse-grid controls.** Two-phase Δ_1 ensembles are the primary resolution control; Δ_2 singles are included for completeness. Point estimates.

Nucleus	Rep.	bits	Clean	ρ	ΔkNN	Post	SR
^{1}H	MRT	8.81	0.1983	0.30	0.0102	0.1881	0.949
	1.0 ppm×2	7.81	0.1973	0.30	0.0000	0.1972	1.000
	0.1 ppm×1	7.14	0.2272	0.30	0.0096	**0.2176**	0.958
	MRT	8.81	0.1983	1.50	0.0308	0.1675	0.845
	1.0 ppm×2	7.81	0.1973	1.50	0.0064	**0.1909**	0.968
	0.1 ppm×1	7.14	0.2272	1.50	0.0520	0.1752	0.771
^{13}C	MRT	9.02	0.2228	0.20	0.0090	0.2138	0.960
	10 ppm×2	9.40	0.2539	0.20	0.0007	0.2532	0.997
	1.0 ppm×1	7.97	0.2631	0.20	0.0070	**0.2562**	0.974
	MRT	9.02	0.2228	1.00	0.0282	0.1946	0.874
	10 ppm×2	9.40	0.2539	1.00	0.0059	**0.2480**	0.977
	1.0 ppm×1	7.97	0.2631	1.00	0.0509	0.2122	0.807

The two-phase Δ_1 ensembles are more jitter-stable than MRT, as expected when $\sigma \ll \Delta_1$. On ^{1}H the clean scores are nearly identical (0.1983 vs. 0.1973), but at $\rho = 1.50$ the coarse ensemble reaches 0.1909 (SR 0.968) versus 0.1675 (SR 0.845) for MRT. On ^{13}C, 10 ppm×2 reaches 0.2480 at $\rho = 1.00$ (SR 0.977) versus 0.1946 (SR 0.874). So MRT beats a *fine* rate-matched codebook near $\rho \approx 1$, but not a coarse grid at a similar bit budget: coarseness, not residual structure alone, explains much of the raw retention. The packed ablation (Appendix E) is the complementary control: same

digits, different exposure. MRT is a stability–discriminability tradeoff against fine codes, not the most stable tokenizer at a given nominal rate.