DP-Splat: Bayesian Nonparametric Complexity Control for Gaussian Splatting

Anonymous authors
Paper under double-blind review

Abstract

3D Gaussian Splatting represents scenes as finite mixtures of anisotropic Gaussians whose number of components K is governed by heuristic adaptive density control or user-specified caps. Variational Bayes Gaussian Splatting (VBGS) recast splat fitting as conjugate variational inference over a finite mixture, but K remains fixed. We replace the finite symmetric Dirichlet over mixture weights with a truncated stick-breaking Dirichlet-process prior—and, as a theory-backed alternative, a sparse overfitted finite Dirichlet—so that the number of *occupied* components adapts to the data while every update remains a closed-form coordinate-ascent step (Beta/Categorical/Normal–Inverse–Wishart); a natural-gradient stochastic variant makes the per-step cost independent of the number of points. We give an exact monotonicity guarantee, a rigorous truncation-error bound that corrects an anti-conservative large-α approximation in common use, and an honest account of what the fitted number of components does and does not estimate. Empirically, (i) the effective complexity $\widehat{K}$ adapts to scene complexity and recovers the true K (within ± 1) on well-separated synthetic data with regime-appropriate concentration; (ii) a deconfounded three-way image comparison shows the DP prior's contribution is complexity *selection*, not per-component efficiency: converged DP fits exceed single-pass fixed-K VBGS at matched budgets by $+2.7$ dB on average yet tie an equally converged fixed-K baseline, while on 3D scenes DP-Splat matches or exceeds VBGS's held-out color prediction with 5.9–7.6$\times$ fewer components (single-pass fits at compute comparable to VBGS's already match with $\approx$5$\times$ fewer); (iii) the posterior-predictive color variance is well calibrated on model-matched synthetic data (regression-ECE $\approx 3 \times 10^{-3}$); and (iv) the ordering suggested by exact-posterior asymptotics reverses under mean-field coordinate ascent at practical N: the Dirichlet-process prior *resists* over-splitting while the sparse finite mixture saturates its truncation—a gap between variational practice and posterior asymptotics that we document across three orders of magnitude in N. Code, exact reproduction commands, and all experiment records accompany the submission.

1 Introduction

3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) represents a scene as a mixture of anisotropic 3D Gaussians and renders it with a differentiable rasterizer, achieving state-of-the-art novel-view synthesis at real-time frame rates. A central but under-examined design choice is the *number of components*: 3DGS grows and prunes Gaussians with adaptive density control—gradient-magnitude splitting and opacity thresholds—that works well in practice but is a stack of heuristics with tunable knobs and no statistical semantics. Follow-up work has made the dynamics more principled (e.g., MCMC-style relocation moves; Kheradmand et al., 2024) yet still requires a user-specified cap on the component count.

Variational Bayes Gaussian Splatting (VBGS; Van de Maele et al., 2024) reframed splat fitting as conjugate variational inference over a finite mixture on colored points: closed-form coordinate-ascent variational inference (CAVI) replaces stochastic gradients, enabling continual learning without replay. But the mixture is finite with K fixed in advance, uncertainty is per-component only, and VBGS inherits a component-recycling heuristic to revive dead components—adaptive density control through the back door.

In the statistics literature, letting a mixture choose its own complexity is a mature problem with three standard routes: Dirichlet-process (DP) mixtures with truncated variational inference (Ferguson, 1973; Sethuraman, 1994; Ishwaran & James, 2001; Blei & Jordan, 2006); sparse overfitted finite mixtures whose superfluous components empty asymptotically (Rousseau & Mengersen, 2011; Malsiner-Walli et al., 2016); and mixtures of finite mixtures with a prior on K (Miller & Harrison, 2018). None of this machinery has, to our knowledge, been applied to splatting-style scene representations.

This paper. We replace VBGS's finite symmetric Dirichlet over mixture weights with a truncated stick-breaking DP prior (primary variant) and a sparse overfitted Dirichlet (secondary, theory-backed variant), keeping the entire pipeline conjugate: every update remains a closed-form CAVI step over Beta, Categorical, Normal–Inverse–Wishart (and optionally Gamma, when the DP concentration α is learned) factors, and a natural-gradient stochastic variant (SVI; Hoffman et al., 2013) has per-step cost independent of the number of points N. The result is a splat-fitting procedure in which the number of *occupied* components $\widehat{K}$ is an output of posterior inference rather than an input.

Our contributions are deliberately statistical rather than benchmark-oriented; we validate claimed properties instead of chasing rendering leaderboards:

1. **Closed-form nonparametric splat fitting.** A "weight-prior switch" (dp / sparse_dir / dir) over an otherwise unchanged conjugate model (§3–§4), with exact CAVI (monotone to machine precision; Proposition 1) and an SVI variant that reproduces CAVI exactly in the full-batch limit and fits $N = 10^7$ points on a laptop CPU.

2. **A corrected truncation-error bound.** We state the rigorous finite-α truncation bound $2N(\alpha/(1+\alpha))^{T-1}$ and show that the widely quoted $4Ne^{-(T-1)/\alpha}$ approximation of Ishwaran & James (2001) is *anti-conservative* at $\alpha \approx 1$—it undershoots even the exactly evaluated Ishwaran–James bound by orders of magnitude in the regime where DP-Splat models actually operate (§5, Figure 3).

3. **An honest account of $\widehat{K}$.** Consistency theory says DP posteriors inflate the number of components (Miller & Harrison, 2014) while sparse overfitted mixtures empty them (Rousseau & Mengersen, 2011). We document that under CAVI *the ordering reverses*: the DP variant shows slow, log-like $\widehat{K}$ growth while the sparse variant saturates its truncation by $N = 10^6$ regardless of its concentration—not a contradiction of the exact-posterior theorems, but evidence that they are a poor guide to variational practice at these scales (§6, Figure 4). We trace both effects to coordinate-ascent dynamics (component death without revival) rather than to the priors, and show that learning α by variational Bayes tracks the local optimum instead of correcting it. $\widehat{K}$ is a useful, pragmatic complexity control; it is not a consistent estimator of a "true" K, and we say so.

4. **Calibrated color uncertainty for free.** The mixture-of-Students posterior predictive yields a closed-form conditional color variance $\mathrm{Var}[c \mid s]$ whose binned calibration error against realized squared error is $\approx 3{\times}10^{-3}$ on model-matched synthetic data (Figure 5), with a misspecified real-scene assessment reported alongside—uncertainty that fixed-precision baselines simply do not expose.

5. **A deconfounded account of matched-budget efficiency.** At the complexity the DP posterior *chooses*, converged CAVI exceeds single-pass fixed-K VBGS by $+2.7\,\mathrm{dB}$ on average—but ties an equally converged fixed-K baseline at the same $\widehat{K}$, isolating the prior's contribution to complexity selection and uncertainty rather than per-component efficiency. On 3D scenes DP-Splat matches or exceeds VBGS's held-out point-color prediction with 5.9–7.6$\times$ fewer effective components; a single-pass fit at compute comparable to VBGS's one-shot protocol already matches it with $\approx$5$\times$ fewer (§6).

Throughout, the likelihood is the mixture density on colored points (as in VBGS): rendering is a deployment map applied to the fitted posterior, not part of the model, and all theory is stated for the point mixture (§5).

2 Related work

Complexity control in splatting. 3DGS (Kerbl et al., 2023) manages its component budget by gradient-magnitude densification and opacity pruning—effective but untheorized heuristics whose thresholds are tuned per scene family. 3DGS-as-MCMC (Kheradmand et al., 2024) replaces cloning/splitting with probability-preserving relocation moves, giving the dynamics a sampling interpretation, but retains a user-specified cap on the number of Gaussians. BOGausS (Pateux et al., 2025) attaches variational-flavored confidence scores to densification decisions. VBGS (Van de Maele et al., 2024) is our direct foundation: it fits the mixture on colored point data with conjugate priors and closed-form CAVI, enabling continual learning, and has been extended to SLAM (Zhu et al., 2026); in all of these the mixture is finite with K fixed, and VBGS additionally relies on a reassignment heuristic that recycles dead components. A separate engineering line compacts *trained* 3DGS models—post-hoc importance pruning (Fan et al., 2024), learned masks (Lee et al., 2024), and Hessian-based uncertainty pruning (Hanson et al., 2025)—demonstrating that standard fits are heavily over-parameterized; these operate on a fixed-budget MLE fit after the fact, whereas we build parsimony into the prior and obtain it during inference, with an account of what the resulting component count means. Our work replaces exactly one ingredient of VBGS—the weight prior—and removes the need to choose K; notably, VBGS's own default concentration ($e_0 = 1/K$) already makes its finite Dirichlet "sparse" in the sense of Rousseau & Mengersen (2011), a fact we make explicit and exploit as a baseline variant.

Automatic complexity in mixtures. The three standard statistical routes are: (i) DP mixtures (Ferguson, 1973; Sethuraman, 1994) with blocked-Gibbs or truncated variational inference (Ishwaran & James, 2001; Blei & Jordan, 2006); (ii) sparse overfitted finite mixtures, where a concentration below a dimension-dependent threshold empties superfluous components asymptotically (Rousseau & Mengersen, 2011; Malsiner-Walli et al., 2016); and (iii) mixtures of finite mixtures with a prior on K, the route with consistent posterior inference on the number of components (Miller & Harrison, 2018; Frühwirth-Schnatter et al., 2021). The cautionary results of Miller & Harrison (2013; 2014)—DP and Pitman–Yor mixture posteriors are inconsistent for the number of components—anchor our discussion of what $\widehat{K}$ estimates. Posterior contraction of the mixing measure in Wasserstein distance is due to Nguyen (2013). Our contribution to this literature is not new asymptotics but a documented case study of how *variational* fitting dynamics interact with (and at practical sample sizes, dominate) these prior-level asymptotics.

Variational methods. We use textbook conjugate mean-field machinery (Bishop, 2006; Blei et al., 2017), stochastic natural-gradient updates (Hoffman et al., 2013; Amari, 1998), and the standard NIW posterior-predictive Student-t (Murphy, 2012); component emptying under variational Bayes for mixtures was observed as early as Ghahramani & Beal (2000). The order-dependence of the truncated stick-breaking family and size-based component reordering were identified by Kurihara et al. (2007); we deliberately implement the plain scheme and quantify its failure modes instead of patching them (§6).

Uncertainty for scene representations. Post-hoc or architectural uncertainty for radiance fields and splatting is an active area (e.g., Fisher-information view selection; Jiang et al., 2024); VBGS exposes per-component parameter uncertainty but fixes its color precision, so it has no predictive color variance. Our mixture-of-Students predictive gives calibrated per-point color variance as a by-product of the posterior (§6), which we validate with regression-style calibration curves (Kuleshov et al., 2018).

3 Model

Data model. Data are N colored points; in 2D-image mode each pixel is a point. Point n has a spatial coordinate $s_n \in \mathbb{R}^{D_s}$ (D_s=3 for scenes, 2 for images) and a color $c_n \in \mathbb{R}^{D_c}$ (D_c=3). Write $x_n = (s_n, c_n)$. Conditional on the assignment $z_n = k$ the two modalities are independent Gaussians,

$$s_n \mid z_n{=}k \sim \mathcal{N}(\mu_{k,s}, \Sigma_{k,s}), \qquad c_n \mid z_n{=}k \sim \mathcal{N}(\mu_{k,c}, \Sigma_{k,c}), \tag{1}$$

as in VBGS, whose likelihood we inherit unchanged except that we place a full conjugate prior on the color covariance where VBGS fixes its color precision (we retain its fixed variant behind a compatibility flag used

for cross-code regression tests, §6). The likelihood is the mixture density on points obtained from RGB-D or lifted depth; rendering—expected color conditional on projected location, with the standard rasterizer handling occlusion—is a deployment map, not part of the likelihood.

Component priors. Per modality $m \in \{s, c\}$, conjugate Normal–Inverse–Wishart:

$$\Sigma_{k,m} \sim \mathcal{IW}(\Psi_{0,m}, \nu_{0,m}), \qquad \mu_{k,m} \mid \Sigma_{k,m} \sim \mathcal{N}\big(m_{0,m},\, \Sigma_{k,m}/\kappa_{0,m}\big), \tag{2}$$

with defaults $m_{0,m}$ the data centroid, $\Psi_{0,m} = \widehat{\sigma}_m^2 I(\nu_0 - D_m - 1)$ with $\widehat{\sigma}_m^2$ the mean per-dimension variance, $\kappa_0 = 10^{-3}$, $\nu_0 = D_m + 2$.

Weight priors (the switch). *Variant A (primary), truncated stick-breaking DP:*

$$v_k \sim \mathrm{Beta}(1, \alpha)\ (k < T), \qquad v_T \coloneqq 1, \qquad \pi_k = v_k \textstyle\prod_{j<k}(1 - v_j), \tag{3}$$

where T is an inference-time truncation (the Gaussian budget for scenes), optionally with a conjugate $\alpha \sim \mathrm{Gamma}(a_0, b_0)$ ($a_0 = b_0 = 1$; flag-controlled). *Variant B (secondary), sparse overfitted finite mixture:* $\pi \sim \mathrm{Dir}(e_0, \ldots, e_0)$ with $K = T$ components and $e_0 \ll 1$; by Rousseau & Mengersen (2011), $e_0 < d/2$ (with d the per-component parameter dimension) makes superfluous components empty asymptotically. *Baseline:* the same Dirichlet with VBGS's shipped $e_0 = 1/K$. All three run through one switch on an otherwise identical model.

Variational family. Truncated mean field,

$$q(z, v, \theta, \alpha) \;=\; \prod_{n=1}^{N} q(z_n) \prod_{k=1}^{T-1} q(v_k) \prod_{k=1}^{T} \prod_{m} q(\mu_{k,m}, \Sigma_{k,m})\; q(\alpha), \tag{4}$$

with $q(z_n) = \mathrm{Cat}(r_{n,1:T})$, $q(v_k) = \mathrm{Beta}(\gamma_{k,1}, \gamma_{k,2})$, $q(\mu, \Sigma) = \mathrm{NIW}(m_k, \kappa_k, \Psi_k, \nu_k)$ per modality, and $q(\alpha) = \mathrm{Gamma}(w_1, w_2)$ when α is learned.

4 Inference

CAVI. With soft counts and weighted moments $N_k = \sum_n r_{nk}$, $\bar{x}_k = N_k^{-1} \sum_n r_{nk} x_n$, $S_k = \sum_n r_{nk}(x_n - \bar{x}_k)(x_n - \bar{x}_k)^\top$ (per modality), coordinate ascent cycles four exact updates until the relative ELBO change is below 10^{-6}:

(i) NIW (each modality independently):

$$\kappa_k = \kappa_0 + N_k, \quad m_k = \frac{\kappa_0 m_0 + N_k \bar{x}_k}{\kappa_k}, \quad \nu_k = \nu_0 + N_k, \quad \Psi_k = \Psi_0 + S_k + \tfrac{\kappa_0 N_k}{\kappa_0 + N_k}(\bar{x}_k - m_0)(\bar{x}_k - m_0)^\top. \tag{5}$$

(ii) Sticks (Variant A): $\gamma_{k,1} = 1 + N_k$, $\gamma_{k,2} = \mathbb{E}_q[\alpha] + \sum_{j>k} N_j$; the Dirichlet variants use $\alpha_k^{\mathrm{post}} = e_0 + N_k$.

(iii) Responsibilities via expected log-weights $\mathbb{E}[\log \pi_k] = \mathbb{E}[\log v_k] + \sum_{j<k} \mathbb{E}[\log(1 - v_j)]$ (digamma identities; Dirichlet form $\psi(\alpha_k^{\mathrm{post}}) - \psi(\sum_j \alpha_j^{\mathrm{post}})$) and the expected Gaussian log-density under the NIW posterior (Bishop, 2006, §10.2):

$$\log \rho_{nk} = \mathbb{E}[\log \pi_k] + \sum_m \Big(\tfrac{1}{2}\mathbb{E}[\log |\Lambda_{k,m}|] - \tfrac{D_m}{2} \log 2\pi - \tfrac{1}{2}\mathbb{E}\big[(x_{n,m} - \mu_{k,m})^\top \Lambda_{k,m} (x_{n,m} - \mu_{k,m})\big] \Big), \tag{6}$$

normalized with log-sum-exp.

(iv) Concentration (flagged): $w_1 = a_0 + T - 1$, $w_2 = b_0 - \sum_{k<T} \mathbb{E}[\log(1 - v_k)]$.

The ELBO is implemented one named function per term and each expectation is verified against Monte Carlo estimates from q; the full loop is verified against an independently written brute-force oracle and against the reference VBGS implementation (§6). Means are initialized by k-means++ seeding on a subsample; all other posteriors start at their priors.

Stochastic variant. For large N, the global parameters $\lambda \in \{\gamma, \text{NIW naturals}, (w_1, w_2)\}$ follow natural-gradient steps (Hoffman et al., 2013): a minibatch B yields intermediate estimates $\hat{\lambda}$ from sufficient statistics scaled by $N/|B|$, blended as $\lambda^{(t)} = (1-\rho_t)\lambda^{(t-1)} + \rho_t \hat{\lambda}$ with $\rho_t = (t+\tau_0)^{-\kappa_{\text{sched}}}$ ($\tau_0 = 64$; $\kappa_{\text{sched}} = 0.7$, distinct from the NIW parameter κ_k; $|B| = 2^{16}$). The NIW blend is performed in natural-parameter space $(\kappa,\ \kappa m,\ \nu,\ \Psi + \kappa m m^\top)$, in which the conjugate update is affine in the sufficient statistics; with $|B| = N$ and $\rho_t = 1$ the scheme reproduces full-batch CAVI to 10^{-8}–10^{-9} relative (unit-tested per parameter block). Per-step cost is independent of N: fitting $N = 10^7$ points takes 12 s on a laptop CPU (§6). The rank-one subtraction recovering Ψ from naturals requires float64 or centered data; we document this and all numerical safeguards in the repository.

Effective complexity and rendering. We report $\widehat{K} = \#\{k : N_k > n_{\min}\}$ with $n_{\min} = 1$ (sensitivity to $n_{\min} \in \{0.5, 1, 2, 5\}$ reported throughout; it never changes a conclusion) and the entropy-based count $\exp(-\sum_k \tilde{\pi}_k \log \tilde{\pi}_k)$, with $\tilde{\pi}_k$ the normalized expected weights. For rendering we map the posterior exactly as VBGS does: component weights $\mathbb{E}_q[\pi_k]$ (posterior expected stick weights for Variant A), splat shape from $\mathbb{E}[\Sigma_k] = \Psi_k/(\nu_k - D - 1)$, components with $N_k \le n_{\min}$ dropped; images use VBGS's pure-JAX expected-color renderer, so image-quality comparisons isolate the weight prior rather than the renderer.

Predictive distribution. Integrating (μ, Σ) over the NIW posterior gives a per-component Student-t, $p(x^*|k) = \mathrm{St}\big(m_k,\ \Psi_k \frac{\kappa_k+1}{\kappa_k \eta_k},\ \eta_k\big)$ with $\eta_k = \nu_k - D + 1$ (Murphy, 2012), and mean-field independence makes the predictive an exact mixture $\hat{p}(x^*) = \sum_k \mathbb{E}[\pi_k]\, \mathrm{St}_s(s^*|k)\, \mathrm{St}_c(c^*|k)$. Conditioning on location gives closed-form color moments

$$w_k(s^*) \propto \mathbb{E}[\pi_k]\, \mathrm{St}_s(s^* \mid k), \quad \mathbb{E}[c \mid s^*] = \sum_k w_k m_{k,c}, \quad \mathrm{Cov}[c \mid s^*] = \sum_k w_k \big(C_k + m_{k,c} m_{k,c}^\top\big) - \mathbb{E}[c \mid s^*]\mathbb{E}[c \mid s^*]^\top, \tag{7}$$

with $C_k = \tilde{S}_{k,c}\, \eta_{k,c}/(\eta_{k,c} - 2)$ the Student-t covariance, where $\tilde{S}_{k,c} = \Psi_{k,c} \frac{\kappa_k+1}{\kappa_k \eta_{k,c}}$ is the color Student-t scale (distinct from the scatter matrix S_k; finite covariance for $\eta_{k,c} > 2$, which also covers existence of the mean, $\eta_{k,c} > 1$; under the default prior $\eta_{k,c} = N_k + 3$, so the condition always holds.) All expectations are with respect to q; the factorized per-component predictive $\mathrm{St}_s \times \mathrm{St}_c$ reflects mean-field independence of the spatial and color blocks. A complete derivation is given in Appendix E; the closed form is additionally verified against independent numerical integration of the joint predictive (agreement $\le 10^{-11}$) and the implementation assembles the covariance in the positive-semidefinite law-of-total-variance form. These moments feed the held-out log-likelihood and calibration experiments.

5 Statistical properties

All statements are for the point-data mixture likelihood (1); the rasterized rendering map is downstream of inference and carries no guarantees here (§7).

Proposition 1 (Exact monotone coordinate ascent)**.** *Each of the four updates of §4 exactly maximizes the ELBO over its mean-field block given the others, hence $\mathcal{L}$ is nondecreasing along the CAVI cycle, which converges to a stationary point.*

Argument. Standard: the model is conjugate exponential-family in every block (Categorical, Beta, NIW, Gamma), so each coordinate update sets the block to the exact conditional maximizer (Bishop, 2006, §10.2); monotone convergence follows from coordinate ascent on a function bounded above. The numerical companion: across all Figure 1 runs (20 seeds $\times$ 4 variants, up to 200 iterations each) the worst observed relative ELBO decrease is -9.7×10^{-16}—a few float64 ulps of the ELBO magnitude—and a separate unit test runs the full 200 iterations $\times$ 20 seeds $\times$ 4 variants with no violation beyond 10^{-10} relative.

Proposition 2 (Truncation error, rigorous finite-α form)**.** *Let m_T and m_∞ denote the marginal densities of N observations under the truncated ($v_T := 1$) and full stick-breaking priors with concentration α, all else equal. Then*

$$\|m_T - m_\infty\|_1 \;\le\; 2\left(1 - \mathbb{E}\Big[\big(\textstyle\sum_{k<T} \pi_k\big)^N\Big]\right) \;\le\; 2N\left(\tfrac{\alpha}{1+\alpha}\right)^{T-1}. \tag{8}$$

Argument. The first inequality is due to Ishwaran & James (2001). For the second, $1 - x^N \leq N(1-x)$ on $[0,1]$ and $\mathbb{E}[1 - \sum_{k<T} \pi_k] = \mathbb{E}[\prod_{k<T}(1-v_k)] = (\alpha/(1{+}\alpha))^{T-1}$ by independence of the v_k. The relation to Ishwaran & James's widely quoted approximation $4Ne^{-(T-1)/\alpha}$ is as follows: since $\log(1+1/\alpha) \leq 1/\alpha$, the per-stick factor obeys $e^{-1/\alpha} \leq \alpha/(1{+}\alpha)$ for *every* $\alpha > 0$ (with near-equality as $\alpha \to \infty$, the gap being $O(\alpha^{-2})$), so the exponential form decays strictly faster in T than the rigorous factor; its constant-4 prefactor compensates only for small T (at $\alpha = 1$, $4Ne^{-(T-1)}$ falls below $2N\,2^{-(T-1)}$ for all $T \geq 4$). Consequently the exponential expression is *anti-conservative* in the small-α regime where splat models operate: at $\alpha \approx 1$ it undershoots even the exactly evaluated intermediate bound $2(1{-}\mathbb{E}[(\sum_{k<T}\pi_k)^N])$ by 3–6 orders of magnitude over the T we sweep (Figure 3, right), and it should not be quoted as a bound there. Figure 3 overlays that exactly evaluated intermediate bound (by Monte Carlo), the closed-form bound (8), and the exponential approximation. Practically, at scene budgets (T in the hundreds to thousands, $\widehat{K} \ll T$) the bound is astronomically small: truncation is a non-issue, and we observe held-out log-likelihood and $\widehat{K}$ to be flat in T for $T \gtrsim 2.5\,\widehat{K}$.

What $\widehat{K}$ estimates (and does not). Even exactly computed DP-mixture posteriors are inconsistent for the number of components: superfluous small clusters persist as $N \to \infty$ (Miller & Harrison, 2013; 2014). Sparse overfitted finite mixtures have the opposite asymptotics—superfluous components empty (Rousseau & Mengersen, 2011)—and mixtures of finite mixtures give the consistent route (Miller & Harrison, 2018; Frühwirth-Schnatter et al., 2021). Our position: $\widehat{K}$ under truncated CAVI is a *pragmatic complexity control*, not an estimator of a "true" K (ill-defined for real scenes in any case), and §6 shows that at practical N the fitting dynamics dominate these prior-level asymptotics—to the point of *inverting* the expected ordering between the DP and sparse variants. We view quantifying this gap between variational practice and posterior asymptotics as a contribution in itself; closing it (merge/split or reordering moves, Kurihara et al., 2007; or MFM-style inference) is future work.

Contraction of the mixing measure. For the fitted mixing measure $\widehat{G} = \sum_k \mathbb{E}[\pi_k]\,\delta_{(m_{k,s}, m_{k,c})}$ we report Wasserstein-2 distance to the true mixing measure on synthetic data (Figure 6); Nguyen (2013) gives posterior contraction rates for mixing measures in Wasserstein distance for exact posteriors, which we cite as context—no claim is made that CAVI attains them.

6 Experiments

Experiments validate claimed statistical properties; we do not tune for rendering benchmarks. All runs are config-driven scripts with fixed seeds; every reported number is written to a machine-readable record by the producing script. Hardware: all results in this paper were produced on a single laptop-class CPU (Apple M-series, float64); the stochastic variant makes this sufficient up to $N = 10^7$.

Verification protocol. Three independent layers guard correctness: (i) every update equation is tested against a brute-force NumPy oracle written directly from the model equations, without reference to the JAX implementation (tolerances 10^{-7}–10^{-12}, parameter-dependent); (ii) every ELBO term is tested against Monte Carlo estimates from q (10^5 samples for scalar terms, SE-calibrated tolerances); (iii) a cross-code regression starts from the reference VBGS implementation's own initialized model and verifies that one step of our CAVI (baseline configuration: Dirichlet weights with VBGS's $e_0 = 1/K$, fixed color precision) reproduces one step of VBGS's `fit_gmm` across all posterior parameters to 10^{-6}–10^{-8} relative (parameter-dependent). The suite comprises 47 tests.

6.1 Exactness and convergence (Figure 1)

Across 20 seeds × 4 variants (dp, dp(learn), sparse_dir, dir) on colored 2D mixtures (K_{true}=10, N=10^4, T=30), the ELBO is monotone to machine precision (worst relative decrease -9.7×10^{-16}, on one seed of one variant; the other three variants show exactly zero violations). DP variants converge in 5–16 iterations versus 30–73 (median ≈38) for the Dirichlet variants. About 10% of DP seeds settle in a visibly lower local optimum—CAVI is a local method and we report this rather than reseeding it away.

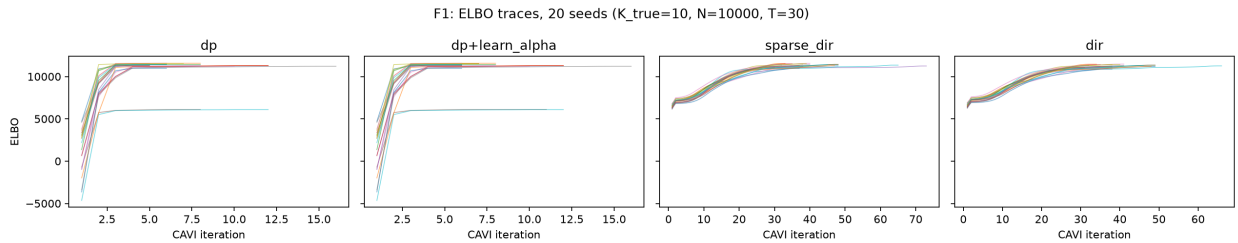

Figure 1: ELBO traces, 20 seeds per panel. All traces are monotone (worst relative decrease -9.7×10^{-16}). DP variants converge 3–5× faster in iteration count; dp(learn) (labeled dp+learn_alpha in the legend) is visually identical to dp. Two DP seeds per panel (10%) converge to a lower local optimum.

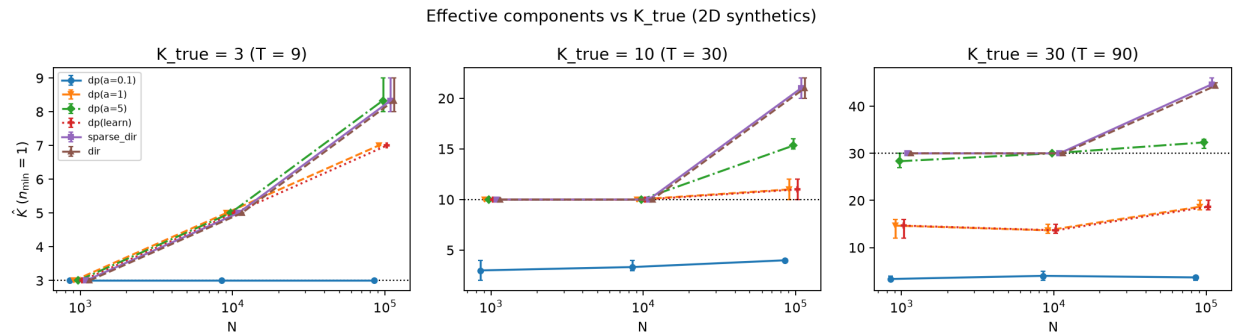

Figure 2: $\widehat{K}$ ($n_{\min}$=1) vs. N for $K_{\text{true}} \in \{3, 10, 30\}$ (dotted line), $T = 3K_{\text{true}}$. Bars: seed min/max (n=3); series are offset horizontally for visibility. dp(learn) coincides with dp(α=1) and sparse_dir with dir (that coincidence is itself a finding; see text).

6.2 $\widehat{K}$ recovery is real but concentration-regime-dependent (Figure 2)

On well-separated colored 2D mixtures with $K_{\text{true}} \in \{3, 10, 30\}$, $N \in \{10^3, 10^4, 10^5\}$, $T = 3K_{\text{true}}$: with a regime-appropriate fixed α, the DP variant recovers K_{true} within ± 1—and often exactly—in every regime (α=0.1 $\leftrightarrow$ K_{true}=3, exact at all N including 10^5; α=1 $\leftrightarrow$ 10; α=5 $\leftrightarrow$ 30 at N=10^4). No single fixed α covers all regimes; small α over-merges large-K data and large α over-splits small-K data at large N. Learning α variationally does *not* escape this: dp(learn) tracks dp(α=1) cell-for-cell—the α-posterior adapts to the local optimum the fit is already in (mode-following), a known but rarely quantified VB pathology. The threshold $n_{\min}$ is immaterial: $\widehat{K}$ is identical across $n_{\min} \in \{0.5, 1, 2, 5\}$ in 159 of 162 fits and differs by at most 1 in the remainder. The components driving over- or under-estimation carry substantial mass—they are not threshold artifacts—and no cell's conclusion changes.

6.3 Truncation is a non-issue—with the right bound (Figure 3)

Sweeping $T \in \{10, 25, 50, 100, 200\}$ at K_{true}=10, N=2×10^4, α=1: held-out log-likelihood and $\widehat{K}$ are flat for $T \geq 25$ ($\widehat{K} = 10.3 \pm 0.5$); at $T = K_{\text{true}}$ two of three seeds land in degraded local optima—an optimization effect, not truncation error. The right panel compares the exactly evaluated intermediate bound $2(1 - \mathbb{E}[(\sum_{k<T} \pi_k)^N])$ with the two closed-form expressions of Proposition 2: the rigorous form $2N(\alpha/(1+\alpha))^{T-1}$ dominates the exact curve as required, while the popular $4Ne^{-(T-1)/\alpha}$ approximation lies *below* that exactly evaluated quantity by 3–6 orders of magnitude at α=1—anti-conservative exactly where these models live.

6.4 The variational inversion of the $\widehat{K}$ asymptotics (Figure 4)

Asymptotic theory predicts DP posteriors *inflate* the number of components (Miller & Harrison, 2014) while sparse overfitted mixtures *empty* them (Rousseau & Mengersen, 2011). Under truncated CAVI on 3D colored

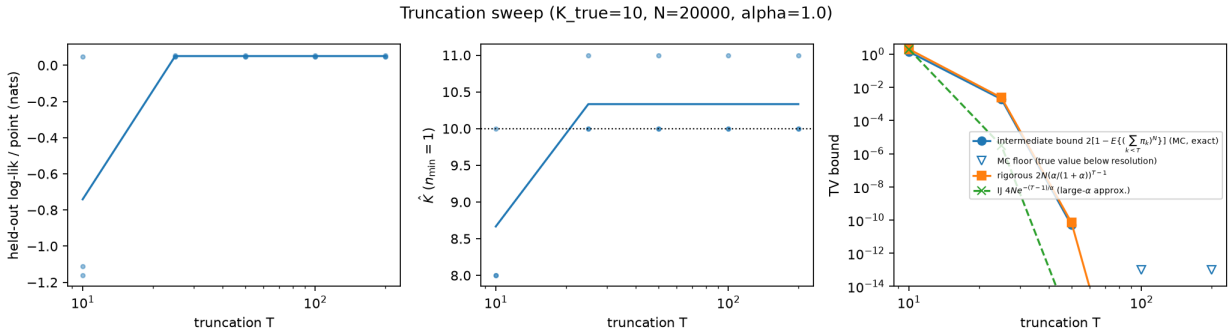

Figure 3: Truncation sweep. Left/middle: held-out log-likelihood and $\widehat{K}$ vs. T (individual seeds shown; T=K_{true} exhibits bimodal local optima). Right: the exactly evaluated intermediate bound (MC; censored below MC resolution), the closed-form bound of Proposition 2, and the anti-conservative large-α approximation.

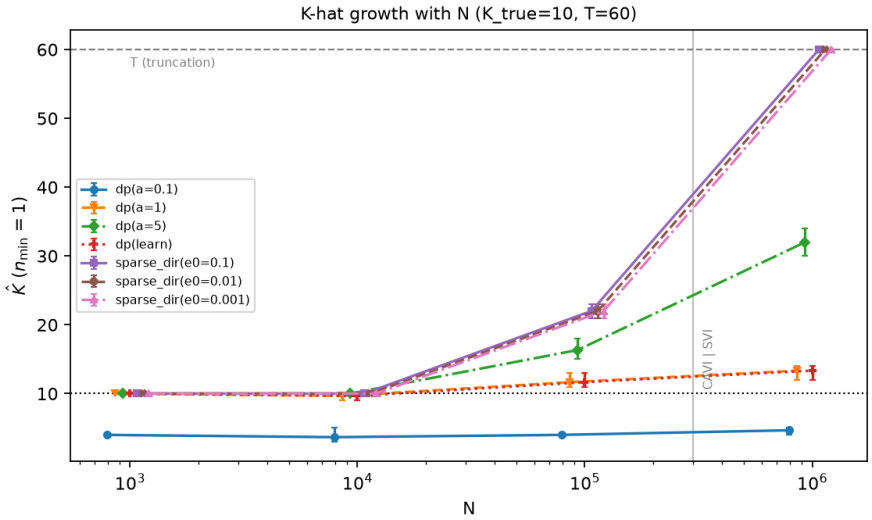

Figure 4: $\widehat{K}$ growth with N (3D, K_{true}=10, T=60; CAVI for $N \leq 10^5$, SVI beyond; bars: seed min/max). The sparse finite mixture saturates T by N=10^6 for all three e_0; the DP variants grow slowly. This inverts the naive reading of the asymptotic theory (see text).

mixtures (K_{true}=10, T=60, N from 10^3 to 10^6) we observe the opposite ordering: dp(α=1) grows slowly ($10 \to 13.3$ over three decades of N, consistent with log-like inflation), while the sparse variant saturates the *entire* truncation ($\widehat{K} = 22$ at 10^5, $60 = T$ at 10^6)—identically for $e_0 \in \{0.1, 0.01, 0.001\}$. The e_0-insensitivity is diagnostic: the posteriors do differ by exactly the e_0 perturbation while $\widehat{K}$ does not, so the over-splitting is driven by likelihood-dominated coordinate-ascent dynamics (splitting a dense cluster costs almost nothing at large N, and CAVI has no merge moves), not by the prior. Conversely, the DP's size-biased stick ordering suppresses late components early—which resists over-splitting at large N but also under-recovers when α is too small for the regime (dp(α=0.1) merges to $\widehat{K} \approx 4$ here). At these sample sizes, *fitting dynamics dominate prior asymptotics.*

6.5 Calibrated color uncertainty (Figure 5)

On 3D synthetics with overlapping clusters (3σ separation), so that predictive color uncertainty genuinely varies in space through component ambiguity, we bin held-out points by predicted $\operatorname{tr}\operatorname{Cov}[c \mid s^*]$ (10 quantile

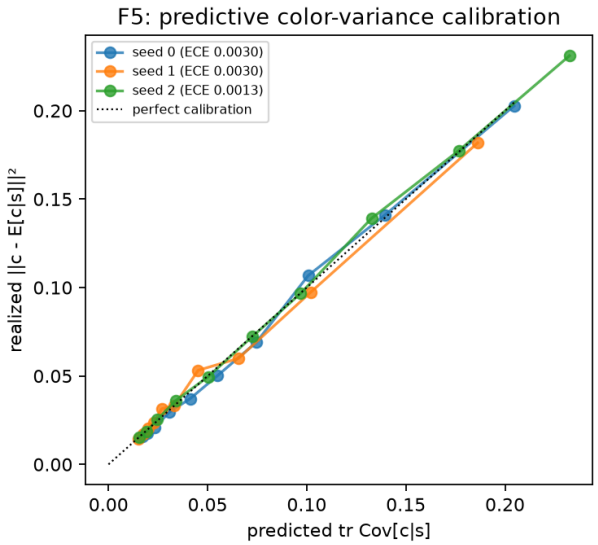

Figure 5: Predictive color-variance calibration: binned realized squared error vs. binned predicted variance (three seeds). "ECE" denotes the variance-calibration analogue $\sum_b f_b \, |\mathrm{MSE}_b - \overline{\mathrm{Var}}_b|$ (Kuleshov et al., 2018).

bins) and compare each bin's mean predicted variance to its realized $\|c^* - \mathbb{E}[c|s^*]\|^2$. The calibration curve sits on the diagonal across the full variance range; the regression expected calibration error is 1.3–3.0×10^{-3} (three seeds), with global mean predicted variance within 2% of realized MSE. This is the well-specified regime, where good calibration is expected; under misspecification—the real scenes of §6.8, same protocol on held-out frames—calibration degrades to ECE 0.011 (*lego*, variance over-estimated by 18%) and 0.016 (*chair*, under-estimated by 21%): still the right order of magnitude and positively correlated with realized error, but no longer sharp. Fixed-color-precision baselines expose no comparable quantity in either regime.

6.6 Mixing-measure contraction (Figure 6)

W_2 between the fitted mixing measure (posterior mean atoms, pruned expected weights) and the truth contracts as a power law in N ($1.56 \to 0.46$ for dp, α=1, over $N = 10^3 \to 10^5$); at $N = 10^5$ the DP variant is closer than the sparse variant (0.46 vs. 0.52 mean), consistent with Figure 4: over-split components carry weight off the true atoms.

6.7 Images at matched effective budgets: a deconfounded comparison (Table 1)

Following VBGS's image protocol (Tiny-ImageNet per Le & Yang, 2015; 64×64, pixels as points, their data pipeline and pure-JAX renderer for *all* methods), we fit DP-Splat with T=2000 (VBGS's demo budget) on 12 validation images and compare against three baselines: (a) VBGS at its full budget K=2000; (b) VBGS at fixed $K = \widehat{K}$ (matched effective budget, VBGS's own single-pass protocol); and (c) a *converged* fixed-K Dirichlet CAVI fit at $K = \widehat{K}$—baseline (c) isolates the weight prior from the effect of running coordinate ascent to convergence. The result is clean: DP-Splat exceeds single-pass VBGS at matched $\widehat{K}$ by +2.7 dB on average, but against the equally converged fixed-K baseline the difference essentially vanishes (−0.01 and +0.17 dB mean at $\alpha \in \{1, 100\}$). *The DP prior's contribution is selecting $\widehat{K}$—and supplying the predictive uncertainty machinery—not per-component rendering efficiency at a given K*; the raw matched-budget gap over VBGS is attributable to convergence. Two further honest observations: (i) $\widehat{K}$ stays two orders of magnitude below the T=2000 budget regardless of α ($\widehat{K} \in [9, 17]$ at α=1, $[29, 53]$ at α=100; component death again—the prior alone cannot force capacity into a fit), so absolute PSNR sits 3–10 dB below VBGS's full-2000-budget render; (ii) at $\alpha = 1$ the DP prior's own $\mathbb{E}[K] \approx \log N \approx 8$ is faithfully realized—the model does what its prior says.

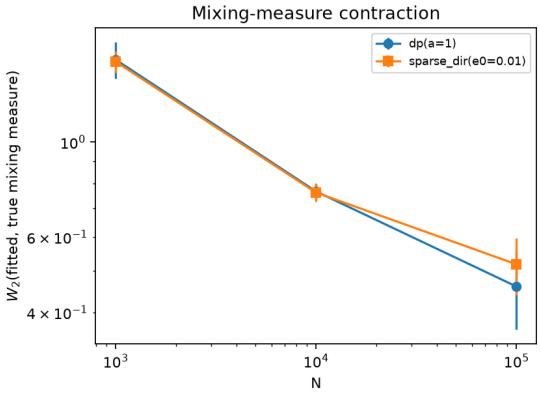

Figure 6: W_2(fitted, true mixing measure) vs. N (three seeds; atoms in joint spatial–color mean space). Context: contraction theory for exact posteriors is given by Nguyen (2013); no claim is made that CAVI attains those rates.

Table 1: Images at matched effective budgets (12 Tiny-ImageNet images, T=2000; mean [min, max] over images; mean only for the full-budget column). All methods rendered identically. "conv. dir@$\widehat{K}$" = converged fixed-K Dirichlet CAVI at the same $\widehat{K}$ (the deconfound). Full-budget VBGS (K=2000): 18.4–33.5 dB across these images.

α	$\widehat{K}$	vs. VBGS@$\widehat{K}$ (single-pass)	vs. conv. dir@$\widehat{K}$	vs. VBGS@2000
1	9–17	+2.80 [+0.64, +4.77]	−0.01 [−0.77, +0.77]	−6.47
100	29–53	+2.70 [+0.51, +5.46]	+0.17 [+0.02, +0.70]	−4.79

6.8 3D scenes, continual fitting, and scaling (Figure 7)

On NeRF-synthetic scenes (Mildenhall et al., 2020, test-split RGB-D point clouds; 40 frames × 20k subsampled points; matched budget $T = K = 2000$), DP-Splat is fit by SVI over the frame stream (10 epochs) and VBGS by its own continual protocol. We report the rasterizer-free held-out metric available to both methods—point-level color prediction $\mathbb{E}[c \mid s]$ against ground truth on an unseen frame—plus $\widehat{K}$ trajectories and DP-Splat's predictive variance maps (Table 2). On *lego*, DP-Splat at α=100 *exceeds* VBGS's held-out prediction (17.42 vs. 17.13 dB) with 5.9× fewer effective components ($\widehat{K}$ = 234 vs. 1374 used); on *chair* it matches within 0.14 dB with 7.6× fewer (181 vs. 1373). Compute is not the driver: a *single-pass* DP-Splat fit (23–25 s, comparable to VBGS's 36 s one-shot protocol) already reaches 17.25 dB on *lego* and 16.03 dB on *chair*; the additional epochs mainly *prune* $\widehat{K}$ (e.g., 276 → 234) rather than improve prediction (Table 2). The fitted complexity also orders the two scenes plausibly—the geometrically richer *lego* receives more components than *chair* at identical settings (an n=2 observation we report as such)—and, unlike the baseline, every prediction carries a predictive variance (Figure 7, middle: uncertainty concentrates on edges and occlusion boundaries). Unlike on 64×64 images, $\widehat{K}$ here rises far past 30 with α—the earlier saturation was data-complexity-limited, not a universal ceiling. Rasterized novel-view PSNR/SSIM require the CUDA rasterizer and full-resolution protocols; we scope them to follow-up (§7) and note that published full-pipeline 3DGS numbers (e.g., *chair* 35.8 dB; Kerbl et al., 2023) are not comparable to this point-level proxy. Finally, the natural-gradient variant's per-step cost is flat in N (0.04–0.05 s/step at $|B| = 2^{16}$, T=60, from $N = 10^5$ to 10^7 on CPU): dataset size is not the binding constraint for DP-Splat.

7 Limitations and discussion

Scope of the guarantees. All theory concerns the mixture likelihood on colored points. The rasterized rendering map (occlusion, alpha compositing) is applied downstream of inference exactly as in VBGS and

Table 2: Scenes at matched budget $T = K = 2000$ (held-out frame; point-level metrics; original color units; fit wall-clock on the same CPU). Single-pass DP-Splat uses compute comparable to VBGS's one-shot continual protocol; extra epochs mainly *prune* $\widehat{K}$ rather than improve prediction. α=100 was fixed a priori across scenes (no per-scene tuning); α=1 shown for sensitivity. DP-Splat additionally provides per-point predictive variance and log-likelihood; VBGS's fixed color precision does not define a comparable quantity.

Scene	Model	Effective K	Point-PSNR (dB)	Held-out LL/pt	Fit (s)
lego	DP-Splat α=1 (10 ep.)	20	14.56	−1.65	223
lego	DP-Splat α=100 (1 ep.)	276	17.25	+0.69	25
lego	DP-Splat α=100 (10 ep.)	**234**	**17.42**	**+1.03**	216
lego	VBGS (K=2000)	1374	17.13	—	36
chair	DP-Splat α=1 (10 ep.)	19	14.55	−4.68	143
chair	DP-Splat α=100 (1 ep.)	264	16.03	−2.06	23
chair	DP-Splat α=100 (10 ep.)	181	16.03	−1.95	144
chair	VBGS (K=2000)	1373	16.17	—	35

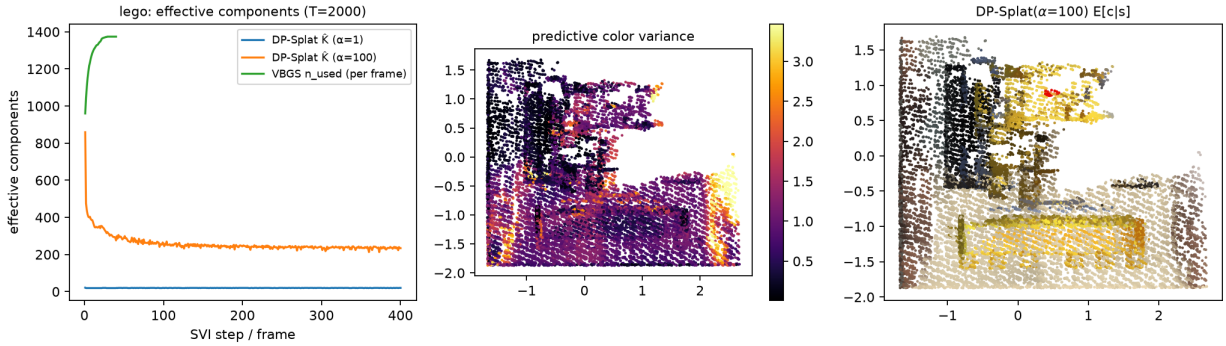

Figure 7: Scene fitting (*lego*). Left: effective components during continual fitting (matched budget T=2000): VBGS pins ∼1374 components of its budget; DP-Splat self-selects ∼234 (α=100, starting from ∼850 and shedding redundant components) or ∼20 (α=1). Middle: held-out-view predictive color variance—uncertainty concentrates on edges and occlusion boundaries. Right: held-out $\mathbb{E}[c|s]$ (denormalized).

carries no statistical guarantees; likewise our scene experiments use RGB-D/depth-lifted points and inherit VBGS's depth dependency.

What the prior can and cannot do under CAVI. Our central empirical message cuts both ways. The DP prior gives genuine, closed-form complexity adaptation—but the realized $\widehat{K}$ is shaped as much by coordinate-ascent dynamics (component death without revival; no merge moves) as by the prior, and hyperparameter learning follows the mode rather than fixing it. We chose to implement the plain scheme and measure these effects rather than patch them with the reassignment heuristics our approach is meant to replace; principled revival—restricted re-initialization between CAVI runs, merge/split proposals, or MFM-style inference (Miller & Harrison, 2018)—is the natural next step, and our results give it a quantitative target.

Rendering-grade evaluation. We deliberately evaluated with the reference implementation's own renderer on images and a rasterizer-free point-level metric on scenes, so that every comparison isolates the weight prior. Full-resolution rasterized novel-view benchmarks (and the corresponding quality-vs-budget frontier against published 3DGS numbers) require the CUDA rasterizer stack and are scoped to follow-up work on GPU hardware; the fitted models and the rendering map are already in place.

Numerical practice. Exactness claims hold in float64. Two float64-critical paths are documented in the repository (digamma/Beta computations; the rank-one recovery of Ψ from NIW naturals in the SVI blend), along with a scale-aware eps-level ridge used only inside Cholesky factorizations. None of these affect reported results; float32 deployments should center data and keep the NIW blend in float64.

Outlook. Because the entire pipeline stays conjugate, the door is open to hierarchical extensions (objects as groups of components via nested constructions) and to uncertainty-driven acquisition (next-best-view from $\mathrm{Var}[c \mid s]$), both of which we leave to future work.

Broader Impact Statement

This work is foundational statistical methodology for 3D scene representation. Calibrated uncertainty in reconstructions can *reduce* downstream harm (e.g., flagging under-observed geometry before decisions are made on it); we are not aware of significant risks of harm specific to this method beyond those generic to 3D reconstruction.

Reproducibility Statement

All results in this paper are produced by config-driven scripts (one per figure) with fixed seeds; every number in the tables is written to a JSON record by the producing script, and a repository document maps each figure to its exact command. The variational updates are tested against an independently written brute-force NumPy oracle, every ELBO term against Monte Carlo estimates from q, and one full CAVI step against the reference VBGS implementation (cross-code regression to 10^{-6}–10^{-8} relative, parameter-dependent). Code and records are provided as anonymized supplementary material and will be released under the MIT license upon publication.

References

Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998. doi: 10.1162/089976698300017746.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006. ISBN 978-0-387-31073-2.

David M. Blei and Michael I. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1(1):121–143, 2006. doi: 10.1214/06-BA104.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017. doi: 10.1080/01621459.2017.1285773.

Peng Ding. On the conditional distribution of the multivariate t distribution. *The American Statistician*, 70 (3):293–295, 2016.

Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. LightGaussian: Unbounded 3D Gaussian compression with 15x reduction and 200+ FPS. In *Advances in Neural Information Processing Systems*, volume 37, pp. 140138–140158. Curran Associates, Inc., 2024. doi: 10.52202/079017-4447.

Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2): 209–230, 1973. doi: 10.1214/aos/1176342360.

Sylvia Frühwirth-Schnatter, Gertraud Malsiner-Walli, and Bettina Grün. Generalized mixtures of finite mixtures and telescoping sampling. *Bayesian Analysis*, 16(4):1279–1307, 2021. doi: 10.1214/21-BA1294.

Zoubin Ghahramani and Matthew J. Beal. Variational inference for Bayesian mixtures of factor analysers. In *Advances in Neural Information Processing Systems 12*, pp. 449–455. MIT Press, 2000.

Alex Hanson, Allen Tu, Vasu Singla, Mayuka Jayawardhana, Matthias Zwicker, and Tom Goldstein. PUP 3D-GS: Principled uncertainty pruning for 3D Gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5949–5958, 2025.

Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14(40):1303–1347, 2013.

Hemant Ishwaran and Lancelot F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001. doi: 10.1198/016214501750332758.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. FisherRF: Active view selection and mapping with radiance fields using Fisher information. In *Computer Vision – ECCV 2024*, Lecture Notes in Computer Science, pp. 422–440. Springer, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, July 2023. doi: 10.1145/3592433.

Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3D Gaussian splatting as Markov chain Monte Carlo. In *Advances in Neural Information Processing Systems*, volume 37, pp. 80965–80986. Curran Associates, Inc., 2024. doi: 10.52202/079017-2573.

Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate uncertainties for deep learning using calibrated regression. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2796–2804. PMLR, 2018.

Kenichi Kurihara, Max Welling, and Yee Whye Teh. Collapsed variational Dirichlet process mixture models. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07)*, pp. 2796–2801, 2007.

Ya Le and Xuan Yang. Tiny ImageNet visual recognition challenge. Technical report, Stanford University, CS231N course report, 2015.

Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3D Gaussian representation for radiance field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21719–21728, 2024.

Gertraud Malsiner-Walli, Sylvia Frühwirth-Schnatter, and Bettina Grün. Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing*, 26(1-2):303–324, 2016. doi: 10.1007/s11222-014-9500-2.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (eds.), *Computer Vision – ECCV 2020*, volume 12346 of *Lecture Notes in Computer Science*, pp. 405–421. Springer, 2020. doi: 10.1007/978-3-030-58452-8_24.

Jeffrey W. Miller and Matthew T. Harrison. A simple example of Dirichlet process mixture inconsistency for the number of components. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Jeffrey W. Miller and Matthew T. Harrison. Inconsistency of Pitman–Yor process mixtures for the number of components. *Journal of Machine Learning Research*, 15:3333–3370, 2014.

Jeffrey W. Miller and Matthew T. Harrison. Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 113(521):340–356, 2018. doi: 10.1080/01621459.2016.1255636.

Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA, 2012. ISBN 978-0-262-01802-9.

XuanLong Nguyen. Convergence of latent mixing measures in finite and infinite mixture models. *The Annals of Statistics*, 41(1):370–400, 2013. doi: 10.1214/12-AOS1065.

Stéphane Pateux, Matthieu Gendrin, Luce Morin, Théo Ladune, and Xiaoran Jiang. BOGausS: Better optimized Gaussian splatting. *arXiv preprint arXiv:2504.01844*, 2025.

Michael Roth. On the multivariate t distribution. Technical Report LiTH-ISY-R-3059, Linköping University, Division of Automatic Control, 2013.

Judith Rousseau and Kerrie Mengersen. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):689–710, 2011. doi: 10.1111/j.1467-9868.2011.00781.x.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.

Toon Van de Maele, Ozan Çatal, Alexander Tschantz, Christopher L. Buckley, and Tim Verbelen. Variational Bayes Gaussian splatting. *arXiv preprint arXiv:2410.03592*, 2024.

Yuhan Zhu, Yanyu Zhang, Jie Xu, and Wei Ren. VBGS-SLAM: Variational Bayesian Gaussian splatting simultaneous localization and mapping. *arXiv preprint arXiv:2604.02696*, 2026.

A ELBO decomposition and verification

The ELBO is $\mathcal{L} = \mathbb{E}[\log p(X|Z,\theta)] + \mathbb{E}[\log p(Z|v)] + \mathbb{E}[\log p(v|\alpha)] + \mathbb{E}[\log p(\theta)] + \mathbb{E}[\log p(\alpha)] - \mathbb{E}[\log q(Z,v,\theta,\alpha)]$, implemented one named function per term. The NIW terms use the Wishart (precision) parameterization $p(\Lambda) = \mathcal{W}(\Psi_0^{-1}, \nu_0)$; per-term values differ from the Inverse-Wishart statement of §3 by the change-of-variables Jacobian $(D+1)\mathbb{E}[\log|\Lambda_k|]$, which cancels between $\mathbb{E}[\log p(\theta)]$ and $\mathbb{E}[\log q(\theta)]$, leaving $\mathcal{L}$ invariant (verified numerically). Every expectation is tested against Monte Carlo draws from q (Wishart/Beta/Gamma/Dirichlet sampling; 10^5 samples for scalar terms; agreement within 5 standard errors), and the assembled cycle is tested for monotonicity over 200 iterations $\times$ 20 seeds $\times$ 4 variants.

B Summary of validation criteria

Table 3: Validation criteria, fixed before the experiments were run, and their outcomes.

Criterion	Outcome
ELBO monotone under full-batch CAVI (200 it. $\times$ 20 seeds, all variants)	Pass; worst relative decrease -9.7×10^{-16}
Updates match independent brute-force oracle ($N \leq 200$, $T \leq 10$)	Pass at 10^{-7}–10^{-12}
ELBO terms match MC from q	Pass within 5 SE
SVI with $\lvert B\rvert{=}N$, $\rho_t{=}1$ reproduces CAVI	Pass at 10^{-8}–10^{-9}
One CAVI step reproduces reference VBGS step (baseline config)	Pass at 10^{-6}–10^{-8}
$\widehat{K}$ recovers $K_{\text{true}} \pm 1$ at $T = 3K_{\text{true}}$	Pass with regime-appropriate α (§6)
Image PSNR vs. fixed-K VBGS at matched effective budget within 0.5 dB	Pass, 24/24 image-α cells (exceeds by +0.5 to +5.5 dB); ties converged fixed-K CAVI at the same $\widehat{K}$ (-0.8 to +0.8 dB)

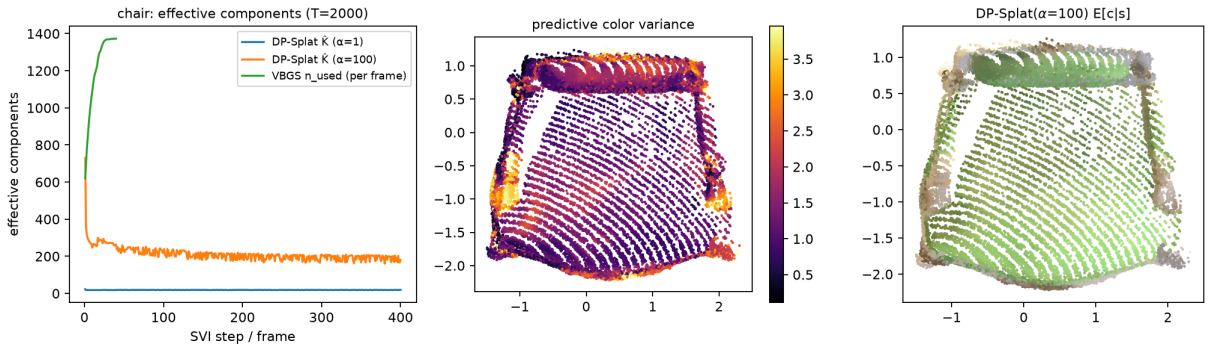

Figure 8: Scene fitting (*chair*), companion to Figure 7: VBGS pins ∼1373 components; DP-Splat self-selects ∼181 (α=100) or ∼19 (α=1).

C Additional scene figure

D Full $\widehat{K}$ tables

Complete per-cell records (including $n_{\min}$ sensitivity and entropy-based effective K) for Figures 2 and 4 ship as machine-readable JSON with the code; the included repository document `RESULTS.md` renders them as tables.

E Derivation of the predictive conditional moments (Eq. 7)

Setup and notation

Data points are $x = (s, c) \in \mathbb{R}^{D_s} \times \mathbb{R}^{D_c}$. The per-component likelihood factorizes across modalities,

$$p(x \mid \theta_k) \;=\; \mathcal{N}\big(s \mid \mu_{k,s}, \Sigma_{k,s}\big)\,\mathcal{N}\big(c \mid \mu_{k,c}, \Sigma_{k,c}\big), \qquad \theta_k = (\theta_{k,s}, \theta_{k,c}), \quad \theta_{k,m} = (\mu_{k,m}, \Sigma_{k,m}),$$

and the mixture likelihood is $p(x \mid \pi, \theta) = \sum_k \pi_k\, p(x \mid \theta_k)$. The variational posterior is mean-field:

$$q(\pi, \theta) \;=\; q(\pi) \prod_k q(\theta_{k,s})\, q(\theta_{k,c}), \qquad q(\theta_{k,m}) = \mathrm{NIW}\big(m_{k,m}, \kappa_{k,m}, \Psi_{k,m}, \nu_{k,m}\big).$$

NIW convention. $\mathrm{NIW}(m, \kappa, \Psi, \nu)$ means $\Sigma \sim \mathrm{IW}(\Psi, \nu)$ with density

$$\mathrm{IW}(\Sigma \mid \Psi, \nu) = \frac{|\Psi|^{\nu/2}}{2^{\nu D/2}\,\Gamma_D(\nu/2)}\,|\Sigma|^{-(\nu+D+1)/2}\,e^{-\frac{1}{2}\operatorname{tr}(\Psi\Sigma^{-1})}, \qquad \nu > D-1,$$

and $\mu \mid \Sigma \sim \mathcal{N}(m, \Sigma/\kappa)$. Here $\Gamma_D(a) = \pi^{D(D-1)/4}\prod_{i=1}^{D}\Gamma\big(a + \frac{1-i}{2}\big)$. This is the convention of Murphy (2012, §4.6.3); equivalently the precision satisfies $\Sigma^{-1} \sim \mathrm{Wishart}(\Psi^{-1}, \nu)$. The conjugate update in `niw.py` ($\Psi_n = \Psi_0 + S + \frac{\kappa_0 N}{\kappa_0 + N}(\bar{x} - m_0)(\bar{x} - m_0)^\top$, $\nu_n = \nu_0 + N$) is stated in exactly this convention, so the predictive formula below applies.

The multivariate Student-t density with location μ, scale matrix $S \succ 0$, and degrees of freedom $\eta > 0$ is

$$\mathrm{St}(x \mid \mu, S, \eta) = \frac{\Gamma\big(\frac{\eta+D}{2}\big)}{\Gamma\big(\frac{\eta}{2}\big)\,(\eta\pi)^{D/2}\,|S|^{1/2}}\Big[1 + \tfrac{1}{\eta}(x-\mu)^\top S^{-1}(x-\mu)\Big]^{-(\eta+D)/2}.$$

Step 1: $\mathbb{E}_q$ of the mixture is the mixture of $\mathbb{E}_q$'s

The posterior predictive under q is $\hat{p}(x^*) := \mathbb{E}_q\big[p(x^* \mid \pi, \theta)\big]$. For fixed x^*,

$$\hat{p}(x^*) = \mathbb{E}_q\Big[\sum_k \pi_k \mathcal{N}(s^* \mid \theta_{k,s})\,\mathcal{N}(c^* \mid \theta_{k,c})\Big] \overset{(i)}{=} \sum_k \mathbb{E}_q\big[\pi_k \mathcal{N}(s^* \mid \theta_{k,s})\,\mathcal{N}(c^* \mid \theta_{k,c})\big]$$
$$\overset{(ii)}{=} \sum_k \mathbb{E}[\pi_k]\;\mathbb{E}\big[\mathcal{N}(s^* \mid \theta_{k,s})\big]\;\mathbb{E}\big[\mathcal{N}(c^* \mid \theta_{k,c})\big].$$

(i) is linearity of expectation; for a finite (truncated) mixture it is trivial, and for the infinite DP sum it is justified by Tonelli's theorem since every term is nonnegative. (ii) holds because under the mean-field q, the random variables π_k, $\mathcal{N}(s^* \mid \theta_{k,s})$, and $\mathcal{N}(c^* \mid \theta_{k,c})$ are functions of the mutually independent blocks π, $\theta_{k,s}$, $\theta_{k,c}$, and the expectation of a product of independent, nonnegative, integrable random variables factorizes (Fubini). Note the k-th term only needs $\pi \perp \theta_{k,s} \perp \theta_{k,c}$; independence across components k is not even required for this identity, though mean-field provides it. Each factor $\mathbb{E}[\mathcal{N}(\cdot \mid \theta_{k,m})]$ is evaluated in Step 2, giving *exactly*

$$\boxed{\hat{p}(s^*, c^*) = \sum_k \mathbb{E}[\pi_k]\;\mathrm{St}_s(s^* \mid k)\;\mathrm{St}_c(c^* \mid k).}$$

This identity is pointwise-exact in x^* under q—no approximation beyond the mean-field posterior itself. (Under a truncated stick-breaking $q(\pi) = \prod_k \mathrm{Beta}(v_k \mid \gamma_{k1}, \gamma_{k2})$, $\mathbb{E}[\pi_k] = \mathbb{E}[v_k]\prod_{j<k}\mathbb{E}[1 - v_j]$ by the same independence argument.)

Step 2: NIW $\Rightarrow$ Student-t predictive

Fix a modality and drop indices: $q(\mu, \Sigma) = \mathrm{NIW}(m, \kappa, \Psi, \nu)$, dimension D. We compute $p(x) = \iint \mathcal{N}(x \mid \mu, \Sigma)\,\mathcal{N}(\mu \mid m, \Sigma/\kappa)\,\mathrm{IW}(\Sigma \mid \Psi, \nu)\,d\mu\,d\Sigma$.

2a: Marginalize μ. Given Σ, write $x = \mu + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \Sigma)$ independent of $\mu \sim \mathcal{N}(m, \Sigma/\kappa)$. Sums of independent Gaussians give

$$\int \mathcal{N}(x \mid \mu, \Sigma)\,\mathcal{N}(\mu \mid m, \Sigma/\kappa)\,d\mu = \mathcal{N}\Big(x \,\Big|\, m,\ \tfrac{\kappa+1}{\kappa}\,\Sigma\Big).$$

2b: Marginalize Σ. First a lemma: if $\Sigma \sim \mathrm{IW}(\Psi, \nu)$ and $b > 0$, then $b\Sigma \sim \mathrm{IW}(b\Psi, \nu)$ (substitute $T = b\Sigma$ in the density; the Jacobian $b^{-D(D+1)/2}$ is absorbed into the normalizer and $\mathrm{tr}(\Psi(T/b)^{-1}) = \mathrm{tr}((b\Psi)T^{-1})$). Hence it suffices to evaluate, with $d := x - m$,

$$p(x) = \int_{\Sigma \succ 0} \mathcal{N}(x \mid m, \Sigma)\,\mathrm{IW}(\Sigma \mid \tilde{\Psi}, \nu)\,d\Sigma, \qquad \tilde{\Psi} := \tfrac{\kappa+1}{\kappa}\Psi.$$

Using $d^\top \Sigma^{-1} d = \mathrm{tr}(dd^\top \Sigma^{-1})$, the integrand is

$$\frac{|\tilde{\Psi}|^{\nu/2}}{2^{\nu D/2}\Gamma_D(\nu/2)\,(2\pi)^{D/2}}\;|\Sigma|^{-\frac{(\nu+1)+D+1}{2}}\;\exp\Big\{-\tfrac{1}{2}\,\mathrm{tr}\big[(\tilde{\Psi} + dd^\top)\Sigma^{-1}\big]\Big\},$$

i.e. an unnormalized $\mathrm{IW}(\tilde{\Psi} + dd^\top,\, \nu + 1)$ kernel. The IW normalization $\int |\Sigma|^{-(\nu'+D+1)/2} e^{-\frac{1}{2}\mathrm{tr}(A\Sigma^{-1})}\,d\Sigma = 2^{\nu' D/2}\,\Gamma_D(\nu'/2)\,|A|^{-\nu'/2}$ with $\nu' = \nu + 1$, $A = \tilde{\Psi} + dd^\top$ yields

$$p(x) = \pi^{-D/2}\,\frac{\Gamma_D\big(\frac{\nu+1}{2}\big)}{\Gamma_D\big(\frac{\nu}{2}\big)}\,\frac{|\tilde{\Psi}|^{\nu/2}}{|\tilde{\Psi} + dd^\top|^{(\nu+1)/2}} \qquad \text{(the powers of 2 cancel: } -\tfrac{\nu D}{2} - \tfrac{D}{2} + \tfrac{(\nu+1)D}{2} = 0\text{)}.$$

The multivariate-gamma ratio telescopes:

$$\frac{\Gamma_D\big(\frac{\nu+1}{2}\big)}{\Gamma_D\big(\frac{\nu}{2}\big)} = \prod_{i=1}^{D} \frac{\Gamma\big(\frac{\nu+2-i}{2}\big)}{\Gamma\big(\frac{\nu+1-i}{2}\big)} = \frac{\Gamma\big(\frac{\nu+1}{2}\big)}{\Gamma\big(\frac{\nu+1-D}{2}\big)},$$

and the matrix determinant lemma gives $|\tilde{\Psi} + dd^\top| = |\tilde{\Psi}|\,(1 + d^\top \tilde{\Psi}^{-1} d)$. Therefore

$$p(x) = \frac{\Gamma\big(\frac{\nu+1}{2}\big)}{\Gamma\big(\frac{\nu+1-D}{2}\big)\,\pi^{D/2}\,|\tilde{\Psi}|^{1/2}}\big(1 + d^\top \tilde{\Psi}^{-1} d\big)^{-(\nu+1)/2}.$$

2c: Identification as Student-t. Match against $\mathrm{St}(x \mid m, S, \eta)$: equate the tail exponents, $\eta + D = \nu + 1 \Rightarrow \eta = \nu - D + 1$, and the quadratic forms, $S^{-1}/\eta = \tilde{\Psi}^{-1} \Rightarrow S = \tilde{\Psi}/\eta$. The normalizers then agree identically: $\Gamma(\frac{\eta+D}{2}) = \Gamma(\frac{\nu+1}{2})$, $\Gamma(\frac{\eta}{2}) = \Gamma(\frac{\nu+1-D}{2})$, and $(\eta\pi)^{D/2}|S|^{1/2} = \eta^{D/2}\pi^{D/2}\,\eta^{-D/2}|\tilde{\Psi}|^{1/2} = \pi^{D/2}|\tilde{\Psi}|^{1/2}$. Substituting $\tilde{\Psi} = \frac{\kappa+1}{\kappa}\Psi$:

$$\boxed{p(x) = \mathrm{St}\Big(x \,\Big|\, m,\ \Psi\,\frac{\kappa+1}{\kappa\,(\nu - D + 1)},\ \nu - D + 1\Big).}$$

This is exactly Murphy (2012, eq. (4.222)): $p(x \mid \mathcal{D}) = t_{\nu_N - D+1}\big(m_N,\ \frac{\kappa_N+1}{\kappa_N(\nu_N - D+1)}S_N\big)$ with $S_N \equiv \Psi_N$ the IW scale. It agrees with Bishop (2006, eq. (10.81)) after translating to the Wishart-on-precision parameterization: $L_k = \frac{(\nu_k+1-D)\beta_k}{1+\beta_k}W_k \Leftrightarrow L_k^{-1} = \frac{\beta_k+1}{\beta_k(\nu_k+1-D)}\Psi_k$ with $W_k = \Psi_k^{-1}$, $\beta_k = \kappa_k$. Note $\eta = \nu - D + 1 > 0$ is automatic from IW properness ($\nu > D - 1$), so the predictive is always a proper density.

Step 3: Conditioning the mixture on the sub-vector $s = s^*$

Write $a_k := \mathbb{E}[\pi_k]$, $f_k(s) := \mathrm{St}_s(s \mid k)$, $g_k(c) := \mathrm{St}_c(c \mid k)$, so $\hat{p}(s, c) = \sum_k a_k f_k(s) g_k(c)$ with $\sum_k a_k = 1$. Since each g_k is a normalized density, the spatial marginal is $\hat{p}(s) = \int \hat{p}(s, c)\,dc = \sum_j a_j f_j(s)$, and $\hat{p}(s^*) > 0$ for every s^* because Student-t densities are strictly positive. Hence

$$\hat{p}(c \mid s^*) = \frac{\hat{p}(s^*, c)}{\hat{p}(s^*)} = \sum_k w_k(s^*)\,g_k(c), \qquad \boxed{w_k(s^*) = \frac{a_k\,f_k(s^*)}{\sum_j a_j\,f_j(s^*)} = \frac{\mathbb{E}[\pi_k]\,\mathrm{St}_s(s^* \mid k)}{\sum_j \mathbb{E}[\pi_j]\,\mathrm{St}_s(s^* \mid j)}.}$$

The critical subtlety. The general rule for conditioning any mixture is $p(c \mid s^*) = \sum_k w_k(s^*)\,p_k(c \mid s^*)$ with the *within-component* conditional $p_k(c \mid s^*) = p_k(s^*, c)/p_k(s^*)$. Here each component's joint predictive is a genuine product $p_k(s, c) = f_k(s)\,g_k(c)$—spatial and color are *independent Student-t's within a component*, because q carries *separate* NIW factors $q(\theta_{k,s})\,q(\theta_{k,c})$ (equivalently: two independent Gamma mixing variables, one per modality). Therefore $p_k(c \mid s^*) = g_k(c)$ with *no* conditioning correction: no location shift, no scale inflation, no dof change.

This would NOT hold for a joint Student-t. If instead $p_k(s, c) = \mathrm{St}\big((s, c) \mid (m_s, m_c), S, \eta\big)$ with blocks S_{ss}, S_{sc}, S_{cc} (as would arise from a single NIW over the full $(D_s{+}D_c)$-dimensional covariance), the conditional is (Roth, 2013; Ding, 2016)

$$c \mid s^* \ \sim\ \mathrm{St}\Big(m_c + S_{cs}S_{ss}^{-1}(s^* - m_s),\ \frac{\eta + \delta_s}{\eta + D_s}\big(S_{cc} - S_{cs}S_{ss}^{-1}S_{sc}\big),\ \eta + D_s\Big), \quad \delta_s = (s^* - m_s)^\top S_{ss}^{-1}(s^* - m_s).$$

Even with block-diagonal scale ($S_{cs} = 0$) the conditional scale still depends on s^* through $\frac{\eta+\delta_s}{\eta+D_s}$ and the dof rises to $\eta + D_s$: for a joint t, uncorrelated $\neq$ independent, because both blocks share one latent Gamma mixing variable. The no-correction step is thus valid *only* in the factorized (product-of-independent-t's) case, which is exactly the case here.

Step 4: Conditional moments via the mixture law of total expectation/variance

Introduce the latent label z with $\Pr(z = k \mid s^*) = w_k(s^*)$ and $c \mid z{=}k \sim g_k = \mathrm{St}(m_{k,c}, S_{k,c}, \eta_{k,c})$. By the law of total expectation (each component mean exists iff $\eta_{k,c} > 1$, and $\mathbb{E}_{g_k}[c] = m_{k,c}$ by symmetry):

$$\boxed{\mathbb{E}[c \mid s^*] = \sum_k w_k(s^*)\,m_{k,c} =: \bar{m}(s^*).}$$

Second moment (each $\mathbb{E}_{g_k}[cc^\top] = C_k + m_{k,c}m_{k,c}^\top$ with $C_k := \mathrm{Cov}_{g_k}[c]$, finite iff $\eta_{k,c} > 2$; see Step 5):

$$\mathbb{E}[cc^\top \mid s^*] = \sum_k w_k\big(C_k + m_{k,c}m_{k,c}^\top\big), \qquad \boxed{\mathrm{Cov}[c \mid s^*] = \sum_k w_k\big(C_k + m_{k,c}m_{k,c}^\top\big) - \bar{m}\,\bar{m}^\top.}$$

Equivalently, in the (algebraically identical, numerically safer) law-of-total-variance form:

$$\mathrm{Cov}[c \mid s^*] = \underbrace{\sum_k w_k\, C_k}_{\text{within-component}} + \underbrace{\sum_k w_k\,(m_{k,c} - \bar{m})(m_{k,c} - \bar{m})^\top}_{\text{between-component}} \succeq 0.$$

Step 5: Student-t covariance factor and its domain

Gaussian scale-mixture representation: if $u \sim \mathrm{Gamma}(\frac{\eta}{2}, \frac{\eta}{2})$ (shape, rate) and $c \mid u \sim \mathcal{N}(m, S/u)$, then marginally $c \sim \mathrm{St}(m, S, \eta)$. Verification: with $\delta = d^\top S^{-1} d$,

$$\int_0^\infty \mathcal{N}(c \mid m, S/u)\,\mathrm{Ga}(u \mid \tfrac{\eta}{2}, \tfrac{\eta}{2})\,du \propto \int_0^\infty u^{\frac{\eta+D}{2}-1} e^{-\frac{u}{2}(\eta+\delta)}\,du = \Gamma\big(\tfrac{\eta+D}{2}\big)\Big(\tfrac{\eta+\delta}{2}\Big)^{-\frac{\eta+D}{2}} \propto \big(1 + \tfrac{\delta}{\eta}\big)^{-\frac{\eta+D}{2}},$$

and tracking constants reproduces the St normalizer exactly. Then by the law of total variance,

$$\mathrm{Cov}[c] = \mathbb{E}\big[\mathrm{Cov}(c \mid u)\big] + \mathrm{Cov}\big(\mathbb{E}[c \mid u]\big) = S\,\mathbb{E}[u^{-1}] + 0.$$

For $u \sim \mathrm{Gamma}(a, b)$, $\mathbb{E}[u^{-1}] = \int u^{-1} \frac{b^a}{\Gamma(a)} u^{a-1} e^{-bu} du = \frac{b^a\,\Gamma(a-1)}{\Gamma(a)\,b^{a-1}} = \frac{b}{a-1}$, finite iff $a > 1$. With $a = b = \eta/2$:

$$\boxed{\mathrm{Cov}[c] = \frac{\eta}{\eta - 2}\,S, \qquad \text{finite iff } \eta > 2.}$$

Domain summary: $\eta \le 1$: mean undefined; $1 < \eta \le 2$: mean $= m$ but covariance infinite; $\eta > 2$: both exist. In DP-Splat, $\eta_{k,c} = \nu_{k,c} - D_c + 1 = \nu_0 + N_k - D_c + 1$; under the default prior $\nu_0 = D_c + 2$ this is $N_k + 3 > 2$ for all $N_k \ge 0$, and `predictive.py` guards non-default priors at runtime.

Correspondence with Eq. (7) and the implementation

Every element of Eq. (7) follows from the derivation: (a) the exact mixture-of-products predictive $\hat{p} = \sum_k \mathbb{E}[\pi_k]\,\mathrm{St}_s\,\mathrm{St}_c$ (Step 1); (b) $S_k = \Psi_k \frac{\kappa_k+1}{\kappa_k \eta_k}$, $\eta_k = \nu_k - D_m + 1$, valid under the $\Sigma \sim \mathrm{IW}(\Psi, \nu)$ convention that `niw.py` demonstrably uses (Step 2, matching Murphy 2012 eq. 4.222); (c) $w_k(s^*) \propto \mathbb{E}[\pi_k]\,\mathrm{St}_s(s^* \mid k)$ with $p_k(c \mid s^*) = \mathrm{St}_c(c \mid k)$ requiring no correction *because and only because* the per-component predictive factorizes (Step 3); (d) $\mathbb{E}[c \mid s^*] = \sum_k w_k m_{k,c}$ and $\mathrm{Cov}[c \mid s^*] = \sum_k w_k(C_k + m_{k,c}m_{k,c}^\top) - \bar{m}\bar{m}^\top$ (Step 4); (e) $C_k = \frac{\eta_{k,c}}{\eta_{k,c}-2} S_{k,c}$ for $\eta_{k,c} > 2$ (Step 5). The code (`niw_predictive`, `student_logpdf`, `conditional_color_moments` in `predictive.py`) implements these formulas exactly, including the $\eta_c > 2$ runtime guard. This establishes Eq. (7) exactly as stated in §4.