Structure from Noise: Confirmation Bias in Particle Picking in Structural Biology

Amnon Balanov*[1], Alon Zabatani[1], and Tamir Bendory[1]

[1]School of Electrical and Computer Engineering, Tel Aviv University, Tel Aviv 69978, Israel

April 12, 2026

Abstract

The computational pipelines of single-particle cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) include an early particle-picking stage, in which a micrograph or tomogram is scanned to extract candidate particles, typically via template matching or deep-learning-based techniques. The extracted particles are then passed to downstream tasks such as classification and 3D reconstruction. Although it is well understood empirically that particle picking can be sensitive to the choice of templates or learned priors, a quantitative theory of the bias introduced by this stage has been lacking.

Here, we develop a mathematical framework for analyzing bias in template matching-based detection with concrete applications to cryo-EM and cryo-ET. We study this bias through two downstream tasks: (i) maximum-likelihood estimation of class means in a Gaussian mixture model and (ii) 3D volume reconstruction from the extracted particle stack. We show that when template matching is applied to pure noise, then under broad noise models, the resulting maximum-likelihood estimates converge asymptotically to deterministic, noise-dependent transforms of the user-specified templates, yielding a structure from noise effect. We further characterize how the resulting bias depends on the noise statistics, sample size, dimension, and detection threshold. Finally, controlled experiments using standard cryo-EM software corroborate the theory, demonstrating reproducible structure from noise artifacts in low-SNR data.

*Corresponding author: amnonba15@gmail.com

Contents

1 Introduction **4**
1.1 Examples of confirmation bias in structural biology . . . 4
1.2 Main results and work structure . . . 7

2 Preliminaries and problem formulation **8**
2.1 The detection and reconstruction process . . . 9
2.2 The template-matching selection . . . 9
2.3 Probabilistic models for the input patches . . . 10
2.4 Confirmation bias analysis via downstream reconstruction . . . 13

3 Theoretical results **14**
3.1 Population analysis of the template-matching selection bias . . . 14
3.2 Finite-sample analysis of template-matching selection bias . . . 16

4 Applications to cryo-EM and cryo-ET **18**
4.1 2D classification . . . 18
4.2 3D volume reconstruction . . . 18

5 Empirical results **20**
5.1 Particle picking based on template-matching . . . 20
5.2 Particle picking based on Topaz neural network . . . 22

6 Discussion and Outlook **24**
6.1 Different forms of confirmation bias in cryo-EM . . . 24
6.2 Strategies for mitigating confirmation bias . . . 26
6.3 Future work . . . 28

Appendices **33**

Appendix A Preliminaries **34**

Appendix B 2D classification in cryo-EM **36**
B.1 2D classification process overview . . . 36
B.2 Bias of per-template picked particles . . . 39
B.2.1 Proof of proposition B.5 . . . 42
B.2.2 Proof of proposition B.6 . . . 44
B.3 GMM maximum likelihood on template-selected particles . . . 47
B.4 Proof of Theorem 3.1 and Theorem 3.2 . . . 52
B.5 Proof of Proposition 3.3 . . . 52
B.6 Template matching with local maxima . . . 56

Appendix C 3D reconstruction **58**
C.1 Unified reconstruction model for cryo-EM and cryo-ET . . . 58
C.2 Proof of Corollary 4.2 . . . 60

Appendix D Experimental methods **62**
D.1 Setup of the numerical experiments 62
D.2 Validation under stationary correlated Gaussian noise 63

1 Introduction

Cryo-electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) have revolutionized structural biology by enabling visualization of macromolecular complexes in near-native states and at increasingly high resolutions. These complementary techniques are now at the heart of the field. Cryo-EM (Figure 1(a)) reconstructs the 3D structures of isolated macromolecules from 2D projection images acquired at unknown orientations, and has become essential for resolving atomic-level structures of proteins that are difficult to crystallize [34, 2, 39, 40, 58]. Cryo-ET (Figure 2(a)) captures 3D volumes of macromolecular assemblies in situ by collecting a tilt series of 2D projections at known angles [15, 42, 59, 52, 56]. While cryo-ET typically yields lower resolution, it preserves the native spatial context and is uniquely suited for probing the complexity of cellular environments.

As these techniques mature, a critical concern has emerged: the presence of biases, both experimental and algorithmic, that can affect the accuracy of the final 3D reconstructions. This concern is well recognized in the structural biology community, and practical strategies have been developed to mitigate known sources of bias, e.g., [50]. However, a deeper theoretical understanding of how such biases arise and propagate through the downstream reconstruction pipeline remains lacking.

Among the algorithmic sources of bias in single-particle reconstruction are incorrect alignment parameters, inaccurate contrast transfer function (CTF) correction, improper image normalization, erroneous masking in real or Fourier space, and the use of incorrect initial templates [50]. This work focuses on the latter and aims to develop a theoretical framework for quantifying its impact. In particular, we seek to establish both theoretical and empirical foundations for analyzing *confirmation bias during the particle-picking stage* in cryo-EM and cryo-ET.

Originally studied in the social sciences, confirmation bias refers to the tendency to interpret information in ways that support pre-existing beliefs or expectations, rather than allowing the data to speak for itself [27, 24, 37]. In the context of structural biology, confirmation bias, viewed as a specific instance of the broader category of model bias, can manifest when researchers, often unconsciously, assume that an unknown structure resembles a previously determined model or a prediction generated by computational tools. These assumptions, introduced early in the pipeline, can bias downstream processing and produce reconstructions that resemble prior models, even when unsupported by the data [45, 50, 4]. This raises a fundamental question that guides this work:

> *To what extent does confirmation bias in particle-picking algorithms shape and potentially distort the downstream pipeline, and in particular the final 3D structure in cryo-EM and cryo-ET?*

1.1 Examples of confirmation bias in structural biology

The Einstein from Noise phenomenon. To illustrate how methodological assumptions can generate structure from noise, we revisit a striking example widely discussed in the cryo-EM community, the Einstein from Noise phenomenon, which serves as the inspiration for this work. In this thought experiment, researchers analyze a dataset believed to contain noisy,

shifted instances of a known image, typically an image of Einstein, but in reality, the data consists entirely of pure noise. Despite the absence of any underlying signal, aligning and averaging the images relative to the template yields a reconstruction that closely resembles the original image [46, 47].

This phenomenon played a key role in a significant scientific controversy over the structure of an HIV molecule [32, 22, 53, 51, 31] and has since served as a cautionary tale for methodological overfitting. For a detailed statistical treatment, see [5, 6]. While often cited in discussions of particle picking using template-matching, it is crucial to note that the Einstein from Noise effect arises from a distinct mechanism: in template-matching particle picking methods, confirmation bias arises from selecting micrograph patches that most closely match the templates (essentially a form of *selection bias*, which is the topic of this work). In contrast, the Einstein from Noise effect emerges from aligning already extracted particles to a reference template. We further elaborate on this distinction in Appendix 6.1.

The risk of confirmation bias. The risk of confirmation bias is especially pronounced in scenarios with low signal-to-noise ratios (SNRs), a defining characteristic of cryo-EM and cryo-ET datasets. These datasets also exhibit significant uncertainty in particle orientations, rendering reconstructions highly sensitive to initialization and heuristic choices [48, 9]. Figures 1 and 2 illustrate the cryo-EM and cryo-ET setups and their respective particle reconstruction workflows, highlighting critical stages where confirmation bias may emerge. The computational workflows for both modalities involve several common stages: particle detection and extraction (i.e., particle picking), alignment and classification, and 3D reconstruction. In cryo-EM, individual 2D particles are picked from micrographs and grouped into 2D classes (Figure 1(a)); in cryo-ET, smaller subvolumes, termed subtomograms, are extracted from a reconstructed tomogram and then aligned and averaged to enhance the SNR (Figure 2(a)) [59, 56]. In both pipelines, downstream reconstruction accuracy depends critically on the fidelity of the initial *particle-picking stage—the main focus of this work.*

Confirmation bias in particle-picking. This study presents a theoretical and empirical framework to systematically investigate confirmation bias during the crucial *particle-picking* stage of cryo-EM and cryo-ET. Particle picking is the process of extracting 2D particle projections from noisy micrographs (in cryo-EM, Figure 1(b–c)) or 3D subtomograms from noisy tomograms (in cryo-ET, Figure 2(b)). While there are many different particle-picking algorithms, most fall into three broad categories: template-matching algorithms (e.g., [16, 23, 33, 13]), template-free methods based on generic shape cues such as blob- or Laplacian-of-Gaussian (LoG)-based picking, and deep learning algorithms (e.g., [12, 54]); see Section 6 for further discussion.

Template-matching methods work by scanning the micrograph or tomogram with a predefined reference structure (or template) and selecting regions that correlate highly with it. Template-free methods, by contrast, avoid explicit structural references and instead rely on coarse geometric or intensity-based cues to identify candidate particles. Deep learning methods train neural networks, often convolutional neural networks (CNNs), to detect particles based on features learned directly from annotated training data. While these approaches often perform well under moderate-to-high SNR, we show that in low-SNR conditions, common

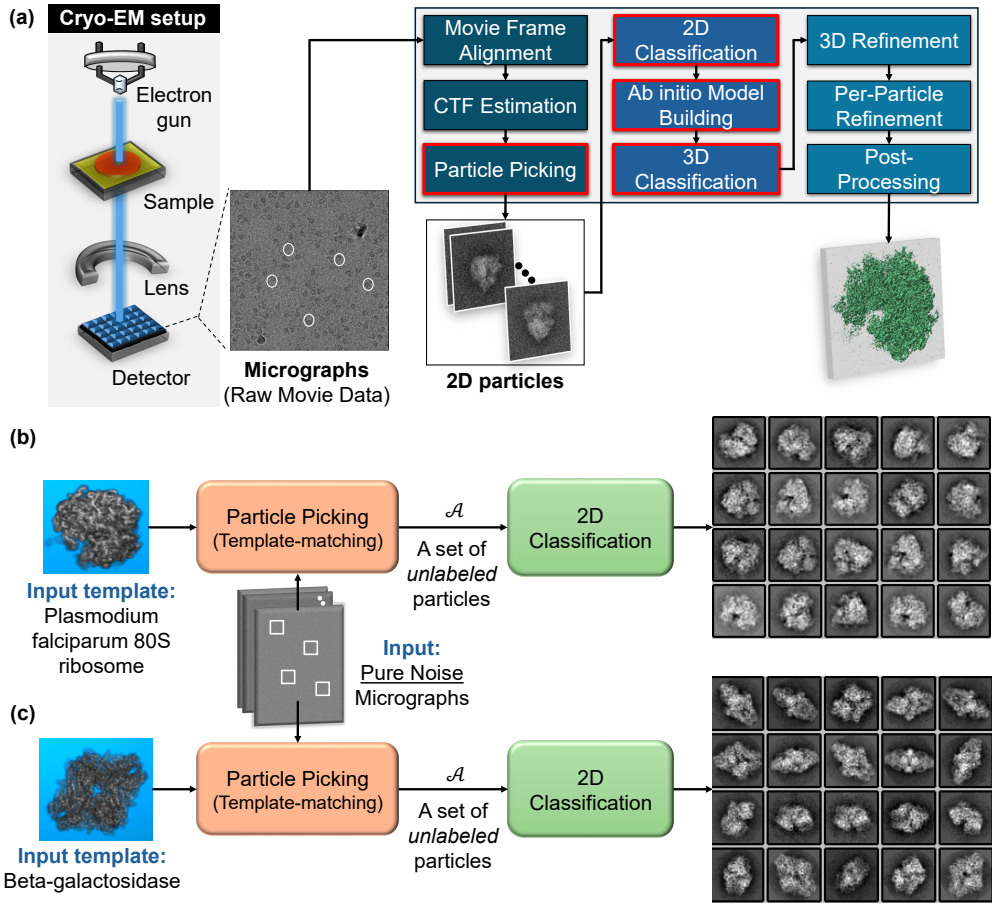

Figure 1: The cryo-electron microscope (cryo-EM) computational pipeline and structure from noise in particle picking. (a) Schematic cryo-EM workflow from 2D micrographs to a 3D reconstruction [9]: preprocessing (motion correction and CTF estimation/correction) is followed by particle picking, 2D classification, and 3D ab initio reconstruction and refinement. Red boxes mark stages particularly vulnerable to confirmation bias. **(b–c) Structure from noise under template-matching picking.** The template-matching picker (Algorithm 1) is applied to *pure-noise* micrographs using two different 3D templates: (b) Plasmodium falciparum 80S ribosome [57] and (c) beta-galactosidase [8]. The extracted particles are then processed by 2D classification (RELION VDAM [25], without in-plane alignment, yielding 2D class centers that closely resemble the picking templates despite the noise-only input. Settings: $L = 20$ template projections (and $L = 20$ 2D classification classes), patch size $d = 48 \times 48$, $M = 2 \times 10^5$ extracted particles (from 1000 micrographs), and threshold $T = 3.8$; ; see Section 5 for an analogous experiment *with in-plane alignment*, together with the full set of parameters.

to both cryo-EM and cryo-ET, they can introduce systematic structural biases.

Figure 1(b–c) demonstrates confirmation bias in cryo-EM particle picking using a template-matching algorithm applied to micrographs composed entirely of pure noise (as in the Einstein from Noise experiment). In these methods, predefined particle templates are correlated with input micrographs to identify potential particle locations. After template matching with different initial templates, the selected particles have been processed using RELION's VDAM 2D classification using a reference-free initialization [25]. Remarkably, despite the absence

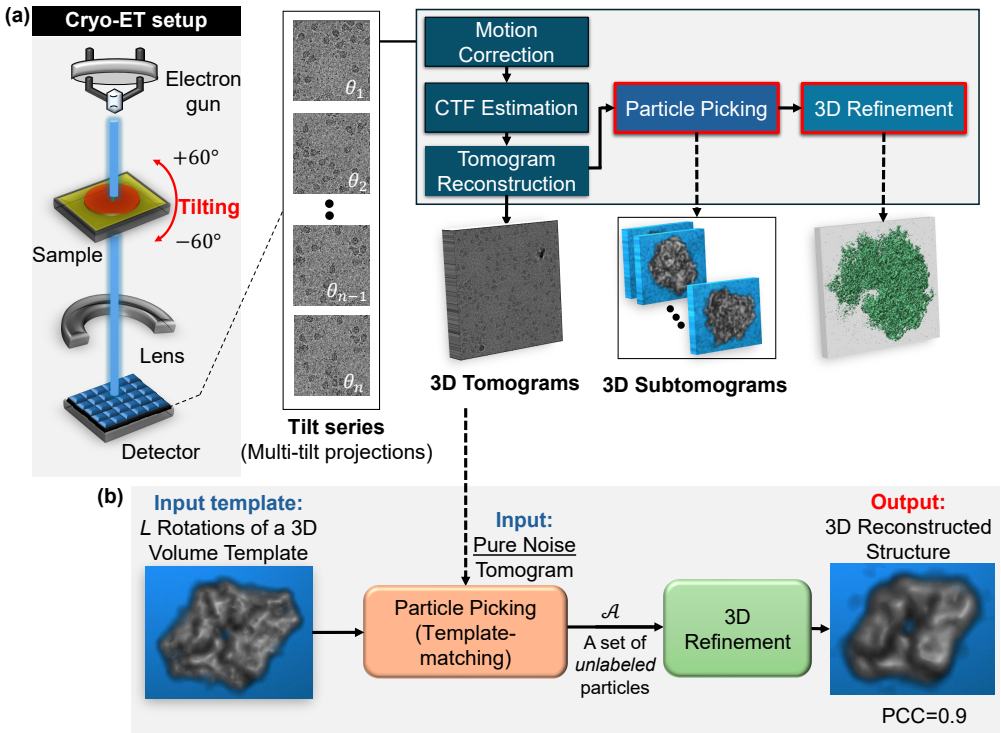

Figure 2: Structure from noise in the particle-picking stage of the cryo-electron tomography (cryo-ET) pipeline. (a) In cryo-ET, a tilt series of 2D projections (typically $-60°$ to $+60°$) is acquired and reconstructed into a 3D tomogram; after standard preprocessing (motion correction, tilt-series alignment, and CTF estimation/correction), particle picking extracts subtomograms that are refined by subtomogram averaging to improve SNR [56]. **(b) Confirmation bias under template-matching picking.** Template matching (Algorithm 1) is applied to a *pure-noise* tomogram using $L = 50$ rotated templates of a beta-galactosidase volume [8]. From $M = 2 \times 10^4$ extracted subtomograms (size $d = 24 \times 24 \times 24$, threshold $T = 3.5$), expectation-maximization 3D refinement algorithm yields a reconstruction highly correlated with the template volume (PCC $= 0.9$).

of any informative signal in the data, the resulting 2D class averages closely resemble projections of the original 3D templates. This behavior is consistent across multiple templates, algorithms, and experimental configurations, and also appears in the cryo-ET pipeline, as shown in Figure 2(b). We refer to this phenomenon, where downstream reconstructions produce biologically plausible structures from pure noise, as *structure from noise*.

1.2 Main results and work structure

While prior work [4, 5, 6] studied confirmation bias arising in downstream classification, where the bias is induced at the level of the estimation and classification procedure itself, the present work focuses on a different statistical mechanism: confirmation bias introduced already at the particle-picking stage, where template matching selects noise patches that are subsequently passed to downstream reconstruction. Thus, the main novelty of this work is the analysis of particle-picking-induced selection bias and how it affects downstream tasks.

Section 6.1 and Figure 8 further compare this mechanism with other forms of confirmation bias arising elsewhere in the cryo-EM/cryo-ET pipeline.

To systematically analyze confirmation bias at the particle-picking stage, we study both classical template matching and deep learning-based particle pickers. To quantify the resulting bias, we introduce a metric that measures the similarity between the extracted particles and the original templates, defined via the outputs of two key downstream tasks: maximum-likelihood estimation of Gaussian mixture model (GMM) class centers (i.e., the statistical model that governs the 2D classification task) and 3D volume reconstruction. We show that, when picking is applied to pure noise (under several models, including white, spherically symmetric, and stationary Gaussian models), the resulting maximum-likelihood estimates converge to deterministic transforms of the picking templates. This reveals a form of model bias: even in the absence of any true particles, the reconstruction is systematically shaped by the user-specified templates, rather than by genuine structure in the data.

We further characterize how the induced bias depends on the noise statistics, the number of extracted particles, the patch or volume dimension, and the detection threshold. On the empirical side, we demonstrate on synthetic cryo-EM and cryo-ET datasets that standard pipelines (RELION VDAM 2D classification and ab initio 3D refinement [25]) exhibit the predicted structure from noise behavior under template-based picking, and that a widely used neural-network picker, Topaz [12], likewise introduces reproducible structure when applied to pure-noise and low-SNR micrographs.

The remainder of the paper is organized as follows. Section 2 formalizes the template-based detection problem, introduces the probabilistic noise models, and describes GMM-based class averaging as a canonical reconstruction task. Section 3 develops the theoretical framework for confirmation bias in template-matching algorithms and analyzes its asymptotic and finite-sample behavior, with specializations to cryo-EM and cryo-ET pipelines given in Section 4. Section 5 presents empirical results using RELION [25] and Topaz [12], and Section 6 discusses implications for routine workflows and strategies for mitigating bias in the particle-picking stage.

2 Preliminaries and problem formulation

In this section, we introduce a general framework for analyzing bias in template matching. We begin by formalizing the detection and reconstruction problem in which template-matching methods are used to extract signals from a large noisy observation. We then define the template-matching selection rule and specify the probabilistic noise models used as input to the template-matching algorithm. Finally, we describe the downstream reconstruction procedure that serves to quantify how the selected patches are biased toward the templates. Concrete instances arising in cryo-EM 2D classification and cryo-EM/cryo-ET 3D reconstruction are presented in Section 4. Throughout the text, $\langle a, b\rangle$ represents the inner product between the vectors a and b.

2.1 The detection and reconstruction process

We begin with a setting in which a researcher observes a single long measurement $y \in \mathbb{R}^N$ (e.g., a time series, a micrograph, or a tomogram). This measurement is believed to contain multiple occurrences (i.e., instances) of a small collection of L unknown signal types $\{s_\ell\}_{\ell=0}^{L-1} \subset \mathbb{R}^d$. Across the observation y, these signal types appear repeatedly at unknown locations, where in total there are $Q \gg L$ copies embedded in y, and the goal is to estimate the underlying patterns $\{s_\ell\}$; see Figure 3 for a conceptual overview.

Formally, the researcher postulates the generative model

$$\text{(Postulated statistics)} \qquad y = \sum_{\ell=0}^{L-1} c_\ell * s_\ell + \xi, \tag{2.1}$$

where $*$ denotes discrete linear convolution, ξ is additive background noise, and each $c_\ell \in \{0,1\}^N$ is a sparse location indicator: $c_\ell(t) = 1$ indicates that an instance of type ℓ starts at position t, and $c_\ell(t) = 0$ otherwise. The total number of embedded instances is $Q \triangleq \sum_{\ell=0}^{L-1} \sum_{t=0}^{N-1} c_\ell(t)$, and it is assumed that signal instances are non-overlapping, that is, no index of y is covered by more than one shifted copy of any signal; equivalently, if $c_{\ell_1}(t_1) = c_{\ell_2}(t_2) = 1$ with $(\ell_1, t_1) \neq (\ell_2, t_2)$, then $|t_1 - t_2| \geq d$.

The goal is to recover the unknown signals $\{s_\ell\}_{\ell=0}^{L-1}$ from the observation y. Many imaging and signal-processing pipelines adopt a two-stage strategy (Figure 3(c)). First, a detection and extraction stage scans y and produces a collection of patches that are believed to contain signal occurrences. Second, a reconstruction stage aggregates these noisy patches to estimate the underlying signals, for example via a downstream procedure, such as maximum-likelihood estimation of class centers in a GMM fitted to the selected patches.

In practice, the detection stage is often implemented by template matching. The researcher specifies a finite bank of reference templates $\{x_\ell\}_{\ell=0}^{L-1} \subset \mathbb{R}^d$ encoding prior expectations or approximate guesses for the signals $\{s_\ell\}$, and a patch is selected whenever its correlation with at least one template exceeds a prescribed threshold.

In our analysis of confirmation bias, we focus on the null regime in which no true signal instances are present, i.e., $c_\ell \equiv 0$ for all ℓ, so that the underlying statistics reduce to

$$\text{(Underlying statistics)} \qquad y = \xi. \tag{2.2}$$

This null hypothesis can also be viewed as the extreme low-SNR limit $\mathrm{SNR} \to 0$ (equivalently $\|s_\ell\| \to 0$), reflecting the highly noise-dominated regime common in many applications, including cryo-EM and cryo-ET. We then ask: if the detection stage is applied to (2.2) rather than to (2.1) and it selects pure-noise patches solely because they correlate with the templates $\{x_\ell\}_{\ell=0}^{L-1}$, what structure does a downstream reconstruction procedure recover from these selected patches?

2.2 The template-matching selection

The researcher generates a collection of candidate patches from the observation y, denoted by $\{y_i\}_{i=0}^{N-1} \subset \mathbb{R}^d$. Then, the *template-matching selector*, as detailed in Algorithm 1, operates by cross-correlating each candidate observation $\{y_i\}_{i=0}^{N-1}$ with a finite bank of normalized

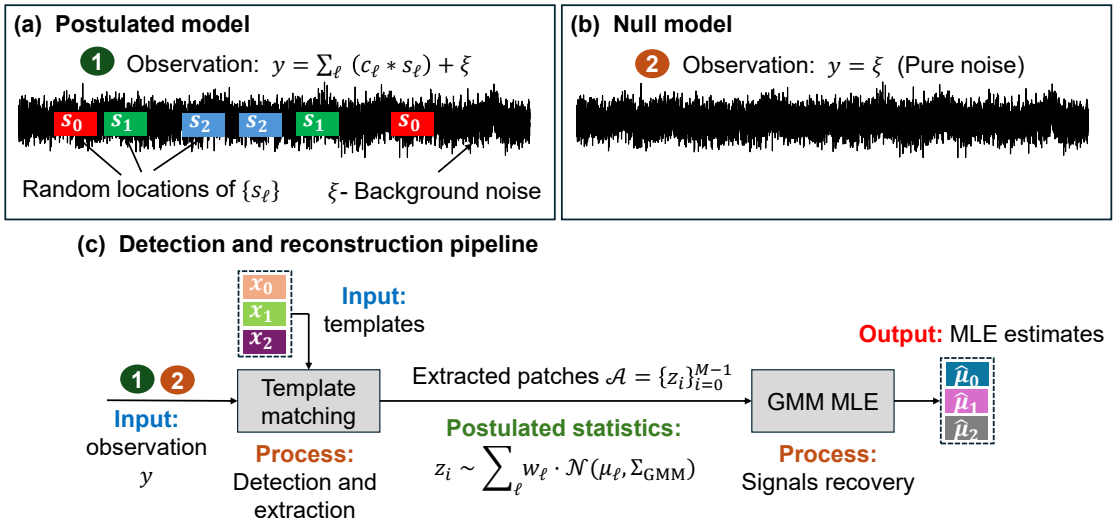

Figure 3: Conceptual overview of the model for analyzing confirmation bias in template matching. (a) Postulated signal-plus-noise model (2.1): the observation contains multiple non-overlapping instances of unknown signals $\{s_\ell\}$ embedded in background noise ξ. **(b)** Null model (2.2): the observation is pure noise. **(c)** Detection-and-reconstruction pipeline: assuming the postulated model, the researcher applies a two-stage procedure. First, template matching selects patches that correlate with a bank of templates $\{x_\ell\}$ (Algorithm 1). Second, a downstream maximum-likelihood estimator (MLE) fits a Gaussian mixture model (GMM) to the selected patches and outputs component-mean estimates $\{\widehat{\mu}_\ell\}$ (see Section 2.4). The central question of this work is whether, and how, this pipeline can yield template-dependent reconstructions even when the true data follow the null model.

templates $\{x_\ell\}_{\ell=0}^{L-1} \subset \mathbb{R}^d$.[1] An observation y_i is selected if its correlation with at least one template exceeds a fixed threshold T, that is, if $\max_{0\le\ell\le L-1}\langle y_i, x_\ell\rangle \ge T$. The selected observations are collected into the set $\mathcal{A}$, with cardinality $|\mathcal{A}| = M$. Importantly, the elements of $\mathcal{A}$ are *unlabeled*: for a given $y \in \mathcal{A}$, the template (if any) with which it has the highest correlation is not recorded by the selection rule. In cryo-EM and cryo-ET, patches $\{y_i\}_{i=0}^{N-1}$ represent candidate boxed images (or subtomograms) extracted from a micrograph (or tomogram), i.e., each patch may or may not contain a true particle instance, and the task is to precisely decide which patches correspond to particles rather than background noise. The templates are typically derived from a reference volume; see Section 4 for these specific applications.

2.3 Probabilistic models for the input patches

Our goal is to quantify confirmation bias induced by template-based detection. Following the Einstein from Noise experiment, and to examine this effect in its most extreme form, we analyze the null regime (2.2) in which the data contain only noise and no true signal

[1] We assume $\|x_\ell\|_2 = 1$ for all ℓ, corresponding to the standard energy normalization used in correlation-based matching. This ensures that correlation scores are directly comparable across templates and observations, avoiding trivial scaling effects.

Algorithm 1 `Template-Matching Selector`

Input: L templates $\{x_\ell\}_{\ell=0}^{L-1}$, N candidate patches $\{y_i\}_{i=0}^{N-1}$, and a threshold T.
Output: Set $\mathcal{A}$ of selected patches.

```
Initialize A ← ∅
for i = 0 to N − 1 do
    if max_{0≤ℓ≤L−1} ⟨y_i, x_ℓ⟩ ≥ T then
        Add y_i to A: A ← A ∪ {y_i}
    end if
end for
return A
```

instances. Motivated by the noise statistics encountered in cryo-EM datasets, we present a hierarchy of probabilistic models for the input patches $\{y_i\}_{i=0}^{N-1}$ presented to the template-matching selector, beginning with white Gaussian noise and progressing to more realistic stationary, strongly mixing noise processes that capture spatial correlations.

Model 2.1 (Independent and identically distributed Gaussian noise patches)**.** *The patches $\{y_i\}_{i=0}^{N-1} \subset \mathbb{R}^d$ are modeled as i.i.d. white Gaussian noise, that is, $y_i \sim \mathcal{N}(0, \sigma^2 I_d)$, where $\mathcal{N}(0, \sigma^2 I_d)$ denotes a d-dimensional Gaussian distribution with zero mean and covariance matrix $\sigma^2 I_d$.*

This idealized model is analytically convenient but not fully realistic: in practice, patches extracted from imaging data are typically dependent (due to spatial overlap), and the background noise is neither perfectly white.

Stationary process with strong-mixing property. To capture these effects, we next consider a more general model based on a stationary noise process. Assume the underlying statistics given by (2.2) so that $y = \xi$, and let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a zero-mean, stationary noise process with covariance function

$$\mathbb{E}[\xi_t] = 0, \qquad \mathrm{Cov}(\xi_t, \xi_{t+h}) = K(h), \tag{2.3}$$

for every $t, h \in \mathbb{Z}$. We construct candidate patches using a sliding window of length d:

$$y_i \triangleq \big(\xi_i, \xi_{i+1}, \ldots, \xi_{i+d-1}\big)^\top \in \mathbb{R}^d, \tag{2.4}$$

for every $i \in \mathbb{Z}$. By stationarity of $\{\xi_t\}$, the marginal distribution of y_i does not depend on i, and each patch has covariance

$$\begin{aligned} \Sigma &\triangleq \mathrm{Cov}(y_i) = \big(\Sigma_{jk}\big)_{1\le j,k\le d}, \\ \Sigma_{jk} &= \mathrm{Cov}(\xi_{i+j-1}, \xi_{i+k-1}) = K(j-k). \end{aligned} \tag{2.5}$$

In many applications, including cryo-EM and cryo-ET, it is natural to assume that distant patches are essentially independent, whereas nearby patches may be correlated but with correlations that decay with distance. We formalize this by assuming that the process $\{\xi_t\}$

is α-mixing (strong mixing) with respect to the index t [14]. Its α-mixing coefficients are defined by

$$\alpha(r) \triangleq \sup_{\substack{A\in\sigma(\xi_t:\, t\le n),\\ B\in\sigma(\xi_t:\, t\ge n+r)}} \big|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\big|, \tag{2.6}$$

where $\sigma(\xi_t : t \le n)$ denotes the σ-algebra generated by $\{\xi_t : t \le n\}$ (events depending only on the past up to time n), and $\sigma(\xi_t : t \ge n+r)$ is the σ-algebra generated by $\{\xi_t : t \ge n+r\}$ (events depending only on the distant future). We further assume that $\alpha(r) \to 0$ as $r \to \infty$ together with the standard summability condition [14]

$$\sum_{r=1}^{\infty} \alpha(r)^{\delta/(2+\delta)} < \infty \qquad \text{for some } \delta > 0. \tag{2.7}$$

Informally, this condition provides a mathematical model for short-range dependence: local regions of a signal (e.g., a micrograph or tomogram) are expected to be correlated, but the dependence weakens with distance, so sufficiently separated patches behave almost independently. Assumption (2.7) strengthens this by requiring that the dependence decay fast enough for empirical averages over many selected patches to exhibit the same qualitative concentration behavior as in the independent case. Such assumptions are standard in the analysis of stochastic processes and are natural here for analyzing the spatially correlated noise model with limited range.

We now specify a marginal model for the patches together with a weak dependence structure across them.

Model 2.2 (Spherically symmetric noise patches)**.** *Let $\{y_i\}_{i=0}^{N-1} \subset \mathbb{R}^d$ be a strictly stationary sequence of random vectors with common marginal distribution $Y \in \mathbb{R}^d$. Assume that $\{y_i\}$ is α-mixing with respect to i, satisfying (2.6)-(2.7) for some $\delta > 0$, and that $\mathbb{E}\|Y\|_2^{2+\delta} < \infty$. Assume that Y admits a probability density function p_Y that is continuous almost everywhere, and that $\mathbb{E}[Y] = 0$. Moreover, assume that Y is rotationally invariant, i.e., $p_Y(y) = \phi(\|y\|_2)$ for some function ϕ. In addition, for every unit vector $u \in \mathbb{S}^{d-1}$,*

$$\lim_{T\to\infty} \frac{\mathbb{E}[\langle u, Y\rangle \mid \langle u, Y\rangle \ge T]}{T} = 1, \tag{2.8}$$

and for every $\varepsilon > 0$,

$$\sup_{u\in\mathbb{S}^{d-1}} \mathbb{P}\left(\left\| \frac{Y}{T} - u \right\|_2 > \varepsilon \,\middle|\, \langle u, Y\rangle \ge T \right) \xrightarrow[T\to\infty]{} 0. \tag{2.9}$$

Conditions (2.8)-(2.9) hold for Gaussian noise and, more generally, for many rotationally invariant light-tailed distributions (e.g., sub-Gaussian and radial log-concave laws). Condition (2.8) requires that, conditioning on a large excursion of $\langle u, Y\rangle$ in direction u, the conditional mean grows linearly with the threshold T. Condition (2.9) strengthens this by imposing directional concentration: given $\langle u, Y\rangle \ge T$, the normalized vector Y/T concentrates around u uniformly over $u \in \mathbb{S}^{d-1}$. The white-Gaussian model in Model 2.1 is a special case (i.i.d. patches with $Y \sim \mathcal{N}(0, \sigma^2 I_d)$).

We next introduce a more realistic correlated-noise model that is generally not spherically symmetric at the patch level. In this setting, each patch remains Gaussian but with a nontrivial covariance structure that reflects the correlations of the underlying noise process.

Model 2.3 (Stationary Gaussian noise patches)**.** *Suppose that the stationary noise process* $\{\xi_t\}_{t\in\mathbb{Z}}$ *in* (2.3) *is Gaussian. Then, each patch* y_i *defined by the sliding window in* (2.4)*, is a* d*-dimensional Gaussian vector with distribution* $y_i \sim \mathcal{N}(0, \Sigma)$*, where the covariance matrix* Σ *is given by* (2.5) *and is assumed to be non-singular. The sequence* $\{y_i\}_{i=0}^{N-1}$ *is therefore a strictly stationary Gaussian vector process that is* α*-mixing with coefficients* $\alpha(r)$ *satisfying the summability condition* (2.7)*.*

2.4 Confirmation bias analysis via downstream reconstruction

Recall that, after template-based extraction, the selected patches are unlabeled. In the downstream analysis, the researcher aims to estimate and recover the underlying signals $\{s_\ell\}$ in (2.1) from these extracted patches. We adopt this same downstream mechanism as a vehicle for quantifying confirmation bias in template matching: rather than applying it to patches extracted under the postulated model (2.1), we apply it to patches extracted under the null statistics (2.2) (pure noise). We then analyze how the resulting reconstructions relate to the templates used in the detection stage.

A central downstream reconstruction process is Gaussian-mixture-based class averaging, which arises, for example, in cryo-EM 2D classification. Let $\mathcal{A} = \{z_i\}_{i=0}^{M-1} \subset \mathbb{R}^d$ denote the collection of patches selected by the template-matching stage. In the downstream analysis, these patches are commonly modeled as independent samples from a GMM [41]:

$$z_0, z_1, \ldots, z_{M-1} \overset{\text{i.i.d.}}{\sim} \sum_{\ell=0}^{K-1} w_\ell \cdot \mathcal{N}(\mu_\ell, \Sigma_{\text{GMM}}), \tag{2.10}$$

where $\mu_\ell \in \mathbb{R}^d$ is the mean of component ℓ, w_ℓ is its mixing weight (with $w_\ell \geq w_{\min} > 0$ and $\sum_{\ell=0}^{K-1} w_\ell = 1$), and Σ_{GMM} is a shared non-singular covariance matrix. In an ideal, unbiased setting under the postulated signal-plus-noise model (2.1), each mixture mean μ_ℓ is interpreted as a class average intended to recover one of the underlying signal types s_ℓ. Accordingly, we set the number of mixture components equal to the number of signal types, $K \triangleq L$. However, in the null setting (2.2), there are no true signal occurrences present, but solely noise. Consequently, any nontrivial resemblance between the reconstructed centers $\{\mu_\ell\}$ and the templates $\{x_\ell\}$ cannot be attributed to genuine signal and must arise from selection bias induced by the template-matching stage.

A standard approach to estimate the means $\{\mu_\ell\}_{\ell=0}^{L-1}$ from the selected patches is maximum-likelihood estimation [55, 43]. The log-likelihood of the GMM in (2.10) is

$$\mathcal{L}\big(\{\mu_\ell\}_{\ell=0}^{L-1}; \{z_i\}_{i=0}^{M-1}\big) = \sum_{i=0}^{M-1} \log\left(\sum_{\ell=0}^{L-1} w_\ell \mathcal{N}\big(z_i; \mu_\ell, \Sigma_{\text{GMM}}\big)\right), \tag{2.11}$$

where $\mathcal{N}(z_i; \mu_\ell, \Sigma_{\text{GMM}})$ denotes the multivariate Gaussian density evaluated at z_i. The GMM maximum-likelihood estimators of the component means are defined by

$$\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1} \triangleq \underset{\{\mu_\ell\}_{\ell=0}^{L-1}}{\arg\max}\, \mathcal{L}\big(\{\mu_\ell\}_{\ell=0}^{L-1}; \{z_i\}_{i=0}^{M-1}\big). \tag{2.12}$$

Within our framework when no signal occurrences appear in the data, the central question is: to what extent do the estimated centers $\{\widehat{\mu}_\ell\}$ resemble the templates $\{x_\ell\}$ used in the template-matching selector? Under an unbiased procedure and in the null regime, one would expect the class averages to bear no systematic correlation with the templates. Any persistent alignment between $\{\widehat{\mu}_\ell\}$ and $\{x_\ell\}$ therefore serves as a quantitative manifestation of confirmation bias.

We emphasize that the GMM in (2.10) provides a *postulated* model for the selected patches. It does not describe the true distribution of the extracted particles by template matching (Algorithm 1), which in general is highly non-Gaussian and strongly biased by the selection rule. In other words, the model is statistically *misspecified*, and this mismatch must be taken into account when interpreting the resulting class averages. A more detailed discussion of this misspecification is provided in Appendix B.1.

3 Theoretical results

In this section we present our main theoretical results on confirmation bias in template-based selection. Section 3.1 analyzes the population-level behavior of bias in the asymptotic regime $N, T \to \infty$, while Section 3.2 studies the effects of finite-sample for $N < \infty$. For completeness, Appendix B.6 introduces a more realistic template-matching mechanism based on local maxima and shows that this refinement does not change the leading-order bias.

3.1 Population analysis of the template-matching selection bias

We begin by quantifying the bias induced by template matching in the setting where the candidate patches are white Gaussian (or, more generally, spherically symmetric) noise. Theorem 3.1 shows that, under Model 2.1 or Model 2.2, the maximum-likelihood centroids of the postulated GMM obtained from the selected patches converge, after rescaling by the threshold T, to the templates themselves. In the following Theorem 3.2 (proved in Appendix B.4) we extend this result to stationary Gaussian noise with spatial correlations (Model 2.3), where the limiting bias is given by an anisotropic rescaling of each template determined by the noise covariance.

Theorem 3.1 (Bias of GMM centers with spherically symmetric noise)**.** *Let $\{y_i\}_{i=0}^{N-1}$ be the candidate patches drawn according to Model 2.1 or Model 2.2. Fix $L \geq 2$, and let $\{x_\ell\}_{\ell=0}^{L-1} \subset \mathbb{R}^d$ denote the normalized input templates, with $x_{\ell_1} \neq x_{\ell_2}$ for every $\ell_1 \neq \ell_2$. Let $T \in \mathbb{R}^+$ be the template-matching threshold and let $\mathcal{A}$ be the set of extracted patches produced by Algorithm 1. In the subsequent reconstruction step, the extracted patches $\mathcal{A}$ are modeled by a GMM with L components as in (2.10), and let $\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1}$ denote the corresponding GMM maximum-likelihood estimators of the component means, as defined in (2.12). Then there exists a permutation $\pi : \{0, \dots, L-1\} \to \{0, \dots, L-1\}$ such that, for every $0 \leq \ell \leq L-1$,*

$$\lim_{T\to\infty} \lim_{N\to\infty} \frac{\widehat{\mu}_{\pi(\ell)}}{T} = x_\ell, \tag{3.1}$$

where the convergence holds in probability.

Theorem 3.1 reveals a pronounced bias introduced by template-based selection: in the spherically symmetric noise setting, the estimated class centroids retain full correlation (Pearson cross-correlation of 1) with the templates used in the selection stage, even when the input data are pure noise. This is in stark contrast to the unbiased expectation that averaging pure noise yields zero as the number of samples grows. The appearance of the permutation π is due to the intrinsic label-switching symmetry of GMMs: the likelihood is invariant under relabeling of the mixture components, so the maximum-likelihood estimator identifies the set of centers only up to a permutation of their indices.

Theorem 3.2 (Bias of GMM centers with stationary Gaussian noise)**.** *Assume the same setting as in Theorem 3.1, except that the input candidate patches $\{y_i\}_{i=0}^{N-1}$ are drawn according to Model 2.3 with a covariance matrix Σ as in (2.5), and suppose that the α-mixing condition (2.7) holds. Then there exists a permutation $\pi : \{0, \ldots, L-1\} \to \{0, \ldots, L-1\}$ such that, for every $0 \le \ell \le L-1$,*

$$\lim_{T\to\infty} \lim_{N\to\infty} \frac{\widehat{\mu}_{\pi(\ell)}}{T} = \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \tag{3.2}$$

where the convergence holds in probability.

Theorem 3.2 shows that the confirmation bias persists even when the noise is non-spherically symmetric. While in the spherically symmetric setting of Theorem 3.1 the estimated centroids align with the raw templates x_ℓ (up to a global scaling by T), under stationary Gaussian noise the limiting direction of the estimated centroids is given by $\frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}$, which is an anisotropic rescaling of the template x_ℓ determined by the noise covariance Σ. Thus, even in the presence of realistic spatial correlations, Gaussian-mixture class averaging of patches selected by template matching remains fully biased toward the templates used in the selection stage, rather than vanishing as one might expect under a noise-only model.

The bias mechanism. We now explain the mechanism underlying the observed bias. The key point is that template-matching is a selection operation that changes the data distribution. Specifically, the extracted patches do not follow the original noise model; rather, they follow the conditional law of Y given the event $\max_\ell \langle Y, x_\ell \rangle \ge T$, so the threshold T induces a truncated (selection-biased) distribution. In particular, when the noise is spherically symmetric, this conditioning breaks rotational invariance and produces a nonzero conditional mean that aligns with the template x_ℓ. More precisely, let $\ell^\star = \arg\max_\ell \langle Y, x_\ell \rangle$. Then, for patches selected with $\ell^\star = \ell$, the conditional mean is aligned with the template x_ℓ:

$$\mathbb{E}\left[Y \,\middle|\, \ell^\star = \ell,\ \max_j \langle Y, x_j \rangle \ge T \right] \ \parallel\ x_\ell. \tag{3.3}$$

For correlated Gaussian noise $Y \sim \mathcal{N}(0, \Sigma)$, the same mechanism holds but the conditional mean aligns with the direction Σx_ℓ.

However, this alignment does not by itself imply that the maximum-likelihood estimators of the GMM means recover these conditional centroids. Indeed, the distribution of the extracted patches is generally non-Gaussian, so the postulated GMM model (2.10) is

misspecified; see Appendix B.3 for further discussion and analysis. In particular, the behavior of one center per template emerges only when T is sufficiently large: as T increases, the true mixture model components induced by the template-matching algorithm become effectively well-separated, so the maximum-likelihood estimation of the GMM centers places one center near each such biased mean. In contrast, at moderate thresholds, these components can overlap substantially, and the maximum-likelihood estimator need not recover these conditional means. Figure 4(a) illustrates this phenomenon: at low threshold values, the centroids do not align with the original templates because the extracted patches are not reliably associated with their corresponding classes. However, as the threshold increases, the estimated centroids progressively align more closely with the original templates. This reveals a trade-off in selecting the threshold during the selection stage, which we discuss further in Section 6.

3.2 Finite-sample analysis of template-matching selection bias

The analytical results derived above are in the asymptotic regime where the number of selected patches satisfies $M \to \infty$. In practice, however, applications involve a large but finite number of patches, so it is important to understand how finite-sample effects influence the observed confirmation bias. To this end, we study the mean squared error $\mathbb{E}_{\mathcal{A}}[\|\widehat{\mu}_\ell(\mathcal{A}) - Tv_\ell\|_2^2]$, which quantifies the discrepancy between the estimated class centroid $\widehat{\mu}_\ell$ and the asymptotic bias direction

$$v_\ell \triangleq \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \tag{3.4}$$

as in the right-hand side of (3.2). The following proposition, proved in Appendix B.5, characterizes the resulting finite-sample impact on the confirmation bias.

Proposition 3.3 (Finite-sample effect on confirmation bias)**.** *Consider the setting of Theorem 3.2. Let v_ℓ be as in (3.4) and $\widehat{\mu}_\ell(\mathcal{A})$ denote the GMM maximum-likelihood estimator of the ℓ-th class mean obtained from the extracted set of patches $\mathcal{A}$ in (2.12), with $|\mathcal{A}| = M$. Then, there exist constants $M_0, T_0, C_1, C_2 > 0$, independent of d, M, and T, such that for every $M \geq M_0$ and $T \geq T_0$,*

$$\mathbb{E}_{\mathcal{A}}\left[\left\|\widehat{\mu}_\ell(\mathcal{A}) - Tv_\ell\right\|_2^2\right] \leq C_1 \frac{d}{M} + \frac{C_2}{T^2}, \tag{3.5}$$

where the expectation is taken over the set of extracted particles $\mathcal{A}$.

The first term in the bound (3.5) has the classical parametric form of d-dimensional component means estimation under a well-separated GMM. The constant C_1 also captures the effect of temporal dependence between patches. Under the summability condition (2.7), one can bound C_1 in terms of $1 + \sum_{r=1}^{\infty} \alpha(r)^{\delta/(2+\delta)}$. When the patches are more strongly correlated, the mixing coefficients $\alpha(r)$ decay more slowly, the sum increases, and C_1 increases accordingly. In the i.i.d. case, $\alpha(r) = 0$ for all $r \geq 1$, the sum vanishes and C_1 reduces to the usual i.i.d. parametric constant. The second term, C_2/T^2, quantifies the residual bias between the population centers and the high-threshold limit Tv_ℓ.

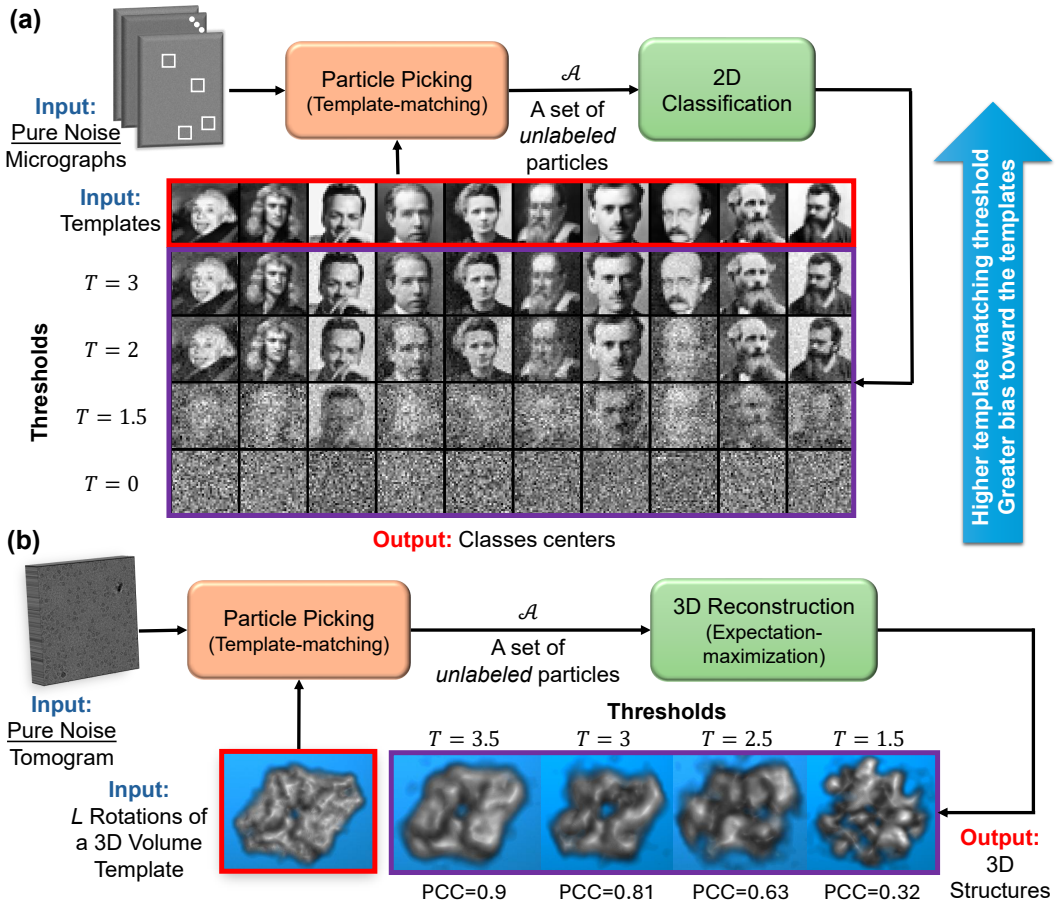

Figure 4: Effect of template-matching thresholds on confirmation bias. (a) GMM means estimation ("2D classification"). Template matching is applied to an observation y from (2.2) under noise Model 2.1 using a bank of $L = 10$ templates of prominent physicists. The resulting (unlabeled) patches are then classified using a standard expectation-maximization algorithm with L components; as the threshold T increases, the estimated centers become increasingly aligned with the templates. Parameters: $M = 10^5$ selected patches, image size $d = 50 \times 50$. **(b) 3D reconstruction (cryo-ET).** Template matching is applied to a pure-noise tomogram from Model 2.1 using $L = 50$ rotated templates of beta-galactosidase [8]. Subtomogram extraction followed by 3D refinement yields reconstructions whose similarity to the template volume increases with T (PCC shown). Parameters: volume size $d = 24 \times 24 \times 24$, $M = 2 \times 10^4$ selected subtomograms, and noise variance $\sigma^2 = 1$.

An immediate consequence is that, for a fixed number of extracted patches M, smaller patches (i.e. lower dimension d) concentrate more rapidly around the biased templates and are therefore more susceptible to confirmation bias than larger patches. Further details and the proof of Proposition 3.3 are given in Appendix B.5.

4 Applications to cryo-EM and cryo-ET

In this section we specialize the framework and results of Section 3 to standard single-particle cryo-EM and cryo-ET workflows.

4.1 2D classification

In single-particle cryo-EM, the signals $\{s_\ell\}$ in (2.1) are not arbitrary: they arise as tomographic projections of a common 3D volume. More precisely, there exists an underlying volume $V : \mathbb{R}^3 \to \mathbb{R}$ such that $s_\ell = \Pi\big(R_\ell \cdot V\big)$, for $\ell = 0, \ldots, L-1$, where Π denotes the tomographic projection operator, $R_\ell \in \mathsf{SO}(3)$ is a 3D rotation acting via $(R_\ell \cdot V)(x) \triangleq V(R_\ell^{-1}x)$, and L is the number of discretized viewing directions. In practice, the set of admissible orientations is continuous, but for the current analysis we work with a finite discretization $\{R_\ell\}_{\ell=0}^{L-1} \subset \mathsf{SO}(3)$. The templates used in the template-matching selector $\{x_\ell\}$, are generated in the same way from a template volume $V_{\text{template}} : \mathbb{R}^3 \to \mathbb{R}$:

$$x_\ell = \Pi\big(R_\ell \cdot V_{\text{template}}\big), \qquad \ell = 0, \ldots, L-1. \tag{4.1}$$

The 2D classification step in cryo-EM is commonly modeled using a GMM, exactly of the form introduced in (2.10). Consequently, the results obtained for GMM-based class averaging developed in Section 3 apply directly to the cryo-EM 2D classification pipeline. We note, however, that the simplified GMM model (2.10) omits two standard aspects of practical cryo-EM workflows. First, in-plane alignment is typically performed, whereby each particle is rotated and translated in the image plane to best match a class average [25]. Second, the number of downstream 2D classes need not equal the number of picking templates. In the theory, we exclude in-plane alignment and take $K = L$ to isolate the particle-picking-induced bias in its clearest analytically tractable form. These simplifications are not essential to the qualitative mechanism: when $K \neq L$, template-based selection still induces a biased conditional distribution on the extracted patches, but the downstream GMM no longer yields a one-to-one correspondence between templates and fitted classes; if $K < L$, several template-induced modes may merge, whereas if $K > L$, a single such mode may split into multiple fitted components. As shown empirically in Section 5, the core confirmation-bias phenomena persist in these more realistic settings.

4.2 3D volume reconstruction

The theoretical framework developed above for 2D classification extends naturally to the standard 3D cryo-EM and cryo-ET reconstruction pipeline. In single-particle cryo-EM, particle picking is performed on 2D micrographs, whereas in cryo-ET it operates directly on 3D tomograms, yielding volumetric subtomograms. Both modalities can be described within a unified forward model: each extracted, discretized observation $z_i \in \mathbb{R}^d$ (a 2D particle image in cryo-EM or a 3D subtomogram in cryo-ET) is treated as a noisy, randomly rotated observation of an underlying 3D volume $V \in \mathbb{R}^{d_{\text{vol}}}$,

$$z_i = \Pi\big(R_i \cdot V\big) + \epsilon_i, \qquad i = 0, \ldots, M-1, \tag{4.2}$$

where $\{R_i\}_{i=0}^{M-1} \subset \mathsf{SO}(3)$ are random rotations and $\{\epsilon_i\}_{i=0}^{M-1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_V)$ are additive Gaussian noise terms with covariance $\Sigma_V \succ 0$. In this unified view, cryo-ET corresponds to $\Pi = I$ (volumetric templates and subtomograms), while cryo-EM corresponds to a tomographic projection operator Π (2D projection templates and particle images). The goal of the reconstruction step is to estimate V in (4.2) from the extracted observations $\{z_i\}_{i=0}^{M-1}$.

Under the postulated model (4.2), integrating over the latent rotations yields the log-likelihood

$$\mathcal{L}\big(V; \{z_i\}_{i=0}^{M-1}\big) = \sum_{i=0}^{M-1} \log \left(\int_{R \in \mathsf{SO}(3)} \mathcal{N}\big(z_i; \Pi(R \cdot V), \Sigma_V\big)\, d\rho(R) \right), \tag{4.3}$$

where ρ is a prior distribution on rotations (often taken to be uniform on $\mathsf{SO}(3)$). The corresponding maximum-likelihood estimator of the volume is

$$\widehat{V} \triangleq \arg\max_V \mathcal{L}\big(V; \{z_i\}_{i=0}^{M-1}\big). \tag{4.4}$$

For analytical tractability, we discretize $\mathsf{SO}(3)$ to a finite set $\{R_\ell\}_{\ell=0}^{L-1}$ and assume that the same set is used both in the template-matching stage and in the subsequent reconstruction. The template bank fed to the particle picker is then $\{x_\ell\}_{\ell=0}^{L-1}$ as defined in (4.1), and Algorithm 1 is applied with these templates and threshold T. In addition, we make the following assumption:

Assumption 4.1. *Under the unified model (4.2), the map $V \mapsto \{\Pi(R_\ell \cdot V)\}_{\ell=0}^{L-1}$ determines V up to a global rotation. That is, if*

$$\big\{\Pi(R_\ell \cdot V_1)\big\}_{\ell=0}^{L-1} = \big\{\Pi(R_\ell \cdot V_2)\big\}_{\ell=0}^{L-1}, \tag{4.5}$$

then there exists $R \in \mathsf{SO}(3)$ such that $V_2 = R \cdot V_1$.

This property holds automatically in the cryo-ET case $\Pi = I$ if there are no internal symmetries to the molecule, and it coincides with standard identifiability assumptions in cryo-EM. The following corollary (proved in Appendix C.2) shows that the same confirmation bias phenomenon persists at the level of 3D reconstruction.

Corollary 4.2 (Bias of 3D volume estimate with spherically symmetric noise)**.** *Let $\{y_i\}_{i=0}^{N-1}$ be the candidate patches drawn according to Model 2.1 or Model 2.2. Fix $L \geq 2$, and let $V_{\text{template}} \in \mathbb{R}^{d_{vol}}$ denote a fixed 3D template volume. Let $\{R_\ell\}_{\ell=0}^{L-1} \subset \mathsf{SO}(3)$ be a collection of distinct rotations, and define the 3D template bank $x_\ell \triangleq \Pi(R_\ell \cdot V_{\text{template}})$ according to (4.1), with $x_{\ell_1} \neq x_{\ell_2}$ for every $\ell_1 \neq \ell_2$. Run Algorithm 1 with templates $\{x_\ell\}_{\ell=0}^{L-1}$ and threshold $T \in \mathbb{R}^+$, and let $\mathcal{A} \subset \{y_i\}_{i=0}^{N-1}$ be the resulting set of extracted particles. In the subsequent 3D reconstruction step, model the extracted particles $\mathcal{A}$ via the likelihood (4.3), and let $\widehat{V}$ denote the corresponding maximum-likelihood estimator as in (4.4), using the same rotations $\{R_\ell\}_{\ell=0}^{L-1}$. Assume that Assumption 4.1 is satisfied for this collection of discretized rotations. Then there exists a rotation $R \in \mathsf{SO}(3)$ such that*

$$\lim_{T \to \infty} \lim_{N \to \infty} \frac{\widehat{V}}{T} = R \cdot V_{\text{template}}, \tag{4.6}$$

where the convergence holds in probability.

In words, under the white or spherically symmetric noise model, when both the threshold T and the number of candidate observations N are large, the reconstructed structure satisfies $\widehat{V} \approx T\,(R \cdot V_{\text{template}})$ for some global rotation $R \in \mathsf{SO}(3)$. Corollary 4.2 relies on two key ingredients: (i) *rotation-grid alignment*: the same discrete set $\{R_\ell\}$ must be used to generate the template bank (4.1) and in the reconstruction likelihood (4.3); and (ii) *identifiability* of V from the collection $\{\Pi(R_\ell \cdot V)\}_{\ell=0}^{L-1}$ up to a global rotation, as stated in Assumption 4.1. When the rotation grids are not aligned, the maximum-likelihood estimator $\widehat{V}$ no longer reproduces the template exactly, but numerical experiments (see Figure 4(b)) indicate that $\widehat{V}$ remains highly correlated with V_{template}.

5 Empirical results

This section presents empirical results that corroborate and extend the theoretical findings from Sections 3–4. We examine how deviations from the simplified model, such as the presence of true particles (i.e., rather than purely noise micrographs), and the use of in-plane alignment, affect the outcomes of 2D classification and 3D reconstruction. Additionally, we evaluate the behavior of modern particle-picking approaches, in particular deep learning-based methods, to assess their susceptibility to confirmation bias in realistic cryo-EM workflows.

Importantly, the 2D classification and 3D reconstruction methods employed in our analysis are reference-free, i.e., they rely on random initialization rather than incorporating prior templates. This design allows us to isolate the confirmation bias introduced during picking. However, we note that the non-convex nature of the downstream reconstruction tasks may also contribute to residual bias.

5.1 Particle picking based on template-matching

In the previous sections, we introduced a simplified model for particle picking based on template matching, assuming that the input micrograph or tomogram consists solely of background noise. While this model offers important insights into the confirmation bias introduced by template-matching particle-picking, it does not fully reflect the complexity of modern cryo-EM reconstruction pipelines. To address this gap, we now turn to empirical experiments that reflect more realistic workflows. Below, we outline the experimental setup and highlight the key differences from the theoretical model. The corresponding results are presented in Figure 5.

Micrograph generation and particle extraction. In the theoretical analysis presented in Section 3, we examined the case where the input micrograph consists solely of noise. This model can be considered an asymptotic case where the SNR of the micrograph approaches zero. However, in practice, true particles are located at unknown positions within the micrograph, with $\text{SNR} > 0$.

We generate synthetic micrographs with non-overlapping true particles by iteratively placing particles at random locations and rejecting any proposed position that would overlap a previously placed particle. Each micrograph has size 2048×2048 and contains 200

projections, randomly drawn from a bank of 50 projections of the same structure, each of size $d = 36 \times 36$. White Gaussian noise with noise variance σ^2 is added to achieve an SNR of $\approx 1/25$, where SNR $\triangleq \frac{\mathrm{Var}(x_i)}{\sigma^2}$ and x_i denotes a normalized particle projection. Using a template-matching threshold of $T = 4$, we collected $M = 2 \times 10^5$ particles in total, corresponding to about 350 picked particles per micrograph, roughly 75% more than the 200 true planted particles. For the particle extraction, we employ template matching with local maxima, as outlined in Section B.6. A detailed simulation of this procedure, including the construction of synthetic micrographs and the behavior of template matching under overlapping conditions, is presented in Appendix D and described explicitly in Algorithm 2.

2D classification. The theoretical model in Section 3, excludes in-plane alignment and assumes that the number of templates used during particle picking matches the number of 2D classes, an assumption that is generally not satisfied in practice. In this section, we consider a more realistic cryo-EM pipeline that includes in-plane alignment and allows the numbers of templates and 2D classes to differ.

To align with standard practice, we use RELION's VDAM algorithm for 2D classification [25]. RELION is widely adopted in the cryo-EM community and VDAM, introduced in version 4.0, is a variable-metric gradient-based optimizer with adaptive moment estimation, inspired by Adam [26] and implemented with mini-batches for computational efficiency. Unless stated otherwise, we run VDAM with $L = 20$ classes, 200 iterations, regularization parameter 3, no CTF correction, and in-plane alignment using an angular sampling of 12°.

3D ab-initio modeling. We also performed ab initio 3D reconstruction using RELION's VDAM algorithm. In this context, ab initio reconstruction refers to generating an initial 3D structure without relying on any external reference, thereby reducing potential sources of bias. We employed RELION's VDAM algorithm for this task with the following parameters: $L = 1$ class, 200 iterations, a regularization parameter of 4, no CTF correction, and no symmetry constraints (i.e., $C1$ symmetry). Unlike the idealized setting of Corollary 4.2, this empirical reconstruction does not use the same discrete rotation grid as the one used during template matching. Rather, the experiment is intentionally carried out in a more realistic regime, in which the picking and reconstruction stages use different angular discretizations. This allows us to assess the extent to which the predicted confirmation-bias phenomenon persists under practical grid mismatch.

Results. Figure 5 presents results from the empirical setup in which the micrograph contains genuine particle projections, but the template used for particle picking corresponds to a different structure. In this scenario, the 2D class averages reflect a mixture of two competing sources: the true underlying particles and the mismatched template. As a result, while some class averages recover features of the actual particles, others resemble the template used for picking. Similarly, the ab initio 3D reconstruction derived from these particles exhibits structural elements from both the true volume and the template. This highlights the risk of introducing structural bias at the particle-picking stage, where even in the presence of a signal, the use of an incorrect template can propagate artifacts into downstream analyses and compromise the reliability of the final reconstruction. In addition, we note that the

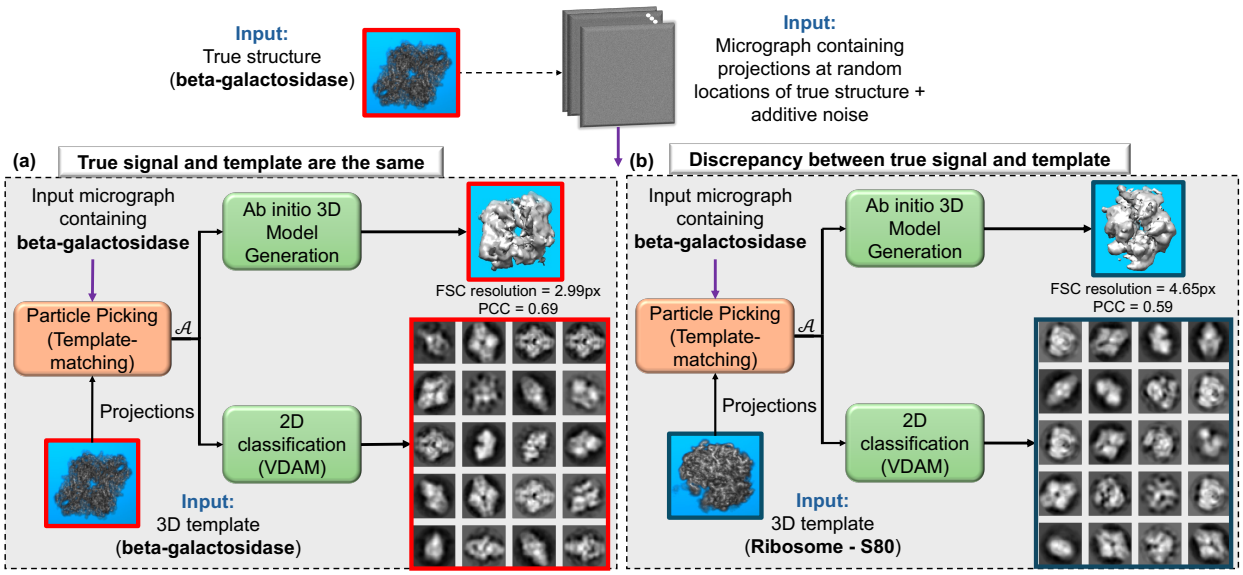

Figure 5: Confirmation bias in particle picking when particles are present. Micrographs contain true particle projections of beta-galactosidase [8] corrupted by additive white Gaussian noise (SNR $\approx 1/25$), and particle picking is performed by template matching using either: **(a)** projections of the true beta-galactosidase structure, or **(b)** projections of an unrelated structure, the Plasmodium falciparum 80S ribosome [57]. In both cases, the extracted particles are processed by 2D classification and ab initio 3D reconstruction in RELION (VDAM) [25]. With matched templates **(a)**, the 2D classes and 3D map resemble the ground truth (PCC = 0.69, Fourier Shell Correlation (FSC) resolution = 2.99 pixels); with mismatched templates **(b)**, spurious ribosome-like classes appear and the 3D reconstruction becomes a hybrid, with degraded agreement (PCC = 0.59, FSC resolution = 4.65 pixels). Settings: patch size $d = 36 \times 36$, $M = 2 \times 10^5$ extracted particles, threshold $T = 4$; VDAM ran for 200 iterations without CTF correction.

use of a markedly dissimilar template in Figure 5 is intended to make the bias mechanism visually transparent. In practice, however, the more dangerous regime may be one in which the incorrect template is structurally similar to the true particle, since the resulting bias can be substantially harder to detect.

5.2 Particle picking based on Topaz neural network

Deep learning–based particle pickers are now widely used in cryo-EM and cryo-ET, with Topaz being a prominent example [12]. Unlike template matching, which introduces bias through more intricate mechanisms, such as explicit similarity to predefined structures, deep neural networks can inherit bias from their training data, architecture, and optimization dynamics. As a result, they may still exhibit confirmation bias by preferentially selecting particles that resemble features emphasized during training. To probe this effect, we consider both the pre-trained Topaz model [12], used without additional training, and Topaz models trained on user-specified particle sets.

Pre-trained Topaz network. For the pre-trained Topaz network, we applied Topaz to micrographs of size 2048×2048, using an internal down-sampling factor of $s = 2$. During extraction, we set the particle radius to $r = 9$, corresponding to a patch size of $d = 36 \times 36$ on the original grid, and used a coordinate scaling factor of $x = 2$ to account for the down-sampling. We used a score threshold of -1 for the Topaz output, corresponding to a predicted particle probability of $p \geq 0.25$. Although the default threshold in Topaz and RELION is typically 0 (that is, $p \geq 0.5$), in practice users often lower this threshold to improve recall, especially for challenging low-SNR datasets. We adopted this more permissive setting to extract a sufficiently large particle stack ($M = 2 \times 10^5$) within a reasonable runtime, enabling a direct comparison with the template-matching experiments in both dataset size and effective SNR. Thus, the observed structure-from-noise artifacts arise under realistic conditions, rather than from forcing the network to select arbitrary low-confidence patches.

Figure 6(a-c) shows that the pre-trained Topaz network induces structured bias. The mean of the selected particles by Topaz in Figure 6(a) displays a faint central circular feature, whereas under an unbiased model the mean of pure-noise patches should converge to zero. A similar central pattern appears in Figure 6(b), which shows the mean of particles selected by template matching with a large, heterogeneous template set, suggesting that both methods imprint structure on the picks. In contrast, the mean of uniformly random patches from the same noise micrographs (Figure 6(c)) is essentially zero.

To assess the impact on 3D refinement, we perform a controlled experiment in which pure-noise micrographs are split into two independent halves. Each half is processed with one of three picking strategies: template matching, Topaz, or random selection. Independent ab initio 3D reconstructions are then computed from each half using RELION's VDAM algorithm, and their consistency is measured by the Fourier Shell Correlation (FSC), shown in Figure 6(d). Both Topaz and template matching yield high FSC curves between the halves, indicating that each method produces reproducible, method-dependent structure from noise. In contrast, random selection yields near-zero FSC, as expected under an unbiased model. Among the three, template matching exhibits the strongest bias, followed by Topaz. These results further demonstrate that the FSC gold standard can be misled by systematic errors, in line with the observations of [50].

Trained Topaz network. To compare deep-learning-based picking with classical template matching under realistic conditions, we tested whether Topaz's learned prior can override weak true signal when real particles are present. Following the setup of Figure 5, we generated synthetic micrographs containing true beta-galactosidase projections at low SNR ($\approx 1/25$). Starting from the same vanilla pre-trained Topaz model used in the previous experiment, we re-trained two networks: one on positive examples from beta-galactosidase, and one on an unrelated structure, the *Plasmodium falciparum* 80S ribosome. In both cases, training was run for 10 epochs on 20 synthetic micrographs, each containing 200 particle projections embedded in white Gaussian noise at the same low-SNR level as the evaluation data.

As shown in Figure 7, when Topaz is trained on the mismatched ribosome data, the extracted particles yield 2D class averages and a 3D reconstruction with clear ribosome-like features. The reconstruction quality is also substantially worse than in the matched-training case (PCC 0.08 versus 0.24, and FSC resolution 17.08 pixels versus 6.84 pixels). This

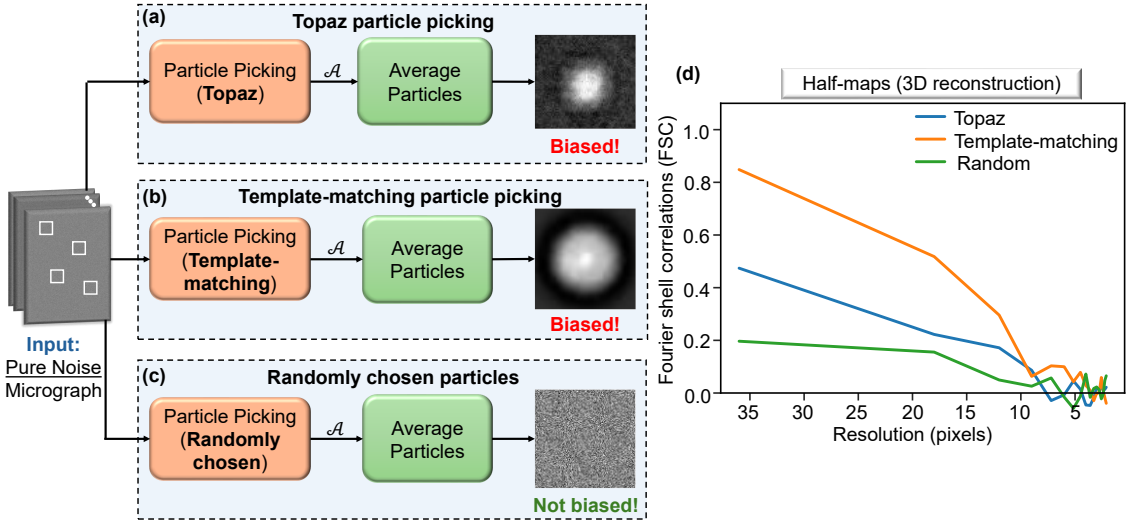

Figure 6: Confirmation bias in pre-trained Topaz particle picking. (a–c) Mean of picked particles from pure-noise micrographs: (a) Topaz picks yield a nonzero mean with a prominent central circular feature; (b) template matching with a large bank of Plasmodium falciparum 80S ribosome projection templates [57] produces a similar pattern; (c) random picks average to zero. **(d)** Bias assessment via half-map reproducibility: noise micrographs are split into two independent halves; on each half we pick $M = 10^5$ particles using Topaz, template matching, or random selection, and reconstruct two half-maps using RELION VDAM. We report the Fourier Shell Correlation (FSC) between the half-maps. Settings: patch size $d = 36 \times 36$; VDAM ran for 200 iterations without CTF correction.

shows that, much like template matching, a dedicated neural-network picker can be strongly affected by the specific structural prior imposed during training. Rather than introducing only a generic bias toward centered, particle-like blobs, the network preferentially selects low-SNR data that resemble its training distribution, thereby propagating structure-specific hallucinations into the downstream pipeline.

6 Discussion and Outlook

To place the bias analyzed in this work in context, we compare it with other sources of bias that arise throughout the cryo-EM and cryo-ET reconstruction pipelines. We then discuss mitigation strategies and conclude with several promising directions for future research.

6.1 Different forms of confirmation bias in cryo-EM

Several stages of the cryo-EM and cryo-ET pipeline are vulnerable to confirmation bias [50, Section 4.1], including particle picking, 2D classification, and 3D reconstruction. The present work focuses on particle picking [44, 21, 12, 20], while related biases in low-SNR EM-based reconstruction have been analyzed in [6]. These stages introduce bias through a different statistical mechanism. Figure 8 compares confirmation bias in particle picking with that in 2D classification. Below, we contrast the Einstein from Noise phenomenon from Section 1

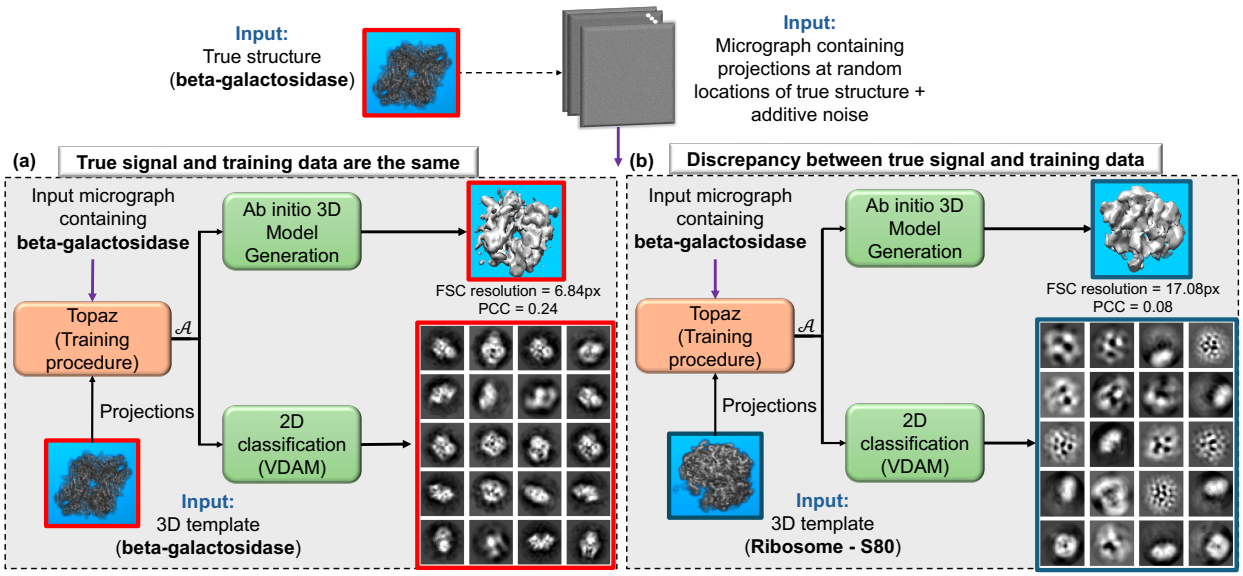

Figure 7: Confirmation bias in trained Topaz particle picking when true particles are present. Micrographs contain true particle projections of beta-galactosidase [8] corrupted by additive white Gaussian noise. Particle picking is performed using a Topaz model trained on either: **(a)** projections of the true beta-galactosidase structure (matched training data), or **(b)** projections of an unrelated structure, the Plasmodium falciparum 80S ribosome [57] (mismatched training data). In both cases, the extracted particles are processed by 2D classification and ab initio 3D reconstruction in RELION (VDAM) [25]. With the matched training data **(a)**, the 2D classes and 3D map successfully recover the beta-galactosidase structure (PCC = 0.24, FSC resolution = 6.84 pixels). Conversely, when trained on the mismatched structure **(b)**, the network preferentially selects noise features that align with its learned prior. As a result, the 3D reconstruction of the true underlying signal is significantly degraded.

and the bias observed in 2D classification [4] with the bias arising from particle picking.

The Einstein from Noise phenomenon. Although frequently mentioned in the context of template-based particle picking, the Einstein from Noise effect arises from a distinct model. In the Einstein from Noise experiment, all observations are aligned and averaged with respect to a fixed template. In contrast, particle picking selects a subset of patches with the highest template correlations. As a result, particle picking produces reconstructions that asymptotically match the template up to global scaling (Theorems 3.1 and 4.2), whereas Einstein from Noise yields an average that is only positively correlated with the template but not identical [5].

Confirmation bias in 2D classification and 3D volume reconstruction. In later stages of cryo-EM processing, such as 2D classification and 3D volume estimation, iterative algorithms often use predefined templates for initialization. When applied to data with no true signal, these methods may still produce output that resembles the initial templates, even though the input data provides no evidence for such structure [4].

Figure 8(b) illustrates this effect in 2D classification. Here, the input consisted en-

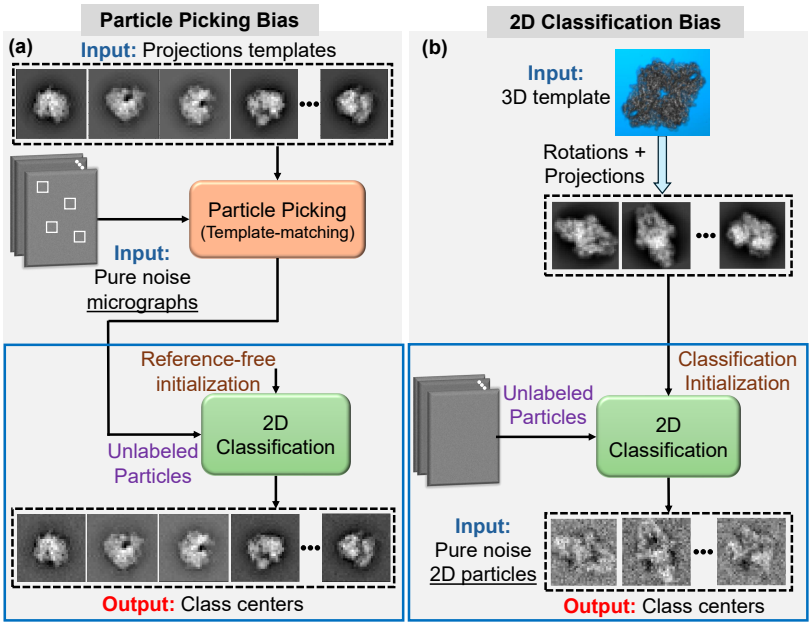

Figure 8: Comparison of confirmation bias in particle picking versus 2D classification. **(a)** Confirmation bias during the particle-picking stage, the primary bias mechanism analyzed in this work. **(b)** Confirmation bias during the 2D classification stage. Here, the input consists solely of pure noise particles, and the classification is initialized using predefined templates. Despite the absence of signal in the data, the resulting class averages exhibit a bias toward the initial templates, indicating a positive correlation (i.e., confirmation bias) [4]. Notably, the class centers in Panel (a) appear to more closely resemble the initial templates than those in Panel (b).

tirely of pure noise, and the initial templates were $L = 30$ distinct projections of the Beta-galactosidase structure of size $d = 36 \times 36$. A total of $M = 5 \times 10^5$ particles were processed by EM-based 2D classification for 20 iterations. Even without signal, the resulting class averages remain positively correlated with the templates. Unlike template-based particle picking, however, they do not reproduce the templates exactly, although they remain visually similar. Many modern 2D classification and 3D ab initio reconstruction methods use reference-free initialization, such as random initialization, to mitigate this bias [25]. Nevertheless, similar confirmation effects can still arise when templates are introduced explicitly or implicitly.

6.2 Strategies for mitigating confirmation bias

Previous studies have explored the detection and mitigation of various forms of bias in cryo-EM. For example, recent work [50] has proposed several practical strategies to reduce reconstruction bias, including identifying bias at the level of estimated parameters, performing multiple independent reconstructions, averaging parameter estimates to reduce variance, applying phase randomization at the image level, and validating algorithms across diverse preprocessing and reconstruction settings. Similarly, an additional work [30] provided an empirical analysis of template bias in high-resolution 2D template matching and proposed practical conditions and diagnostics to reduce template-induced bias in reconstruction. Despite these advances, the particle-picking stage remains particularly vulnerable. Here, we

focus specifically on computational strategies to mitigate bias during particle picking, incorporating both practical considerations and new theoretically grounded approaches.

Statistical control of selection thresholds. Despite its importance, the threshold T is typically chosen heuristically or manually, without rigorous statistical justification. A more principled alternative is to set thresholds using multiple hypothesis testing frameworks that control the false discovery rate (FDR) [11]. These techniques guarantee that the expected proportion of false positives remains below a predefined level, while typically achieving high statistical power. As a result, they help reduce the selection of spurious particles that arise due to random correlations with the templates. However, these methods are often specifically tailored to well-specified statistical models and may not generalize easily to more complex or poorly understood settings [36]. Recent developments, such as knockoff-based procedures, offer a promising direction for extending FDR control to more flexible and model-agnostic scenarios by introducing synthetic variables that mimic the data's dependence structure [7].

Template filtering and design. Several tools in current cryo-EM pipelines aim to reduce bias by preprocessing the templates themselves [43]. For example, low-pass filtering the templates reduces their high-frequency features, which are most susceptible to overfitting by noise. This encourages the selection of broader structural motifs rather than fine-grained details, mitigating high-frequency confirmation bias. Our theoretical model (see Section 3) shows that this practice leads to reconstructions that converge toward low-pass-filtered versions of the templates. While not a complete remedy, such filtering is routinely employed and is supported by empirical evidence, including the analysis presented in [5].

Template-free picking and alternative similarity measures. A natural way to mitigate confirmation bias is to reduce or avoid structure-specific templates during particle picking. In practical cryo-EM workflows, this is often done by starting with template-free candidate generation, such as blob- or Laplacian-of-Gaussian (LoG)-based picking, followed by 2D classification. The resulting class averages can then be reused as internal references in later picking rounds. Unlike classical template matching with an external structural reference, blob/LoG methods rely only on weak, generic, low-resolution shape priors rather than molecule-specific templates. Accordingly, based on the results of this work, one expects these methods to favor coarse, blob-like, low-frequency structure rather than detailed structural features.

More broadly, several template-free approaches have been proposed for unsupervised particle detection without external references [21]. Such methods reduce the direct influence of externally imposed structural priors and therefore offer a potentially safer alternative to template-based picking. However, they are also expected to become unreliable below some SNR threshold.

A complementary mitigation strategy within the template-based paradigm is to refine the similarity measure used during matching. Rather than relying on standard cross-correlation, one may consider more robust criteria, such as cross-entropy-based measures [49]. While promising, the design, analysis, and empirical validation of such alternatives remain largely open challenges.

Bypassing particle picking. Below a certain SNR threshold, any particle-picking method is fundamentally doomed to fail [17]. One way to circumvent this limitation is to eliminate particle picking altogether. Recently, it was shown, both theoretically and empirically, that one can reconstruct the 3D structure directly from raw micrographs at arbitrarily low SNR by bypassing particle picking and treating the unknown particle locations as nuisance variables [3]. Earlier direct-structure methods, including approaches based on the method of moments [10] and approximations to EM [28], likewise aim to recover 3D volumes directly from micrographs without explicit particle selection. In principle, these approaches avoid selection bias altogether.

6.3 Future work

Several directions remain open for future work.

Beyond the present picking model. A natural next step is to extend the present theory beyond the pure-noise regime to the practically relevant low-SNR setting, where genuine particles are present. Such an analysis could quantify how genuine signal and template-induced bias interact in downstream reconstruction, and how it depends on the SNR, the density of the true particles, the picking threshold, and the mismatch between the true structure and the picking template.

Another important direction is to develop a theoretical framework for deep particle pickers. Unlike hand-crafted template-matching rules, deep pickers rely on learned scoring functions whose inductive biases are shaped by the training data, architecture, and optimization procedure. A rigorous analysis of such learned selection rules would broaden the scope of the present work and help clarify how modern data-driven picking pipelines may amplify or mitigate bias in low-SNR regimes.

A related extension is to move beyond controlled synthetic micrographs to realistic experimental cryo-EM data. Experimental micrographs contain structured background effects, such as ice gradients, carbon support edges, contamination, and variable ice thickness, that are absent from the idealized pure-noise model studied here.

CTF-aware modeling. For analytical tractability, the present work does not model the CTF explicitly. In the pure-noise regime considered here, a fixed CTF acts as a linear filter, so Gaussian background noise remains Gaussian after CTF filtering, but with a modified covariance structure. From this viewpoint, CTF effects are partially absorbed into the correlated-noise model through the covariance matrix in Theorem 3.2. In practice, however, template matching is often CTF-aware and the CTF may vary across micrographs due to defocus changes. Extending the framework to explicit CTF-aware scoring and heterogeneous defocus is therefore an important direction for future work.

Cryo-ET validation and missing-wedge effects. An important direction for future work is broader empirical and theoretical validation in a practical cryo-ET setting. While the present work includes controlled cryo-ET illustrations on synthetic pure-noise tomograms, these experiments are intended as proof-of-concept demonstrations of the bias mechanism

rather than fully realistic subtomogram-averaging studies. Incorporating key experimental features such as missing-wedge effects [15], tomographic reconstruction artifacts, and validation on real cryo-ET datasets therefore remains for future work.

Data Availability

The detailed implementation and code are available at https://github.com/AmnonBa/particle-picking-confirmation-bias.

Acknowledgment

T.B. is supported in part by BSF under Grant 2020159, in part by NSF-BSF under Grant 2019752, in part by ISF under Grant 1924/21, and in part by a grant from The Center for AI and Data Science at Tel Aviv University (TAD). We thank Amitay Eldar for insightful discussions and for his contributions to the initial formulation of the framework.

References

[1] TW Anderson. An introduction to multivariate statistical analysis. *Wiley series in probability and statistics Show all parts in this series*, 2003.

[2] Xiao-Chen Bai, Greg McMullan, and Sjors HW Scheres. How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences*, 40(1):49–57, 2015.

[3] Amnon Balanov, Tamir Bendory, and Dan Edidin. Orbit recovery under the rigid motions group. *arXiv preprint arXiv:2512.07405*, 2025.

[4] Amnon Balanov, Tamir Bendory, and Wasim Huleihel. Confirmation bias in Gaussian Mixture Models. *IEEE Transactions on Information Theory*, 71(11):8871–8898, 2025.

[5] Amnon Balanov, Wasim Huleihel, and Tamir Bendory. Einstein from noise: Statistical analysis. *arXiv preprint arXiv:2407.05277*, 2024.

[6] Amnon Balanov, Wasim Huleihel, and Tamir Bendory. Expectation-maximization for low-SNR multi-reference alignment. *arXiv preprint arXiv:2505.21435*, 2026.

[7] Rina Foygel Barber and Emmanuel J Candès. Controlling the false discovery rate via knockoffs. *The Annals of statistics*, pages 2055–2085, 2015.

[8] Alberto Bartesaghi, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam. Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences*, 111(32):11709–11714, 2014.

[9] Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE signal processing magazine*, 37(2):58–76, 2020.

[10] Tamir Bendory, Nicolas Boumal, William Leeb, Eitan Levin, and Amit Singer. Toward single particle reconstruction without particle picking: Breaking the detection limit. *SIAM Journal on Imaging Sciences*, 16(2):886–910, 2023.

[11] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[12] Tristan Bepler, Andrew Morin, Micah Rapp, Julia Brasch, Lawrence Shapiro, Alex J Noble, and Bonnie Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods*, 16(11):1153–1160, 2019.

[13] Jochen Böhm, Achilleas S Frangakis, Reiner Hegerl, Stephan Nickell, Dieter Typke, and Wolfgang Baumeister. Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. *Proceedings of the National Academy of Sciences*, 97(26):14245–14250, 2000.

[14] Richard C Bradley. Basic properties of strong mixing conditions. a survey and some open questions. 2005.

[15] Muyuan Chen, James M Bell, Xiaodong Shi, Stella Y Sun, Zhao Wang, and Steven J Ludtke. A complete data processing workflow for cryo-ET and subtomogram averaging. *Nature methods*, 16(11):1161–1168, 2019.

[16] Sergio Cruz-León, Tomáš Majtner, Patrick Hoffmann, Jan P Kreysing, Maarten W Tuijtel, Stefan L Schaefer, Katharina Geißler, Martin Beck, Beata Turoňová, and Gerhard Hummer. High-confidence 3D template matching for cryo-electron tomography. *Biophysical Journal*, 123(3):183a, 2024.

[17] Marom Dadon, Wasim Huleihel, and Tamir Bendory. Detection and recovery of hidden submatrices. *IEEE Transactions on Signal and Information Processing over Networks*, 10:69–82, 2024.

[18] Sanjoy Dasgupta. Learning mixtures of gaussians. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 634–644. IEEE, 1999.

[19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[20] Amitay Eldar, Boris Landa, and Yoel Shkolnisky. KLT picker: Particle picking using data-driven optimal templates. *Journal of structural biology*, 210(2):107473, 2020.

[21] Ayelet Heimowitz, Joakim Andén, and Amit Singer. APPLE picker: Automatic particle picking, a low-effort cryo-EM framework. *Journal of structural biology*, 204(2):215–227, 2018.

[22] Richard Henderson. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041, 2013.

[23] Zhong Huang and Pawel A Penczek. Application of template matching technique to particle detection in electron micrographs. *Journal of Structural Biology*, 145(1-2):29–40, 2004.

[24] Saul M Kassin, Itiel E Dror, and Jeff Kukucka. The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of applied research in memory and cognition*, 2(1):42–52, 2013.

[25] Dari Kimanius, Liyi Dong, Grigory Sharov, Takanori Nakane, and Sjors HW Scheres. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal*, 478(24):4169–4185, 2021.

[26] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[27] Joshua Klayman. Varieties of confirmation bias. *Psychology of learning and motivation*, 32:385–418, 1995.

[28] Shay Kreymer, Amit Singer, and Tamir Bendory. A stochastic approximate expectation-maximization for structure determination directly from cryo-EM micrographs. *arXiv preprint arXiv:2303.02157*, 2023.

[29] Malcolm R Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Science & Business Media, 2012.

[30] Bronwyn A Lucas, Benjamin A Himes, and Nikolaus Grigorieff. Baited reconstruction with 2D template matching for high-resolution structure determination in vitro and in vivo without template bias. *eLife*, 12:RP90486, nov 2023.

[31] Youdong Mao, Luis R Castillo-Menendez, and Joseph G Sodroski. Reply to subramaniam, van heel, and henderson: Validity of the cryo-electron microscopy structures of the HIV-1 envelope glycoprotein complex. *Proceedings of the National Academy of Sciences*, 110(45):E4178–E4182, 2013.

[32] Youdong Mao, Liping Wang, Christopher Gu, Alon Herschhorn, Anik Désormeaux, Andrés Finzi, Shi-Hua Xiang, and Joseph G Sodroski. Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. *Proceedings of the National Academy of Sciences*, 110(30):12438–12443, 2013.

[33] Antonio Martinez-Sanchez. Template matching and machine learning for cryo-electron tomography. *Current Opinion in Structural Biology*, 93:103058, 2025.

[34] Jacqueline LS Milne, Mario J Borgnia, Alberto Bartesaghi, Erin EH Tran, Lesley A Earl, David M Schauder, Jeffrey Lengyel, Jason Pierson, Ardan Patwardhan, and Sriram Subramaniam. Cryo-electron microscopy–a primer for the non-microscopist. *The FEBS journal*, 280(1):28–45, 2013.

[35] Ankur Moitra and Gregory Valiant. Settling the polynomial learnability of mixtures of gaussians. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 93–102. IEEE, 2010.

[36] Keren Mor Waknin, Amitay Eldar, Samuel Davenport, Tamir Bendory, Armin Schwartzman, and Yoel Shkolnisky. Object detection under the linear subspace model with application to cryo-em images. *The Annals of Statistics*, 54(2):910–928, 2026.

[37] Clifford R Mynatt, Michael E Doherty, and Ryan D Tweney. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 29(1):85–95, 1977.

[38] Whitney K Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994.

[39] Eva Nogales. The development of cryo-EM into a mainstream structural biology technique. *Nature methods*, 13(1):24–27, 2016.

[40] Jean-Paul Renaud, Ashwin Chari, Claudio Ciferri, Wen-ti Liu, Hervé-William Rémigy, Holger Stark, and Christian Wiesmann. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature reviews Drug discovery*, 17(7):471–492, 2018.

[41] Douglas A Reynolds et al. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009.

[42] Miroslava Schaffer, Stefan Pfeffer, Julia Mahamid, Stephan Kleindiek, Tim Laugks, Sahradha Albert, Benjamin D Engel, Andreas Rummel, Andrew J Smith, Wolfgang Baumeister, et al. A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native Caenorhabditis elegans tissue. *Nature methods*, 16(8):757–762, 2019.

[43] Sjors HW Scheres. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology*, 180(3):519–530, 2012.

[44] Sjors HW Scheres. Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of structural biology*, 189(2):114–122, 2015.

[45] Sjors HW Scheres and Shaoxia Chen. Prevention of overfitting in cryo-EM structure determination. *Nature methods*, 9(9):853–854, 2012.

[46] Maxim Shatsky, Richard J Hall, Steven E Brenner, and Robert M Glaeser. A method for the alignment of heterogeneous macromolecules from electron microscopy. *Journal of structural biology*, 166(1):67–78, 2009.

[47] Fred J Sigworth. A maximum-likelihood approach to single-particle image refinement. *Journal of structural biology*, 122(3):328–339, 1998.

[48] Amit Singer and Fred J Sigworth. Computational methods for single-particle electron cryomicroscopy. *Annual review of biomedical data science*, 3:163–190, 2020.

[49] Carlos Oscar S Sorzano, JR Bilbao-Castro, Y Shkolnisky, M Alcorlo, R Melero, G Caffarena-Fernández, M Li, G Xu, R Marabini, and JM Carazo. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *Journal of structural biology*, 171(2):197–206, 2010.

[50] COS Sorzano, Amaya Jiménez-Moreno, David Maluenda, Marta Martínez, Erney Ramírez-Aportela, James Krieger, Roberto Melero, Ana Cuervo, Javier Conesa, J Filipovic, et al. On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy. *Biological Crystallography*, 78(4):410–423, 2022.

[51] Sriram Subramaniam. Structure of trimeric HIV-1 envelope glycoproteins. *Proceedings of the National Academy of Sciences*, 110(45):E4172–E4174, 2013.

[52] Martin Turk and Wolfgang Baumeister. The promise and the challenges of cryo-electron tomography. *FEBS letters*, 594(20):3243–3261, 2020.

[53] Marin van Heel. Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, 110(45):E4175–E4177, 2013.

[54] Thorsten Wagner, Felipe Merino, Markus Stabrin, Toshio Moriya, Claudia Antoni, Amir Apelbaum, Philine Hagel, Oleg Sitsel, Tobias Raisch, Daniel Prumbaum, et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications biology*, 2(1):218, 2019.

[55] Abraham Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949.

[56] Abigail JI Watson and Alberto Bartesaghi. Advances in cryo-ET data processing: meeting the demands of visual proteomics. *Current Opinion in Structural Biology*, 87:102861, 2024.

[57] Wilson Wong, Xiao-chen Bai, Alan Brown, Israel S Fernandez, Eric Hanssen, Melanie Condron, Yan Hong Tan, Jake Baum, and Sjors HW Scheres. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife*, 3:e03080, 2014.

[58] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-em. *Nature*, 587(7832):157–161, 2020.

[59] Peijun Zhang. Advances in cryo-electron tomography and subtomogram averaging and classification. *Current opinion in structural biology*, 58:249–258, 2019.

Appendix

Appendix organization. Appendix A provides the preliminaries for the appendix. Appendix B develops the theoretical framework and proofs for our general bias analysis of the GMM mean maximum-likelihood estimators, and its relation to the templates used in the template-matching selector. Appendix C extends this framework to the 3D reconstruction process in both single-particle cryo-EM and cryo-ET. In Appendix D, we describe in detail the empirical setup and procedures used for the simulations presented in Section 5. Appendix 6.1 compares the confirmation bias induced by template-matching selection, as analyzed in this work, with other sources of bias arising in single-particle cryo-EM and cryo-ET.

Notations. Throughout, $\xrightarrow{\mathcal{D}}$, $\xrightarrow{\mathcal{P}}$, $\xrightarrow{\text{a.s.}}$, and $\xrightarrow{\mathcal{L}^p}$ denote convergence in distribution, in probability, almost surely, and in $\mathcal{L}^p$, respectively, for sequences of random variables. The Euclidean inner product is written as either $\langle a, b\rangle$ or $a^\top b$ when $a, b \in \mathbb{R}^d$, and as $\langle a, b\rangle \triangleq a^* b$ when $a, b \in \mathbb{C}^d$. We denote by $\mathsf{SO}(3)$ the special orthogonal group representing 3D rotations. All norms $\|\cdot\|$ are Frobenius norms. $\mathbb{1}$ represents the indicator function.

We employ standard asymptotic notation. For nonnegative sequences a_n and b_n, we write $a_n = O(b_n)$ if there exists a constant $C < \infty$ and n_0 such that $a_n \le C b_n$ for all $n \ge n_0$, and $a_n = o(b_n)$ if $a_n/b_n \to 0$. We write $a_n = \omega(b_n)$ if $a_n/b_n \to \infty$, and $a_n = \Theta(b_n)$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$ hold.

For random sequences, let $\{X_n\}$ be random variables and $\{b_n\}$ a positive deterministic sequence. We write $X_n = O_{\mathbb{P}}(b_n)$ if the family $\{X_n/b_n\}$ is bounded in probability, that is, for every $\varepsilon > 0$ there exists $M < \infty$ such that $\sup_n \mathbb{P}\big(|X_n/b_n| > M\big) \le \varepsilon$. We write $X_n = o_{\mathbb{P}}(b_n)$ if $X_n/b_n \to 0$ in probability, that is, for every $\varepsilon > 0$, $\mathbb{P}\big(|X_n/b_n| > \varepsilon\big) \longrightarrow 0$ as $n \to \infty$.

A Preliminaries

The particle picker's role is to classify each patch as either a particle or noise assuming no overlap between patches. Each micrograph patch, or equivalently, candidate particle is represented as $\{y_i\}_{i=0}^{N-1} \in \mathbb{R}^d$, where N is the number of candidate particles, and d is their dimension (which may correspond to the number of pixels or voxels in 2D image or 3D volume). Accordingly, the particle-picking task is defined by a binary function $f : \mathbb{R}^d \to \{0, 1\}$:

$$f(y_i) = \begin{cases} 1, & \text{if } y_i \text{ is a particle,} \\ 0, & \text{otherwise.} \end{cases} \tag{A.1}$$

To analyze confirmation bias in particle picking, we consider two noise models for the candidate particles $\{y_i\}_{i=0}^{N-1}$. In the Gaussian setting, we assume $y_i \sim \mathcal{N}(0, \Sigma)$, which covers both Model 2.1 (i.i.d. with $\Sigma = \sigma^2 I_d$) and Model 2.3 (stationary Gaussian process with covariance Σ and α-mixing coefficients satisfying Assumption (2.7)). Alternatively, in the spherically symmetric i.i.d. setting (Model 2.2), we assume y_i are i.i.d. with density $p_{\text{sph}}(y)$,

which by the spherical symmetry has the radial form

$$p_{\text{sph}}(y) = \phi(\|y\|_2), \qquad y \in \mathbb{R}^d. \tag{A.2}$$

We consider a particle picker based on *template matching*, as outlined in Algorithm 1. This picker operates by cross-correlating each candidate particle $\{y_i\}_{i=0}^{N-1}$ with L predefined templates, denoted $\{x_\ell\}_{\ell=0}^{L-1}$. Templates are assumed to be normalized so that $\|x_\ell\|_2 = 1$ for all $\ell \in \{0, \ldots, L-1\}$. If the correlation with at least one template exceeds a threshold T, the observation is classified as a particle and added to the set of selected particles, $\mathcal{A}$. We denote by $M = |\mathcal{A}|$ the number of selected particles.

To examine the confirmation bias introduced during the particle selection stage, it is necessary to establish a metric to compare the extracted particles with the initial templates used in the template selection process. Specifically, we analyze two approaches applied to the particles extracted during this stage: one is based on the 2D classification output (relevant for cryo-EM) and the other is based on the 3D reconstruction process (relevant for sub-tomogram averaging in cryo-ET and 3D reconstruction process in single-particle cryo-EM). In Appendix B, we focus on the 2D classification output, modeled using Gaussian mixture models (GMM). We analyze the relationship between the maximum-likelihood estimates of the GMM means and the predefined templates employed during particle picking. In Appendix C, we turn to 3D reconstruction process and analyze the connection between the 3D template volume used in the particle-picking stage and the maximum-likelihood estimator derived from the 3D reconstruction process output. This analysis aims to uncover how the initial templates shape the reconstruction results and whether they contribute to biases in the reconstruction workflow.

We state the following known result for truncated Gaussian distributions which will be used through the proofs.

Lemma A.1 ([19])**.** *Let $X \sim \mathcal{N}(0, \sigma^2)$. Then, for any threshold $T \in \mathbb{R}$,*

$$\mathbb{E}[X \mid X \geq T] = \sigma \cdot \frac{\varphi(T/\sigma)}{Q(T/\sigma)}, \tag{A.3}$$

where $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is the standard normal density function, $Q(z) = \int_z^\infty \varphi(t)\, dt = 1 - \Phi(z)$ is the standard Gaussian upper tail probability, and Φ is the CDF of the standard normal distribution. In addition, the variance of X conditioned on $X \geq T$ is given by:

$$\mathrm{Var}(X \mid X \geq T) = \sigma^2 \left(1 + \frac{T}{\sigma} \cdot \frac{\varphi(T/\sigma)}{Q(T/\sigma)} - \left(\frac{\varphi(T/\sigma)}{Q(T/\sigma)} \right)^2 \right). \tag{A.4}$$

Asymptotically, as $T \to \infty$, we have, [2]

$$\mathbb{E}[X \mid X \geq T] \sim T + \frac{\sigma^2}{T}, \tag{A.5}$$

$$\mathrm{Var}(X \mid X \geq T) \sim \sigma^2 \left(1 - \frac{\sigma^2}{T^2} \right). \tag{A.6}$$

[2]Here, the symbol $f(t) \sim g(t)$ denotes asymptotic equivalence, i.e., $\lim_{t\to\infty} \frac{f(t)}{g(t)} = 1$.

B 2D classification in cryo-EM

B.1 2D classification process overview

In cryo-EM, following the particle-picking stage, the extracted particles are typically subjected to a 2D classification process. This step groups particles into clusters based on their similarity, with each cluster represented by a centroid, which is the average of the observations assigned to that cluster. These centroids serve two primary purposes: they help assess the overall quality of the dataset and provide a basis for iteratively sorting out false positives or poorly aligned particles. In practice, the appearance of particle-like features in the centroids is often interpreted by researchers as evidence that the selected particles contain meaningful structural information.

2D Classification based on Gaussian mixture models. During the 2D classification process, particles are often modeled using a GMM, a widely used and effective probabilistic framework for describing particle heterogeneity. While alternative models exist, we focus on the GMM due to its analytical tractability. In this framework, the observed data are assumed to be generated from a mixture of Gaussian distributions, with each component representing a distinct subpopulation within the dataset. Formally, the realizations of the extracted particles are modeled as:

$$\text{(Postulated statistics)} \qquad z_0, z_1, \ldots, z_{M-1} \overset{\text{i.i.d.}}{\sim} \sum_{\ell=0}^{L-1} w_\ell \cdot \mathcal{N}(\mu_\ell, \Sigma_\ell), \tag{B.1}$$

where $z_0, z_1, \ldots, z_{M-1}$ are the realizations of the *postulated* GMM distribution of the extracted particles, w_ℓ, μ_ℓ, and Σ_ℓ are the weight, mean, and covariance matrix of the ℓ-th Gaussian component, respectively. The weights $\{w_\ell\}_{\ell=0}^{L-1}$ satisfy $w_\ell \geq 0$, and $\sum_{\ell=0}^{L-1} w_\ell = 1$, and are assumed to be known. For the purpose of theoretical modeling, we assume that the number of templates used in the particle-picking process is equal to the number of classes in the GMM, denoted by L. this assumption. This assumption is formally stated below as Assumption B.1.

Assumption B.1. *The number of templates used in the particle extraction algorithm (Algorithm 1) and the number of classes assumed in the GMM during the 2D classification process* (B.1) *are equal to L.*

Remark B.2. *Throughout the Appendix, we use the notation $\{y_i\}_{i=0}^{N-1}$ to denote the candidate particles provided as input to the template-matching particle picker. The resulting set of selected particles, which serves as the input for downstream tasks such as 2D classification and 3D reconstruction, is denoted by $\{z_i\}_{i=0}^{M-1}$.*

The log-likelihood function of GMM. In cryo-EM, it is common to assume that all components of the GMM share the same covariance matrix Σ_{GMM}. The objective of the 2D classification process is to estimate the maximum-likelihood centroids $\{\mu_\ell\}_{\ell=0}^{L-1}$ in (B.1). In

this context, the likelihood of the GMM specified in (B.1) is given by:

$$\mathcal{L}\left(\{\mu_\ell\}_{\ell\in[L]}\,;\mathcal{A}=\{z_i\}_{i\in[M]}\right) = \prod_{i=0}^{M-1} f\left(z_i;\{\mu_\ell\}_{\ell=0}^{L-1}\right)$$
$$= \prod_{i=0}^{M-1}\sum_{\ell=0}^{L-1} w_\ell\cdot\mathcal{N}\left(z_i;\mu_\ell,\Sigma_{\text{GMM}}\right), \tag{B.2}$$

where we have defined,

$$f\left(z;\{\mu_\ell\}_{\ell=0}^{L-1}\right) \triangleq \sum_{\ell=0}^{L-1} w_\ell f_\ell(z) = \sum_{\ell=0}^{L-1} w_\ell\cdot\mathcal{N}\left(z;\mu_\ell,\Sigma_{\text{GMM}}\right). \tag{B.3}$$

Here, $\mathcal{N}(z_i;\mu_\ell,\Sigma_{\text{GMM}})$ represents the multivariate Gaussian density for the ℓ-th component, and the weights w_ℓ are the mixture probabilities for each component. The objective of the 2D classification process is to estimate the centroids $\{\mu_\ell\}_{\ell=0}^{L-1}$, using the maximum-likelihood estimator that maximizes the likelihood (B.2). The maximum-likelihood centroids, denoted by $\widehat{\mu}_\ell$, are formally defined as

$$\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1} = \arg\max_{\{\mu_\ell\}_{\ell=0}^{L-1}} \mathcal{L}(\{\mu_\ell\}_{\ell=0}^{L-1}\,;\mathcal{A}). \tag{B.4}$$

Thus, the aim of the 2D classification is to estimate the centroids of the GMM, defined in (B.1) using the extracted unlabeled particles $\mathcal{A}=\{z_i\}_{i=0}^{M-1}$ (i.e., using the output particles of Algorithm 1). The main question we address in this subsection is how the estimated centroids $\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1}$ relate to the templates $\{x_\ell\}_{\ell=0}^{L-1}$ used in the particle-picker template-matching (Algorithm 1), when the input micrograph is pure noise.

Remark B.3. *We use the notation Σ_{GMM} to denote the covariance matrix of the postulated GMM model described in (B.1), and Σ to denote the covariance matrix of the noise in the input to the particle picker, as described in Algorithm 1. These two covariance matrices are not necessarily equal in our analysis.*

The extracted particles form a non-Gaussian mixture model. Although the extracted particles are often assumed to follow a Gaussian distribution, as defined in (B.1), the statistical properties of the particle picking process described in Algorithm 1 deviate from this assumption. Formally, the law of a particle Z extracted by Algorithm 1 is a finite mixture

$$g(z) = \sum_{\ell=0}^{L-1} \pi_\ell\, g_\ell^{(T)}(z), \qquad z\in\mathbb{R}^d, \tag{B.5}$$

where L is the number of templates, the mixing weights satisfy $\pi_\ell\geq 0$ and $\sum_{\ell=0}^{L-1}\pi_\ell=1$, and $g_\ell^{(T)}:\mathbb{R}^d\to[0,\infty)$ denotes the density associated with template x_ℓ.

In the Gaussian setting, where the underlying patches satisfy $y_i \sim \mathcal{N}(0, \Sigma)$, each component has the truncated-Gaussian form

$$g_\ell^{(T)}(z) = C_\ell^{(T)} \exp\left(-\frac{1}{2} z^\top \Sigma^{-1} z\right) \mathbb{1}_{\{\langle z, x_\ell \rangle \geq T\}}, \tag{B.6}$$

where T is the template–matching threshold and $C_\ell^{(T)}$ is the normalizing constant chosen so that $\int_{\mathbb{R}^d} g_\ell^{(T)}(z)\, dz = 1$, that is,

$$C_\ell^{(T)} = \left(\int_{\{\langle z, x_\ell \rangle \geq T\}} \exp\left(-\frac{1}{2} z^\top \Sigma^{-1} z\right) dz \right)^{-1}. \tag{B.7}$$

In the spherically symmetric i.i.d. noise model (Model 2.2), with base density p_{sph}, the corresponding components are

$$g_\ell^{(T)}(z) = C_\ell^{(T)}\, p_{\text{sph}}(z)\, \mathbb{1}_{\{\langle z, x_\ell \rangle \geq T\}}, \tag{B.8}$$

with $C_\ell^{(T)}$ again defined by the normalization condition $\int_{\mathbb{R}^d} g_\ell^{(T)}(z)\, dz = 1$.

Remark B.4. *In both cases, we denote the resulting component densities by the same symbols $g_\ell^{(T)}$ and $C_\ell^{(T)}$. Thus, $g_\ell^{(T)}$ and $C_\ell^{(T)}$ should be understood as model-dependent: in the Gaussian setting, they are defined via the Gaussian base density, whereas in the spherically symmetric setting, they are defined using p_{sph}. We keep the same notation to avoid introducing separate symbols for each model.*

In other words, $g_\ell^{(T)}$ represents the distribution from which a particle associated with the ℓ-th template is selected, meaning its correlation with the template x_ℓ exceeds the threshold T. Clearly, $g_\ell^{(T)}(z)$ does not follow a Gaussian distribution as it is strictly positive only for $z \in \mathbb{R}^d$ values which satisfy $\langle z, x_\ell \rangle \geq T$. Figure 9 illustrates the probability density function of this truncated Gaussian distribution in one and two dimensions for the case $\Sigma = I_d$.

Misspecification of true statistics and the GMM model. It is important to recognize that the statistical assumptions underlying the particles, specifically, the GMM statistics defined in (B.1), do not accurately reflect the true statistics of the particle picking process based on template matching, as described in (B.5) and (B.6). Consequently, *the models are misspecified*, and this must be considered when estimating the centroids.

Relationship between 2D classification centers and templates. The model presented here raises two central questions:

1. What is the relationship between the mean of each component $g_\ell^{(T)}$ in the true mixture model (B.6) and the corresponding template x_ℓ when the number of particles $M \to \infty$? (answered in Proposition B.5 and Proposition B.6)

2. In the 2D classification process based on a GMM model, how do the templates $\{x_\ell\}_{\ell=0}^{L-1}$ relate to the GMM maximum-likelihood estimators of the means $\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1}$, as specified in (B.4)? (answered in Proposition B.10)

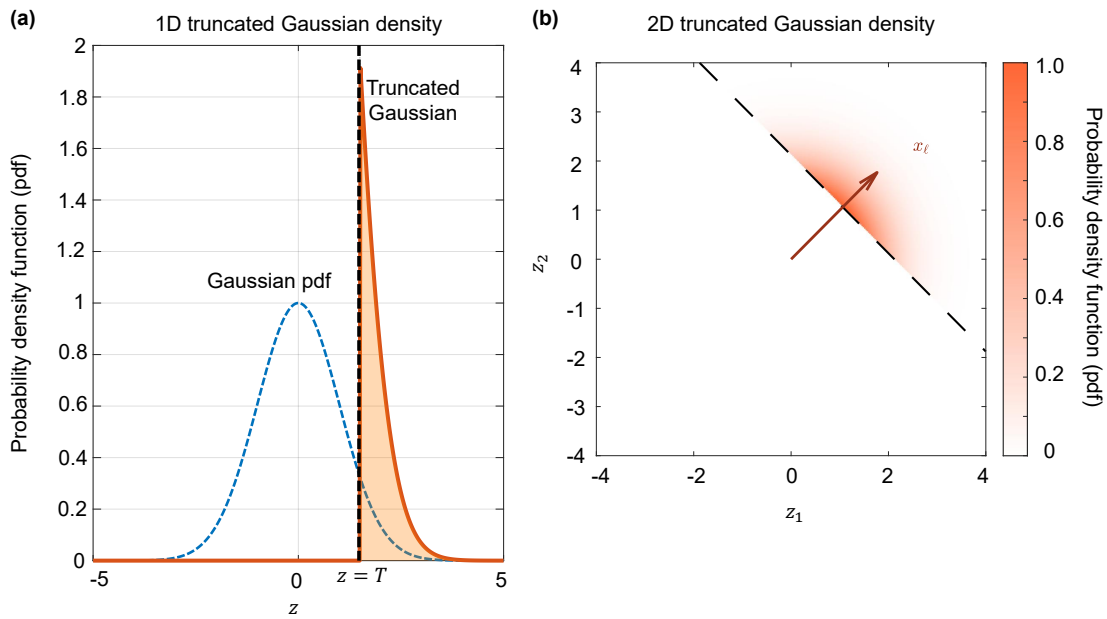

Figure 9: Template-matching-induced truncation of a Gaussian distribution. (a) In one dimension, the original Gaussian density (dashed) is truncated by a hard threshold T, resulting in a re-normalized density $g_\ell(z)$ supported only on the region where $z \geq T$. **(b)** In two dimensions, the original isotropic Gaussian is truncated to the half-space defined by $\langle z, x_\ell \rangle \geq T$, where x_ℓ is the template direction (shown as a red arrow) and the dashed line denotes the threshold boundary.

B.2 Bias of per-template picked particles

We begin by presenting a theoretical result that establishes the relationship between the mean of the ℓ-th component of the mixture model, as defined in (B.6), and the corresponding template x_ℓ.

The centers of the template matching process. Using the definitions in Algorithm 1, let $\widehat{m}_\ell^{(T)}$ be defined as:

$$\widehat{m}_\ell^{(T)} = \frac{1}{|\mathcal{A}_\ell^{(T)}|} \sum_{z_i \in \mathcal{A}_\ell^{(T)}} z_i, \tag{B.9}$$

where

$$\mathcal{A}_\ell^{(T)} = \{z_i \in \mathcal{A} \mid \langle z_i, x_\ell \rangle \geq T\}. \tag{B.10}$$

Here, $\widehat{m}_\ell^{(T)}$ represents the centroid of the particles whose correlation with the template x_ℓ exceeds the threshold T. The key question we address is: what is the relationship between $\widehat{m}_\ell^{(T)}$ and the corresponding template x_ℓ?

To this end, it is convenient to define, for every $\ell \in [L]$, the following sets:

$$\mathcal{V}_\ell^{(T)} \triangleq \left\{ v \in \mathbb{R}^d : \langle v, x_\ell \rangle \geq T \right\}, \tag{B.11}$$

which collect all vectors whose correlation with x_ℓ exceeds the threshold T. Note that $\{\mathcal{V}_\ell^{(T)}\}_{\ell\in[L]}$ are deterministic and are not necessarily disjoint.

We now characterize the limiting behavior of the empirical centers $\widehat{m}_\ell^{(T)}$ under two noise models: general stationary Gaussian noise and spherically symmetric noise.

Stationary Gaussian noise. We first consider the case where the candidate particles arise from a stationary Gaussian process. The next proposition shows that, under mild mixing conditions, the empirical centers align with the templates Σx_ℓ; the proof is given in Appendix B.2.1.

Proposition B.5. *Let $\{y_i\}_{i=0}^{N-1}$ denote the input candidate particles to Algorithm 1, drawn according to either Model 2.1 or Model 2.3. In both cases, each y_i is marginally Gaussian with mean zero and non-singular covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$, and in the stationary model the process is α-mixing with coefficients satisfying the summability condition (2.7).*

Let $\{x_\ell\}_{\ell=0}^{L-1}$ denote the normalized templates, $T \in \mathbb{R}^+$ the template-matching threshold, and $\mathcal{A} \subset \{y_i\}_{i=0}^{N-1}$ the output set of selected particles of Algorithm 1. For each ℓ, let $\mathcal{A}_\ell^{(T)}$ be the corresponding subset defined in (B.10), and let $\widehat{m}_\ell^{(T)}$ be its empirical mean as in (B.9). Then:

1. *For every fixed $T < \infty$ and every $0 \le \ell \le L-1$,*

$$\widehat{m}_\ell^{(T)} = \frac{1}{|\mathcal{A}_\ell^{(T)}|} \sum_{y_i \in \mathcal{A}_\ell^{(T)}} y_i \xrightarrow[N\to\infty]{\text{a.s.}} m_\ell^{(T)}, \tag{B.12}$$

where

$$m_\ell^{(T)} = \mathbb{E}\big[y_0 \,\big|\, \langle y_0, x_\ell\rangle \ge T\big] = \lambda_\ell(T)\,\frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \tag{B.13}$$

for some scalar $\lambda_\ell(T) > 0$. In particular, $m_\ell^{(T)}$ lies in the one-dimensional subspace spanned by Σx_ℓ.

2. *As $T \to \infty$, we have*

$$\lim_{T\to\infty} \frac{m_\ell^{(T)}}{T} = \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \tag{B.14}$$

and hence

$$\lim_{T\to\infty}\lim_{N\to\infty} \frac{\widehat{m}_\ell^{(T)}}{T} = \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \tag{B.15}$$

where the convergence holds almost surely, for every $\ell \in \{0,\dots,L-1\}$.

Spherically symmetric noise. We next specialize to the spherically symmetric setting, where the conditional means align exactly with the templates themselves; the proof of the following proposition is given in Appendix B.2.2.

Proposition B.6. *Let $\{y_i\}_{i=0}^{N-1}$ denote the input candidate particles to Algorithm 1, drawn according to Model 2.2, distributed with common distribution $Y \in \mathbb{R}^d$. Let $\{x_\ell\}_{\ell=0}^{L-1}$ denote the normalized templates, $\|x_\ell\|_2 = 1$, $T \in \mathbb{R}^+$ the template–matching threshold, and $\mathcal{A} \subset \{y_i\}_{i=0}^{N-1}$ the output set of selected particles of Algorithm 1. For each ℓ, let $\mathcal{A}_\ell^{(T)}$ be the corresponding subset defined in (B.10), and let $\widehat{m}_\ell^{(T)}$ be its empirical mean as in (B.9). Then:*

1. *For every fixed $T < \infty$ and every $0 \le \ell \le L-1$,*

$$\widehat{m}_\ell^{(T)} \xrightarrow[N\to\infty]{\text{a.s.}} m_\ell^{(T)} \triangleq \mathbb{E}\big[Y \,\big|\, \langle Y, x_\ell\rangle \ge T\big], \tag{B.16}$$

   and there exists a scalar $\beta(T) \ge T$, independent of ℓ, such that for all $\ell \in \{0, 1, \ldots, L-1\}$,

$$m_\ell^{(T)} = \beta(T)\, x_\ell. \tag{B.17}$$

   In particular, each conditional mean lies in the one-dimensional subspace spanned by the corresponding template x_ℓ.

2. *As $T \to \infty$, we have*

$$\lim_{T\to\infty} \frac{m_\ell^{(T)}}{T} = x_\ell, \tag{B.18}$$

   and hence

$$\lim_{T\to\infty} \lim_{N\to\infty} \frac{\widehat{m}_\ell^{(T)}}{T} = x_\ell, \tag{B.19}$$

   where the convergence holds almost surely, for every $\ell \in \{0, \ldots, L-1\}$.

This result shows that, under Model 2.2, averaging the noise observations $\{y_i\}_{i=0}^{N-1}$ whose correlation with the template x_ℓ exceeds the threshold T converges almost surely to the template x_ℓ, scaled by a factor $\beta(T) \ge T$. Moreover, the second part of the result implies that for asymptotically large thresholds T, the ratio $\beta(T)/T$ approaches 1, so that $\beta(T)$ attains its lower bound asymptotically.

In essence, Propositions B.5 and B.6 establish that the centers of the mixture components induced by template matching resemble the templates (or filtered templates) up to a scalar factor. This is in contrast with the "Einstein from noise" phenomenon and the associated confirmation bias in 2D classification, where the resulting estimates are positively correlated with the initialization but do not fully reproduce it [5].

B.2.1 Proof of proposition B.5

We treat both noise models in a unified way. In either Model 2.1 or Model 2.3, each candidate particle y_i is marginally Gaussian with mean zero and covariance Σ, where Σ is non-singular. In the white-noise model the y_i are i.i.d.; in the stationary model the sequence $\{y_i\}$ is strictly stationary and α-mixing with coefficients satisfying the summability condition (2.7).

Fix ℓ and T, and recall that

$$\mathcal{A}_\ell^{(T)} = \{y_i : \langle y_i, x_\ell \rangle \geq T\}, \qquad \widehat{m}_\ell^{(T)} = \frac{1}{|\mathcal{A}_\ell^{(T)}|} \sum_{y_i \in \mathcal{A}_\ell^{(T)}} y_i. \tag{B.20}$$

We can rewrite $\widehat{m}_\ell^{(T)}$ as a ratio of empirical averages:

$$\widehat{m}_\ell^{(T)} = \frac{\frac{1}{N}\sum_{i=0}^{N-1} y_i \, \mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}}}{\frac{1}{N}\sum_{i=0}^{N-1} \mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}}}. \tag{B.21}$$

Step 1: Law of large numbers (i.i.d. and α-mixing cases). In the white-noise model, the terms $y_i \, \mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}}$ and $\mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}}$ are i.i.d. and integrable, so the classical SLLN applies.

In the stationary case, both sequences $\{y_i \, \mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}}\}$ and $\{\mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}}\}$ are strictly stationary with α-mixing satisfying $\sum_{r=1}^{\infty} \alpha(r)^{\delta/(2+\delta)} < \infty$, and they have finite moments of all orders because the underlying field is Gaussian. Hence, by the strong law of large numbers for α-mixing sequences (see, e.g., [14]), we have almost surely, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=0}^{N-1} y_i \, \mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}} \xrightarrow{\text{a.s.}} \mathbb{E}\big[y_0 \, \mathbb{1}_{\{\langle y_0, x_\ell \rangle \geq T\}}\big], \tag{B.22}$$

$$\frac{1}{N}\sum_{i=0}^{N-1} \mathbb{1}_{\{\langle y_i, x_\ell \rangle \geq T\}} \xrightarrow{\text{a.s.}} \mathbb{P}\big(\langle y_0, x_\ell \rangle \geq T\big) > 0, \tag{B.23}$$

for every finite T. Combining these limits with (B.21) and applying the continuous mapping theorem yields

$$\widehat{m}_\ell^{(T)} \xrightarrow[N\to\infty]{\text{a.s.}} m_\ell^{(T)} \triangleq \frac{\mathbb{E}\big[y_0 \, \mathbb{1}_{\{\langle y_0, x_\ell \rangle \geq T\}}\big]}{\mathbb{P}\big(\langle y_0, x_\ell \rangle \geq T\big)} = \mathbb{E}\big[y_0 \,\big|\, \langle y_0, x_\ell \rangle \geq T\big]. \tag{B.24}$$

This proves the first convergence in part (1).

Step 2: Conditional mean for Gaussian distribution. We now compute $m_\ell^{(T)}$ explicitly for a Gaussian vector $y_0 \sim \mathcal{N}(0, \Sigma)$, with Σ non-singular. Set

$$s \triangleq \langle y_0, x_\ell \rangle = x_\ell^\top y_0. \tag{B.25}$$

Then (y_0, s) is jointly Gaussian with $\mathbb{E}[y_0] = 0$, $\mathbb{E}[s] = 0$, and

$$\mathrm{Var}(s) = x_\ell^\top \Sigma x_\ell, \tag{B.26}$$

$$\mathrm{Cov}(y_0, s) = \mathbb{E}[y_0 s] = \mathbb{E}[y_0 y_0^\top] x_\ell = \Sigma x_\ell. \tag{B.27}$$

By the standard formula for the conditional expectation of a Gaussian vector given a linear observation [1], we have

$$\mathbb{E}[y_0 \mid s] = \Sigma x_\ell \, (x_\ell^\top \Sigma x_\ell)^{-1} s. \tag{B.28}$$

Therefore, by iterated conditioning,

$$m_\ell^{(T)} = \mathbb{E}[y_0 \mid s \geq T] = \mathbb{E}\big[\mathbb{E}[y_0 \mid s] \mid s \geq T\big] = \Sigma x_\ell \, (x_\ell^\top \Sigma x_\ell)^{-1} \mathbb{E}[s \mid s \geq T]. \tag{B.29}$$

Since $s \sim \mathcal{N}\big(0, x_\ell^\top \Sigma x_\ell\big)$, we may write $s = \sqrt{x_\ell^\top \Sigma x_\ell}\, Z$, where $Z \sim \mathcal{N}(0,1)$. Thus

$$\mathbb{E}[s \mid s \geq T] = \sqrt{x_\ell^\top \Sigma x_\ell}\, \mathbb{E}\left[Z \,\middle|\, Z \geq \frac{T}{\sqrt{x_\ell^\top \Sigma x_\ell}} \right]. \tag{B.30}$$

Combining this with (B.29), we obtain

$$m_\ell^{(T)} = \lambda_\ell(T) \, \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \tag{B.31}$$

where

$$\lambda_\ell(T) \triangleq \mathbb{E}[s \mid s \geq T] = \sqrt{x_\ell^\top \Sigma x_\ell}\, \mathbb{E}\left[Z \,\middle|\, Z \geq \frac{T}{\sqrt{x_\ell^\top \Sigma x_\ell}} \right] > 0. \tag{B.32}$$

This shows that $m_\ell^{(T)}$ lies in the one-dimensional subspace spanned by Σx_ℓ, completing part (1).

Step 3: Asymptotics as $T \to \infty$. It remains to analyze the behavior of $m_\ell^{(T)}/T$ as $T \to \infty$. Let $Z \sim \mathcal{N}(0,1)$ with density φ and distribution function Φ. Then, by Lemma A.1

$$\mathbb{E}[Z \mid Z \geq t] = \frac{\int_t^\infty z \, \varphi(z) \, dz}{\int_t^\infty \varphi(z) \, dz} = \frac{\varphi(t)}{1 - \Phi(t)}. \tag{B.33}$$

and using the Mills ratio asymptotics,

$$\frac{\varphi(t)}{1 - \Phi(t)} = t + \frac{1}{t} + O\left(\frac{1}{t^3}\right) \qquad (t \to \infty), \tag{B.34}$$

we obtain, in particular,

$$\mathbb{E}[Z \mid Z \geq t] = t + \frac{1}{t} + O\left(\frac{1}{t^3}\right) \sim t \qquad \text{as } t \to \infty. \tag{B.35}$$

Taking $t = T/\sqrt{x_\ell^\top \Sigma x_\ell}$ in the expression for $\lambda_\ell(T)$, we obtain

$$\frac{\lambda_\ell(T)}{T} = \frac{1}{T} \sqrt{x_\ell^\top \Sigma x_\ell}\, \mathbb{E}\left[Z \,\middle|\, Z \geq \frac{T}{\sqrt{x_\ell^\top \Sigma x_\ell}} \right] = \frac{1}{t} \, \mathbb{E}[Z \mid Z \geq t], \tag{B.36}$$

with $t = T/\sqrt{x_\ell^\top \Sigma x_\ell} \to \infty$ as $T \to \infty$. Thus

$$\lim_{T\to\infty} \frac{\lambda_\ell(T)}{T} = 1. \tag{B.37}$$

Substituting into (B.29) yields

$$\lim_{T\to\infty} \frac{m_\ell^{(T)}}{T} = \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}. \tag{B.38}$$

Combining this with the almost sure convergence $\widehat{m}_\ell^{(T)} \to m_\ell^{(T)}$ from (B.24) gives

$$\lim_{T\to\infty} \lim_{N\to\infty} \frac{\widehat{m}_\ell^{(T)}}{T} = \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \tag{B.39}$$

which completes the proof.

B.2.2 Proof of proposition B.6

Step 1: Law of large numbers. Fix ℓ and T. As in (B.24), we rewrite $\widehat{m}_\ell^{(T)}$ as a ratio of empirical averages. By the strong law of large numbers for α-mixing sequences with a finite $(2+\delta)$-moment (see (2.7) and Model 2.2), we obtain, for each fixed T,

$$\widehat{m}_\ell^{(T)} \xrightarrow[N\to\infty]{\text{a.s.}} m_\ell^{(T)} \triangleq \frac{\mathbb{E}\big[Y\, \mathbb{1}_{\{\langle Y, x_\ell\rangle \geq T\}}\big]}{\mathbb{P}\big(\langle Y, x_\ell\rangle \geq T\big)} = \mathbb{E}\big[Y \,\big|\, \langle Y, x_\ell\rangle \geq T\big]. \tag{B.40}$$

Since the one-dimensional marginal $\langle Y, x_\ell\rangle$ has a continuous density and non-compact support (by spherical symmetry and the assumptions in Model 2.2), we have $\mathbb{P}\big(\langle Y, x_\ell\rangle \geq T\big) > 0$ for every finite T, so the denominator in (B.40) is strictly positive.

Step 2: Direction of the conditional mean. We now show that $m_\ell^{(T)}$ is colinear with x_ℓ. This condition can be equivalently expressed algebraically: for any vector $u \in \mathbb{R}^d$ orthogonal to x_ℓ (i.e., $\langle u, x_\ell\rangle = 0$), the mean $m_\ell^{(T)}$ must also be orthogonal to u i.e.,

$$\langle u, m_\ell^{(T)}\rangle = 0. \tag{B.41}$$

To prove (B.41), we state the following Lemma.

Lemma B.7. *Let $\mathcal{P}(y) = p_{\text{sph}}(y)$ be the distribution of the random variable Y according to Model 2.2. Assume $u \in \mathbb{R}^d$, satisfying $\langle u, x_\ell\rangle = 0$. Then, for every T,*

$$\langle u, m_\ell^{(T)}\rangle\, \mathbb{P}\left[\langle Y, x_\ell\rangle \geq T\right] = \int_{\mathbb{R}^d} \langle y, u\rangle \mathcal{P}(y)\, \mathbb{1}_{\langle y, x_\ell\rangle \geq T}\, dy = 0. \tag{B.42}$$

In other words, Lemma B.7 shows that x_ℓ and $m_\ell^{(T)}$ are collinear, which is equivalent to

$$m_\ell^{(T)} = \alpha_\ell^{(T)}\, x_\ell, \tag{B.43}$$

where,

$$\alpha_\ell^{(T)} \triangleq \langle m_\ell^{(T)}, x_\ell \rangle \in \mathbb{R}. \tag{B.44}$$

In Lemma B.8, stated below, we prove that $\alpha_\ell^{(T)} \geq T$, and that $\alpha_\ell^{(T)}$ is independent of ℓ, i.e., $\alpha_{\ell_1}^{(T)} = \alpha_{\ell_2}^{(T)}$, for every $\ell_1, \ell_2 \in [L]$, which completes the proof of the first part of the proposition.

Lemma B.8. *Recall the definition of $m_\ell^{(T)}$ in (B.40). For every $\ell_1, \ell_2 \in [L]$, the following holds,*

$$\alpha_{\ell_1}^{(T)} = \langle m_{\ell_1}^{(T)}, x_{\ell_1} \rangle = \langle m_{\ell_2}^{(T)}, x_{\ell_2} \rangle = \alpha_{\ell_2}^{(T)} \geq T. \tag{B.45}$$

Following Lemma B.8, we denote by $\beta(T) \triangleq \alpha_\ell^{(T)}$, which is independent of ℓ.

Step 3: Asymptotics as $T \to \infty$. From (B.40) and the discussion above, we have

$$\beta(T) = \mathbb{E}[Z \mid Z > T], \tag{B.46}$$

where Z is the common one–dimensional marginal of $\langle Y, u \rangle$ for any unit vector u. By assumption in Model 2.2,

$$\lim_{T\to\infty} \frac{\beta(T)}{T} = \lim_{T\to\infty} \frac{\mathbb{E}[Z \mid Z > T]}{T} = 1. \tag{B.47}$$

Therefore,

$$\lim_{T\to\infty} \frac{m_\ell^{(T)}}{T} = \lim_{T\to\infty} \frac{\beta(T)}{T} x_\ell = x_\ell, \tag{B.48}$$

for every ℓ. Combining this with the almost sure convergence $\widehat{m}_\ell^{(T)} \to m_\ell^{(T)}$ from (B.40), we obtain

$$\lim_{T\to\infty} \lim_{N\to\infty} \frac{\widehat{m}_\ell^{(T)}}{T} = x_\ell, \tag{B.49}$$

for each $0 \leq \ell \leq L-1$. This proves part (2) and completes the proof. It remains to prove the lemmas.

Proof of Lemma B.7. Fix $u \in \mathbb{R}^d$, such that $\langle u, x_\ell \rangle = 0$. First, we show that for every $v \in \mathcal{V}_\ell^{(T)}$, there is a unique corresponding $v' \in \mathcal{V}_\ell^{(T)}$ such that $\langle u, v \rangle = -\langle u, v' \rangle$, and $\mathcal{P}(v) = \mathcal{P}(v')$. Every $v \in \mathbb{R}^d$ can be decomposed uniquely by $v = \gamma_v u + w$, where $w \in \mathbb{R}^d$ orthogonal to u, $\langle u, w \rangle = 0$. Define $v' = -\gamma_v u + w$. Clearly, the map $v \to v'$ is unique, and due to the orthogonality between u and w, we have, $\langle u, v \rangle = \gamma_v$, and $\langle u, v' \rangle = -\gamma_v$; thus, $\langle u, v \rangle = -\langle u, v' \rangle$. In addition, $v \in \mathcal{V}_\ell^{(T)}$ if and only if $v' \in \mathcal{V}_\ell^{(T)}$, since,

$$\begin{aligned} v \in \mathcal{V}_\ell^{(T)} &\iff \langle v, x_\ell \rangle \geq T \iff \langle \gamma_v u + w, x_\ell \rangle \geq T \\ &\iff \langle w, x_\ell \rangle \geq T \end{aligned} \tag{B.50}$$

$$\iff \langle -\gamma_v u + w, x_\ell \rangle \geq T \iff \langle v', x_\ell \rangle \geq T \iff v' \in \mathcal{V}_\ell^{(T)},$$

where the transition in (B.50) follows from the orthogonality between u and x_ℓ.

Next, we show that $\mathcal{P}(v) = \mathcal{P}(v')$. As $\mathcal{P}(v)$ is a spherically symmetric function that depends solely on the norm of v. Thus, we prove below that $\|v\| = \|v'\|$, which would prove that $\mathcal{P}(v) = \mathcal{P}(v')$:

$$\|v\|^2 = \|\gamma_v u + w\|^2 = \gamma_v^2 \|u\|^2 + \|w\|^2 = \|-\gamma_v u + w\|^2 = \|v'\|^2, \tag{B.51}$$

where in the second equality we have used the orthogonality between u and w. Therefore, $\|v\| = \|v'\|$, which shows that $\mathcal{P}(v) = \mathcal{P}(v')$.

Next, note that (B.42) can be rewritten as follows,

$$\begin{aligned}\int_{\mathbb{R}^d} \langle y, u \rangle \mathcal{P}(y) \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy &= \int_{y \in \mathcal{V}_\ell^{(T)}} \langle y, u \rangle \mathcal{P}(y) \, dy \\ &= \int_{y \in \mathcal{V}_\ell^{(T)} : \langle y, u \rangle > 0} \langle y, u \rangle \mathcal{P}(y) \, dy + \int_{y \in \mathcal{V}_\ell^{(T)} : \langle y, u \rangle < 0} \langle y, u \rangle \mathcal{P}(y) \, dy. \end{aligned} \tag{B.52}$$

Thus, since for every $v \in \mathcal{V}_\ell^{(T)}$ there is a unique corresponding $v' \in \mathcal{V}_\ell^{(T)}$ such that $\langle v, u \rangle = -\langle v', u \rangle$ and $\mathcal{P}(v) = \mathcal{P}(v')$, it is clear that the r.h.s. of (B.52) is zero, which completes the proof. □

Proof of Lemma B.8. First, we show that for every $\ell_1, \ell_2 \in [L]$, $\langle x_{\ell_1}, m_{\ell_1}^{(T)} \rangle = \langle x_{\ell_2}, m_{\ell_2}^{(T)} \rangle$, which proves that α is independent of ℓ. As the templates $\{x_\ell\}_{\ell=0}^{L-1}$ are normalized, there is a orthogonal matrix O_{ℓ_1,ℓ_2}, such that $O_{\ell_1,\ell_2} x_{\ell_1} = x_{\ell_2}$. Then, following from the right-hand-side of (B.42), we have,

$$\langle x_{\ell_1}, m_{\ell_1}^{(T)} \rangle \mathbb{P}[\langle Y, x_{\ell_1} \rangle \geq T] = \int_{\mathbb{R}^d} \langle y, x_{\ell_1} \rangle \mathcal{P}(y) \mathbb{1}_{\langle y, x_{\ell_1} \rangle \geq T} \, dy \tag{B.53}$$

$$= \int_{\mathbb{R}^d} \langle y, O_{\ell_2,\ell_1} x_{\ell_2} \rangle \mathcal{P}(y) \mathbb{1}_{\langle y, O_{\ell_2,\ell_1} x_{\ell_2} \rangle \geq T} \, dy \tag{B.54}$$

$$= \int_{\mathbb{R}^d} \langle O_{\ell_2,\ell_1} y, O_{\ell_2,\ell_1} x_{\ell_2} \rangle \mathcal{P}(O_{\ell_2,\ell_1} y) \mathbb{1}_{\langle O_{\ell_2,\ell_1} y, O_{\ell_2,\ell_1} x_{\ell_2} \rangle \geq T} \, dy \tag{B.55}$$

$$= \int_{\mathbb{R}^d} \langle y, x_{\ell_2} \rangle \mathcal{P}(O_{\ell_2,\ell_1} y) \mathbb{1}_{\langle y, x_{\ell_2} \rangle \geq T} \, dy \tag{B.56}$$

$$= \int_{\mathbb{R}^d} \langle y, x_{\ell_2} \rangle \mathcal{P}(y) \mathbb{1}_{\langle y, x_{\ell_2} \rangle \geq T} \, dy \tag{B.57}$$

$$= \langle x_{\ell_2}, m_{\ell_2}^{(T)} \rangle \mathbb{P}[\langle Y, x_{\ell_2} \rangle \geq T], \tag{B.58}$$

where (B.53) follows from the definition of $m_\ell^{(T)}$; (B.54) follows from the orthonormal operator definition of O_{ℓ_2,ℓ_1}; (B.55) follows from substitution of variables $y \to O_{\ell_2,\ell_1} y$; (B.56) follows from the property of orthonormal operator O, which satisfies, $\langle Oa, Ob \rangle = \langle a, b \rangle$ for every $a, b \in \mathbb{R}^d$; (B.57) follows from the spherical symmetry property of $\mathcal{P}(y) = \mathcal{P}(Oy)$, that is for every orthonormal operator O and for every $y \in \mathbb{R}^d$, $\|Oy\| = \|y\|$; and (B.58) follows from the definition of $m_\ell^{(T)}$. In addition, by the spherical symmetry of Y, we have,

$$\mathbb{P}[\langle Y, x_{\ell_1} \rangle \geq T] = \mathbb{P}[\langle Y, x_{\ell_2} \rangle \geq T] \tag{B.59}$$

Next, we show that $\alpha_\ell \geq T$. Similar to Lemma B.7, for every $y \in \mathcal{V}_\ell^{(T)}$, we can decompose it into $y = \beta_y x_\ell + \gamma_y w_y$, where $\langle w_y, x_\ell \rangle = 0$, and $\beta_y \geq T$, by the definition of the domain of the integral. Thus, we can decompose the integral in (B.42) as follows:

$$\int_{\mathbb{R}^d} \langle y, x_\ell \rangle \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy = \int_{\mathbb{R}^d} \langle \beta_y x_\ell, x_\ell \rangle \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy + \int_{\mathbb{R}^d} \langle \gamma_y w_y, x_\ell \rangle \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy \tag{B.60}$$

By Lemma B.7, there exists $y' = \beta_y x_\ell - \gamma_y w_y$, such that $\|y'\| = \|y\|$, and $\langle y', x_\ell \rangle \geq T$. Thus, as $\mathcal{P}$ is spherically symmetric, that is, $\mathcal{P}(y) = \mathcal{P}(y')$ for $\|y\| = \|y'\|$, the last term in (B.60) vanishes:

$$\int_{\mathbb{R}^d} \langle \gamma_y w_y, x_\ell \rangle \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy = 0. \tag{B.61}$$

Substituting (B.61) into (B.60) results,

$$\begin{aligned} \int_{\mathbb{R}^d} \langle y, x_\ell \rangle \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy &= \int_{\mathbb{R}^d} \langle \beta_y x_\ell, x_\ell \rangle \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy \\ &= \int_{\mathbb{R}^d} \beta_y \|x_\ell\|^2 \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy \\ &= \int_{\mathbb{R}^d} \beta_y \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy, \end{aligned} \tag{B.62}$$

where we have used $\|x_\ell\| = 1$. As $\beta_y \geq T$ for every $y \in \mathcal{V}_\ell^{(T)}$, we have,

$$\alpha_\ell = \frac{\int_{\mathbb{R}^d} \beta_y \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy}{\int_{\mathbb{R}^d} \mathcal{P}(y) \, \mathbb{1}_{\langle y, x_\ell \rangle \geq T} \, dy} \geq T, \tag{B.63}$$

which completes the proof of the lemma. □

B.3 GMM maximum likelihood on template-selected particles

In the previous section, we showed that the particle-picking process induces a mixture model with L components, where L is the number of templates. The corresponding component means $\{m_\ell^{(T)}\}_{\ell=0}^{L-1}$ reproduce the templates (up to a scaling factor that depends on the threshold T and the noise model). However, this does not automatically imply that the output of the maximum-likelihood estimation for a GMM, recovers these centroids. The key difficulty is that the GMM is only a *postulated* model: the true distribution of the picked particles is not Gaussian.

In general, under model correctness, maximum-likelihood estimators converge to the true parameter values [38]. In particular, if the extracted particles were in fact drawn from a GMM with component means $\{m_\ell^{(T)}\}$, then the maximum-likelihood estimators $\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1}$ in (2.12) would converge to $\{m_\ell^{(T)}\}$. In our setting, however, the postulated model is an L-component GMM, while the *actual* statistics of the picked particles are described by the

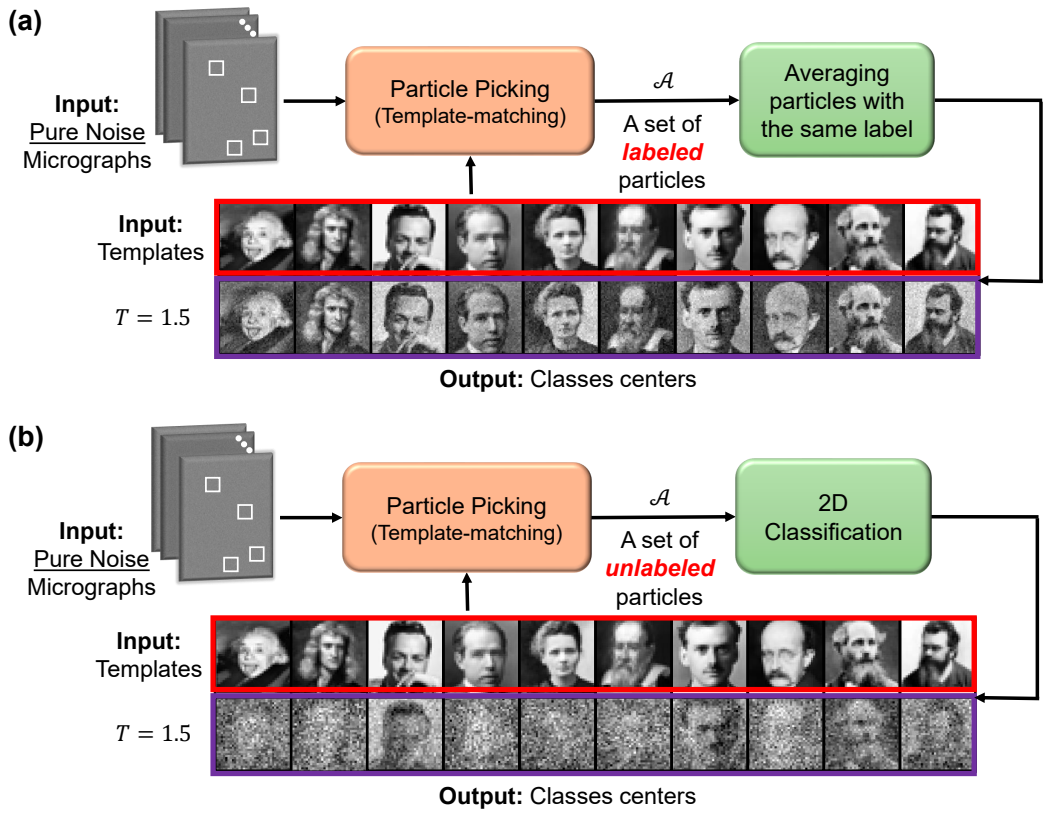

Figure 10: Illustration of the relationship between the means for labeled particles (Proposition B.5) and unlabeled particles (Proposition B.10). Starting from pure-noise observations, template-based particle picking selects highly correlated candidates. **(a)** When the particles are labeled, that is, each selected particle is assigned to a known template, Proposition B.5 characterizes the asymptotic behavior of the mean of particles within each label. These means converge to the corresponding templates up to a scaling factor, and this convergence holds for any threshold T. **(b)** When the particles are unlabeled, i.e., the problem involves latent variables, the 2D classification step treats the data as a GMM. Proposition B.10 shows that as the threshold increases, the estimated class centers increasingly align with the original templates (see Figure 4). However, at low thresholds, this property does not necessarily occur as can be seen in the example. This simulation uses $M = 10^5$ particles collected during the particle-picking process and image sizes of $d = 50 \times 50$. The data generation model for this experiment is the same as presented in Model 2.1.

mixture in (B.5)-(B.6), which is non-Gaussian. We must therefore analyze the behavior of the GMM maximum-likelihood estimator under model misspecification.

The connection between Proposition B.5 and the GMM mean estimators is illustrated in Figure 10. In the labeled setting of Proposition B.5, where the particle-template assignments are known, the empirical means converge directly to scaled versions of the templates. In contrast, in the unlabeled setting underlying GMM-based mean estimation, the component means converge to the templates only in the high-threshold regime.

Recall the postulated GMM model from (B.3)

$$f\big(z;\{\mu_\ell\}_{\ell=0}^{L-1}\big) = \sum_{\ell=0}^{L-1} w_\ell f_\ell(z) = \sum_{\ell=0}^{L-1} w_\ell \mathcal{N}\big(z;\mu_\ell,\Sigma_{\mathrm{GMM}}\big), \tag{B.64}$$

and the true mixture model of the extracted particles

$$g\big(z;\{m_\ell^{(T)}\}_{\ell=0}^{L-1}\big) = \sum_{\ell=0}^{L-1} \pi_\ell\, g_\ell^{(T)}(z), \tag{B.65}$$

where $g_\ell^{(T)}$ is defined by (B.6) in the Gaussian case and by (B.8) in the spherical symmetric case. The vectors $m_\ell^{(T)} \in \mathbb{R}^d$ are the component means of the true mixture,

$$m_\ell^{(T)} \triangleq \mathbb{E}_{g_\ell^{(T)}}[Z] = \int_{\mathbb{R}^d} z\, g_\ell^{(T)}(z)\, dz. \tag{B.66}$$

By Proposition B.5 and Proposition B.6,

$$\lim_{T\to\infty} \frac{m_\ell^{(T)}}{T} = v_\ell \neq 0, \tag{B.67}$$

where

$$v_\ell = \begin{cases} x_\ell, & \text{under Model 2.2,} \\ \dfrac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, & \text{under } y_i \sim \mathcal{N}(0,\Sigma). \end{cases} \tag{B.68}$$

Moreover, for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left\| \frac{Y}{T} - v_\ell \right\|_2 > \varepsilon \;\middle|\; \langle x_\ell, Y\rangle \geq T \right) \xrightarrow[T\to\infty]{} 0, \tag{B.69}$$

which follows from (2.9) for the spherically symmetric model and can be verified directly in the Gaussian case. Intuitively, conditional on exceeding a large threshold in direction x_ℓ, the picks concentrate in a narrow cone around the direction v_ℓ with radius $O(T)$.

The output of Algorithm 1 is the set of extracted particles $\mathcal{A} = \{z_i\}_{i=0}^{M-1} \subset \mathbb{R}^d$, which we model as a stationary sequence with marginal density g. The GMM maximum-likelihood estimators of the component means are defined by

$$\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1} \triangleq \operatorname*{arg\,max}_{\{\mu_\ell\}_{\ell=0}^{L-1}} \prod_{i=0}^{M-1} f\big(z_i;\{\mu_\ell\}_{\ell=0}^{L-1}\big). \tag{B.70}$$

For fixed $T < \infty$, it is convenient to introduce the empirical average log-likelihood

$$\mathcal{L}_{M,T}(\{\mu_\ell\}) \triangleq \frac{1}{M} \sum_{i=0}^{M-1} \log f\big(z_i;\{\mu_\ell\}\big), \tag{B.71}$$

and the corresponding population objective

$$\mathcal{H}_T(\{\mu_\ell\}) \triangleq \mathbb{E}_{Z\sim g(\cdot;\{m_\ell^{(T)}\})}\big[\log f(Z;\{\mu_\ell\})\big]. \tag{B.72}$$

The next lemma is the standard misspecified maximum-likelihood estimation fact: the GMM maximum-likelihood estimator converges to the KL minimizer between g and the GMM family $\{f(\cdot;\{\mu_\ell\})\}$.

Lemma B.9 (maximum-likelihood estimator under model misspecification)**.** *Let $\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1}$ be the maximum-likelihood estimators defined in (B.70) for the postulated GMM model with density f in (B.64), and let g be the true density of the extracted particles, given by (B.65). Assume that the underlying process $\{z_i\}$ is strictly stationary and α-mixing with summable coefficients (2.7). Then, as $M\to\infty$,*

$$\{\widehat{\mu}_\ell\}_{\ell=0}^{L-1} \xrightarrow[M\to\infty]{\text{a.s.}} \operatorname*{arg\,min}_{\{\mu_\ell\}_{\ell=0}^{L-1}} D_{\mathrm{KL}}\Big(g(\cdot;\{m_\ell^{(T)}\}) \,\Big\|\, f(\cdot;\{\mu_\ell\})\Big), \tag{B.73}$$

where D_{KL} denotes the Kullback–Leibler divergence. In particular, any sequence of maximizers of $\mathcal{L}_{M,T}$ converges almost surely to a maximizer of $\mathcal{H}_T$.

Lemma B.9 shows that, for each fixed threshold T, the GMM maximum-likelihood estimator finds in the limit $M\to\infty$ the Gaussian mixture that is closest in KL divergence to the true mixture $g(\cdot;\{m_\ell^{(T)}\})$. We now analyze this KL minimizer in the high-threshold regime.

Proposition B.10 (High-threshold alignment of GMM centers)**.** *For each (N,T), let $\{z_i\}_{i=0}^{M-1}$ be the extracted particles, modeled as a stationary sequence with marginal density (B.65), where $M=M(N,T)$ is the number of picks produced by Algorithm 1. Recall the definition of v_ℓ from (B.68). Let*

$$\{\widehat{\mu}_\ell(N,T)\}_{\ell=0}^{L-1} \in \operatorname*{arg\,max}_{\{\mu_\ell\}} \prod_{i=0}^{M-1} f\big(z_i;\{\mu_\ell\}\big) \tag{B.74}$$

denote the corresponding GMM maximum-likelihood estimators of the component means. Assume that the underlying process is α-mixing with summable coefficients (2.7). Then there exists a permutation π of $\{0,\dots,L-1\}$ such that

$$\lim_{T\to\infty}\lim_{N\to\infty}\frac{\widehat{\mu}_{\pi(\ell)}(N,T)}{T} = v_\ell, \qquad 0\le \ell\le L-1, \tag{B.75}$$

where the convergence holds in probability.

Proof of Proposition B.10. We prove in steps.

Step 1: Population limit for fixed T. By Lemma B.9 and the assumption $M(N,T)\to\infty$ as $N\to\infty$, any sequence of GMM maximum-likelihood estimators $\{\widehat{\mu}_\ell(N,T)\}$ converges almost surely, for each fixed T, to a maximizer $\{\mu_\ell^\star(T)\}$ of the population objective $\mathcal{H}_T$, i.e.

$$\{\mu_\ell^\star(T)\} \in \operatorname*{arg\,max}_{\{\mu_\ell\}} \mathcal{H}_T(\{\mu_\ell\}). \tag{B.76}$$

Step 2: High-threshold structure of the population objective. We next show that there exists a permutation π, such that the following holds as $T \to \infty$:

$$\max_{0 \le \ell \le L-1} \|\mu^\star_{\pi(\ell)}(T) - m_\ell^{(T)}\|_2 = o(T). \tag{B.77}$$

Using the mixture representation (B.65), we can decompose

$$\mathcal{H}_T(\{\mu_\ell\}) = \sum_{\ell=0}^{L-1} \pi_\ell \, \mathbb{E}_{Z \sim g_\ell^{(T)}} \big[\log f(Z; \{\mu_\ell\})\big]. \tag{B.78}$$

By (B.69), if $Z \sim g_\ell^{(T)}$ then

$$\frac{Z}{T} \xrightarrow[T\to\infty]{\mathbb{P}} v_\ell, \qquad \frac{Z - m_\ell^{(T)}}{T} \xrightarrow[T\to\infty]{\mathbb{P}} 0. \tag{B.79}$$

The log-density of the k-th Gaussian component of f satisfies

$$\log \mathcal{N}(Z; \mu_k, \Sigma_{\mathrm{GMM}}) = -\frac{1}{2}(Z - \mu_k)^\top \Sigma_{\mathrm{GMM}}^{-1}(Z - \mu_k) + C, \tag{B.80}$$

for a constant C independent of Z and μ_k. Expanding the quadratic form at $Z = m_\ell^{(T)} + R_T$ with $R_T/T \to 0$ in probability, we obtain

$$\begin{aligned}(Z - \mu_k)^\top \Sigma_{\mathrm{GMM}}^{-1}(Z - \mu_k) &= (m_\ell^{(T)} - \mu_k)^\top \Sigma_{\mathrm{GMM}}^{-1}(m_\ell^{(T)} - \mu_k) \\ &\quad + 2R_T^\top \Sigma_{\mathrm{GMM}}^{-1}(m_\ell^{(T)} - \mu_k) + R_T^\top \Sigma_{\mathrm{GMM}}^{-1} R_T.\end{aligned} \tag{B.81}$$

Since the cross term is $O_{\mathbb{P}}(\|R_T\|_2 T) = o_{\mathbb{P}}(T^2)$ and the remainder term is $O_{\mathbb{P}}(\|R_T\|_2^2) = o_{\mathbb{P}}(T^2)$, we have,

$$(Z - \mu_k)^\top \Sigma_{\mathrm{GMM}}^{-1}(Z - \mu_k) = (m_\ell^{(T)} - \mu_k)^\top \Sigma_{\mathrm{GMM}}^{-1}(m_\ell^{(T)} - \mu_k) + o_{\mathbb{P}}(T^2), \tag{B.82}$$

uniformly over configurations with $\|m_\ell^{(T)} - \mu_k\|_2 = O(T)$. Consequently, the leading contribution to $\mathbb{E}_{Z \sim g_\ell^{(T)}}[\log f(Z; \{\mu_\ell\})]$ is of order $-T^2$ and is minimized (i.e., the log-likelihood is maximized) by choosing at least one center μ_k close to $m_\ell^{(T)}$ on the scale T. Thus, for T large enough, any maximizer $\{\mu_\ell^\star(T)\}$ of $\mathcal{H}_T$ must satisfy

$$\min_k \|\mu_k^\star(T) - m_\ell^{(T)}\|_2 = o(T), \qquad 0 \le \ell \le L-1. \tag{B.83}$$

Since the directions $\{v_\ell\}$ are distinct, the high-threshold cones around $\{m_\ell^{(T)}\}$ become disjoint on the scale T, and we can index the centers so that there exists a permutation π with

$$\max_{0 \le \ell \le L-1} \|\mu^\star_{\pi(\ell)}(T) - m_\ell^{(T)}\|_2 = o(T). \tag{B.84}$$

Dividing (B.84) by T and using $m_\ell^{(T)}/T \to v_\ell$ yields

$$\max_{0 \le \ell \le L-1} \left\| \frac{\mu^\star_{\pi(\ell)}(T)}{T} - v_\ell \right\|_2 \xrightarrow[T\to\infty]{} 0. \tag{B.85}$$

Thus, we obtain for every $0 \le \ell \le L-1$,

$$\frac{\mu^\star_{\pi(\ell)}(T)}{T} \xrightarrow[T\to\infty]{} v_\ell. \tag{B.86}$$

Step 3: Combining the limits. Fix ℓ. For each T, (B.76) and Lemma B.9 imply that

$$\widehat{\mu}_{\pi(\ell)}(N,T) \xrightarrow[N\to\infty]{\text{a.s.}} \mu^{\star}_{\pi(\ell)}(T), \tag{B.87}$$

while (B.86) gives

$$\frac{\mu^{\star}_{\pi(\ell)}(T)}{T} \xrightarrow[T\to\infty]{} v_\ell. \tag{B.88}$$

Combining these yields the double limit

$$\lim_{T\to\infty}\lim_{N\to\infty}\frac{\widehat{\mu}_{\pi(\ell)}(N,T)}{T} = v_\ell, \tag{B.89}$$

with convergence in probability, which is (B.75). □

B.4 Proof of Theorem 3.1 and Theorem 3.2

The proof is a direct consequence of Proposition B.5, Proposition B.6 and Proposition B.10. For each (N,T), let $\{\widehat{\mu}_\ell(N,T)\}_{\ell=0}^{L-1}$ denote the GMM maximum-likelihood estimators of the component means based on the extracted set $\mathcal{A}$, as in (B.74). By Proposition B.5 and Proposition B.6, the means $\{m_\ell^{(T)}\}$ of the true mixture components satisfy

$$\lim_{T\to\infty}\frac{m_\ell^{(T)}}{T} = v_\ell \neq 0, \tag{B.90}$$

with

$$v_\ell = \begin{cases} x_\ell, & \text{under Model 2.2,} \\ \dfrac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, & \text{under } y_i \sim \mathcal{N}(0,\Sigma). \end{cases} \tag{B.91}$$

In the spherically symmetric setting of the theorem (Model 2.1 with $\Sigma \propto I_d$ or Model 2.2), we have $v_\ell = x_\ell$ for all ℓ.

By Proposition B.10, that there exists a permutation π of $\{0,\dots,L-1\}$ such that

$$\lim_{T\to\infty}\lim_{N\to\infty}\frac{\widehat{\mu}_{\pi(\ell)}(N,T)}{T} = v_\ell, \qquad 0 \le \ell \le L-1, \tag{B.92}$$

where the convergence holds in probability. This proves the two theorems for the spherically symmetric case with $v_\ell = x_\ell$ and for the stationary Gaussian case with $v_\ell = \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}$.

B.5 Proof of Proposition 3.3

Fix $L \ge 2$ and templates $\{x_\ell\}_{\ell=0}^{L-1} \subset \mathbb{R}^d$ with $x_{\ell_1} \neq x_{\ell_2}$ for $\ell_1 \neq \ell_2$. Let $Y \sim \mathcal{N}(0,\Sigma)$ with $\Sigma \succ 0$, and consider the template–matching selection rule at threshold $T > 0$ as in

Algorithm 1. Recall the definition of $g_T(z)$ from (B.6), and recall the postulated Gaussian mixture model $f(z;\{\mu_k\})$ from (B.3). Recall the definition of $\{\mu_k^\star(T)\}$ from (B.76)

$$\{\mu_k^\star(T)\} \in \arg\min_{\{\mu_k\}} D_{\mathrm{KL}}\big(g_T \,\|\, f(\cdot;\{\mu_k\})\big). \tag{B.93}$$

and recall that $\widehat{\mu}_\ell(\mathcal{A})$ denotes the mixture-model maximum-likelihood estimator of the ℓ-th class mean obtained from the extracted set of patches $\mathcal{A}$. Set

$$v_\ell \triangleq \frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}, \qquad \sigma_\ell^2 \triangleq x_\ell^\top \Sigma x_\ell. \tag{B.94}$$

We first establish auxiliary lemmas and then prove the proposition. Lemma B.11 is proved in the end of the end of the subsection.

Lemma B.11. *There exist constants $T_0 > 0$ and $C_{\mathrm{bias}} > 0$ (depending only on Σ, $\{x_\ell\}$, and $\{w_k\}$) and a permutation π of $\{0,\dots,L-1\}$ such that, for all $T \geq T_0$ and all $0 \leq \ell \leq L-1$,*

$$\big\|\mu_{\pi(\ell)}^\star(T) - T\,v_\ell\big\|_2 \leq \frac{C_{\mathrm{bias}}}{T}. \tag{B.95}$$

Lemma B.12. *Under the assumption of Proposition 3.3 There exist constants $M_0 \geq 1$ and $C_{\mathrm{var}} > 0$, independent of d, M, T, such that for all $M \geq M_0$ and $T \geq T_0$,*

$$\mathbb{E}\big\|\widehat{\mu}_\ell(\mathcal{A}) - \mu_\ell^\star(T)\big\|_2^2 \leq \frac{C_{\mathrm{var}}\, d}{M}, \qquad 0 \leq \ell \leq L-1. \tag{B.96}$$

Lemma B.12 is a standard result in the literature on well-separated GMMs; see, for example, [38, 35, 18]. In our setting, the high-threshold regime $T \to \infty$ ensures that the mixture components are well separated, so these results apply for sufficiently large T.

Proof of Proposition 3.3. Fix $\ell \in \{0,\dots,L-1\}$ and $T \geq T_0$, and let π be the permutation from Lemma B.11. We decompose

$$\widehat{\mu}_{\pi(\ell)}(\mathcal{A}) - Tv_\ell = \big(\widehat{\mu}_{\pi(\ell)}(\mathcal{A}) - \mu_{\pi(\ell)}^\star(T)\big) + \big(\mu_{\pi(\ell)}^\star(T) - Tv_\ell\big). \tag{B.97}$$

Using $\|a+b\|_2^2 \leq 2\|a\|_2^2 + 2\|b\|_2^2$ and taking expectations in (B.97) gives

$$\mathbb{E}\Big[\big\|\widehat{\mu}_{\pi(\ell)}(\mathcal{A}) - Tv_\ell\big\|_2^2\Big] \leq 2\,\mathbb{E}\Big[\big\|\widehat{\mu}_{\pi(\ell)}(\mathcal{A}) - \mu_{\pi(\ell)}^\star(T)\big\|_2^2\Big] + 2\,\big\|\mu_{\pi(\ell)}^\star(T) - Tv_\ell\big\|_2^2. \tag{B.98}$$

By the variance bound of Lemma B.12,

$$\mathbb{E}\Big[\big\|\widehat{\mu}_{\pi(\ell)}(\mathcal{A}) - \mu_{\pi(\ell)}^\star(T)\big\|_2^2\Big] \leq \frac{C_{\mathrm{var}}\, d}{M}, \qquad M \geq M_0,\ T \geq T_0, \tag{B.99}$$

while the high-threshold bias control of Lemma B.11 yields

$$\big\|\mu_{\pi(\ell)}^\star(T) - Tv_\ell\big\|_2^2 \leq \frac{C_{\mathrm{bias}}^2}{T^2}, \qquad T \geq T_0. \tag{B.100}$$

Substituting (B.99) and (B.100) into (B.98), we obtain

$$\mathbb{E}\Big[\big\|\widehat{\mu}_{\pi(\ell)}(\mathcal{A}) - Tv_\ell\big\|_2^2\Big] \le 2\,\frac{C_{\text{var}}\,d}{M} + 2\,\frac{C_{\text{bias}}^2}{T^2}$$
$$\le C_1\,\frac{d}{M} + \frac{C_2}{T^2}, \tag{B.101}$$

for all $M \ge M_0$ and $T \ge T_0$, where $C_1 \triangleq 2C_{\text{var}}$ and $C_2 \triangleq 2C_{\text{bias}}^2$ are independent of d, M, T, completing the proof. □

Proof of Lemma B.11. Fix ℓ and consider the scalar projection $U_\ell \triangleq \langle x_\ell, Y\rangle$. Then $U_\ell \sim \mathcal{N}(0, \sigma_\ell^2)$ with σ_ℓ^2 as in (B.94). For the one–dimensional Gaussian tail, the conditional mean admits the standard Mills–ratio expansion (Lemma A.1):

$$\mathbb{E}[U_\ell \mid U_\ell \ge T] = \sigma_\ell\,\frac{\varphi(a_T)}{1-\Phi(a_T)}, \qquad a_T \triangleq \frac{T}{\sigma_\ell}, \tag{B.102}$$

where φ and Φ denote the standard normal pdf and cdf, respectively. As $T \to \infty$,

$$\frac{\varphi(a_T)}{1-\Phi(a_T)} = a_T + \frac{1}{a_T} + O\!\left(\frac{1}{a_T^3}\right), \tag{B.103}$$

so that

$$\mathbb{E}[U_\ell \mid U_\ell \ge T] = \sigma_\ell\left(a_T + \frac{1}{a_T} + O\!\left(\frac{1}{a_T^3}\right)\right)$$
$$= T + \frac{\sigma_\ell^2}{T} + O\!\left(\frac{1}{T^3}\right). \tag{B.104}$$

Next, decompose Y along v_ℓ and its orthogonal complement. By Gaussian regression,

$$\mathbb{E}[Y \mid U_\ell = u] = \frac{u}{\sigma_\ell^2}\,\Sigma x_\ell = u\,v_\ell. \tag{B.105}$$

Recall the definition of $m_\ell^{(T)}$ from (B.66). Conditioning first on $\{U_\ell \ge T\}$ and then on U_ℓ, we obtain from (B.105)–(B.104)

$$m_\ell^{(T)} = \mathbb{E}[Y \mid U_\ell \ge T] = \mathbb{E}[\,\mathbb{E}[Y \mid U_\ell] \mid U_\ell \ge T]$$
$$= \mathbb{E}[U_\ell \mid U_\ell \ge T]\,v_\ell$$
$$= \left(T + \frac{\sigma_\ell^2}{T} + O\!\left(\frac{1}{T^3}\right)\right) v_\ell, \qquad T \to \infty. \tag{B.106}$$

In particular, there exist constants $T_1 > 0$ and $C_1 > 0$ such that for all $T \ge T_1$,

$$\big\|m_\ell^{(T)} - Tv_\ell\big\|_2 \le \frac{C_1}{T}. \tag{B.107}$$

By definition, the KL–minimizer $\{\mu_k^\star(T)\}$ in (B.93) also maximizes the population log–likelihood $\mathcal{H}_T(\{\mu_k\}) \triangleq \mathbb{E}_{Z\sim g_T}\big[\log f(Z;\{\mu_k\})\big]$. The gradient of $\mathcal{H}_T$ with respect to μ_k is

$$\nabla_{\mu_k}\mathcal{H}_T(\{\mu_j\}) = \mathbb{E}_{Z\sim g_T}\Big[\gamma_k(Z;\{\mu_j\})\,\Sigma_{\text{GMM}}^{-1}(Z-\mu_k)\Big], \tag{B.108}$$

where

$$\gamma_k(z;\{\mu_j\}) \triangleq \frac{w_k\,\mathcal{N}(z;\mu_k;\Sigma_{\mathrm{GMM}})}{\sum_{m=0}^{L-1} w_m\,\mathcal{N}(z;\mu_m;\Sigma_{\mathrm{GMM}})}, \tag{B.109}$$

is the posterior responsibility of component k under the model f. At the true parameter we have the stationarity condition

$$\nabla_{\mu_k}\mathcal{H}_T(\{\mu_j^\star(T)\}) = 0, \qquad k = 0,\dots,L-1. \tag{B.110}$$

Decompose the expectation in (B.108) over the true mixture g_T:

$$\begin{aligned} 0 &= \mathbb{E}_{Z\sim g_T}\Big[\gamma_k(Z;\{\mu_j^\star(T)\})\,\Sigma_{\mathrm{GMM}}^{-1}(Z-\mu_k^\star(T))\Big] \\ &= \sum_{\ell=0}^{L-1}\pi_\ell\mathbb{E}_{Z\sim g_\ell^{(T)}}\Big[\gamma_k(Z;\{\mu_j^\star(T)\})\,\Sigma_{\mathrm{GMM}}^{-1}(Z-\mu_k^\star(T))\Big]. \end{aligned} \tag{B.111}$$

For large T, the component distributions $\{g_\ell^{(T)}\}$ are concentrated in disjoint high-dimensional caps around $\{m_\ell^{(T)}\}$, and the distance between points of $g_\ell^{(T)}$ and $g_k^{(T)}$ for $k\neq\ell$ is of order T. Consequently, there exist constants $c, C_2 > 0$ such that, for all $k\neq\ell$,

$$\mathbb{E}_{Z\sim g_\ell^{(T)}}\big[\gamma_k(Z;\{\mu_j^\star(T)\})\big] \le C_2 e^{-cT^2}, \tag{B.112}$$

while

$$\mathbb{E}_{Z\sim g_\ell^{(T)}}\big[\gamma_\ell(Z;\{\mu_j^\star(T)\})\big] = 1 + O(e^{-cT^2}). \tag{B.113}$$

Plugging (B.112)–(B.113) into (B.111) and isolating the dominant term with $\ell = k$ yields

$$\pi_k\mathbb{E}_{Z\sim g_k^{(T)}}\big[\Sigma_{\mathrm{GMM}}^{-1}(Z-\mu_k^\star(T))\big] = O(e^{-cT^2}). \tag{B.114}$$

Equivalently, as $\pi_k > 0$ for every $k\in\{0,1,\dots L-1\}$,

$$\mu_k^\star(T) - m_k^{(T)} = O(e^{-cT^2}). \tag{B.115}$$

Thus there exist constants $T_2 > 0$ and $C_2' > 0$ such that for all $T \ge T_2$,

$$\big\|\mu_k^\star(T) - m_k^{(T)}\big\|_2 \le C_2' e^{-cT^2}. \tag{B.116}$$

Combining (B.107) and (B.116) via the triangle inequality, we obtain for $T \ge T_0 \triangleq \max\{T_1, T_2\}$,

$$\begin{aligned} \big\|\mu_\ell^\star(T) - Tv_\ell\big\|_2 &\le \big\|\mu_\ell^\star(T) - m_\ell^{(T)}\big\|_2 + \big\|m_\ell^{(T)} - Tv_\ell\big\|_2 \\ &\le C_2' e^{-cT^2} + \frac{C_1}{T}. \end{aligned} \tag{B.117}$$

Since $e^{-cT^2} \le 1/T$ for all sufficiently large T, we may absorb the exponential term into the $1/T$ term: there exists $C_{\mathrm{bias}} > 0$ and possibly a larger T_0 such that, for all $T \ge T_0$,

$$\big\|\mu_\ell^\star(T) - Tv_\ell\big\|_2 \le \frac{C_{\mathrm{bias}}}{T}, \tag{B.118}$$

for each ℓ. This completes the proof.

□

B.6 Template matching with local maxima

In practical implementations, template matching and the subsequent extraction are typically restricted to local maxima of the correlation score: rather than selecting every location where the cross-correlation exceeds a threshold T, one first constructs a correlation heatmap and then retains only those locations that are *local maxima* in a prescribed neighborhood.

To connect this to our theoretical model, for each template x_ℓ and patch center $s \in \mathbb{Z}$, define the template-matching score

$$C_\ell(s) \triangleq \langle x_\ell, y_s \rangle, \tag{B.119}$$

where y_s is the patch extracted at location s as in (2.4). The corresponding heatmap is

$$C(s) \triangleq \max_{0 \le \ell \le L-1} C_\ell(s), \tag{B.120}$$

and a patch is retained at s if $C(s) \ge T$ and s is a local maximum of C in a prescribed neighborhood.

To formalize local maxima, fix an integer radius $r \ge 0$ and, for each $s \in \mathbb{Z}$, define

$$N(s) \triangleq \{t \in \mathbb{Z} : |t - s| \le r\}. \tag{B.121}$$

We say that s is a local maximum of C if $C(s) \ge C(s')$ for all $s' \in N(s)$. Intuitively, in the high-threshold regime $T \to \infty$, exceedances above T become rare. In this regime, the local-maximum constraint has a negligible effect on the leading-order behavior: conditional on having at least one exceedance in a finite window, it is overwhelmingly likely that there is exactly one high exceedance, and hence at most one high local maximum. Consequently, in the high-threshold regime, the non-local-maximum suppression step does not change the leading-order bias induced by template matching.

Proposition B.13. *Let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a stationary Gaussian noise process satisfying (2.3) and the α-mixing condition (2.7), and let $C(s)$ be the template-matching heatmap defined in (B.119)-(B.120) for a finite collection of templates $\{x_\ell\}_{\ell=0}^{L-1}$. Fix a finite index window $W \subset \mathbb{Z}$ and define*

$$N_T(W) \triangleq \big|\{s \in W : C(s) \ge T\}\big| \tag{B.122}$$

to be the number of exceedances of level T in W. Then, as $T \to \infty$,

$$\mathbb{P}\big(N_T(W) \ge 2 \,\big|\, N_T(W) \ge 1\big) = \frac{\mathbb{P}(N_T(W) \ge 2)}{\mathbb{P}(N_T(W) \ge 1)} \longrightarrow 0. \tag{B.123}$$

Proof of Proposition B.13. By stationarity of $\{C(s)\}_{s\in\mathbb{Z}}$,

$$\lambda_T \triangleq \mathbb{E}[N_T(W)] = \sum_{s\in W} \mathbb{P}\big(C(s) \ge T\big) = |W|\,\mathbb{P}\big(C(0) \ge T\big). \tag{B.124}$$

Since $C(0)$ is the maximum of finitely many centered Gaussian random variables $\{C_\ell(0)\}_{\ell=0}^{L-1}$ (see (B.119)–(B.120)), it has a sub-Gaussian tail: there exist constants $c_1 > 0$ and $K_1 < \infty$ such that

$$\mathbb{P}\big(C(0) \ge T\big) \le K_1 e^{-c_1 T^2}, \qquad T \ge 0, \tag{B.125}$$

by a union bound and standard Gaussian tail estimates [19]. Hence $\lambda_T \to 0$ exponentially fast as $T \to \infty$.

Because W is finite, we can bound the probability of multiple exceedances by a union bound over pairs of indices:

$$\mathbb{P}\big(N_T(W) \geq 2\big) \leq \sum_{\substack{s,t \in W \\ s \neq t}} \mathbb{P}\big(C(s) \geq T,\ C(t) \geq T\big). \tag{B.126}$$

Fix $s \neq t$. By definition $C(s) = \max_{0 \leq \ell \leq L-1} C_\ell(s)$ and $C(t) = \max_{0 \leq k \leq L-1} C_k(t)$ with $C_\ell(s) = \langle x_\ell, y_s \rangle$, where each $C_\ell(s)$ is a centered Gaussian linear functional of the underlying field $\{\xi_u\}$. Thus,

$$\mathbb{P}\big(C(s) \geq T,\ C(t) \geq T\big) \leq \sum_{\ell,k=0}^{L-1} \mathbb{P}\big(C_\ell(s) \geq T,\ C_k(t) \geq T\big). \tag{B.127}$$

For each fixed pair (ℓ, k) and $s \neq t$, the vector $(C_\ell(s), C_k(t))$ is a non-degenerate bivariate Gaussian. In particular, its correlation coefficient satisfies $|\rho_{\ell,k}(s,t)| < 1$ (a consequence of strict positive-definiteness of the covariance in (2.3)). Standard bivariate Gaussian tail bounds [29] then yield

$$\mathbb{P}\big(C_\ell(s) \geq T,\ C_k(t) \geq T\big) \leq K_2 e^{-c_2 T^2}, \qquad T \geq 0, \tag{B.128}$$

for some constants $K_2 < \infty$ and $c_2 > c_1$ that are uniform over (ℓ, k) and $s \neq t$ in the finite window W. Combining this with (B.127) gives

$$\mathbb{P}\big(C(s) \geq T,\ C(t) \geq T\big) \leq L^2 K_2 e^{-c_2 T^2}. \tag{B.129}$$

Substituting into (B.126), we obtain

$$\mathbb{P}\big(N_T(W) \geq 2\big) \leq |W|(|W|-1) L^2 K_2 e^{-c_2 T^2} = O\big(e^{-c_2 T^2}\big). \tag{B.130}$$

On the other hand, by Markov's inequality and the definition of λ_T,

$$\mathbb{P}\big(N_T(W) \geq 1\big) \leq \lambda_T = |W|\, \mathbb{P}\big(C(0) \geq T\big) \leq |W| K_1 e^{-c_1 T^2}, \tag{B.131}$$

while a lower bound is obtained by the union bound with a correction for double counting:

$$\begin{aligned} \mathbb{P}\big(N_T(W) \geq 1\big) &\geq \sum_{s \in W} \mathbb{P}\big(C(s) \geq T\big) - \sum_{\substack{s,t \in W \\ s \neq t}} \mathbb{P}\big(C(s) \geq T,\ C(t) \geq T\big) \\ &= \lambda_T - O\big(e^{-c_2 T^2}\big). \end{aligned} \tag{B.132}$$

Since $c_2 > c_1$, the error term $e^{-c_2 T^2}$ is $o(e^{-c_1 T^2})$, and hence $O(e^{-c_2 T^2}) = o(\lambda_T)$ as $T \to \infty$. Combining (B.131) and (B.132) gives

$$\mathbb{P}\big(N_T(W) \geq 1\big) \sim \lambda_T. \tag{B.133}$$

Therefore, as $T \to \infty$,

$$\mathbb{P}\big(N_T(W) \geq 2 \,\big|\, N_T(W) \geq 1\big) = \frac{\mathbb{P}(N_T(W) \geq 2)}{\mathbb{P}(N_T(W) \geq 1)} = \frac{O(e^{-c_2 T^2})}{\lambda_T} \longrightarrow 0, \tag{B.134}$$

because $\lambda_T = \Theta(e^{-c_1 T^2})$ with $c_1 < c_2$. This proves (B.123). □

C 3D reconstruction

C.1 Unified reconstruction model for cryo-EM and cryo-ET

Forward model. We describe cryo-EM and cryo-ET within a common framework. Let $V \in \mathbb{R}^{d_{\text{vol}}}$ denote the underlying 3D volume, and let $\{R_i\}_{i=0}^{M-1} \subset \mathsf{SO}(3)$ be random rotations. Each extracted datum $z_i \in \mathbb{R}^d$ (with d the number of pixels or voxels of the observed object) is modeled as

$$z_i = \Pi\big(R_i \cdot V\big) + \epsilon_i, \qquad i = 0, \ldots, M-1, \tag{C.1}$$

where Π is a linear imaging operator and $\{\epsilon_i\}_{i=0}^{M-1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_V)$ are additive Gaussian noise terms with covariance $\Sigma_V \succ 0$.

- *Cryo-ET:* Π is the identity operator on volumes, so z_i is a 3D subtomogram and $d = d_{\text{vol}}$.
- *Cryo-EM:* Π is the tomographic projection operator (optionally including CTF and other imaging effects), so z_i is a 2D particle image and $d < d_{\text{vol}}$.

In both modalities, a particle-picking stage first detects and extracts z_i from the raw data (micrographs in cryo-EM, tomograms in cryo-ET), and a subsequent reconstruction stage estimates V from the extracted observations $\{z_i\}_{i=0}^{M-1}$.

Remark C.1. *We use Σ_V to denote the noise covariance in the* postulated *reconstruction model* (C.1)*, and Σ to denote the covariance of the noise in the input to the particle picker, as in Algorithm 1. These covariance matrices need not coincide in our analysis.*

Template generation and particle picking. During particle picking, a bank of templates is generated from a single template volume V_{template} by applying a finite set of rotations $\{R_\ell\}_{\ell=0}^{L-1} \subset \mathsf{SO}(3)$:

$$x_\ell \triangleq \Pi\big(R_\ell \cdot V_{\text{template}}\big), \qquad \ell = 0, \ldots, L-1. \tag{C.2}$$

In cryo-ET, these are 3D templates (since $\Pi = I$); in cryo-EM, they are 2D projection templates. The templates $\{x_\ell\}$ are used as inputs to Algorithm 1, with a correlation threshold T. When Algorithm 1 is applied to pure noise whose candidate patches satisfy $y_i \sim \mathcal{N}(0, \Sigma)$ (or more generally Model 2.2), the law of a picked datum Z is a finite mixture

$$g^{(T)}(z) \triangleq \sum_{\ell=0}^{L-1} \pi_\ell \, g_\ell^{(T)}(z), \qquad z \in \mathbb{R}^d, \tag{C.3}$$

where $\pi_\ell \geq 0$, $\sum_{\ell=0}^{L-1} \pi_\ell = 1$ reflect the relative frequencies of picks associated with template x_ℓ, and $g_\ell^{(T)}$ is the component density corresponding to template $x_\ell = \Pi(R_\ell \cdot V_{\text{template}})$.

In the Gaussian setting $y_i \sim \mathcal{N}(0, \Sigma)$, each component is a truncated Gaussian,

$$g_\ell^{(T)}(z) = C_\ell^{(T)} \exp\left(-\frac{1}{2} z^\top \Sigma^{-1} z\right) \mathbb{1}_{\{\langle z, \Pi(R_\ell \cdot V_{\text{template}})\rangle \geq T\}}, \tag{C.4}$$

where $C_\ell^{(T)}$ is the normalizing constant chosen so that $\int_{\mathbb{R}^d} g_\ell^{(T)}(z)\,dz = 1$. In the spherically symmetric i.i.d. noise model (Model 2.2) with base density p_{sph}, the components take the form

$$g_\ell^{(T)}(z) = C_\ell^{(T)}\, p_{\text{sph}}(z)\, \mathbb{1}_{\{\langle z, \Pi(R_\ell \cdot V_{\text{template}})\rangle \geq T\}}, \tag{C.5}$$

with $C_\ell^{(T)}$ again defined by the normalization $\int_{\mathbb{R}^d} g_\ell^{(T)}(z)\,dz = 1$.

Thus, in both cryo-EM and cryo-ET, the output of the template-matching particle picker with templates $x_\ell = \Pi(R_\ell \cdot V_{\text{template}})$ and threshold T is a stationary sequence of observations whose marginal distribution is the mixture (C.3).

Reconstruction likelihood. The goal of the reconstruction step is to estimate V in the postulated model (C.1) from the extracted observations $\mathcal{A} = \{z_i\}_{i=0}^{M-1}$. Let $f(z; V)$ denote the likelihood of a single *postulated* observation:

$$f(z; V) = \frac{1}{(2\pi)^{d/2} \det(\Sigma_V)^{1/2}} \int_{R \in \mathsf{SO}(3)} \exp\left(-\frac{(\, z - \Pi(R \cdot V)\,)^\top \Sigma_V^{-1} (\, z - \Pi(R \cdot V)\,)}{2} \right) d\rho(R), \tag{C.6}$$

where ρ is a prior distribution on rotations.

The corresponding log-likelihood of V given the extracted data is

$$\mathcal{L}(V; \mathcal{A}) \triangleq \frac{1}{M} \sum_{i=0}^{M-1} \log \left[\int_{R \in \mathsf{SO}(3)} \exp\left(-\frac{(\, z_i - \Pi(R \cdot V)\,)^\top \Sigma_V^{-1} (\, z_i - \Pi(R \cdot V)\,)}{2} \right) d\rho(R) \right] + \text{const.} \tag{C.7}$$

In practice, the integral over $\mathsf{SO}(3)$ is approximated by a finite set of rotations. For analytical convenience (and to match the picking stage), we consider a discretization $\{R_\ell\}_{\ell=0}^{L-1} \subset \mathsf{SO}(3)$ with associated weights $\{w_\ell\}_{\ell=0}^{L-1}$ satisfying $w_\ell \geq 0$ and $\sum_{\ell=0}^{L-1} w_\ell = 1$. Then, the single-particle density is approximated by

$$f(z; V) = \frac{1}{(2\pi)^{d/2} \det(\Sigma_V)^{1/2}} \sum_{\ell=0}^{L-1} w_\ell \exp\left(-\frac{(\, z - \Pi(R_\ell \cdot V)\,)^\top \Sigma_V^{-1} (\, z - \Pi(R_\ell \cdot V)\,)}{2} \right), \tag{C.8}$$

and the discretized likelihood becomes

$$\mathcal{L}(V; \mathcal{A}) \triangleq \frac{1}{M} \sum_{i=0}^{M-1} \log \left[\sum_{\ell=0}^{L-1} w_\ell \exp\left(-\frac{(\, z_i - \Pi(R_\ell \cdot V)\,)^\top \Sigma_V^{-1} (\, z_i - \Pi(R_\ell \cdot V)\,)}{2} \right) \right] + \text{const.} \tag{C.9}$$

The maximum-likelihood estimator of the volume under this discretized model is

$$\widehat{V} \triangleq \arg\max_V \mathcal{L}(V; \mathcal{A}). \tag{C.10}$$

Note that the *postulated* statistics of the extracted data, $f(z; V)$ in (C.8), differ from the *actual* statistics of the picked observations, $g^{(T)}(z)$ in (C.3). Thus, in both cryo-EM and cryo-ET, the reconstruction model is misspecified, and the high-threshold analysis developed in the GMM setting applies with the unified forward operator Π (identity for cryo-ET, tomographic projection for cryo-EM).

Assumption C.2. *The same finite set of rotations* $\{R_\ell\}_{\ell=0}^{L-1} \subset \mathsf{SO}(3)$ *is used both to generate the templates in the particle-picking stage,*

$$x_\ell = \Pi\big(R_\ell \cdot V_{\text{template}}\big), \qquad \ell = 0, \ldots, L-1,$$

and as the discretization of $\mathsf{SO}(3)$ *in the reconstruction model* (C.8).

The main question is how the reconstructed volume $\widehat{V}$ in (C.10) relates to the template volume V_{template} used for picking, when the underlying candidate observations are drawn from a zero-mean noise model (Model 2.1 or Model 2.2). At first glance, one might expect $\widehat{V}$ to converge to zero, since the input data have zero mean before selection. However, Corollary 4.2 shows that, as the number of extracted observations $M \to \infty$ and the threshold $T \to \infty$, the reconstructed volume $\widehat{V}$ instead converges (up to scaling) to the template volume V_{template}, rotated by an element of $\mathsf{SO}(3)$.

C.2 Proof of Corollary 4.2

Let $\widehat{V}(N,T)$ denote any maximum-likelihood estimator of the volume under the discretized unified model (C.8)–(C.10), based on the $M = M(N,T)$ extracted observations produced by Algorithm 1. We reduce the statement to the high-threshold alignment of GMM centers (Proposition B.10).

Step 1: Unified likelihood as a constrained GMM. Under the unified model (C.8), a single observation z has density

$$f(z;V) = \frac{1}{(2\pi)^{d/2}\det(\Sigma_V)^{1/2}} \sum_{\ell=0}^{L-1} w_\ell \exp\left(-\frac{(\,z - \Pi(R_\ell \cdot V)\,)^\top \Sigma_V^{-1} (\,z - \Pi(R_\ell \cdot V)\,)}{2}\right), \tag{C.11}$$

which is exactly a Gaussian mixture with component means

$$\mu_\ell(V) \;\triangleq\; \Pi\big(R_\ell \cdot V\big), \qquad \ell = 0, \ldots, L-1, \tag{C.12}$$

shared covariance Σ_V, and fixed weights $\{w_\ell\}$. Thus the reconstruction log-likelihood $\mathcal{L}(V;\mathcal{A})$ in (C.9) is the GMM log-likelihood

$$\mathcal{L}(V;\mathcal{A}) = \frac{1}{M}\sum_{i=0}^{M-1} \log f(z_i;V) = \mathcal{L}_{\text{GMM}}\big(\{\mu_\ell(V)\};\mathcal{A}\big), \tag{C.13}$$

evaluated on the special parametrization $\mu_\ell = \Pi(R_\ell \cdot V)$.

For each (N,T), define the induced centers

$$\widehat{\mu}_\ell(N,T) \triangleq \Pi\big(R_\ell \cdot \widehat{V}(N,T)\big), \qquad \ell = 0, \ldots, L-1. \tag{C.14}$$

By construction, $\{\widehat{\mu}_\ell(N,T)\}$ is a maximizer of the empirical GMM log-likelihood over the restricted parameter set

$$\mathcal{M}_{\text{orbit}} \triangleq \Big\{(\mu_0, \ldots, \mu_{L-1}) : \exists\, V \in \mathbb{R}^{d_{\text{vol}}} \text{ such that } \mu_\ell = \Pi(R_\ell \cdot V),\ 0 \le \ell \le L-1\Big\}. \tag{C.15}$$

Step 2: Population limit for fixed T. By the same misspecified maximum-likelihood estimation argument used in Lemma B.9, for each fixed T, as $N \to \infty$ we have

$$\widehat{V}(N,T) \xrightarrow[N\to\infty]{\text{a.s.}} V^\star(T), \tag{C.16}$$

where $V^\star(T)$ maximizes the population objective

$$\mathcal{H}_T^{\text{unified}}(V) \triangleq \mathbb{E}_{Z\sim g^{(T)}}\big[\log f(Z;V)\big], \tag{C.17}$$

with $g^{(T)}$ the picking-induced mixture (C.3). Equivalently, the induced centers

$$\mu_\ell^\star(T) \triangleq \Pi\big(R_\ell \cdot V^\star(T)\big) \tag{C.18}$$

form a maximizer of the GMM population objective

$$\mathcal{H}_T(\{\mu_\ell\}) \triangleq \mathbb{E}_{Z\sim g^{(T)}}\big[\log f(Z;\{\mu_\ell\})\big], \tag{C.19}$$

restricted to $\mathcal{M}_{\text{orbit}}$:

$$\{\mu_\ell^\star(T)\} \in \operatorname*{arg\,max}_{\{\mu_\ell\}\in\mathcal{M}_{\text{orbit}}} \mathcal{H}_T(\{\mu_\ell\}). \tag{C.20}$$

Step 3: High-threshold structure and alignment with the template orbit. By Proposition B.5 and Proposition B.6, the component means $m_\ell^{(T)}$ of the true mixture $g^{(T)}$ satisfy, under spherical symmetric noise,

$$\frac{m_\ell^{(T)}}{T} \xrightarrow[T\to\infty]{} x_\ell = \Pi\big(R_\ell \cdot V_{\text{template}}\big), \qquad 0 \le \ell \le L-1. \tag{C.21}$$

Thus, in the notation of Proposition B.10, we have

$$v_\ell = x_\ell = \Pi\big(R_\ell \cdot V_{\text{template}}\big). \tag{C.22}$$

Applying Proposition B.10 to the same mixture $g^{(T)}$, we obtain that there exists a permutation π of $\{0,\dots,L-1\}$ such that any unconstrained population maximizer $\{\tilde{\mu}_\ell^\star(T)\}$ of $\mathcal{H}_T$ satisfies

$$\frac{\tilde{\mu}_{\pi(\ell)}^\star(T)}{T} \xrightarrow[T\to\infty]{} v_\ell = \Pi\big(R_\ell \cdot V_{\text{template}}\big), \qquad 0 \le \ell \le L-1. \tag{C.23}$$

The high-threshold limiting configuration

$$\lim_{T\to\infty} \frac{m_\ell^{(T)}}{T} = \Pi\big(R_\ell \cdot V_{\text{template}}\big) \tag{C.24}$$

lies in the restricted manifold $\mathcal{M}_{\text{orbit}}$, since

$$\big(\Pi(R_0 \cdot V_{\text{template}}),\dots,\Pi(R_{L-1} \cdot V_{\text{template}})\big) = \{\mu_\ell(V_{\text{template}})\}. \tag{C.25}$$

Therefore, the constrained maximizers $\{\mu_\ell^\star(T)\}$ cannot do better asymptotically (at scale T^2) than the unconstrained maximizers $\{\tilde{\mu}_\ell^\star(T)\}$, and the same high-threshold argument as in Step 2 of the proof of Proposition B.10 implies that there exists a rotation $R \in \mathsf{SO}(3)$ such that

$$\frac{\Pi(R_\ell \cdot V^\star(T))}{T} = \frac{\mu_\ell^\star(T)}{T} \xrightarrow[T\to\infty]{} \Pi\big(R_\ell \cdot (R \cdot V_{\text{template}})\big), \qquad 0 \le \ell \le L-1. \tag{C.26}$$

Under Assumption 4.1, the map $V \mapsto \{\Pi(R_\ell \cdot V)\}_{\ell=0}^{L-1}$ identifies V up to a global rotation. Hence the above convergence is only possible if

$$\frac{V^\star(T)}{T} \xrightarrow[T\to\infty]{} R \cdot V_{\text{template}}, \tag{C.27}$$

for some $R \in \mathsf{SO}(3)$.

Step 4: Combining the limits. Fix T. From (C.16) we have

$$\widehat{V}(N,T) \xrightarrow[N\to\infty]{\text{a.s.}} V^\star(T), \tag{C.28}$$

and hence

$$\frac{\widehat{V}(N,T)}{T} \xrightarrow[N\to\infty]{\text{a.s.}} \frac{V^\star(T)}{T}. \tag{C.29}$$

Combining this with the population limit (C.27) and taking the iterated limits $N \to \infty$, then $T \to \infty$, yields

$$\lim_{T\to\infty} \lim_{N\to\infty} \frac{\widehat{V}(N,T)}{T} = R \cdot V_{\text{template}}, \tag{C.30}$$

with convergence in probability, which is exactly (4.6).

D Experimental methods

D.1 Setup of the numerical experiments

In the theoretical model presented in Section 3, candidate particles $\{y_i\}_{i=0}^{N-1}$ were assumed to be drawn i.i.d. from an underlying distribution. However, in practical particle picking from micrographs, candidates are not sampled i.i.d.; instead, they are extracted using a sliding-window approach that introduces spatial overlap between neighboring regions and induces statistical dependencies. Furthermore, when selecting particles, it is desirable to avoid redundancy by ensuring that the selected patches do not overlap. To achieve this, the particle detection procedure must enforce a non-overlapping constraint: only one candidate patch should be selected from any set of overlapping regions. The algorithm used to implement this constraint is described in detail in Algorithm 2.

This algorithm identifies candidate particles in a collection of 2D micrographs $\{Y_i\}_{i=0}^{R-1} \subset \mathbb{R}^{k\times k}$ by computing their cross-correlation with a set of template images $\{x_\ell\}_{\ell=0}^{L-1} \subset \mathbb{R}^{d\times d}$. For

Algorithm 2 `Template-Matching Particle Picker with Overlap`

Input: Templates $\{x_\ell\}_{\ell=0}^{L-1} \subset \mathbb{R}^{d\times d}$, Micrographs $\{Y_i\}_{i=0}^{R-1} \subset \mathbb{R}^{k\times k}$, Threshold $T \in \mathbb{R}$
Output: Set $\mathcal{A}$ of extracted particle patches

$\mathcal{A} \leftarrow \emptyset$
for $i = 0$ to $R-1$ **do** ▷ Loop over micrographs
  $C_i \leftarrow$ zero matrix in $\mathbb{R}^{k\times k}$ ▷ Heatmap storing the maximum cross-correlation
  for $\ell = 0$ to $L-1$ **do** ▷ Loop over templates
    $C_{\ell,i} \leftarrow x_\ell \star Y_i$ ▷ 2D circular cross-correlation
    for all pixels (u, v) **do**
      $C_i[u, v] \leftarrow \max(C_i[u, v], C_{\ell,i}[u, v])$ ▷ Maximum of the heatmap
    end for
  end for
  $(S_i, I_i) \leftarrow$ SortDescend(Flatten(C_i)) ▷ Sort the heatmap
  $M_i \leftarrow$ zero mask in $\{0, 1\}^{k\times k}$ ▷ Binary map tracking already selected regions
  for $m = 0$ to $k^2 - 1$ **do** ▷ Loop over pixels
    $(u, v) \leftarrow$ IndexToCoords($I_i[m], k$) ▷ Convert flat index to 2D coordinates
    if $S_i[m] > T$ **and** patch around (u, v) of size $d \times d$ does not overlap in M_i **then**
      $P \leftarrow Y_i[u - d/2 : u + d/2,\ v - d/2 : v + d/2]$
      $\mathcal{A} \leftarrow \mathcal{A} \cup \{P\}$
      Fill with ones the region of size $d \times d$ around (u, v) in M_i
    end if
  end for
end for
return $\mathcal{A}$

each micrograph-template pair, a 2D circular cross-correlation (assuming periodic boundary conditions) is computed to produce a heatmap $C_{\ell,i}$, where each pixel indicates the similarity between a local patch of the micrograph and the template. Each $C_{\ell,i}$ has the same spatial resolution as the micrograph. To consolidate the response across templates, the algorithm computes a combined heat map C_i by taking the maximum pixel over all L templates. This results in a score map where each entry reflects the strongest template match at that location. The pixels in C_i are then flattened and sorted in descending order by score. The algorithm then iterates through the ranked pixels and selects particle candidates that (i) exceed a predefined threshold T, and (ii) do not spatially overlap with any previously selected patch. A binary mask is maintained to track used regions and enforce non-overlap. Only the highest-scoring, non-overlapping patches are added to the output set $\mathcal{A}$. This non-overlapping template matching strategy is commonly used in cryo-EM; see, for example, [36].

D.2 Validation under stationary correlated Gaussian noise

To complement the white-noise experiments in the main text, we carried out an additional supplementary simulation study under stationary correlated Gaussian noise, with the goal of empirically validating the prediction of Theorem 3.2. This experiment should be viewed as the correlated-noise analogue of the white-noise validation presented in the main text: there,

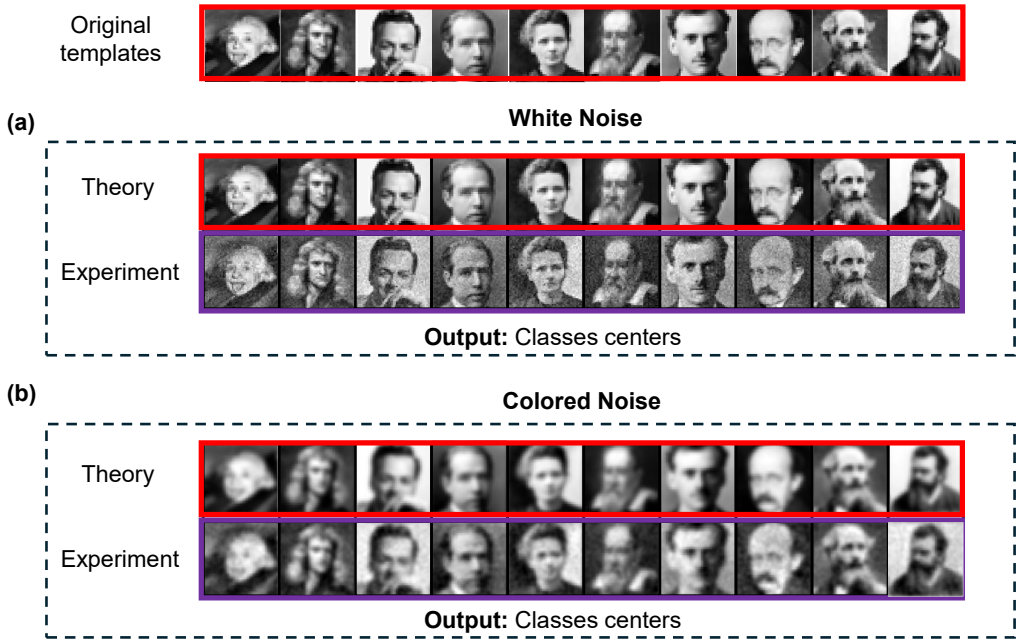

Figure 11: Validation of template-induced bias under white and stationary correlated Gaussian noise. Top row: the original templates. Panel (a) shows the white-noise setting. In this case, the theoretical prediction for the normalized class centers coincides with the templates themselves, and the empirical class centers obtained from the selected noise samples closely match this prediction. Panel (b) shows the stationary correlated-noise setting. Here, the theory predicts that the normalized class centers are no longer equal to the original templates, but rather to their covariance-shaped transforms under the noise covariance operator, given by (D.3). The empirical class centers again closely agree with the theoretical prediction.

the selected pure-noise samples yield class centers that align with the templates themselves, whereas here the theory predicts that the selected class centers should align with a covariance-shaped transform of the templates.

In this experiment, the input patches are drawn from a zero-mean stationary Gaussian field with a prescribed non-flat two-dimensional power spectrum. Specifically, if W denotes white Gaussian noise and $S(k_x, k_y)$ is a nonnegative power spectrum, then each realization is generated as

$$Y = \mathcal{F}^{-1}\big(\sqrt{S}\,\mathcal{F}(W)\big), \tag{D.1}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the two-dimensional discrete Fourier transform and its inverse, respectively. In the colored-noise experiment, we choose S explicitly as a Gaussian low-pass spectrum,

$$S(k_x, k_y) = \exp\left(-\frac{k_x^2 + k_y^2}{2\ell^2}\right), \tag{D.2}$$

where $\ell > 0$ is a bandwidth parameter and (k_x, k_y) range over the discrete Fourier frequencies. We then normalize S by its average value so that the resulting spatial variance is

approximately one. Equivalently, the covariance operator of Y is a stationary convolution operator Σ, which acts by pointwise multiplication by S in the Fourier domain. In contrast to the white-noise case, Σ is not proportional to the identity, and therefore the induced template-matching bias depends explicitly on the correlation structure of the noise.

We then apply the same template-matching selection rule as in the main text to a large collection of pure-noise samples, using a bank of normalized templates $\{x_\ell\}_{\ell=1}^{L}$. For each selected sample, we record the template attaining the largest correlation score, and for each template class ℓ we compute the empirical conditional mean of the selected samples. Theorem 3.2 predicts that, after normalization by the selection threshold, this empirical class mean should align with

$$\frac{\Sigma x_\ell}{x_\ell^\top \Sigma x_\ell}. \tag{D.3}$$

In the white-noise case, where $\Sigma \propto I$, this expression reduces to the template itself. In the correlated-noise setting, however, it becomes a covariance-shaped transform of the template.

The results, shown in Figure 11, confirm this prediction. Panel (a) reproduces the white-noise benchmark: the theoretical and empirical normalized class centers both closely match the original templates. Panel (b) shows the stationary correlated-noise setting: although the theoretical outputs are now visibly distorted relative to the original templates, the empirical class centers remain in close agreement with the corresponding theoretical predictions. Quantitatively, this agreement is also reflected in a PCC close to one between the empirical and theoretical class centers. Thus, as in the white-noise case studied in the main text, template-based selection from pure noise still produces a structured biased output; the main difference is that under correlated noise the limiting bias is shaped not only by the templates, but also by the covariance operator Σ.