Selective Disk Bispectrum: A Complete and Rotation Invariant Image Descriptor

Adele Myers Lantow
Department of Physics
University of California, Santa Barbara

Nina Miolane
Department of Electrical and Computer Engineering
University of California, Santa Barbara

Abstract

The classical bispectrum is a signal representation with two fundamental properties desirable for machine learning. First, it is invariant to translation of the signal, and second the signal can be reconstructed (inverted) from its bispectrum coefficients, up to a translation. These properties, along with its ability to capture higher-order frequency correlations, have made the bispectrum a fixture for signal processing across science and engineering. For image processing, the relevant symmetry is rotation, and the disk bispectrum has been proposed for this domain. However, its size and computation time grow almost cubically with the number of basis functions, making its tractability difficult on downstream tasks. Furthermore, no inversion or completeness proof had previously been proposed. We introduce the *Selective Disk Bispectrum* (SDB), a structured subset of disk bispectrum coefficients whose size grows linearly with the number of basis functions. We derive an explicit inversion algorithm proving that the SDB is complete, since an image can be reconstructed, up to a rotation, from its SDB coefficients alone. We verify the inversion numerically and demonstrate the SDB's practical benefits for (i) classification and (ii) multi-reference alignment on translated and noisy MNIST, FashionMNIST, and cell microscopy images. (i) Classification: Across datasets, the SDB fed to an MLP classifier matches learned baselines. (ii) Rotational MRA: Across datasets, the SDB inversion recovers an image from N noisy rotated copies. Our results bring to images what the classical bispectrum has long provided for 1D signals: an invariant that is complete, efficient, and invertible, and has proven advantages across downstream tasks.

1 Introduction

Introduced in the mid-20th century in signal processing [44, 5, 31], the bispectrum is a third-order statistic constructed from triple products of Fourier coefficients of an input signal. Its ability to encode higher-order frequency correlations has led to its widespread use across the sciences, from ocean waves [12] and seismic data [10], to cardiac [28] and brain signals [51] (see Koukiou [22] for a review).

The bispectrum has two fundamental properties desirable in machine learning tasks: translation-invariance and invertibility. Translation-invariance guarantees that bispectrum statistics will not be affected by the global phase of the signal, and will instead only be affected by shape changes. Bispectrum invertibility means a generic signal can be reconstructed from its bispectrum coefficients, up to translation [14], guaranteeing the bispectrum encodes all information about the original signal, other than its global phase. This makes it a so-called *complete* invariant.

In *classification*, completeness guarantees that two signals that differ by more than a translation receive different descriptors, hence no discriminative information is lost before learning starts. Integrating bispectra in neural networks leads to up to $+30\%$ accuracy over baselines under fixed

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

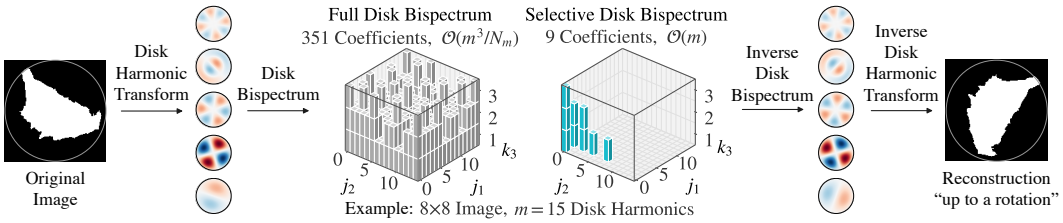

Figure 1: **Our selective disk bispectrum grows linearly with the number m of disk harmonic basis functions.** For a small 8×8 image at the library default bandlimit ($m = 15$ DH coefficients), the full disk bispectrum has 351 coefficients, whereas the SDB has 9. For a 28×28 image, these numbers rise to 152,025 and 97, respectively, for $m = 185$ DH coefficients.

parameter budgets, equal accuracy with 10× fewer parameters [29], and provable resistance to the invariance-based adversarial attacks that fool standard models [37]. In *multi-reference alignment* (MRA), where a signal is recovered from many translated, noisy observations of itself by taking their mean in bispectrum space, invertibility guarantees that the resulting mean can be used to reconstruct the average, de-noised signal [4].

For images, the relevant transformation is often rotation rather than translation. Many images carry no meaningful orientation, like a cell under a microscope, a particle in cryo-electron microscopy, or a galaxy in a telescope. The *disk bispectrum* [53] extends the bispectrum construction for rotation-invariance on images. Its coefficients achieve rotation-invariance by replacing the Fourier basis with the disk harmonic basis. However, it has seen limited use for three reasons. First, it is too big. The number of coefficients grows almost cubically with the number of basis functions. Second, it is too slow, for the same reason. Third, no inversion has been derived, and its completeness is unproven. Therefore, its tractability and discriminative power for classification is not guaranteed, and it cannot be used for MRA.

Contributions. We propose a computationally tractable selective disk bispectrum, along with a disk bispectrum inversion. We provide an open sourced implementation for the forward and inverse SDB maps at `https://github.com/geometric-intelligence/bispectrum/`, and demonstrate their utility in downstream tasks.

- **Disk bispectrum inversion (main theoretical contribution).** We derive the first analytic inversion of the disk bispectrum. This inversion proves that the disk bispectrum contains all information in an image, up to a rotation. We verify this inversion numerically by mapping MNIST, FashionMNIST, and cell microscopy images to and from their SDB coefficients.
- **The Selective Disk Bispectrum (SDB).** Our inversion reveals that most disk bispectrum coefficients are redundant: only a minimal subset is needed for complete image reconstruction. We call this subset the SDB. Because an image can be reconstructed from its SDB coefficients, the SDB is a *complete invariant* — two images have the same SDB if and only if they differ by a planar rotation alone. Using the SDB instead of the full disk bispectrum reduces spatial complexity from $\mathcal{O}(m^3/N_m)$ to $\mathcal{O}(m)$ and time complexity from $\mathcal{O}(L^3 + m^3/N_m)$ to $\mathcal{O}(L^3 + m)$, for an $L \times L$ image represented by m disk harmonic basis functions with maximum angular frequency N_m. Time complexity can be further reduced to $\mathcal{O}(L^2 \log L + m)$ with a recent approximation of disk harmonic coefficients [27]. Spatial complexity improvements are shown visually in Fig. 1.
- **Accuracy guarantees and noise analysis.** We provide SDB accuracy guarantees when using [27]'s disk harmonic coefficient approximation. We also derive its expectation and variance under i.i.d. noise and prove that most coefficients are asymptotically unbiased, the guarantee required to apply the descriptor in lower SNR regimes.
- **Classification of noisy and rotated MNIST, FashionMNIST, and cell microscopy images.** Using the SDB as input to an MLP classifier achieves comparable accuracy as pooled G-CNNs, harmonic networks, $SE(2)$ Networks or frozen DINOv3, EVA-02, and OpenCLIP features. Relative success between representations seems to be dataset-dependent.
- **MRA on noisy and rotated MNIST, FashionMNIST, and cell microscopy images.** Bispectrum-based MRA requires an analytic inversion, which did not exist for any rotation-invariant bispectrum

until this work. We demonstrate that our inversion makes rotation-MRA possible, and propose an alteration to make the inversion more robust against noise propagation during averaging.

Together, these contributions characterize a new disk bispectrum whose newfound tractability opens doors for utilization in signal processing, science, and engineering.

2 Background

We first review existing bispectra that motivated the definition of our new selective disk bispectrum.

Classical Bispectrum. The *classical 1D translation-invariant bispectrum* [44] is defined as

$$B_{\nu_1,\nu_2} = F_{\nu_1} F_{\nu_2} F^*_{\nu_1+\nu_2}, \tag{1}$$

where $F_\nu = |F_\nu| e^{-i\nu\phi_\nu}$ are Fourier coefficients of $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$, with frequencies ν. Under cyclic translation $g \in \mathbb{Z}/n\mathbb{Z}$, f transforms as $\tilde{f}(x) = f(x-g)$, with Fourier coefficients transforming equivariantly as $\tilde{F}_\nu = e^{-i\nu g} F_\nu$, so phases cancel in the bispectrum product:

$$\tilde{B}_{\nu_1,\nu_2} = e^{-i\nu_1 g} F_{\nu_1} \cdot e^{-i\nu_2 g} F_{\nu_2} \cdot e^{i(\nu_1+\nu_2)g} F^*_{\nu_1+\nu_2} = B_{\nu_1,\nu_2}. \tag{2}$$

Unlike the power spectrum $|F_\nu|^2$, which discards phase entirely, the bispectrum preserves relative phase ϕ_ν between coefficients, making it a *complete* invariant. This was proven by [14], who demonstrated f can be reconstructed from B up to a global translation, as shown in Appendix A.4.

G-Bispectrum Mathematically, a group is a pair $(G, \circ)$, where G is a set and $\circ$ a binary operation on that set. Groups must satisfy axioms of closure, associativity, inverse, and identity. The classical bispectrum is invariant to the group of 1D translations $(\mathbb{Z}/n\mathbb{Z}, +)$, where $+$ is addition. This concept is extended by the *group-invariant G-Bispectrum* [18], which is constructed from the Group-Fourier Transform (G-FT) [35]. For discrete groups, the G-FT is defined $F_\rho = \sum_{g\in G} \rho(g) f(g)$, where $\rho(g)$ is an irreducible representation of a group element $g \in G$, defined in Appendix A. Note that for the G-FT, $f : G \to \mathbb{C}$, meaning the domain of f is an abstract group. Consequently, utilization of this bispectrum is not always straightforward, as most signals are defined over domains such as the line (1D signals), plane (2D images), sphere (earth models), or volume (3D voxels).

Disk Bispectrum Because of the domain restrictions of the G-Bispectrum, the rotation-invariant disk bispectrum was developed for signals whose domain is planar, enabling utilization directly on images. In particular, it is convenient to consider an image as being a square integrable real-valued function $f : \mathbb{D} \to \mathbb{R}$ defined on the disk $\mathbb{D}$. This can be implemented for any $L \times L$ image by masking any pixels outside radius $L/2$. The disk bispectrum [53], defined as

$$b_{j_1,j_2,k_3} = a_{n_{j_1},k_{j_1}} \cdot a_{n_{j_2},k_{j_2}} \cdot a^*_{n_{j_1}+n_{j_2},k_3} \in \mathbb{C}, \tag{3}$$

is a product of disk harmonic coefficients [45],

$$a_{nk} = \int_0^{2\pi} \int_0^1 f(r,\theta)\, \psi^*_{nk}(r,\theta)\, r\, dr\, d\theta, \tag{4}$$

where ψ_{nk} are *disk harmonics* (DH), defined as $\psi_{nk}(r,\theta) = c_{nk} J_n(\lambda_{nk} r) e^{in\theta}$. Here, $(n,k) \in \mathbb{Z} \times \mathbb{Z}_{>0}$, J_n is the n-th order Bessel function of the first kind, λ_{nk} is its k-th root of the n-th Bessel function, and c_{nk} is a normalization constant.

Disk harmonics are a natural basis for a rotation-invariant bispectrum. Just as Fourier modes transform equivariantly under 1D translation, DH coefficients transform equivariantly under rotation, i.e. $a_{nk} \longrightarrow e^{in\phi} a_{nk}$ for $\phi \in SO(2)$, the group of 2D rotations. Moreover, just as Fourier modes e^{-ivx} (see classical bispectrum) and frequencies v are eigenfunctions and eigenvalues of the Dirichlet Laplacian over S^1, the domain of continuous cyclic signals, disk harmonics and their "frequency" λ_{nk} are eigenfunctions and eigenvalues of the Dirichlet Laplacian over $\mathbb{D}$.

Because disk harmonics have a natural notion of frequency, the Nyquist criterion dictates a frequency cutoff or *bandlimit* of $\frac{\lambda}{2\pi} \leq \frac{L}{4}$ for an $L \times L$ image, where higher-frequency components correspond to features beyond the image resolution [52]. When [27] implemented the first tractable DH coefficients,

Rotation-Equivariant (Equiv.) & Invariant (Invar.) Descriptors	Equiv.	Local Invar.	Global Invar.	Magnitude is Invar.	Invar. if Pooled	Invertible
Selective Disk Bispectrum (SDB) (Ours) Full Disk Bispectrum [53]	✗	✗	✓	✓	✗	✓
G-CNNs [7] Spherical Convolutions [19] $SE(2)$ Steerable Networks [46] Capsule Networks [36]	✓	✗	✗	✗	✓	✗
Disk Harmonic Transform [27] Fourier Mellin Transform [33] Tensor Field Networks [42] Harmonic Networks [48]	✓	✗	✗	✓	✗	✗
SIFT: Scale–Invariant Feature Transform [25] SURF: Speeded–Up Robust Features [3] ORB: oFAST & Rotated BRIEF [34] AKAZE: Accelerated KAZE [1]	✗	✓	✗	✗	✗	✗
Hu Moments [16]	✗	✗	✓	✗	✗	✗
Zernike Moments [41]	✗	✗	✗	✓	✗	✗

Table 1: **The SDB (ours) is the only rotation-invariant descriptor that is simultaneously globally invariant, magnitude-preserving, *and* invertible — meaning it retains all image information up to a global rotation.** Equivariant methods achieve invariance only after pooling or taking a norm, costing of structural information. Local invariants discard global structure. Other global invariants are not invertible and therefore not complete.

they truncate according to this frequency cutoff. They order λ_{nk} from lowest frequency to highest $\lambda_0 \leq \cdots \lambda_j \cdots \leq \lambda_{m-1}$, listing $-n$ before n when $\lambda_{-n,k} = \lambda_{n,k}$. The index $j = 0, ..., m-1$ represents this order. Here, notation shifts to accommodate indexing by j. Concretely, since DH frequency is unique to a (n, k) pair, $\lambda_{n,k}$ can be equivalently written as λ_j or $\lambda_{n_j k_j}$, DHs $\psi_{n,k}$ can be written as ψ_j or $\psi_{n_j k_j}$, and DH coefficients $a_{n,k}$ as a_j or $a_{n_j k_j}$, where n_j, k_j are the (n, k) corresponding to the jth frequency λ_j.

Then, the number of DH coefficients m in an expansion $f(r, \theta) = \sum_{n \in \mathbb{Z}} \sum_{k \in \mathbb{Z}_{>0}} a_{nk}\, \psi_{nk}(r, \theta)$ is the number of DHs within the bandlimit, $m = \#\{(n, k) : \lambda_{nk} \leq \lambda\}$. Similarly, $n \in \{-N_m, \ldots, N_m\}$ where $N_m = \max\{n_j \in \mathbb{Z} : j \in \{0, \ldots, m-1\}\}$ and $k \in \{1, ..., K_n\}$ where $K_n = \max\{k \in \mathbb{Z}_{>0} : \lambda_{nk} \leq \lambda\}$. Table 3 in Appendix B shows an example of this correspondence between j and (n, k) for an 8×8 image expanded in the truncated DH basis. In the disk bispectrum definition, $-N_m \leq n_{j_1} + n_{j_2} \leq N_m$, and $k_3 \leq K_{n_{j_1} + n_{j_2}}$.

3 Related Works

We review the existing bispectra in Tab. 2 and compare common rotation equivariant and invariant descriptors in Tab. 1.

G-Bispectra and Bispectra on G-Homogeneous Spaces. As discussed in Sec. 2, the primary challenge of the G-Bispectrum is its assumption that the signal domain is the space of the group itself G. Thus, the $SO(2)$ invariant G-bispectrum is only well-defined for signals $f : SO(2) \to \mathbb{C}$, whose domain is the group $SO(2)$. This makes the G-Bispectrum ill-defined as a rotation-invariant image descriptor, because images $f : \mathbb{R}^2 \to \mathbb{C}$ lie in the plane $\mathbb{R}^2$. [29] circumvented this problem by taking $SO(2)$ group convolutions over an image, thus taking the G-Bispectrum of the convoluted signal whose domain is the group, rather than the image itself. [20] worked towards addressing the domain problem with a "Homogeneous G-Bispectrum", relevant for signals defined over spaces homogeneous for the group, $\mathcal{H}$. A space that is homogeneous for a group is one where for any x, y there is a group element $g \in G$ s.t. $g \circ x = y$. This is helpful for $SO(3)$ invariant bispectra whose signals $f : S^2 \to \mathbb{C}$ are defined on the sphere S^2.

Bispectrum Name	Invariant to	Signal Domain	Constructed From	Inversion
Classical Bispectrum	$SO(2)$	$\mathbb{R}$	Fourier Coefficients	[14]
Classical 2D Bispectrum	$SO(2) \times SO(2)$	$\mathbb{R}^2$	2D Fourier Coefficients	[14]
G-Bispectrum	G	G	G-Fourier Coeffs. [35]	[29]
Selective G-bispectrum [29]	G	G	G-Fourier Coeffs.[35]	[29]
G-Bisp. on Homogeneous Spaces $\mathcal{H}$ [20]	G	$\mathcal{H}$	G-Fourier Coeffs.[35]	None
Full Disk Bispectrum [53]	$SO(2)$	Disk $\mathbb{D}$	Disk Harmonics [45, 27]	**(Ours)**
Selective Disk Bispectrum (SDB) **(Ours)**	$SO(2)$	Disk $\mathbb{D}$	Disk Harmonics [45, 27]	**(Ours)**

Table 2: **The SDB (ours) is the first bispectrum on a non-homogeneous space to admit an analytic inversion, establishing it as a complete invariant.**

Bispectra on Non-Homogeneous Spaces 2D domains like planes and disks are not homogeneous for the group $SO(2)$, because $SO(2)$ group action can only "move a point" along the circle S^1. This necessitated the creation of the disk bispectrum [53], defined in Sec. 2, which is the first bispectrum defined for a signal not homogeneous to the invariant group. Unfortunately, the "full" disk bispectrum is both spatially and computationally expensive. To reduce spatial complexity and enable utility on classification tasks, [53] resorted to PCA on bispectrum coefficients. Authors reported a feature-vector length of $\mathcal{O}(m^3/N_m)$, where m is the number of disk harmonic coefficients, and N_m the maximum angular frequency. Furthermore, they offered no disk bispectrum inversion, meaning that there was no guarantee that the disk bispectrum was a *complete invariant*.

Rotation-invariant descriptors Existing rotation-invariant descriptors (Tab. 1) do not possess the disk bispectrum's mathematical properties of completeness and inversion. Hand-crafted descriptors such as SIFT [25], ORB [34] are locally invariant to rotation but Zernike moment magnitudes [41], and Fourier–Mellin magnitudes [33] are fast and require no training, but they are not complete representations; the original signal cannot be recovered from their coefficients. This means that two images with very different visual content could map to identical descriptors, making them potentially less reliable for fine-grained recognition tasks. G-CNNs [7] and related equivariant architectures [8, 42] have achieved great success in image processing. However, they obtain rotation-invariance through pooling, which is a lossy operation, or through data augmentation, which spends data and compute to approximate a symmetry that is known a priori. Our experiments will compare our selective disk bispectrum to these descriptors and show that its properties of completeness and inversion provide a learning-free way to achieve rotation-invariance and discriminative power in classification, and they are indispensable in other tasks such as MRA, which requires an invertible rotation-invariant descriptor.

4 Theory: The Selective Disk Bispectrum and Its Inversion

We propose the selective disk bispectrum (SDB), defined as a specific subset of disk bispectrum coefficients. Proofs for all theorems are in Appendices D, E, and K.

Definition 1 (Selective Disk Bispectrum (SDB))**.** *Let $f : \mathbb{D} \to \mathbb{R}$ be a real-valued square-integrable function and let $a_{n,k}$ be its disk harmonic coefficients, bandlimited by λ. The selective disk bispectrum is defined as the following subset:*

$$b^f = \begin{bmatrix} b_{0,0,1} \cdots b_{0,0,K_0} \\ b_{2,0,1} \cdots b_{2,0,K_1} \\ \cdots \\ b_{2,N_m-1,1} \cdots b_{2,N_m-1,K_{N_m}} \end{bmatrix} = \begin{cases} b_{0,0,k} = a_{0,1}^2 \cdot a_{0,k}^*, \text{ for } k \in \{1, \ldots, K_0\}, \\ b_{2,n,k} = a_{1,1} \cdot a_{n,1} \cdot a_{n+1,k}^*, \\ \qquad \text{for } k \in \{1, \ldots, K_{n+1}\}, n \in \{0, \ldots, N_m - 1\}. \end{cases} \tag{5}$$

The SDB coefficients fix $k_{j_1} = k_{j_2} = 1$ and limit $j_1 \in \{0, 2\}$, allowing simplified indexing by j_1, n, k_3, as opposed to j_1, j_2, k_3. Choosing only a subset of disk bispectrum coefficients leads to dramatic improvements in space and time complexities. Both descriptors first require the same disk harmonic transform, costing $\mathcal{O}(L^3)$. The full disk bispectrum then forms $\Theta(m^3/N_m)$ products, for a space complexity of $\mathcal{O}(m^3/N_m)$ and a time complexity of $\mathcal{O}(L^3 + m^3/N_m)$. In contrast, the SDB forms only $\Theta(m)$ products.

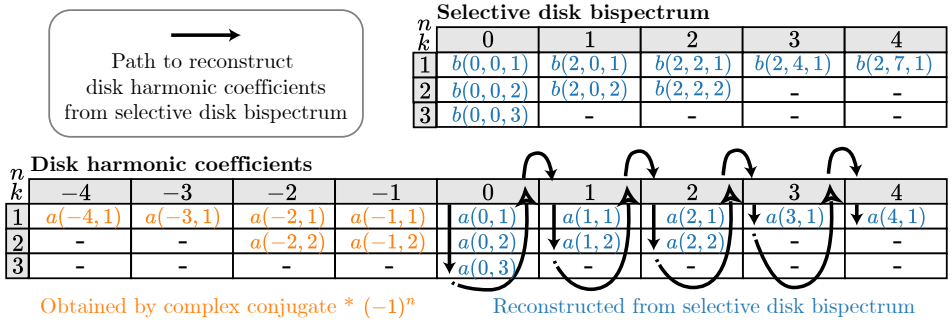

Figure 2: **The SDB inversion recovers each disk harmonic coefficient $a_{n,k}$ from exactly one SDB coefficient and previously recovered ones, column by column.** Arrows give the recovery order.

Proposition 1 (Complexity of the SDB)**.** *Consider an image of size $L \times L$, represented with m DH coefficients $a_{n,k}$. With N_m denoting the maximum angular frequency n, the space complexity of the SDB is $O(m)$, and the time complexity is $O(L^3 + m)$.*

Despite this reduction in complexity, the selected coefficients retain enough information to completely represent the image, up to a rotation and bandlimit, as proven by Th. 1. This means that most of the coefficients in the original disk bispectrum are, in fact, redundant.

Theorem 1 (Inversion)**.** *Let $f : \mathbb{D} \to \mathbb{R}$ be a real-valued square-integrable function and let $a_{n,k}$ be its disk harmonic coefficients, bandlimited by λ. Assume that $a_{n,1} \neq 0$ for $n \in \{0, ..., N_m - 1\}$. Then, its selective disk bispectrum b^f can be inverted, that is f can be reconstructed from b^f up to a planar rotation and frequency λ. With f discretized on an $L \times L$ image, the time complexity of the inversion is $O(m + L^3)$.*

An analysis of round trip error, from image to SDB and back, is in Appendix H, with examples shown in Appendix Fig. 7. The full inversion proof is in Appendix D. We summarize here and provide a graphical illustration of the reconstruction path in Fig. 2. First, reconstruct $a_{n,k}$ with $n = 0$ with

$$\|a_{0,1}\| = \|b_{0,0,1}\|^{1/3},\ \arg(a_{0,1}) = \arg(b_{0,0,1});\ a_{0,k} = \left(\frac{b_{0,0,k}}{a_{0,1}^2}\right)^* \text{ for } k \in \{2, \dots, K_0\}. \tag{6}$$

Next, reconstruct $a_{n,k}$ with $n = 1$ with

$$\|a_{1,1}\| = \left\|\frac{b_{2,0,1}}{a_{0,1}}\right\|^{1/2},\ \arg(a_{1,1}) = 0;\ a_{1,k} = \left(\frac{b_{2,0,k}}{a_{0,1} \cdot a_{1,1}}\right)^* \text{ for } k \in \{2, \dots, K_1\}. \tag{7}$$

Then, reconstruct $a_{n,k}$ with $n > 1$ with

$$a_{n,k} = \left(\frac{b_{2,n-1,k}}{a_{1,1} \cdot a_{n-1,1}}\right)^* \text{ for } k \in \{1, \dots, K_n\}. \tag{8}$$

Notationally, $b_{2,n-1,k} \to b_{2,i(n-1,1),k}$ for the full disk bispectrum inversion. Finally, reconstruct negative coefficients and invert the DH coefficients to a reconstructed image [45]

$$a_{-n,k} = (-1)^n a^*_{n,k} \text{ for } n \in \{1, \dots, N_m\}; \quad f(r,\theta) = \sum_{n=-N_m}^{N_m} \sum_{k=1}^{K_n} a_{n,k}\, \psi_{n,k}(r,\theta). \tag{9}$$

As the SDB coefficients are the exact coefficients used in our inversion Th. 1, our main result shows that despite discarding many coefficients from the full disk bispectrum, the SDB coefficients are sufficient to completely represent the image, up to a global rotation and bandlimit. With this, we proceed to establish the SDB as a *complete invariant*, meaning it retains all information about the image, while being invariant to global rotation.

Theorem 2 (Complete Invariance)**.** *Let f and f' be a pair of real-valued, square-integrable functions on the disk. Let the bispectrum be defined as in Definition 1. Assume that the DH coefficients $a_{n,1}$ of f are non-zero for all $n \in \{0, \dots, N_m - 1\}$. Then $b^f = b^{f'}$ if and only if $f' = f \circ R_\phi$ for some 2D rotation $\phi \in SO(2)$. Thus, the selective bispectrum is a complete invariant.*

The SDB's completeness comes from the inversion proof. Its invariance comes from the equivariance of the DH coefficients, as shown in Appendix D.

4.1 Fast Selective Disk Bispectrum Approximation

To further reduce the time complexity of the SDB, we propose using a recent numerical approximation of the DH coefficients [27] that computes the disk harmonic transform with time complexity $O(L^2 \log L)$ instead of $\mathcal{O}(L^3)$.

Proposition 2 (Complexity of the approximate SDB)**.** *Denote $L \times L$ the size of the image, m the number of DH coefficients. The time complexity of the approximate SDB, computed using the approximate DH coefficients of Marshall et al. [27], is $O(L^2 \log L + m)$.*

We provide guarantees on the accuracy of the approximate SDB coefficients and their invariance.

Theorem 3 (Guarantee for Approximate Accuracy)**.** *Let $0 < \varepsilon \leq 1$ be a target accuracy, and assume that $\lambda \leq \sqrt{\pi L^2}$ and $|\log \varepsilon| \leq L$, where λ is the DH coefficient bandlimit for an image f of size $L \times L$. Let b^f and $\tilde{b}^f$ denote the exact and approximate SDB of f. Then,*

$$\|\tilde{b}^f - b^f\|_\infty \leq C\varepsilon \|f\|_1,$$

for a constant C depending only on a bound of f.

As a corollary, we find that the bispectrum coefficients are approximately invariant with controlled guarantees.

Corollary 1 (Guarantee for Approximate Invariance)**.** *Let f, f' be a pair of real-valued square integrable functions on the disk such that there exists a $\phi \in SO(2)$ such that $f' = f \circ R_\phi$. Under the notations and assumptions of Th. 3, we have:*

$$\|\tilde{b}^f - \tilde{b}^{f'}\|_\infty \leq 2C\varepsilon \|f\|_1, \tag{10}$$

for a constant C depending only on a bound of f.

Linking the accuracy to f is useful, as it allows us to control accuracy through number of pixels p. If $|f|$ and C are large, but we still want to reach a given accuracy, then we can choose p accordingly such that $\log(\epsilon) < \sqrt{p}$ to match the desired accuracy.

4.2 Noisy Selective Disk Bispectrum Statistics

Real-world images are often corrupted by noise, raising the question of whether the SDB remains informative in this setting. We derive closed-form expressions for the expectation and variance of each SDB coefficient when additive i.i.d. Gaussian noise is added to the image, and show that most SDB coefficients are asymptotically unbiased. This makes the mean SDB a reliable and noise-robust representation. We also propose an adaptation to our recursive analytic inversion to make it more robust to noise. In the theorems that follow, consider a noiseless signal f with bispectrum b and DH coefficients $a_{n,k}$. Upon corruption with Gaussian noise, $\tilde{f}(x) = f(x) + \eta_x$ for $\eta_x \sim \mathcal{N}(0, \sigma^2)$ added i.i.d. at each pixel.

Theorem 4 (Expectation)**.** *The expectation of the noise-corrupted SDB $\tilde{b}$ is given by the following.*

$$\mathbb{E}[\tilde{b}_{00k}] = b_{00k} + \sigma^2 a^*_{0,k} + \delta_{k,1} 2\sigma^2 a_{0,1}; \quad \mathbb{E}[\tilde{b}_{201}] = b_{201} + \sigma^2 a_{0,1}; \quad \textit{Else, } \mathbb{E}[\tilde{b}_{2nk}] = b_{2nk}, \tag{11}$$

where δ denotes the Kronecker delta.

The variance of $\tilde{b}$ is given by the following.

Theorem 5 (Variance)**.** *The variance of $\tilde{b}$ is given by*

$$Var(b_{j_1 n_2 k_3}) = A_2\sigma^2 + A_4\sigma^4 + A_6\sigma^6, \tag{12}$$

where A_6 is a scalar, A_4 is a linear combination of $|a_{n,k}|^2$, and A_2 is a linear combination $|a_{n,k}|^4$. Exact A_2, A_4, A_6 are provided in Appendix K and depend on j_1, n_2, k_3 of the SDB coefficient.

Together, the expectation and variance tell us precisely how noise distorts the SDB. The expectation shows where the noisy SDB distribution is centered relative to the ground truth, and the variance shows how much noisy SDB observations fluctuate around that center. In practical terms, given an estimate of the noise level σ^2 and the signal's DH power spectrum, a practitioner can predict the reliability of any SDB coefficient before using it downstream.

Analytic inversion of the SDB proceeds by the recursion of Th. 1, in which each DH coefficient $a_{n,k}$ is recovered by dividing the SDB coefficient $b_{2,n-1,k}$ by the previously recovered product $d_n = a_{1,1}\, a_{n-1,1}$. In the noiseless setting this division is exact even when d_n is small, because the numerator $b_{2,n-1,k} = d_n\, a^*_{n,k}$ contains the divisor as a factor and the two cancel. In multi-reference alignment, numerator and divisor are instead estimated from finitely many noisy observations and carry separate errors that do not cancel, so a divisor that falls near zero makes the quotient arbitrarily large. The same instability is known for the analogous recursion in one-dimensional multi-reference alignment, leading Bendory et al. [4] to abandon the recursion for a least-squares fit of all bispectrum coefficients. We keep the recursion and instead use Tikhonov regularization (ridge regression) [43] at each recursion step.

Theorem 6 (Damped inversion)**.** *Let $\hat{b}$ be the SDB averaged over N noisy copies $\tilde{f}_i = R_{\theta_i} f + \eta_i$, $\eta_i \sim \mathcal{N}(0, \sigma^2)$, of a template f with unknown rotations θ_i, and let $\hat{d}_n = \hat{a}_{1,1}\hat{a}_{n-1,1}$ be the estimated divisor in the recursion of Th. 1. For $\varepsilon > 0$, the damped inversion replaces each division $\hat{b}_{2,n-1,k}/\hat{d}_n$ by*

$$\hat{a}^{\varepsilon}_{n,k} = \left(\frac{\hat{b}_{2,n-1,k}\, \hat{d}^*_n}{|\hat{d}_n|^2 + \varepsilon^2} \right)^* . \tag{13}$$

The damped quotient has the same phase as $\hat{b}_{2,n-1,k}/\hat{d}_n$, and its magnitude is $|\hat{b}_{2,n-1,k}|$ times $|\hat{d}_n|/(|\hat{d}_n|^2 + \varepsilon^2)$ instead of $1/|\hat{d}_n|$. The two agree when $|\hat{d}_n| \gg \varepsilon$, but the damped factor never exceeds $1/(2\varepsilon)$, whereas the exact one is unbounded as $\hat{d}_n \to 0$.

The derivation is in Appendix G, where we also set the frequency-dependent scale ε_n using the variance (Th. 5) of SDB coefficients in this setting, and concepts from Wiener filtering [47].

5 Experiments: Image Classification

We classify noisy and rotated MNIST [23], FashionMNIST [49] and real mouse osteosarcoma cell images ($N = 650$) [2, 24] to see where theoretical guarantees of Th. 1–5 (completeness, rotation invariance, and noise stability) translate into practical advantage. We use the SDB as a fixed, parameter-free descriptor and benchmark it against the DH power spectrum, Harmonic Networks [48], and convolutional and transformer baselines. In each setting, features are classified by an MLP. We vary MLP parameter counts $p \in \{10k, 50k, 500k\}$ and image resolutions $L \times L \in \{28 \times 28, 56 \times 56, 112 \times 112\}$. We train each classifier pipeline on one NVIDIA `A100` GPU.

Convolutional baselines are $SE(2)$ steerable networks [46], max and average pooled G-CNNs [7], the Selective G-Bispectrum [29], and the G-Triple Correlation [37, 29], all with 24 filters. For G-CNNs, we pool the final convolutional layer before passing features to the MLP classifier. Convolutional filters are trained end-to-end jointly with the MLP classifier. Vision encoder baselines are pre-trained transformers `DINOv3` [39] (`ViT-B/16`), `EVA-02` [13] (`ViT-B/16`), and `OpenCLIP` [17] (`ViT-L/14`) with backbones `vit_base_patch16_dinov3.lvd1689m`, `eva02_base_patch14_224.mim_in22k`, and `laion2b_s32b_b82k` respectively, used as frozen feature extractors. For `DINOv3`, `EVA-02`, and `OpenCLIP` we must up-sample each image to 224×224.

In the cell dataset, each image has a label indicating control, or one of two cancer drug treatments (`jasp` or `cytd`), for three classes in total. The train/test sets explicitly test rotation invariance. Train images are randomly rotated between $0°$ and $90°$ and test images between $90°$ and $180°$, ensuring that performance reflects true rotational invariance rather than interpolation within the training distribution. We add Gaussian noise $\mathcal{N}(0, \sigma^2)$ to each image pixel in both train and test sets with $\sigma^2 \in \{0, 0.2, 0.4\}$. We use canonical train/test sets for MNIST and FashionMNIST and 70/30 split for the osteosarcoma dataset. We repeat across 5 random seeds.

On the cell dataset, we also compare against the full disk bispectrum descriptor reduced by PCA, as was used in [53]. We don't include this descriptor in the MNIST or FashionMNIST experiments

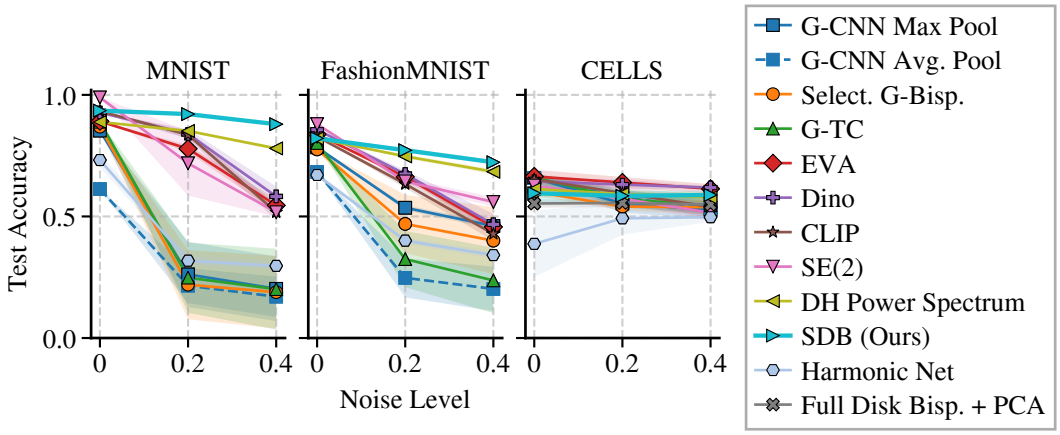

Figure 3: **Using the SDB as a rotation-invariant descriptor, MLPs can accurately classify noisy rotated MNIST digits, FashionMNIST classes, and cancer cell treatments.** The SDB descriptor achieves accuracy comparable to transformer descriptors and SE(2) network representations. It is more robust to noise than convolutional descriptors. Here, results are shown for $L = 28$ with a 500,000-parameter MLP, and markers show the mean over five seeds, with standard deviation shaded. Appendix Fig. 5 shows results for every combination of MLP and image size.

because computing it for every image across seed, image size, and noise levels would take about 220 GPU-days per dataset.

Results. The SDB outperforms convolution-based descriptors when any amount of noise is added to the image. This is true across dataset, image size, and MLP parameter budget, as shown in Fig. 3. Their relatively low accuracy reflects the difficulty of generalizing rotation invariance to a set of unknown angles, on noisy data. It is known that max and average pooled convolutional layers are not robust to noise [50]. On MNIST and FashionMNIST, the SDB outperforms transformer representations when significant up-sampling is required on very noisy images, mimicking realistic conditions for practitioners wishing to use pre-trained transformers on small, noisy datasets. The advantage narrows when transformer representations are classified using sufficiently large MLPs, and disappears at image size 112×112 and for all image resolutions of the cell dataset.

The DH power spectrum is a strong baseline because it retains the magnitude of every disk harmonic coefficient $|a_{n,k}|^2$. This is a fine-grained map of how the image's "energy" is distributed over angular frequency and radius, and discards only the phases that encode how these components are oriented relative to one another. The power spectrum seems to match the SDB on FashionMNIST and CELLS, where classes differ mainly in coarse shape and radial energy profile. Its classification accuracy is slightly worse on MNIST, where digits share similar strokes and differ in how those strokes are arranged. The SDB, which is complete up to rotation, thus matches the power spectrum where magnitudes suffice and outperforms it where classes differ in arrangement.

The SDB requires zero learned parameters, while convolutional and transformer encoders require thousands and hundreds of millions of learned parameters respectively. Despite this asymmetry in total system complexity, the SDB achieves competitive accuracy, validating that its theoretical guarantees translate into practical usage.

6 Experiments: Multi-Reference Alignment

Multi-reference alignment (MRA) is an active field of research for the 1D translation bispectrum [4, 6, 15]. Given multiple noisy, randomly translated observations of a ground truth signal, MRA seeks to recover the ground truth without explicit alignment. All noisy translated images are mapped to their bispectra. Then, the mean bispectrum is computed and inverted to signal space. Bispectrum-based MRA has previously been impossible on rotated images, as no inversion existed for a rotation-invariant bispectrum.

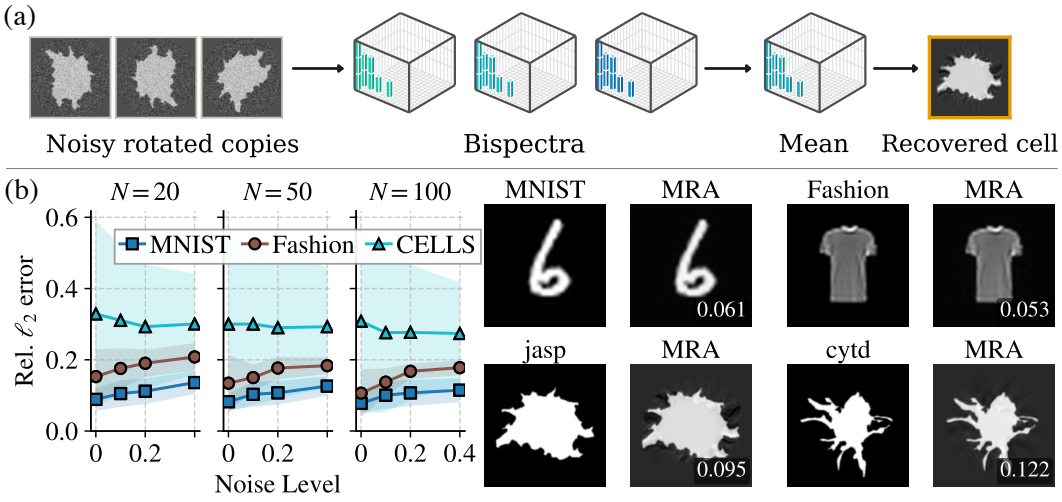

Figure 4: **The damped SDB inversion recovers templates from averaged noisy, rotated observations on all three datasets.** (a) The MRA task considers noisy images in unknown orientations and seeks to recover their common underlying signal by mapping each image to its bispectrum coefficients and inverting the images' average bispectrum. (b) Reconstructions for $L = 256$. In Appendix Fig. 6, we report these results for image sizes $L \in \{56, 112\}$.

Here, we show that SDB MRA is possible via our inversion algorithm (Th. 1) and suggest an MRA-specific damping technique to stabilize the inversion in this case (Th. 6). Starting from an initial noiseless 56×56, 112×112, or 256×256 image, we generate $n_x \in \{20, 50, 100\}$ randomly rotated and noisy $\sim \mathcal{N}(0, \sigma^2)$ observations under varying noise levels $\sigma^2 \in \{0.0, 0.1, 0.2, 0.4\}$ and evaluate reconstruction quality by comparing the relative error between the original image and the reconstructed image. The bispectrum inversion yields the original signal in an arbitrary orientation, so we register the reconstructed image to the original before computing the relative error.

Results. Our damped inversion (Th. 6) successfully recovers templates from averaged noisy and rotated observations, as shown in Fig. 4 across all three datasets. Simple reconstruction results in Appendix H directly verify our SDB inversion and thus our choice of SDB coefficients. MRA is a more complex setting, and successful reconstruction relies on our expectation (Th. 4) and variance (Th. 5) derivations, in order to properly account for bias and coefficient spread induced by noise. These results constitute the first demonstrated instance of bispectrum-based multi-reference alignment on rotationally-varied images, a task that was impossible prior to the inversion derived in this work.

7 Conclusion

We introduce a compact rotation-invariant image descriptor, the SDB, by deriving a minimal number of disk bispectrum coefficients necessary to reconstruct the image. We also provide accuracy and invariance guarantees for its approximation, as well as summary statistics and inversion stabilization techniques under Gaussian noise. Classification tasks and SDB MRA on rotated images demonstrate rotation-invariance and noise robustness of the SDB descriptor and viability of our SDB inversion (Th. 1), which in turn validates our choice of SDB coefficients.

Limitations. The SDB, like all bispectra, is invariant to one (global) group action: planar rotation. The SDB map is not invariant to rotation of local features, global translation, or reflection. Therefore, SDB is best used on centered single-object images, however we do discuss the effect of global translation in Appendix J. In addition, SDB seems to correctly group signal deformations *to an extent*, as it can classify MNIST and FashionMNIST, but a more thorough analysis of object deformations in bispectrum space would be interesting future work.

Acknowledgements

A. M. L. and N. M. acknowledge partial support from NSF grant 2313150, NSF Career grant 2240158 and from the Noyce foundation.

A. M. L. also acknowledges support from an NSF GRFP Fellowship and NSF Grant No. DGE-2125644.

N. M. also acknowledges a Goodfire Research Grant for credits using Silico, Goodfire's interpretability research agent.

References

[1] Pablo F. Alcantarilla, Adrien Bartoli, and Andrew J. Davison. Kaze features. In *European Conference on Computer Vision (ECCV)*, pages 214–227. Springer, 2012.

[2] Elaheh Alizadeh, Wenlong Xu, Jordan Castle, Jacqueline Foss, and Ashok Prasad. TISMorph: A tool to quantify texture, irregularity and spreading of single cells. *PLOS ONE*, 14(6):e0217346, 2019.

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision (ECCV)*, pages 404–417. Springer, 2006.

[4] Tamir Bendory, Nicolas Boumal, Chao Ma, Zhizhen Zhao, and Amit Singer. Bispectrum inversion with application to multireference alignment. *IEEE Transactions on Signal Processing*, 66(4):1037–1050, 2018.

[5] David R. Brillinger. Some history of higher-order statistics and spectra. *Statistica Sinica*, 1(2):465–476, 1991.

[6] Hua Chen, Mona Zehni, and Zhizhen Zhao. A spectral method for stable bispectrum inversion with application to multireference alignment. *IEEE Signal Processing Letters*, 25(7):911–915, 2018.

[7] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2990–2999, New York, New York, USA, 2016. PMLR.

[8] Taco Cohen, Mario Geiger, Jonas Koehler, and Max Welling. Spherical CNNs. *International Conference on Learning Representations (ICLR)*, 2018.

[9] David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3): 425–455, 1994.

[10] Wen-Xuan Du, Clifford H. Thurber, and Donna Eberhart-Phillips. Earthquake relocation using cross-correlation time delay estimates verified with the bispectrum method. *Bulletin of the Seismological Society of America*, 94(3):856–866, 2004.

[11] Dan Edidin and Joshua Katz. The reflection-invariant bispectrum: Signal recovery in the dihedral model: The reflection-invariant bispectrum: Signal recovery in the dihedral modeld. edidin, j. katz. *Journal of Fourier Analysis and Applications*, 32, 2025.

[12] Steve Elgar and RT Guza. Observations of bispectra of shoaling surface gravity waves. *Journal of Fluid Mechanics*, 161:425–448, 1985.

[13] Yuxin Fang, Quan Sun, Xinggang Wang, Tiejun Huang, Xinlong Wang, and Yue Cao. Eva-02: A visual representation for neon genesis. *Image and Vision Computing*, 149:105171, 2024.

[14] Georgios B. Giannakis. Signal reconstruction from multiple correlations: frequency- and time-domain approaches. *J. Opt. Soc. Am. A*, 6(5):682–697, 1989.

[15] James Herring, James Nagy, and Lars Ruthotto. Gauss–newton optimization for phase recovery from the bispectrum. *IEEE Transactions on Computational Imaging*, pages 1–1, 2019.

[16] M.K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2): 179–187, 1962.

[17] Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. Openclip, 2021. If you use this software, please cite it as below.

[18] Ramakrishna Kakarala. Completeness of bispectrum on compact groups. In *arXiv preprint arXiv:0902.0196*, 2009.

[19] Ramakrishna Kakarala. The bispectrum as a source of phase-sensitive invariants for fourier descriptors: A group-theoretic approach. *Journal of Mathematical Imaging and Vision*, 44, 2009.

[20] R. Kakarala. The bispectrum as a source of phase-sensitive invariants for fourier descriptors: a group-theoretic approach. *Journal of Mathematical Imaging and Vision*, 44(3):341–353, 2012.

[21] R. Kakarala, B. M. Bennett, M. D'Zmura, and G. J. Iverson. Bispectral techniques for spherical functions. In *Proceedings of ICASSP '93*, pages 237–240, Minneapolis, MN, 1993. IEEE.

[22] Georgia Koukiou. Identifying system non-linearities by fusing signal bispectral signatures. *Electronics*, 13 (7):1287, 2024.

[23] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. *ATT Labs*, 2, 2010.

[24] Wanxin Li, Ashok Prasad, Nina Miolane, and Khanh Dao Duc. Using a Riemannian elastic metric for statistical analysis of tumor cell shape heterogeneity. In *Geometric Science of Information (GSI 2023)*, pages 583–592. Springer Nature Switzerland, 2023.

[25] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[26] J. C. Marron, P. P. Sanchez, and R. C. Sullivan. Unwrapping algorithm for least-squares phase recovery from the modulo 2π bispectrum phase. *Journal of the Optical Society of America A*, 7(1):14–20, 1990.

[27] Nicholas F. Marshall, Oscar Mickelin, and Amit Singer. Fast expansion into harmonics on the disk: A steerable basis with fast radial convolutions. *SIAM Journal on Scientific Computing*, 45(5):A2431–A2457, 2023.

[28] Adrián Martín-Montero, Gonzalo C Gutiérrez-Tobal, David Gozal, Verónica Barroso-García, Daniel Álvarez, Félix Del Campo, Leila Kheirandish-Gozal, and Roberto Hornero. Bispectral analysis of heart rate variability to characterize and help diagnose pediatric sleep apnea. *Entropy*, 23(8):1016, 2021.

[29] Simon Mataigne, Johan Mathe, Sophia Sanborn, Christopher Hillar, and Nina Miolane. The selective g-bispectrum and its inversion: Applications to g-invariant networks. In *Advances in Neural Information Processing Systems*, pages 115682–115711. Curran Associates, Inc., 2024.

[30] Charles L. Matson. Weighted-least-squares phase reconstruction from the bispectrum. *Journal of the Optical Society of America A*, 8(12):1905–1913, 1991.

[31] Chrysostomos L Nikias and Jerry M Mendel. Signal processing with higher-order spectra. *IEEE Signal processing magazine*, 10(3):10–37, 1993.

[32] L. Pinilla, A. Mishra, and B. M. Sadler. Unique bispectrum inversion for signals with finite spectral/temporal support. In *Proceedings of ICASSP 2023*. IEEE, 2023.

[33] B. S. Reddy and B. N. Chatterji. An fft-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8):1266–1271, 1996.

[34] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2564–2571, 2011.

[35] Walter Rudin. *Fourier Analysis on Groups*. Courier Dover Publications, 1962.

[36] Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 3856–3866, 2017.

[37] Sophia Sanborn and Nina Miolane. A general framework for robust g-invariance in g-equivariant networks. In *Advances in Neural Information Processing Systems*, pages 67103–67124. Curran Associates, Inc., 2023.

[38] Sophia Sanborn, Christian Shewmake, Bruno Olshausen, and Christopher Hillar. Bispectral neural networks. In *International Conference on Learning Representations (ICLR)*, 2023.

[39] Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. Dinov3, 2025.

[40] G. Sundaramoorthy, M. R. Raghuveer, and S. A. Dianat. Bispectral reconstruction of signals in noise: amplitude reconstruction issues. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(7): 1297–1306, 1990.

[41] M. R. Teague. Image analysis via the general theory of moments. *JOSA*, 70(8):920–930, 1980.

[42] Nathaniel Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, P. Riley, and K. Kohlhoff. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. In *NeurIPS*, 2018.

[43] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of Ill-Posed Problems*. V. H. Winston & Sons, Washington, DC, 1977. Translation editor Fritz John; distributed by Halsted Press (Wiley), New York.

[44] J. Tukey. The spectral representation and transformation properties of the higher moments of stationary time series. *Reprinted in The Collected Works of John W. Tukey*, 1:165–184, 1953.

[45] Steven C. Verrall and Ramakrishna Kakarala. Disk-harmonic coefficients for invariant pattern recognition. *Journal of the Optical Society of America A*, 15(2):389, 1998.

[46] Maurice Weiler and Gabriele Cesa. General E(2)-equivariant steerable CNNs. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc.

[47] Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*. Technology Press of the Massachusetts Institute of Technology and John Wiley & Sons, Cambridge, MA and New York, 1949.

[48] Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7168–7177, 2017.

[49] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms, 2017.

[50] Afia Zafar, Noushin Saba, Ali Arshad, Amerah Alabrah, Saman Riaz, Mohsin Suleman, Shahneer Zafar, and Muhammad Nadeem. Convolutional neural networks: A comprehensive evaluation and benchmarking of pooling layer variants. *Symmetry*, 16(11), 2024.

[51] Ji-Wu Zhang, Chong-Xun Zheng, and An Xie. Bispectrum analysis of focal ischemic cerebral eeg signal using third-order recursion method. *IEEE transactions on biomedical engineering*, 47(3):352–359, 2000.

[52] Zhizhen Zhao and Amit Singer. Fourier–bessel rotational invariant eigenimages. *Journal of the Optical Society of America A*, 30(5):871–877, 2013.

[53] Zhizhen Zhao and Amit Singer. Rotationally invariant image representation for viewing direction classification in cryo-em. *Journal of structural biology*, 186(1):153–166, 2014.

A Formal Definitions

A.1 Groups and Group Action

A group $\mathcal{G}$ is a set $\{a, b, c, ...\} \in G$ equipped with an operation $\circ : G \times G \to G$ that combines two elements of the set to produce a third element within the same set. The operation must be associative, and have an identity element and inverse. A group can also act on an external set S via the group action $g : G \times S \to S$. For example, if $G = (\mathbb{Z}, +)$, group action could be addition acting on $\mathbb{R}$.

A.2 Equivariance, Invariance, and Irreducible Representations

We adopt [38]'s description. Consider a group G acting on a space X. Next, consider a function $\phi : X \to Y$. ϕ is *G-equivariant* if $\phi(g \circ x) = g' \circ \phi(x)$ for all $g \in G, x \in X$, where $g' \in G'$ is homomorphic to G. ϕ is *G-invariant* if $\phi(g \circ x) = \phi(x)$ for all $g \in G, x \in X$. ϕ is a *complete invariant* if $\phi(x_1) = \phi(x_2) \iff x_2 = g \circ x_1$, guaranteeing that ϕ maps all elements of an *orbit* $\{\phi(gx) : g \in G\}$ to a unique point in Y. A *representation* of a group is a map $\rho : G \to GL(V)$ that assigns elements of G to elements of the group of linear transformations (e.g. matrices) over a vector space V such that $\rho(g_1 g_2) = \rho(g_1)\rho(g_2)$. A representation is *reducible* if a change of basis can decompose the representation into a direct sum of other representations. An *irreducible representation* is one that cannot be further reduced in this manner. Put simply, *representation theory* describes group elements as matrices so that group action can be described by matrix multiplication.

A.3 Fourier Transform, Power Spectrum, and Translation-Invariant Bispectrum

The Fourier transform provides a canonical example of an equivariant representation. A signal $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ can be expanded in the Fourier basis as

$$f(x) = \sum_{k=0}^{n-1} F_k e^{i2\pi kx/n}. \tag{14}$$

Using shorthand $\nu = 2\pi k/n$,

$$F_\nu = |F_\nu| e^{-i\nu\phi_\nu} = \frac{1}{n} \sum_{x=0}^{n-1} f(x) e^{-i\nu x} \tag{15}$$

Respectively, $|F_\nu|$ and ϕ_ν are the magnitude and phase of F_ν, and both are crucial for signal reconstruction from Fourier coefficients using Eq. 14. From a group theoretic perspective, a Fourier transform is a projection of a signal onto the irreducible representations of $(\mathbb{Z}/n\mathbb{Z}, +)$, indexed by integer frequencies k. Under translation $g \in \mathbb{Z}/n\mathbb{Z}$, f transforms as $(g \circ f)(x) = f(x - g) = \tilde{f}(x)$ and the Fourier coefficients transform as

$$\tilde{F}_\nu = e^{-i\nu g} F_\nu = |F_\nu| e^{-i\nu(\phi_\nu + g)}, \tag{16}$$

making them equivariant to $(\mathbb{Z}/n\mathbb{Z}, +)$.

Next, invariant quantities can be constructed from equivariant ones. For example, the power spectrum

$$P_\nu = F_\nu F_\nu^* = |F_\nu| e^{-i\nu\phi_\nu} |F_\nu| e^{+i\nu\phi_\nu} = |F_\nu|^2 \tag{17}$$

does not change under translation $g \in \mathbb{Z}/n\mathbb{Z}$, since the phase cancels entirely. However, without ϕ_ν, signal reconstruction is impossible.

The *translation-invariant bispectrum* [44, 5, 31], developed in the mid-20th century for signal processing, resolves this limitation by constructing invariants that preserve relative phase information while disregarding global phase information altered by $(g \circ f)$. It is defined as

$$B_{\nu_1,\nu_2} = F_{\nu_1} F_{\nu_2} F^*_{\nu_1+\nu_2}. \tag{18}$$

Under translation,

$$\tilde{B}_{\nu_1,\nu_2} \to e^{-i\nu_1 g} F_{\nu_1} \cdot e^{-i\nu_2 g} F_{\nu_2} \cdot e^{i(\nu_1+\nu_2)g} F^*_{\nu_1+\nu_2} = B_{\nu_1,\nu_2}. \tag{19}$$

The key observation is that the bispectrum combines Fourier coefficients in a way that cancels global phase, while preserving relative phase relationships ϕ_ν between coefficients F_ν. As a result, it is invariant to translation but retains sufficient information for reconstruction under mild conditions, as shown in [14]. For this reason, it is referred to as a *complete invariant*. This construction illustrates a general principle: *invariants can be built by combining equivariant features that transform according to irreducible representations*.

$n \rightarrow$ $k \downarrow$	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
1	44	38	30	25	19	15	10	6	3	1	0	2	4	7	11	16	20	26	31	39	45
2	-	-	-	48	40	32	23	17	12	8	5	9	13	18	24	33	41	49	-	-	-
3	-	-	-	-	-	-	42	34	27	21	14	22	28	35	43	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	46	36	29	37	47	-	-	-	-	-	-	-	-

Table 3: **The SDB keeps one coefficient for each DH coefficient with $n \geq 0$, the right half of this index map, so its size is $(m + K_0)/2$.** DH coefficient index map $j \leftrightarrow (n, k)$, for an 8×8 image, truncated to the $\lfloor \pi L^2/4 \rfloor = 50$ lowest frequencies, i.e. one basis function per pixel inside the disk, the maximum truncation of Marshall et al. [27]. This is a larger truncation than the library default bandlimit used in Fig. 1, which keeps $m = 15$ coefficients at $L = 8$. Columns and rows are different values of n and k respectively, and table entries are j. Here, $m = 50$, $N_m = 10$, and $K_{\max} = 4$.

A.4 Bispectrum Inversion

To reconstruct the original signal from the bispectrum, Fourier coefficients can be reconstructed from bispectral coefficients according to the algorithm [14]:

$$1)|F_0| = |b_{00}|^{1/3}, arg(F_0) = arg(b_{00}); 2)|F_1| = \sqrt{\frac{b_{10}}{F_0}}; 3)F_{i>1} = \left(\frac{b_{1(i-1)}}{F_1 F_{i-1}}\right)^*. \quad (20)$$

The global phase of the reconstructed signal is set by the phase of F_1. [14] also developed an inversion algorithm for the 2D translation-invariant bispectrum, which is constructed using 2D Fourier coefficients and can therefore be defined directly on images. Because the signal can be fully reconstructed (up to a global translation) from its bispectrum coefficients, the bispectrum is referred to as a *complete invariant*. This also means that two signals, identical up to a global translation, will map to the same set of bispectrum coefficients.

Beyond [14]'s 1D translation bispectrum inversion, [40] developed another analytical inversion algorithm that is more robust to noise, [32] revisited this inversion problem, providing a gradient-based inversion procedure, and [4] derived an inversion optimized for multi-reference alignment. [14] also extended his 1D inversion scheme to invert the 2D translation-invariant bispectrum, invariant to the product of cyclic groups $C_n \times C_n$. [29] generalized that proof for the product of any number of cyclic groups, *i.e.*, to any G-bispectrum, for signals defined over a finite and commutative group G. A reflection-invariant bispectrum was also proposed by [11], who formulated a gradient descent optimization for its inversion. Methods to invert the G-bispectrum of signals defined on a (continuous) Lie group G have been proposed in [20, 21].

B Example Disk Harmonic Coefficient Truncation

Table 3 shows an example of this correspondence between j and (n, k) for an 8×8 image expanded in the truncated DH basis. In the disk bispectrum definition, $-N_m \leq n_{j_1} + n_{j_2} \leq N_m$, and $k_3 \leq K_{n_{j_1}+n_{j_2}}$.

C Proofs for Theorems on the Disk Bispectrum

Proposition 3 (Complexity of the disk bispectrum)**.** *Denote L the size of the image, m the number of disk harmonic coefficients, N_m the maximum angular frequency. The space complexity of the disk bispectrum is $O(m^3/N_m)$. The time complexity of the disk bispectrum is $O(L^3 + m^3/N_m)$.*

Proof. **Space complexity:** The space complexity is equal to the number of bispectral coefficients b_{j_1,j_2,k_3} for $j_1 \in \{1, ..., m\}$, $j_2 \in \{j \in \{1, ..., m\} \mid -N_m \leq n_{j_1} + n_{j_2} \leq N_m\}$ and $k_3 \in \{1, ..., K_{n_{j_1}+n_{j_2}}\}$:

$$\begin{aligned} N &= \sum_{j_1=1}^{m} \sum_{\substack{j_2=1 \\ -N_m \leq n_{j_1}+n_{j_2} \leq N_m}}^{m} \sum_{k_3=1}^{K_{n_{j_1}+n_{j_2}}} 1 \\ &= O(m^2.K_{\max}) = O(m^3/N_m), \end{aligned}$$

where $K_{\max} = \max\{K_n \mid n \in \{-N_m, ..., N_m\}\}$.

Time complexity: The time complexity is the number of operations required to compute the disk harmonic transform, and then 3 operations per bispectrum coefficient. The time complexity required to compute the disk harmonic transform is $O(L^3)$ [53], hence the time complexity is $O(L^3 + m^3/N_m)$. □

D Proofs for Theorems on the Selective Disk Bispectrum

Theorem 7 (Inversion). *Let $f : \mathbb{D} \to \mathbb{R}$ be a real-valued square-integrable function and let $a_{n,k}$ be its disk harmonic coefficients, bandlimited by λ. Assume that $a_{n,1} \neq 0$ for $n \in \{0, ..., N_m - 1\}$. Then, its disk bispectrum b can be inverted, that is f can be reconstructed from b up to a planar disk rotation and frequency λ. The time complexity of the inversion is $O(m + L^3)$.*

Proof. We show how to recover the DH coefficients of an image f, up to a global phase shift, from its disk bispectrum b coefficients. Afterwards, the disk harmonic expansion [45] in Sec. 2 can be used to reconstruct f, up to a global orientation shift and bandlimit.

Column $n = 0$ We first reconstruct the disk coefficients with Bessel order $n = 0$. We have:

$$b_{0,0,1} = a_{n_0,k_0}.a_{n_0,k_0}.a^*_{n_0+n_0,1} = a_{0,1}.a_{0,1}.a^*_{0,1} = a_{0,1}.\|a_{0,1}\|^2 \tag{21}$$

Hence: $\|a_{0,1}\| = \|b_{0,0,1}\|^{1/3}$ and $\arg(a_{0,1}) = \arg(b_{0,0,1})$. For the rest of this column, i.e., for $k_3 \leq K_{n_{j_1}+n_{j_2}} = K_0$:

$$b_{0,0,k_3} = a_{n_0,k_0}.a_{n_0,k_0}.a^*_{n_0+n_0,k_3} = a_{0,1}.a_{0,1}.a^*_{0,k_3} = a^2_{0,1}.a^*_{0,k_3} \tag{22}$$

Hence, $a_{0,k_3} = \left(\frac{b_{0,0,k_3}}{a^2_{0,1}}\right)^*$ for $k_3 \in \{1, ..., K_0\}$.

Column $n = 1$

Next, we reconstruct the disk coefficients with Bessel order $n = 1$. We have:

$$b_{2,0,1} = a_{n_2,k_2}.a_{n_0,k_0}.a^*_{n_0+n_2,1} = a_{1,1}.a_{0,1}.a^*_{1,1} = a_{0,1}.\|a_{1,1}\|^2 \tag{23}$$

Hence: $\|a_{1,1}\| = \|\frac{b_{2,0,1}}{a_{0,1}}\|^{1/2}$ and we fix its phase argument to 0, which fixes the orientation of the reconstructed image to something arbitrary: $\arg(a_{1,1}) = 0$. For the rest of this column, i.e., for every $k_3 \leq K_{n_{j_1}+n_{j_2}} = K_1$:

$$b_{2,0,k_3} = a_{n_2,k_2}.a_{n_0,k_0}.a^*_{n_0+n_2,k_3} = a_{1,1}.a_{0,1}.a^*_{1,k_3} \tag{24}$$

Hence $a_{1,k_3} = \left(\frac{b_{2,0,k_3}}{a_{0,1}.a_{1,1}}\right)^*$ for $k_3 \in \{1, ..., K_1\}$.

Column n from column $n - 1$

Assume that we have recovered the disk harmonic coefficients up to $n - 1$, i.e., all $a_{n',k'}$ for $n' \in \{0, ..., n - 1\}$ and $k' \in \{1, ..., K_{n'}\}$. We show that we can recover the nth column. We have, for every $k_3 \leq K_{n_{j_1}+n_{j_2}} \equiv K_n$:

$$b_{2,i(n-1,1),k} = a_{n_2,k_2}.a_{n_{i(n-1,1)},k_{i(n-1,1)}}.a^*_{n_2+n_{i(n-1,1)},k} = a_{1,1}.a_{n-1,1}.a^*_{n,k}, \tag{25}$$

where $i : (n, k) \to j$. Hence $a_{n,k} = \left(\frac{b_{2,i(n-1,1),k}}{a_{1,1}.a_{n-1,1}}\right)^*$ for $k \in \{1, ..., K_n\}$. With this, we can recover all disk harmonics up to the predefined bandlimit λ, assuming $a_{n,1} \neq 0$, which is standard in bispectrum inversions and satisfied with any amount of noise on the images. Choosing an arbitrary phase argument for $a_{1,1}$ means the disk harmonic expansion will reconstruct f up to a global rotation.

We note that reconstructing each disk harmonic coefficient requires exactly one disk bispectrum coefficient. Hence, the time complexity of this step is $O(m)$. Second, the time complexity of the inverse disk harmonic transform is $O(L^3)$, hence the final complexity of $O(m + L^3)$. □

Proposition 4 (Complexity of the selective disk bispectrum). *Denote L the size of the image, m the number of disk harmonic coefficients, N_m the maximum angular frequency. The number of selective bispectrum coefficients is $N = \sum_{n=0}^{N_m} K_n$. The space complexity of the selective disk bispectrum is $O(m)$. Its time complexity is $O(L^3 + m)$.*

Proof. The number of selective bispectrum coefficients is:

$$N = K_0 + \sum_{n_+=0}^{N_m-1} K_{n_++1} = K_0 + \sum_{n=1}^{N_m} K_n = \sum_{n=0}^{N_m} K_n. \tag{26}$$

Space complexity: We have $N = O(N_m.K_{\max}) = O(m)$ for space complexity.

Time complexity: The time complexity is the number of operations required to compute the disk harmonic transform, and then 3 operations per selective bispectrum coefficient. The time complexity required to compute the disk harmonic transform is $O(L^3)$ [53], hence the time complexity is $O(L^3 + m)$. □

Proposition 5 (Time complexity of the approximated selective disk bispectrum)**.** *Denote $L \times L$ the size of the image, m the number of DH coefficients. The time complexity of the approximate selective disk bispectrum, using DH coefficients from [27], is $O(L^2 \log L + m)$.*

Theorem 8 (Complete Invariance)**.** *Let f and f' be a pair of real-valued, square-integrable functions on the disk, and let the bispectrum be defined as in Selective Disk Bispectrum Definition. Assume that the disk harmonic coefficients $a_{n,1}$ of f are non-zero for all $n \in \{0, \dots, N_m - 1\}$. Then $f' = f \circ R_\phi$ for some 2D rotation $\phi \in SO(2)$ if and only if $b^f = b^{f'}$. Thus, the selective bispectrum is a complete invariant.*

Proof. **If $f' = f \circ R_\phi$ then $b^f = b^{f'}$.** We first prove that if $f' = f \circ R_\phi$ then $b^f = b^{f'}$, i.e., prove the invariance of the selective bispectrum. We compute the bispectral coefficients of f', as:

$$\begin{aligned}
b^{f'}_{j_1,j_2,k_3} &= a^{f'}_{n_{j_1},k_{j_1}} \cdot a^{f'}_{n_{j_2},k_{j_2}} \cdot {a^{f'}_{n_{j_1}+n_{j_2},k_3}}^* \\
&= e^{in_{j_1}\phi} \cdot a^{f}_{n_{j_1},k_{j_1}} \cdot e^{in_{j_2}\phi} \cdot a^{f}_{n_{j_2},k_{j_2}} \cdot \\
&\times e^{-i(n_{j_1}+n_{j_2})\phi} \cdot {a^{f}_{n_{j_1}+n_{j_2},k_3}}^* \\
&= a^{f}_{n_{j_1},k_{j_1}} \cdot a^{f}_{n_{j_2},k_{j_2}} \cdot {a^{f}_{n_{j_1}+n_{j_2},k_3}}^* \\
&= b^{f}_{j_1,j_2,k_3}.
\end{aligned}$$

Hence, two functions that differ from a rotation have the same full disk bispectrum and thus the same selective disk bispectrum.

If $b^f = b^{f'}$ then there exists $\phi \in SO(2)$ such that $f' = f \circ R_\phi$. Let $\tilde{f}$ be the function reconstructed from our inversion algorithms from $b^f = b^{f'}$. We have that there exists ϕ, ϕ' such that $\tilde{f} = f \circ R_\phi$ and $\tilde{f} = f' \circ R_{\phi'}$. Hence, $f' = f \circ R_{\phi - \phi'}$, where $\phi - \phi'$ is itself a rotation in $SO(2)$, completing the proof that f' and f differ by a planar rotation. □

E Proofs for Theorems on Approximating the Selective Disk Bispectrum

To further reduce the time complexity of the selective bispectrum, we propose leveraging a recent numerical approximation of the DH coefficients [27] which computes the disk harmonic transform in $O(L^2 \log L)$. We prove that using this DH coefficient approximation, we still have precise guarantees on the accuracy of the SDB coefficient approximations, the SDB rotation invariance, and reconstruction of the original image from SDB coefficients.

Theorem 9 (Accuracy Guarantee)**.** *Let $0 < \varepsilon \leq 1$ be a target accuracy, and assume that $\lambda \leq \sqrt{\pi p}$ and $|\log \varepsilon| \leq \sqrt{p}$, where λ is the bandlimit for the disk harmonic coefficients and $p = L^2$ the number of pixels in the image f of size L. Let b^f and $\tilde{b}^f$ denote the exact and approximate selective disk bispectra of f, where $\tilde{b}^f$ is computed using the approximate disk harmonic transform of Marshall et al. [27]. Then,*

$$\|\tilde{b}^f - b^f\|_\infty \leq C\varepsilon \|f\|_1,$$

Proof. Consider one coefficient of the selective disk bispectrum b and its approximate version $\tilde{b}$ as:

$$b = a_1 \cdot a_2 \cdot a_3, \quad \text{and:} \quad \tilde{b} = \tilde{a}_1 \cdot \tilde{a}_2 \cdot \tilde{a}_3, \tag{27}$$

where we index the disk harmonic coefficients by $1, 2, 3$ for simplicity, without loss of generality.

The approximate disk harmonic coefficients, $\tilde{a}$, are given by Marshall et al. [27]. Using their Theorem 4.1., we know that:

$$\|\tilde{a} - a\|_{\ell^\infty} \leq \varepsilon \|f\|_{\ell^1}. \tag{28}$$

Hence, we can compute:

$$\begin{aligned}
&|\tilde{b} - b| \\
&= |\tilde{a}_1 \cdot \tilde{a}_2 \cdot \tilde{a}_3 - a_1 \cdot a_2 \cdot a_3| \\
&= |\tilde{a}_1 \cdot \tilde{a}_2 \cdot \tilde{a}_3 - a_1 \cdot \tilde{a}_2 \cdot \tilde{a}_3 + a_1 \cdot \tilde{a}_2 \cdot \tilde{a}_3 - a_1 \cdot a_2 \cdot \tilde{a}_3 \\
&\quad + a_1 \cdot a_2 \cdot \tilde{a}_3 - a_1 \cdot a_2 \cdot a_3| \\
&= |(\tilde{a}_1 - a_1) \cdot \tilde{a}_2 \cdot \tilde{a}_3 + a_1 \cdot (\tilde{a}_2 - a_2) \cdot \tilde{a}_3 \\
&\quad + a_1 \cdot a_2 \cdot (\tilde{a}_3 - a_3)| \\
&\leq |\tilde{a}_1 - a_1| \cdot |\tilde{a}_2| \cdot |\tilde{a}_3| + |a_1| \cdot |\tilde{a}_2 - a_2| \cdot |\tilde{a}_3| \\
&\quad + |a_1| \cdot |a_2| \cdot |\tilde{a}_3 - a_3| \text{ (triangle inequality)} \\
&\leq \varepsilon \|f\|_{\ell^1} \cdot |\tilde{a}_2| \cdot |\tilde{a}_3| + |a_1| \cdot \varepsilon \|f\|_{\ell^1} \cdot |\tilde{a}_3| \\
&\quad + |a_1| \cdot |a_2| \cdot \varepsilon \|f\|_{\ell^1} \text{ ([27] Th. 4.1)} \\
&= \varepsilon \|f\|_{\ell^1} \cdot (|\tilde{a}_2| \cdot |\tilde{a}_3| + |a_1| \cdot |\tilde{a}_3| + |a_1| \cdot |a_2|) \\
&\quad \text{(reorganizing terms)}
\end{aligned}$$

To conclude, we must bound $|a_j|$ and $|\tilde{a}_j|$. We start with the exact disk harmonic coefficient. By definition:

$$a_j = \int_{\mathbb{D}} f(x).\psi_j^*(x)dx = \langle f, \psi_j \rangle, \tag{29}$$

where ψ_j is the disk harmonic of index j. Applying the Cauchy–Schwarz inequality, we get:

$$|a_j| \leq \|f\|_2 \cdot \|\psi_j\|_2 = \|f\|_2, \tag{30}$$

where we use the fact that the disk harmonics form an orthonormal basis and thus have norm 1.

Then, we have:

$$|\tilde{a}_j| = |\tilde{a}_j - a_j + a_j| \leq |a_j| + |\tilde{a}_j - a_j| \leq \|f\|_2 + \varepsilon \|f\|_1 \tag{31}$$

We plug this in the inequality derivations above.

$$\begin{aligned}
|\tilde{b} - b| &\leq \varepsilon \|f\|_{\ell^1} \cdot (|\tilde{a}_2| \cdot |\tilde{a}_3| + |a_1| \cdot |\tilde{a}_3| + |a_1| \cdot |a_2|) \\
&\leq \varepsilon \|f\|_{\ell^1} \cdot \{(\|f\|_2 + \varepsilon \|f\|_1)^2 \\
&\quad + \|f\|_2 \cdot (\|f\|_2 + \varepsilon \|f\|_1) + \|f\|_2 \cdot \|f\|_2\}.
\end{aligned}$$

Since the disk is a compact domain, f can be bounded from above by a constant, and hence its L_1 and L_2 norms are also bounded from above. Hence, we can bound the terms in the parenthesis by a constant C.

$$\begin{aligned}
|\tilde{b} - b| &\leq \varepsilon \|f\|_{\ell^1} \cdot (|\tilde{a}_2| \cdot |\tilde{a}_3| + |a_1| \cdot |\tilde{a}_3| + |a_1| \cdot |a_2|) \\
&\leq \varepsilon \|f\|_{\ell^1} \cdot C
\end{aligned}$$

This concludes the proof. □

As a corollary, we find that the bispectrum coefficients are approximately invariant with controlled guarantees.

Corollary 2 (Invariance Guarantee)**.** *Let f, f' be a pair of real-valued square-integrable functions on the disk such that there exists a $\phi \in SO(2)$ such that $f' = f \circ R_\phi$. Under the notations and assumptions of Theorem 3, we have:*

$$\|\tilde{b}^f - \tilde{b}^{f'}\|_\infty \leq 2C\varepsilon \|f\|_1,$$

for a constant C depending only on a bound of f on its compact disk support.

Proof. Consider one coefficient of $\tilde{b}^f$ and $\tilde{b}^{f'}$, respectively. We have

$$\begin{aligned}
|\tilde{b}^f - \tilde{b}^{f'}| &= |\tilde{b}^f - b^f + b^f - b^{f'} + b^{f'} - \tilde{b}^{f'}| \\
&\leq |\tilde{b}^f - b^f| + |b^f - b^{f'}| + |b^{f'} - \tilde{b}^{f'}| \\
&\quad \text{(triangle inequality)} \\
&= |\tilde{b}^f - b^f| + |b^{f'} - \tilde{b}^{f'}| \\
&\quad \text{(invariance of bispectrum)} \\
&\leq 2C\varepsilon\|f\|_1 \\
&\quad (f \text{ and } f' \text{ have the same bound on the disk}).
\end{aligned}$$

If all coefficients are bounded, then their L_∞ norm is also bounded. This concludes the proof. □

Linking the accuracy to f is useful, as it allows us to control accuracy through number of pixels p. If $|f|$ and C are large, but we still want to reach a given accuracy, then we can choose p accordingly such that $|\log \varepsilon| \leq \sqrt{p}$ to match the desired accuracy.

F Details of Classification MLP and Hyperparameters

We vary MLP parameter counts $p \in \{10k, 50k, 500k\}$ and image resolutions $L \times L \in \{28 \times 28, 56 \times 56, 112 \times 112\}$. The MLP architecture (0–2 hidden layers) is determined by the input feature size and parameter budget. We train for 15 epochs with learning rate 0.005, dropout 0.1, and batch size 512. We use the classic train/test split of MNIST and FashionMNIST. 20 percent of the train dataset is used for validation.

G Stable MRA Under Noise

Setting. We observe N copies $x_i = R_{\theta_i} f + \eta_i$, $i = 1, \ldots, N$, of a template f with unknown rotations $\theta_i \sim \mathrm{Unif}[0, 2\pi)$ and white pixel noise $\eta_i \sim \mathcal{N}(0, \sigma^2 I)$. Since the SDB is rotation-invariant, the SDB of each observation approximates the SDB of f. Without inversion damping, we would average them, subtract the bias of Th. 4, and invert the result $\hat{b}$ with the recursion of Th. 1. The template is recovered only up to a global rotation, so reconstruction error is measured after minimization over $SO(2)$.

Why bias correction is not enough. After the bias is subtracted, $\hat{b}$ is an unbiased estimate of b, so $\hat{b} \to b$ as $N \to \infty$ and the recursion recovers f in the limit. At any finite N, however, $\hat{b}$ differs from b by a random fluctuation (Th. 5), and the recursion divides by estimated quantities. Bias correction fixes where $\hat{b}$ is centred; it does not prevent a divisor from being a random number close to zero. That is the problem addressed here.

The recursion under noise. For $n \geq 2$ the recursion recovers

$$\hat{a}_{n,k} = \left(\frac{\hat{b}_{2,n-1,k}}{\hat{d}_n} \right)^*, \qquad \hat{d}_n = \hat{a}_{1,1}\, \hat{a}_{n-1,1}, \tag{32}$$

with $\hat{a}_{1,1}$ and $\hat{a}_{n-1,1}$ recovered at earlier steps. With exact inputs, $b_{2,n-1,k} = a_{1,1} a_{n-1,1} a^*_{n,k}$ contains the divisor $d_n = a_{1,1} a_{n-1,1}$ as a factor, so it cancels and (32) returns $a_{n,k}$ without issue, however small d_n is. However, if the estimations of $b_{2,n-1,k}$ and d_n are corrupted by noise after averaging many noisy signals, the appropriate terms do not cancel exactly. If d_n is close to zero and does not cancel, this causes instabilities that propagate through the recursive inversion.

Noise floor. Each observation is the rotated template plus pixel noise of standard deviation σ. Because the DH basis is orthonormal, the same is true of its coefficients. Every $a_{n,k}$ is the template coefficient, rotated, plus noise of standard deviation σ. For the SDB coefficieint $b_{2,n-1,k} = a_{1,1}\, a_{n-1,1}\, a^*_{n,k}$ with $n \geq 2$, Th. 13 gives,

$$\begin{aligned}
\mathrm{Var}(b_{2,n-1,k}) = \sigma^2 \big(&|a_{1,1}|^2 |a_{n-1,1}|^2 + |a_{1,1}|^2 |a_{n,k}|^2 + |a_{n-1,1}|^2 |a_{n,k}|^2 \big) \\
&+ \sigma^4 \big(|a_{1,1}|^2 + |a_{n-1,1}|^2 + |a_{n,k}|^2 \big) + \sigma^6,
\end{aligned} \tag{33}$$

and the average over N independent observations has variance $\mathrm{Var}(b_{2,n-1,k})/N$. The σ^4 and σ^6 terms are negligible against the σ^2 term when $\sigma \ll |a_{1,1}|$ and $\sigma \ll |a_{n-1,1}|$ (a condition on ratios, since σ and the coefficients have the same units). We assume this in what follows.

The recursion is seeded by $\hat{a}_{0,1} = \hat{b}_{001}^{1/3}$ and $|\hat{a}_{1,1}| = (\hat{b}_{201}/\hat{a}_{0,1})^{1/2}$ (Th. 1). By (201), the averaged $\hat{b}_{001}$ has standard deviation $\sqrt{9\sigma^2 a_{0,1}^4/N}$ at leading order, so, to first order, $\hat{a}_{0,1}$ has standard deviation $\sqrt{9\sigma^2 a_{0,1}^4/N}/(3a_{0,1}^2) = \sigma/\sqrt{N}$. This error only changes the mean brightness of the reconstruction, and since $a_{0,1}$ is the largest coefficient it is negligible in relative terms, so we take $\hat{a}_{0,1} = a_{0,1}$ in what follows. The coefficient that matters is $a_{1,1}$, which enters every divisor d_n. By (203), the averaged $\hat{b}_{201}$ has standard deviation $\sqrt{\sigma^2(|a_{1,1}|^4 + 2a_{0,1}^2|a_{1,1}|^2)/N}$ at leading order. Since $|\hat{a}_{1,1}|^2 = \hat{b}_{201}/a_{0,1}$ with $a_{0,1}$ fixed, $|\hat{a}_{1,1}|^2$ has standard deviation $\mathrm{std}(\hat{b}_{201})/a_{0,1}$; and since $|\hat{a}_{1,1}| = (|\hat{a}_{1,1}|^2)^{1/2}$ with $\frac{d}{dy}\, y^{1/2} = 1/(2y^{1/2})$, to first order $|\hat{a}_{1,1}|$ has standard deviation $\mathrm{std}(|\hat{a}_{1,1}|^2)/(2|a_{1,1}|)$. Together,

$$\begin{aligned} \mathrm{std}(|\hat{a}_{1,1}|) &= \frac{\mathrm{std}(\hat{b}_{201})}{2a_{0,1}|a_{1,1}|} = \frac{\sqrt{\sigma^2(|a_{1,1}|^4 + 2a_{0,1}^2|a_{1,1}|^2)/N}}{2a_{0,1}|a_{1,1}|} \\ &= \frac{\sigma}{\sqrt{N}}\sqrt{\frac{1}{2} + \frac{|a_{1,1}|^2}{4a_{0,1}^2}} \approx \frac{\sigma}{\sqrt{2N}}, \end{aligned} \tag{34}$$

where the last step follows since $a_{0,1}$ is the largest coefficient. The phase of $a_{1,1}$ is fixed to zero, which sets the global rotation of the reconstruction.

For a general step, write $e_b = \hat{b}_{2,n-1,k} - b_{2,n-1,k}$ for the error of the averaged bispectrum. If the divisor were exact, the recursion (32) would pass this error through as e_b/d_n, of standard deviation

$$\frac{\sqrt{\mathrm{Var}(b_{2,n-1,k})/N}}{|a_{1,1}||a_{n-1,1}|} \approx \frac{\sigma}{\sqrt{N}}\sqrt{1 + \frac{|a_{n,k}|^2}{|a_{n-1,1}|^2} + \frac{|a_{n,k}|^2}{|a_{1,1}|^2}}. \tag{35}$$

The divisor is not exact, but its error is built from the same noise as e_b and, to first order, cancels part of (35) rather than adding to it, just as d_n cancels exactly in the noiseless case; (35) is therefore a conservative estimate of the error in each recovered coefficient. The square root is of order one, since the coefficients decay with n and $a_{1,1}$, $a_{n-1,1}$ are leading radial modes. Up to the global rotation, which the error metric removes, each recovered coefficient therefore carries an error of order

$$s = \frac{\sigma}{\sqrt{N}}, \tag{36}$$

the usual $1/\sqrt{N}$ of averaging N measurements. We call s the noise floor: a recovered coefficient smaller than s cannot be distinguished from noise.

Why the recursion fails. This holds only while the divisors are well above the noise. Because the coefficients of any template decay with n while s does not, there is always a frequency beyond which $|a_{n-1,1}| \lesssim s$. There $\hat{a}_{n-1,1}$ is mostly noise, so $\hat{d}_n$ is a random number of scale $|a_{1,1}|s$ centred near zero, and the next step divides $\hat{b}_{2,n-1,k}$, whose own noise does not cancel against it, by this number. The result can be arbitrarily large, and through the chain it corrupts every higher frequency. More observations lower s as $1/\sqrt{N}$, which moves this frequency up but does not remove it. This failure mode is known. For the analogous recursion in one-dimensional multi-reference alignment, called frequency marching, Bendory et al. [4] note that it is exact without noise but that, because of its recursive nature, error in the low frequencies propagates to the high frequencies; in bispectral phase recovery for speckle imaging the recursion was replaced by weighted least squares over all bispectrum coefficients for the same reason [26, 30], and Bendory et al. [4] likewise turn to least-squares and semidefinite formulations. We keep the recursion and regularize the division.

Damped inversion. We replace the division in (32) by

$$\hat{a}_{n,k}^{\varepsilon} = \left(\frac{\hat{b}_{2,n-1,k}\,\hat{d}_n^*}{|\hat{d}_n|^2 + \varepsilon^2}\right)^* = \frac{|\hat{d}_n|^2}{|\hat{d}_n|^2 + \varepsilon^2}\,\hat{a}_{n,k}, \tag{37}$$

which is the exact recovery multiplied by a real factor between 0 and 1. This is the Tikhonov-regularized inverse of $\hat{d}_n$ [43]. When $|\hat{d}_n| \gg \varepsilon$ the factor is one and nothing changes. When $|\hat{d}_n| \ll \varepsilon$ the factor tends to zero, and the coefficient is set to zero rather than to the arbitrarily large value of the exact recursion. For every $\hat{d}_n$, the damped quotient has magnitude $|\hat{b}_{2,n-1,k}|\,|\hat{d}_n|/(|\hat{d}_n|^2+\varepsilon^2) \le |\hat{b}_{2,n-1,k}|/(2\varepsilon)$. With ε at the noise scale of $\hat{d}_n$, the coefficients suppressed are exactly those whose divisor cannot be distinguished from noise, where nothing was recoverable in the first place.

Choice of ε. The noise in $\hat{d}_n = \hat{a}_{1,1}\hat{a}_{n-1,1}$ is mainly that of $\hat{a}_{n-1,1}$, of standard deviation s, multiplied by $|a_{1,1}|$, so the natural scale is $\varepsilon = |a_{1,1}|s$. At exactly this scale a divisor at the noise level is only halved, whereas a predecessor far below the noise should be removed almost entirely. We therefore enlarge the scale by the noise-to-signal ratio of the predecessor, as a Wiener filter does [47],

$$\varepsilon_n^2 = |\hat{a}_{1,1}|^2\, s^2 \left(1 + \frac{s^2}{|a_{n-1,1}|^2}\right), \tag{38}$$

so that the bracket is near one when $|a_{n-1,1}| \gg s$ and grows without bound when $|a_{n-1,1}| \ll s$.

Estimating σ and $|a_{n-1,1}|^2$. Both follow from the second moments of Th. 12. For a noisy coefficient $\tilde{a}_{n,k} = a_{n,k} + N_{n,k}$, $\mathbb{E}[N_{n,k}] = 0$ and $\mathbb{E}|N_{n,k}|^2 = \sigma^2$ give $\mathbb{E}|\tilde{a}_{n,k}|^2 = |a_{n,k}|^2 + \sigma^2$, and a rotation of the observation multiplies $a_{n,k}$ by a unit phase, so the identity holds for every observation regardless of its unknown rotation. The mean squared norm over observations, $\hat{S}_{n,k} = N^{-1}\sum_i |a_{n,k}(x_i)|^2$, therefore estimates $|a_{n,k}|^2 + \sigma^2$. At the highest angular frequencies the signal term is negligible, leaving σ^2, so we take

$$\hat{\sigma}^2 = \operatorname{median}_{n \ge \tau n_{\max}} \hat{S}_{n,k}, \qquad |a_{n-1,1}|^2 \approx \max\big(0,\ \hat{S}_{n-1,1} - \hat{\sigma}^2\big), \tag{39}$$

with $\tau = 0.7$, the median guarding against residual signal in the same way as the noise-level estimator of Donoho and Johnstone [9]. The predecessor power is taken from $\hat{S}_{n-1,1}$ rather than from the recovered $\hat{a}_{n-1,1}$ because $\hat{S}_{n-1,1}$ never passes through a division, so it remains more reliable.

All multi-reference alignment results, shown in Fig. 6, use the damped recursion with ε_n from (38).

H SDB Round Trip Error

Reconstruction accuracy is tied to the accuracy of the bispectrum map. Theorem 3 bounds the error between the exact bispectrum and the bandlimited implemented bispectrum by a term that shrinks as the number of pixels grows. We demonstrate this on full evaluation sets across six datasets at $L \in \{28, 56, 112, 256, 500\}$, using all N=650 osteosarcoma cell silhouettes [2, 24] and the full 10,000-image test sets for MNIST, Fashion-MNIST, rotated and translated (RT) MNIST, rotated and translated (RT) Fashion-MNIST, and CIFAR10. Errors are median relative ℓ_2 after rotation registration. Each entry is SDB inversion error (with DHC round-trip error in parentheses). Example reconstructions for each dataset and image size can be found in Fig. 7.

dataset	L=28	L=56	L=112	L=256	L=500
osteosarcoma cells	0.259 (0.051)	0.153 (0.073)	0.076 (0.030)	0.086 (0.056)	0.079 (0.038)
CIFAR10	0.281 (0.133)	0.207 (0.160)	0.144 (0.121)	0.077 (0.061)	0.055 (0.044)
Fashion-MNIST	0.342 (0.123)	0.245 (0.178)	0.153 (0.115)	0.059 (0.019)	0.038 (0.009)
RT Fashion-MNIST	0.360 (0.138)	0.305 (0.242)	0.206 (0.171)	0.088 (0.044)	0.053 (0.015)
MNIST	0.446 (0.134)	0.250 (0.158)	0.136 (0.085)	0.064 (0.017)	0.043 (0.007)
RT MNIST	0.481 (0.163)	0.315 (0.203)	0.163 (0.102)	0.069 (0.022)	0.046 (0.009)

At L=500, a common microscopy image size, every dataset has median SDB error at most 0.08. The highest is osteosarcoma cells at 0.079 and the lowest is Fashion-MNIST at 0.038. Osteosarcoma cells already reach 0.076 at L=112. At this resolution the reconstructed cell morphology is visually not distinguishable from the original. Most residual numerical error is an intensity fade, for example white turning to grey, while structure is preserved.

The key point is that SDB error stays close to DHC round-trip error at large L. Much of the reconstruction error is already present when we project onto disk harmonics and reconstruct the image, without ever forming the SDB. Selective inversion does not introduce a large new failure mode on top of bandlimiting.

L	λ	SDB fwd (s)	SDB inv (s)	DHC fwd (s)	DHC inv (s)
28	20	0.104 ± 0.004	0.091 ± 0.004	0.075 ± 0.001	0.077 ± 0.004
56	20	0.104 ± 0.002	0.122 ± 0.008	0.075 ± 0.001	0.099 ± 0.002
112	28	0.149 ± 0.035	0.146 ± 0.024	0.121 ± 0.032	0.107 ± 0.006
256	64	0.122 ± 0.003	0.422 ± 0.003	0.095 ± 0.003	0.273 ± 0.002
500	125	0.167 ± 0.002	1.537 ± 0.008	0.146 ± 0.004	0.967 ± 0.004

Table 4: **The SDB forward pass stays comparable to a DHC transform at every resolution.** Forward and inverse wall-clock time versus image size. Inversion cost grows with L because the sequential recovery walks every selected coefficient, reaching 1.5 s per image at $L = 500$.

I SDB Timing

Table 4 reports wall-clock time to compute the SDB (forward) and to invert it (inverse), compared with the disk-harmonic coefficient (DHC) transform of the same image. Times are mean$\pm$s.d. over five seeds of five timed repetitions on a single image, at the library default bandlimit $\lambda = \max(20, L/4)$, matching the round-trip errors of Sec. H.

J SDB Under Image Translation

SDB is rotation-invariant but not translation-invariant, so translations change the descriptor. We therefore study the effect of translations on the SDB representation. A translation is called *admissible* when the object's support remains inside the computational disk after the translation (no clipping at the disk boundary).

Practical context. At high SNR (for example, histopathology or clean microscopy), centering is cheap and reliable. An intensity center-of-mass or a single registration pass costs $O(N)$ in the number of pixels and is already part of most pipelines. At low SNR (for example, cryo-EM), centering can be harder, and particles may retain residual translations. SDB is suited to settings where reliable centering is available; the experiments below show that it also tolerates residual translations.

Experiment (i): SDB separability for translated MNIST. We take one center-of-mass-centered prototype per digit. For each maximum translation radius $r \in \{0, 1, 2, 4, 8\}$ pixels (0%, 1.8%, 3.6%, 7.1%, and 14.3% of the disk diameter 56), we draw twenty random admissible translations with $|\Delta_x|, |\Delta_y| \le r$ and compute pairwise ℓ_2 distances on the real and imaginary parts of the SDB vectors. *Separability* is the mean between-digit SDB ℓ_2 divided by the mean within-digit distance. A ratio $\gg 1$ means digit identity remains recoverable; a ratio near 1 means collapse. At $r = 0$, the ratio is ∞.

max translation r	% of disk diameter	SDB sep. after translation
0 px	0%	∞
1 px	1.8%	22.6
2 px	3.6%	12.8
4 px	7.1%	7.2
8 px	14.3%	3.7

Even without perfect centering, same-digit SDBs stay clustered at moderate residual translations. Digit identity therefore remains recoverable from ℓ_2 distances in SDB space.

Experiment (ii): classification of translated and rotated MNIST. We evaluate classification on images that are both rotated and translated with an admissible translation, and include SE(2) and Harmonic baselines. We use $L \in \{28, 56, 112\}$ and an MLP sized $\approx 50{,}000$ parameters. We report mean$\pm$std over seeds (typically n=5; a few cells use n=4), for L and $\sigma \in \{0, 0.2, 0.4\}$.

L	σ	SDB	G-CNN (max)	`select_gbisp`	`tc`	SE(2)	Harmonic	DINOv3	EVA	OpenCLIP
28	0	.837±.002	.720±.006	.712±.005	.761±.002	.969±.000	.681±.091	.887±.002	.865±.005	.904±.003
	0.2	.743±.039	.225±.080	.199±.094	.217±.093	.613±.203	.196±.017	.762±.009	.722±.011	.764±.006
	0.4	.551±.072	.177±.075	.163±.085	.169±.087	.171±.107	.145±.018	.483±.030	.440±.029	.442±.019
56	0	.850±.001	.740±.003	.732±.003	.788±.003	.978±.001	.897±.045	.916±.003	.896±.007	.938±.001
	0.2	.812±.021	.278±.074	.233±.078	.275±.073	.903±.061	.283±.047	.862±.008	.851±.008	.896±.004
	0.4	.735±.042	.175±.064	.158±.062	.165±.065	.659±.296	.188±.025	.744±.006	.727±.006	.728±.006
112	0	.808±.002	.724±.006	.693±.004	.757±.004	.981±.000	.872±.016	.914±.002	.901±.003	.944±.002
	0.2	.772±.034	.399±.060	.332±.057	.398±.064	.919±.048	.562±.035	.896±.006	.870±.004	.928±.005
	0.4	.697±.032	.243±.055	.201±.048	.224±.051	.907±.053	.273±.052	.848±.008	.827±.006	.891±.002

Under translated and rotated MNIST, SDB stays far above G-CNN, `select_gbisp`, `tc`, and Harmonic as noise grows. SE(2) is strongest at larger L across noise levels; at L=28 it is strong when noiseless but drops sharply as σ grows. Among pretrained encoders, OpenCLIP and DINOv3 lead at low σ, while SDB is competitive at higher noise, especially at L=56.

Translation equivariance. SDB is not translation-invariant. More precisely, under an in-disk support condition, it is translation-equivariant at the disk-harmonic and SDB level. This matters because the change under small admissible translations is structured and continuous in the translation, which explains the clustering in Experiment (i).

Take the unit disk $D = \{x : |x| \leq 1\}$ with orthonormal basis $\{\psi_{nk}\}$ and coefficients $a(f) = (a_{nk}(f))$, where $a_{nk}(f) = \langle f, \psi_{nk}\rangle$. Let the object be supported in a smaller disk $K_0 = \{x : |x| \leq \rho\}$ with $\rho < 1$, and zero outside K_0. An admissible translation τ satisfies $|\tau| + \rho \leq 1$, so the translated support K_τ stays inside D. Define the translated image by $(T_\tau f)(x) = f(x - \tau)$ on K_τ and 0 elsewhere. On this class of images, both a and T_τ are linear. Expanding f in the basis and using linearity gives $a(T_\tau f) = U_\tau a(f)$, where U_τ is a fixed matrix depending on τ and D (but not on f), with entries given by inner products of translated basis functions against the original basis. Each SDB coefficient is a fixed triple product of disk-harmonic coefficients, so applying SDB after translation is the same as applying the triple-product map to $U_\tau a(f)$. Thus SDB is equivariant under admissible translations, not invariant. Center rotations cancel in the triple products, yielding rotation invariance. As $\tau \to 0$, U_τ approaches the identity, so slightly translated SDBs stay near the centered SDB and Experiment (i) separability remains $\gg 1$ at small r.

K Proofs for the Selective Disk Bispectrum Distribution Under Gaussian Noise

Theorem 10 (SDB Expectation)**.** *Consider a noiseless signal f with bispectrum b and DH coefficients $a_{n,k}$. Upon corruption with Gaussian noise, $\tilde{f} = f + n$ for $n \in \mathcal{N}(0, \sigma^2)$. The expectation of $\tilde{b}$ is given by*

$$\mathbb{E}[\tilde{b}_{00k}] = b_{00k} + \sigma^2 a^*_{0,k} + \delta_{k,1} 2\sigma^2 a_{0,1}; \quad \mathbb{E}[\tilde{b}_{201}] = b_{201} + \sigma^2 a_{0,1}; \quad \textit{Else, } \mathbb{E}[\tilde{b}_{2nk}] = b_{2nk}, \tag{40}$$

which can be re-written as

$$\begin{cases} \mathbb{E}[\tilde{b}_{00(k=1)}] = b_{001} + \sigma^2 a^*_{0,1} + 2\sigma^2 a_{0,1} \\ \mathbb{E}[\tilde{b}_{00(k\neq 1)}] = b_{00k} + \sigma^2 a^*_{0,k} \\ \mathbb{E}[\tilde{b}_{201}] = b_{201} + \sigma^2 a_{0,1} \\ \textit{Else, no correction: } \mathbb{E}[\tilde{b}_{2nk}] = b_{2nk}. \end{cases} \tag{41}$$

Proof. Consider a signal f corrupted with Gaussian noise such that $\tilde{f} = f + n$, and $n \sim \mathcal{N}(0, \sigma^2)$. The DHCs of the noiseless signal f are given by

$$a_{n_j k_j} = \int_0^{2\pi} \int_0^1 c_{nk} f(r, \theta) J_n(\lambda_{nk} r) e^{-in\theta} r dr d\theta. \tag{42}$$

The DHCs of the noisy signal $\tilde{f}$ are given by

$$\tilde{a}_{n_j k_j} = \int_0^{2\pi} \int_0^1 c_{nk} [f(r, \theta) + n(r, \theta)] J_n(\lambda_{nk} r) e^{-in\theta} r dr d\theta \tag{43}$$

$$= a_{n_j, k_j} + N_{n_j, k_j} \tag{44}$$

where we denote N_{n_j,k_j} as the DHC of the noise. The SDB coefficients are thus given by

$$\tilde{b}_{j_1j_2k_3} = (a_{n_{j_1},k_{j_1}} + N_{n_{j_1},k_{j_1}})(a_{n_{j_2},k_{j_2}} + N_{n_{j_2},k_{j_2}})(a^*_{n_{j_1}+n_{j_2},k_3} + N^*_{n_{j_1}+n_{j_2},k_3}). \tag{45}$$

Upon expansion, we see a polynomial of the form

$$\begin{aligned}
&\tilde{b}_{j_1j_2k_3} = a_{n_{j_1},k_{j_1}} a_{n_{j_2},k_{j_2}} a^*_{n_{j_1}+n_{j_2},k_3} && (46)\\
&+ \{2a \text{ term}\}\{1N \text{ term}\} + \{1a \text{ term}\}\{2N \text{ term}\} && (47)\\
&+ N_{n_{j_1},k_{j_1}} N_{n_{j_2},k_{j_2}} N^*_{n_{j_1}+n_{j_2},k_3}. && (48)
\end{aligned}$$

Taking the expectation

$$\begin{aligned}
&\mathbb{E}[\tilde{b}_{j_1j_2k_3}] && (49)\\
&= b_{j_1j_2k_3} + \mathbb{E}[N_{n_{j_1},k_{j_1}} N_{n_{j_2},k_{j_2}} N^*_{n_{j_1}+n_{j_2},k_3}] && (50)\\
&+ \{2a \text{ term}\}\mathbb{E}[\{1N \text{ term}\}] + \{1a \text{ term}\}\mathbb{E}[\{2N \text{ term}\}] && (51)\\
&= b_{j_1j_2k_3} + \{1a \text{ term}\}\mathbb{E}[\{2N \text{ term}\}] && (52)
\end{aligned}$$

where b_{123} denotes the bispectrum of f. Prop. 6 proves that Isserlis' Theorem 11 applies to $\mathbb{E}[N_1, ..., N_n]$. Therefore, the odd-order N moments disappear, leaving only even-order N moments.

Writing out such even moment terms, and using shorthand $\tilde{b}_{123} = \tilde{a}_{n_1,k_1}\tilde{a}_{n_2,k_2}\tilde{a}^*_{n_3,k_3}$ in place of $\tilde{b}_{j_1j_2k_3} = \tilde{a}_{n_{j_1},k_{j_1}}\tilde{a}_{n_{j_2},k_{j_2}}\tilde{a}^*_{n_{j_1}+n_{j_2},k_3}$, we find

$$\begin{aligned}
&\mathbb{E}[\tilde{b}_{123}] = b_{123} + a^*_{n_3,k_3}\mathbb{E}[N_{n_1,k_1}N_{n_2,k_2}] && (53)\\
&+ a_{n_2,k_2}\mathbb{E}[N_{n_1,k_1}N^*_{n_3,k_3}] + a_{n_1,k_1}\mathbb{E}[N_{n_2,k_2}N^*_{n_3,k_3}]. && (54)
\end{aligned}$$

Next, we substitute the second-order moment results of Th. 12

$$\begin{aligned}
\mathbb{E}[N_{nk}N_{n'k'}] &= \sigma^2\delta_{n,-n'}\delta_{k',k}\gamma_{n,k'}\\
\mathbb{E}[N_{nk}N^*_{n'k'}] &= \sigma^2\delta_{k,k'}\delta_{n,n'},
\end{aligned}$$

where $\gamma_{nk'} \equiv (-1)^n \frac{|J_{1+n}(\lambda_{nk'})|}{|J_{1-n}(\lambda_{nk'})|}$, to find

$$\begin{aligned}
&\mathbb{E}[\tilde{b}_{123}] = b_{123} && (55)\\
&+ \sigma^2\{a^*_{n_3,k_3}\delta_{n_1,-n_2}\delta_{k_2,k_1}\gamma_{n_1,k_2} && (56)\\
&+ a_{n_2,k_2}\delta_{k_1,k_3}\delta_{n_1,n_3} + a_{n_1,k_1}\delta_{k_2,k_3}\delta_{n_2,n_3}\}. && (57)
\end{aligned}$$

Removing the shorthand,

$$\begin{aligned}
&\mathbb{E}[\tilde{b}_{j_1j_2k_3}] && (58)\\
&= b_{j_1j_2k_3} + \sigma^2\{a^*_{n_{j_1}+n_{j_2},k_3}\delta_{n_{j_1},-n_{j_2}}\delta_{k_{j_2},k_{j_1}}\gamma_{n_{j_1},k_{j_2}} && (59)\\
&+ a_{n_{j_2},k_{j_2}}\delta_{k_{j_1},k_3}\delta_{n_{j_1},n_{j_1}+n_{j_2}} + a_{n_{j_1},k_{j_1}}\delta_{k_{j_2},k_3}\delta_{n_{j_2},n_{j_1}+n_{j_2}}\} && (60)\\
&= b_{j_1j_2k_3} + \sigma^2\{a^*_{0,k_3}\delta_{n_{j_1},-n_{j_2}}\delta_{k_{j_2},k_{j_1}}\gamma_{n_{j_1},k_{j_2}} && (61)\\
&+ a_{0,k_{j_2}}\delta_{k_{j_1},k_3}\delta_{n_{j_2},0} + a_{0,k_{j_1}}\delta_{k_{j_2},k_3}\delta_{n_{j_1},0}\}. && (62)
\end{aligned}$$

This is the expectation of the full disk bispectrum of a signal corrupted by zero-mean gaussian noise.

For the selective disk bispectrum, we consider

$$\begin{cases}
b_{0,0,k} = a^2_{0,1}\cdot a^*_{0,k}, \\
\qquad \text{for } k \in \{1,\dots,K_0\}, \\
b_{2,n,k} = a_{1,1}\cdot a_{n,1}\cdot a^*_{n+1,k}, \\
\qquad \text{for } n \in \{0, N_m - 1\}, \text{ and } k \in \{1,\dots,K_{n+1}\},
\end{cases} \tag{63}$$

where the selective bispectrum coefficients are indexed by j_1, n_2, k_3 (limiting $j_1 \in \{0, 2\}$), as opposed to j_1, j_2, k_3. This means we only consider cases where $n_{j_1} = n_{j_2} = 0, k_{j_1} = k_{j_2} = 1$ (first set of terms) and $n_{j_1} = 1, k_{j_1} = k_{j_2} = 1$ (second set of terms).

With this, we can simplify $\mathbb{E}[\tilde{b}_{j_1 j_2 k_3}]$ in the SDB case as

$$\begin{aligned}
&\mathbb{E}[\tilde{b}_{j_1 j_2 k_3}] && (64)\\
&= b_{j_1 j_2 k_3} + \sigma^2\{a^*_{0,k_3}\delta_{n_{j_1},0}\delta_{n_{j_2},0}\gamma_{0,1} && (65)\\
&+ a_{0,1}\delta_{1,k_3}\delta_{n_{j_2},0} + a_{0,1}\delta_{1,k_3}\delta_{n_{j_1},0}\}, && (66)
\end{aligned}$$

noting $\gamma_{0,k} = 1$. Thus, bias terms only appear for the following terms

$$\begin{cases}
\mathbb{E}[\tilde{b}_{00(k=1)}] = b_{001} + \sigma^2 a^*_{0,1} + 2\sigma^2 a_{0,1}\\
\mathbb{E}[\tilde{b}_{00(k\neq 1)}] = b_{00k} + \sigma^2 a^*_{0,k}\\
\mathbb{E}[\tilde{b}_{201}] = b_{201} + \sigma^2 a_{0,1}\\
\text{Else, no correction: } \mathbb{E}[\tilde{b}_{2nk}] = b_{2nk}.
\end{cases} \tag{67}$$

□

Corollary 3. *The only SDB coefficients that are biased in estimation are the ones for which the bispectrum has zero-valued phase.*

Proof. According to Th. 10, $\mathbb{E}[\tilde{b}]$ is biased for the following terms

$$\begin{cases}
\mathbb{E}[\tilde{b}_{00(k=1)}] = b_{001} + \sigma^2 a^*_{0,1} + 2\sigma^2 a_{0,1}\\
\mathbb{E}[\tilde{b}_{00(k\neq 1)}] = b_{00k} + \sigma^2 a^*_{0,k}\\
\mathbb{E}[\tilde{b}_{201}] = b_{201} + \sigma^2 a_{0,1}\\
\text{Else, no correction: } \mathbb{E}[\tilde{b}_{2nk}] = b_{2nk}.
\end{cases} \tag{68}$$

Recall that similarly to Fourier coefficients, disk harmonic coefficients can be written as $a_{n,k} = |a_{n,k}|e^{-in\phi_{n,k}}$, so

$$\begin{aligned}
b_{0,0,k} &= a^2_{0,1}\cdot a^*_{0,k} && (69)\\
&= |a_{0,1}|^2 e^{-i2\cdot 0\phi_{0,1}}|a_{0,k}|e^{-i0\phi_{0,k}} && (70)\\
&= |a_{0,1}|^2|a_{0,k}| && (71)\\
&\text{for } k \in \{1,\dots,K_0\} && (72)\\
& && (73)
\end{aligned}$$

$$\begin{aligned}
b_{2,0,1} &= a_{1,1}\cdot a_{0,1}\cdot a^*_{1,1} && (74)\\
&= |a_{1,1}|^2 a_{0,1} && (75)\\
&= |a_{1,1}|^2|a_{0,1}|e^{-i0\phi_{0,1}} && (76)\\
&= |a_{1,1}|^2|a_{0,1}|. && (77)
\end{aligned}$$

□

Theorem 11 (Isserlis' Theorem / Wick's Theorem)**.** *If $(X_1, X_2, X_3, ...X_n)$ is a real or complex zero-mean multivariate normal random vector, even-order moments (even n) are given by*

$$\mathbb{E}[X_1, X_2, X_3, ...X_n] = \mathbb{E}[\prod_{i=1}^{n=2m} X_i] = \sum_{p\in P_n^2}\prod_{\{i,j\}\in p} E[X_i X_j] \tag{78}$$

where P_n^2 is permutation "n choose two", so the sum is over all the pairings of $\{1,\dots,n\}$, i.e. all distinct ways of partitioning $\{1,\dots,n\}$ into pairs $\{i,j\}$, and the product is over the pairs contained in p. Odd order moments (odd n) are equal to zero.

Proposition 6 (Isserlis Theorem Applies to zero-mean Gaussian noise in DH space)**.** *Isserlis' Theorem 11 applies to $\mathbb{E}[N_{n_1,k_1}, ..., N_{n_j,k_j}]$, where N_{n_j,k_j} is the DH transform of i.i.d. zero-mean Gaussian noise.*

Proof. Isserlis' Theorem applies to zero-mean multivariate normal vectors. A random vector is said to be *k*-variate normally distributed if every linear combination of its *k* components has a univariate normal distribution. Here, we will prove that the vector $(N_{n_1,k_1}, N_{n_2,k_2}, ..., N_{n_j,k_j})$ found in our second-, fourth- and sixth-order moment proofs is a zero-mean multivariate normal vector.

In line with all our proofs, consider a signal f corrupted with i.i.d. Gaussian noise such that $\tilde{f} = f+n$, and $n \sim \mathcal{N}(0, \sigma^2)$. The DHCs of the noiseless signal f are given by

$$a_{n_j k_j} = \int_0^{2\pi} \int_0^1 c_{nk} f(r,\theta) J_n(\lambda_{nk} r) e^{-in\theta} r dr d\theta. \tag{79}$$

The DHCs of the noisy signal $\tilde{f}$ are given by

$$\tilde{a}_{n_j k_j} = \int_0^{2\pi} \int_0^1 c_{nk}[f(r,\theta) + n(r,\theta)] J_n(\lambda_{nk} r) e^{-in\theta} r dr d\theta \tag{80}$$

$$= a_{n_j,k_j} + N_{n_j,k_j}. \tag{81}$$

Recall n has distribution $n \sim \mathcal{N}(0, \sigma^2)$, and the DH transform is a linear operation. Therefore, N_{n_j,k_j} will also be normally distributed with mean zero for all j. The linear combination of such random variables will be a univariate zero-mean normal distribution. Therefore, $(N_{n_1,k_1}, N_{n_2,k_2}, ..., N_{n_j,k_j})$, is a zero-mean multivariate normal vector for all j. □

Theorem 12 (Second Order Moments of Noise in DHC Space)**.** *The second order moments found in the bispectrum expectation and variance proofs are given by*

$$\mathbb{E}[N_{nk} N_{n'k'}] = \mathbb{E}[N^*_{nk} N^*_{n'k'}] = \sigma^2 \delta_{n,-n'} \delta_{k',k} \gamma_{n,k'}$$
$$\mathbb{E}[N_{nk} N^*_{n'k'}] = \mathbb{E}[N^*_{nk} N_{n'k'}] = \sigma^2 \delta_{k,k'} \delta_{n,n'}$$

where $\gamma_{nk'} \equiv (-1)^n \frac{|J_{1+n}(\lambda_{nk'})|}{|J_{1-n}(\lambda_{nk'})|}$.

Proof.

$$\begin{aligned}
&\mathbb{E}[N_{nk} N_{n'k'}] \\
&= \mathbb{E}[\int_0^{2\pi} \int_0^1 c_{nk} n(r_1,\theta_1) J_n(\lambda_{nk} r_1) e^{-in\theta_1} r_1 dr_1 d\theta_1 \\
&\times \int_0^{2\pi} \int_0^1 c_{n'k'} n(r_2,\theta_2) J_{n'}(\lambda_{n'k'} r_2) e^{-in'\theta_2} r_2 dr_2 d\theta_2] \\
&= \int_0^{2\pi} \int_0^1 \int_0^{2\pi} \int_0^1 c_{nk} c_{n'k'} \mathbb{E}[n(r_1,\theta_1) n(r_2,\theta_2)] \\
&\times J_n(\lambda_{nk} r_1) e^{-in\theta_1} J_{n'}(\lambda_{n'k'} r_2) e^{-in'\theta_2} r_1 r_2 dr_1 dr_2 d\theta_1 d\theta_2
\end{aligned}$$

Noise at one point is independent from noise at another point, so

$$\mathbb{E}[n(r_1,\theta_1) n(r_2,\theta_2)] = \begin{cases} \sigma^2 & \text{if } (r_1,\theta_1) = (r_2,\theta_2) \\ \mathbb{E}[n(r_1,\theta_1)]\mathbb{E}[n(r_2,\theta_2)] = 0 & \text{otherwise.} \end{cases}$$

Thus, $\mathbb{E}[n(r_1,\theta_1) n(r_2,\theta_2)] = \sigma^2 \frac{1}{r_1} \delta(r_1 - r_2) \delta(\theta_1 - \theta_2)$, yielding

$$\begin{aligned}
&\mathbb{E}[N_{nk} N_{n'k'}] \\
&= \int_0^{2\pi} \int_0^1 \int_0^{2\pi} \int_0^1 c_{nk} c_{n'k'} \sigma^2 \delta(r_1 - r_2) \delta(\theta_1 - \theta_2) \\
&\times e^{-i(n\theta_1 + n'\theta_2)} J_n(\lambda_{nk} r_1) J_{n'}(\lambda_{n'k'} r_2) \\
&\times r_1 r_2 dr_1 dr_2 d\theta_1 d\theta_2
\end{aligned}$$

where δ is the dirac delta function. Integrating over r_2, θ_2 and setting $r_1 \to r, \theta_1 \to \theta$,

$$\mathbb{E}[N_{nk}N_{n'k'}] = c_{nk}c_{n'k'}\sigma^2 \int_0^{2\pi} e^{-i(n+n')\theta} d\theta \times \int_0^1 J_n(\lambda_{nk}r)J_{n'}(\lambda_{n'k'}r)rdr$$

Note that

$$\int_0^{2\pi} e^{-i(n+n')\theta} d\theta = \begin{cases} 2\pi & \text{if } n+n'=0 \\ 0 & \text{otherwise} \end{cases}$$

since the Fourier transform of a scalar is only nonzero at the DC component. Thus, the "frequency" $n+n'$ must equal zero for this Fourier transform integral to be nonzero. Using the shorthand $\delta_{n,-n'}$ for $\delta(n+n')$,

$$\begin{aligned} &\mathbb{E}[N_{nk}N_{n'k'}] \\ &= c_{nk}c_{n'k'}\sigma^2\delta_{n,-n'}(2\pi)\int_0^1 J_n(\lambda_{nk}r)J_{n'}(\lambda_{n'k'}r)rdr \\ &= c_{nk}c_{-nk'}\sigma^2\delta_{n,-n'}(2\pi) \\ &\quad \times \int_0^1 J_n(\lambda_{n,k}r)J_{-n}(\lambda_{-n,k'}r)rdr \end{aligned}$$

Next, we can utilize a common orthogonality relation [4, 45]

$$(2\pi)\int_0^1 c_{nk'}J_n(\lambda_{n,k}r)J_n(\lambda_{n,k'}r)rdr = \frac{1}{c_{nk}}\delta_{k,k'}.$$

Specifically, by recalling that $J_{-n}(r) = (-1)^n J_n(r)$, and $\lambda_{nk} = \lambda_{-nk}$ since J_{-n} has the exact same roots as J_n, we can write

$$\begin{aligned} \mathbb{E}[N_{nk}N_{n'k'}] &= c_{-nk'}\sigma^2\delta_{n,-n'} \\ &\quad \times (-1)^n(2\pi)\int_0^1 c_{nk}J_n(\lambda_{n,k}r)J_n(\lambda_{n,k'}r)rdr \\ &= (-1)^n\sigma^2\frac{c_{-nk'}}{c_{nk'}}\delta_{n,-n'}\delta_{k',k} \\ &= (-1)^n\sigma^2\frac{|J_{1+n}(\lambda_{nk'})|}{|J_{1-n}(\lambda_{nk'})|}\delta_{n,-n'}\delta_{k',k} \end{aligned}$$

where the last equality comes from the fact that

$$c_{nk} = \frac{1}{\sqrt{\pi}|J_{n+1}(\lambda_{nk})|}$$

from Eq. 2.2 in [4], along with the fact that $\lambda_{n,k} = \lambda_{-n,k}$ as mentioned previously.

Finally, naming $\gamma_{nk'} \equiv (-1)^n\frac{|J_{1+n}(\lambda_{nk'})|}{|J_{1-n}(\lambda_{nk'})|}$, we have

$$\mathbb{E}[N_{nk}N_{n'k'}] = \sigma^2\delta_{n,-n'}\delta_{k',k}\gamma_{n,k'} \tag{82}$$

The proof for $\mathbb{E}[N_{nk}N^*_{n'k'}]$ is nearly identical, so we will outline it in less detail, only highlighting key points of difference.

$$\begin{aligned}
&\mathbb{E}[N_{nk}N^*_{n'k'}] \\
&= \mathbb{E}[\int_0^{2\pi}\int_0^1 c_{nk}n(r_1,\theta_1)J_n(\lambda_{nk}r_1)e^{-in\theta_1}r_1dr_1d\theta_1 \\
&\times \int_0^{2\pi}\int_0^1 c_{n'k'}n(r_2,\theta_2)J_{n'}(\lambda_{n'k'}r_2)e^{in'\theta_2}r_2dr_2d\theta_2] \\
&= c_{nk}c_{n'k'}\sigma^2\int_0^{2\pi}e^{-i(n-n')\theta)}d\theta \\
&\times \int_0^1 J_n(\lambda_{nk}r)J_{n'}(\lambda_{n'k'}r)rdr \\
&= c_{nk'}\sigma^2\delta_{n,n'}(2\pi)\int_0^1 c_{nk}J_n(\lambda_{n,k}r)J_n(\lambda_{n,k'}r)rdr \\
&= c_{nk'}\sigma^2\delta_{n,n'}\frac{1}{c_{nk}}\delta_{k,k'} \\
&= \sigma^2\delta_{n,n'}\delta_{k,k'}.
\end{aligned}$$

Finally, we see that $\mathbb{E}[N_{nk}N_{n'k'}] = \mathbb{E}[N^*_{nk}N^*_{n'k'}]$ and $\mathbb{E}[N_{nk}N^*_{n'k'}] = \mathbb{E}[N^*_{nk}N_{n'k'}]$

$$\begin{aligned}
\mathbb{E}[N^*_{nk}N^*_{n'k'}] &= \mathbb{E}[N_{nk}N_{n'k'}]^* && (83)\\
&= \{\sigma^2\delta_{n,-n'}\delta_{k',k}\gamma_{n,k'}\}^* && (84)\\
&= \sigma^2\delta_{n,-n'}\delta_{k',k}\gamma_{n,k'} && (85)\\
&= \mathbb{E}[N_{nk}N_{n'k'}] && (86)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[N^*_{nk}N_{n'k'}] &= \mathbb{E}[N_{nk}N^*_{n'k'}]^* && (87)\\
&= \{\sigma^2\delta_{k,k'}\delta_{n,n'}\}^* && (88)\\
&= \sigma^2\delta_{k,k'}\delta_{n,n'} && (89)\\
&= \mathbb{E}[N_{nk}N^*_{n'k'}]. && (90)
\end{aligned}$$

□

Proposition 7 (Fourth Order Moments in Bispectrum Expectation and Variance)**.** *Selected fourth order moments, seen in the variance proofs, are given by*

$$\begin{aligned}
&\mathbb{E}[N_1N_1^*N_2N_2^*] = \mathbb{E}[|N_1|^2|N_2|^2] = \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2] + \mathbb{E}[N_1^*N_2]^2 + \mathbb{E}[N_1N_2]^2 && (91)\\
&\mathbb{E}[N_1N_1^*N_2N_3] = \mathbb{E}[|N_1|^2N_2N_3] && (92)\\
&\quad = \mathbb{E}[|N_1|^2]\mathbb{E}[N_2N_3] + \mathbb{E}[N_1N_3]\mathbb{E}[N_1^*N_2] + \mathbb{E}[N_1N_2]\mathbb{E}[N_1^*N_3] && (93)\\
& && (94)\\
& && (95)
\end{aligned}$$

where in the second example, N_2 or N_3 can be conjugated, and the statement will still hold.

Proof. Prop. 6 proves that Isserlis' Theorem 11 applies to $\mathbb{E}[N_1, ..., N_n]$. Therefore,

$$\begin{aligned}
\mathbb{E}[N_1N_2N_3N_4] &= \mathbb{E}[N_1N_2]\mathbb{E}[N_3N_4] && (96)\\
&+ \mathbb{E}[N_1N_3]\mathbb{E}[N_2N_4] && (97)\\
&+ \mathbb{E}[N_1N_4]\mathbb{E}[N_2N_3]. && (98)
\end{aligned}$$

Via direct substitution,

$$\begin{aligned}
&\mathbb{E}[|N_1|^2N_2N_3] = \mathbb{E}[|N_1|^2]\mathbb{E}[N_2N_3] + \mathbb{E}[N_1N_3]\mathbb{E}[N_1^*N_2] + \mathbb{E}[N_1N_2]\mathbb{E}[N_1^*N_3] && (99)\\
& && (100)
\end{aligned}$$

and

$$\mathbb{E}[|N_1|^2|N_2|^2] = \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2] + \mathbb{E}[N_1N_2^*]\mathbb{E}[N_1^*N_2] + \mathbb{E}[N_1N_2]\mathbb{E}[N_1^*N_2^*] \tag{101}$$

$$= \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2] + \mathbb{E}[N_1^*N_2]^2 + \mathbb{E}[N_1N_2]^2, \tag{102}$$

where the last equivalence relies on the complex conjugate relations in Th. 12. □

Proposition 8 (Sixth Order Moments in Bispectrum expectation)**.** *The sixth order moments found in the variance proof is given by*

$$\mathbb{E}[|N_1|^2|N_2|^2|N_3|^2] = \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2]\mathbb{E}[|N_3|^2] \tag{103}$$

$$\tag{104}$$

$$+ \mathbb{E}[N_1N_2] \tag{105}$$

$$\times \{\mathbb{E}[N_1N_2]\mathbb{E}[|N_3|^2] + 2\mathbb{E}[N_1^*N_3]\mathbb{E}[N_2N_3] + 2\mathbb{E}[N_1N_3]\mathbb{E}[N_2^*N_3]\} \tag{106}$$

$$\tag{107}$$

$$+ \mathbb{E}[N_1N_3] \tag{108}$$

$$\times \{\mathbb{E}[N_1N_3]\mathbb{E}[|N_2|^2] + 2\mathbb{E}[N_1^*N_2]\mathbb{E}[N_2N_3]\} \tag{109}$$

$$\tag{110}$$

$$+ \mathbb{E}[N_2N_3] \tag{111}$$

$$\times \{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2N_3]\} \tag{112}$$

$$\tag{113}$$

$$+ \mathbb{E}[N_2^*N_3] \tag{114}$$

$$\times \{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^*N_3] + 2\mathbb{E}[N_1^*N_2]\mathbb{E}[N_1^*N_3]\} \tag{115}$$

$$\tag{116}$$

$$+ \mathbb{E}[N_1^*N_3] \tag{117}$$

$$\times \{\mathbb{E}[N_1^*N_3]\mathbb{E}[|N_2|^2]\} \tag{118}$$

$$\tag{119}$$

$$+ \mathbb{E}[N_1^*N_2] \tag{120}$$

$$\times \{\mathbb{E}[N_1^*N_2]\mathbb{E}[|N_3|^2]\} \tag{121}$$

$$\tag{122}$$

Proof. Prop. 6 proves that Isserlis' Theorem 11 applies to $\mathbb{E}[N_1, ..., N_n]$. Therefore,

$$\mathbb{E}[N_1N_2N_3N_4N_5N_6] \tag{123}$$

$$= \mathbb{E}[N_1N_2]\mathbb{E}[N_3N_4]\mathbb{E}[N_5N_6] + \mathbb{E}[N_1N_2]\mathbb{E}[N_3N_5]\mathbb{E}[N_4N_6] + \mathbb{E}[N_1N_2]\mathbb{E}[N_3N_6]\mathbb{E}[N_4N_5] \tag{124}$$

$$+ \mathbb{E}[N_1N_3]\mathbb{E}[N_2N_4]\mathbb{E}[N_5N_6] + \mathbb{E}[N_1N_3]\mathbb{E}[N_2N_5]\mathbb{E}[N_4N_6] + \mathbb{E}[N_1N_3]\mathbb{E}[N_2N_6]\mathbb{E}[N_4N_5] \tag{125}$$

$$+ \mathbb{E}[N_1N_4]\mathbb{E}[N_2N_3]\mathbb{E}[N_5N_6] + \mathbb{E}[N_1N_4]\mathbb{E}[N_2N_5]\mathbb{E}[N_3N_6] + \mathbb{E}[N_1N_4]\mathbb{E}[N_2N_6]\mathbb{E}[N_3N_5] \tag{126}$$

$$+ \mathbb{E}[N_1N_5]\mathbb{E}[N_2N_3]\mathbb{E}[N_4N_6] + \mathbb{E}[N_1N_5]\mathbb{E}[N_2N_4]\mathbb{E}[N_3N_6] + \mathbb{E}[N_1N_5]\mathbb{E}[N_2N_6]\mathbb{E}[N_3N_4] \tag{127}$$

$$+ \mathbb{E}[N_1N_6]\mathbb{E}[N_2N_3]\mathbb{E}[N_4N_5] + \mathbb{E}[N_1N_6]\mathbb{E}[N_2N_4]\mathbb{E}[N_3N_5] + \mathbb{E}[N_1N_6]\mathbb{E}[N_2N_5]\mathbb{E}[N_3N_4] \tag{128}$$

$$\tag{129}$$

In the variance proof we encounter a sixth order moment terms of the form $\mathbb{E}[N_1N_1^*N_2N_2^*N_3N_3^*] = \mathbb{E}[|N_1|^2|N_2|^2|N_3|^2]$. We see from substitution that

$$
\begin{aligned}
&\mathbb{E}[|N_1|^2|N_2|^2|N_3|^2] && (130)\\
&\quad = \mathbb{E}[N_1N_1^*]\mathbb{E}[N_2N_2^*]\mathbb{E}[N_3N_3^*] && (131)\\
&\quad + \mathbb{E}[N_1N_1^*]\mathbb{E}[N_2N_3]\mathbb{E}[N_2^*N_3^*] && (132)\\
&\quad + \mathbb{E}[N_1N_1^*]\mathbb{E}[N_2N_3^*]\mathbb{E}[N_2^*N_3] && (133)\\
& && (134)\\
&\quad + \mathbb{E}[N_1N_2]\mathbb{E}[N_1^*N_2^*]\mathbb{E}[N_3N_3^*] && (135)\\
&\quad + \mathbb{E}[N_1N_2]\mathbb{E}[N_1^*N_3]\mathbb{E}[N_2^*N_3^*] && (136)\\
&\quad + \mathbb{E}[N_1N_2]\mathbb{E}[N_1^*N_3^*]\mathbb{E}[N_2^*N_3] && (137)\\
& && (138)\\
&\quad + \mathbb{E}[N_1N_2^*]\mathbb{E}[N_1^*N_2]\mathbb{E}[N_3N_3^*] && (139)\\
&\quad + \mathbb{E}[N_1N_2^*]\mathbb{E}[N_1^*N_3]\mathbb{E}[N_2N_3^*] && (140)\\
&\quad + \mathbb{E}[N_1N_2^*]\mathbb{E}[N_1^*N_3^*]\mathbb{E}[N_2N_3] && (141)\\
& && (142)\\
&\quad + \mathbb{E}[N_1N_3]\mathbb{E}[N_1^*N_2]\mathbb{E}[N_2^*N_3^*] && (143)\\
&\quad + \mathbb{E}[N_1N_3]\mathbb{E}[N_1^*N_2^*]\mathbb{E}[N_2N_3^*] && (144)\\
&\quad + \mathbb{E}[N_1N_3]\mathbb{E}[N_1^*N_3^*]\mathbb{E}[N_2N_2^*] && (145)\\
& && (146)\\
&\quad + \mathbb{E}[N_1N_3^*]\mathbb{E}[N_1^*N_2]\mathbb{E}[N_2^*N_3] && (147)\\
&\quad + \mathbb{E}[N_1N_3^*]\mathbb{E}[N_1^*N_2^*]\mathbb{E}[N_2N_3] && (148)\\
&\quad + \mathbb{E}[N_1N_3^*]\mathbb{E}[N_1^*N_3]\mathbb{E}[N_2N_2^*]. && (149)\\
& && (150)
\end{aligned}
$$

First, we sort in an order that will be advantageous for future use, as

$$\mathbb{E}[|N_1|^2|N_2|^2|N_3|^2] = \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2]\mathbb{E}[|N_3|^2] \tag{151}$$
$$\tag{152}$$
$$+ \mathbb{E}[N_1 N_2] \tag{153}$$
$$\times \{\mathbb{E}[N_1^* N_2^*]\mathbb{E}[|N_3|^2] + \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2^* N_3^*] + \mathbb{E}[N_1^* N_3^*]\mathbb{E}[N_2^* N_3]\} \tag{154}$$
$$\tag{155}$$
$$+ \mathbb{E}[N_1^* N_2^*] \tag{156}$$
$$\times \{\mathbb{E}[N_1 N_3]\mathbb{E}[N_2 N_3^*] + \mathbb{E}[N_1 N_3^*]\mathbb{E}[N_2 N_3]\} \tag{157}$$
$$\tag{158}$$
$$+ \mathbb{E}[N_1 N_3] \tag{159}$$
$$\times \{\mathbb{E}[N_1^* N_2]\mathbb{E}[N_2^* N_3^*] + \mathbb{E}[N_1^* N_3^*]\mathbb{E}[|N_2|^2]\} \tag{160}$$
$$\tag{161}$$
$$+ \mathbb{E}[N_1^* N_3^*] \tag{162}$$
$$\times \{\mathbb{E}[N_1 N_2^*]\mathbb{E}[N_2 N_3]\} \tag{163}$$
$$\tag{164}$$
$$+ \mathbb{E}[N_2 N_3] \tag{165}$$
$$\times \{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^* N_3^*]\} \tag{166}$$
$$\tag{167}$$
$$+ \mathbb{E}[N_2^* N_3] \tag{168}$$
$$\times \{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2 N_3^*] + \mathbb{E}[N_1 N_3^*]\mathbb{E}[N_1^* N_2]\} \tag{169}$$
$$\tag{170}$$
$$+ \mathbb{E}[N_2 N_3^*] \tag{171}$$
$$\times \{\mathbb{E}[N_1 N_2^*]\mathbb{E}[N_1^* N_3]\} \tag{172}$$
$$\tag{173}$$
$$+ \mathbb{E}[N_1 N_2^*] \tag{174}$$
$$\times \{\mathbb{E}[N_1^* N_2]\mathbb{E}[|N_3|^2]\} \tag{175}$$
$$\tag{176}$$
$$+ \mathbb{E}[N_1 N_3^*] \tag{177}$$
$$\times \{\mathbb{E}[N_1^* N_3]\mathbb{E}[|N_2|^2]\} \tag{178}$$
$$\tag{179}$$

Next, recall the complex conjugate relations in Th. 12, which say that

$$\mathbb{E}[N_{nk} N_{n'k'}] = \mathbb{E}[N_{nk}^* N_{n'k'}^*] = \sigma^2 \delta_{n,-n'} \delta_{k',k} \gamma_{n,k'}$$
$$\mathbb{E}[N_{nk} N_{n'k'}^*] = \mathbb{E}[N_{nk}^* N_{n'k'}] = \sigma^2 \delta_{k,k'} \delta_{n,n'}.$$

Finally, we substitute these relations, standardizing to $\mathbb{E}[N_{nk}N_{n'k'}]$ and $\mathbb{E}[N^*_{nk}N_{n'k'}]$ format, and condensing as

$$\mathbb{E}[|N_1|^2|N_2|^2|N_3|^2] = \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2]\mathbb{E}[|N_3|^2] \tag{180}$$
$$\tag{181}$$
$$+ \mathbb{E}[N_1N_2] \tag{182}$$
$$\times \{\mathbb{E}[N_1N_2]\mathbb{E}[|N_3|^2] + 2\mathbb{E}[N_1^*N_3]\mathbb{E}[N_2N_3] + 2\mathbb{E}[N_1N_3]\mathbb{E}[N_2^*N_3]\} \tag{183}$$
$$\tag{184}$$
$$+ \mathbb{E}[N_1N_3] \tag{185}$$
$$\times \{\mathbb{E}[N_1N_3]\mathbb{E}[|N_2|^2] + 2\mathbb{E}[N_1^*N_2]\mathbb{E}[N_2N_3]\} \tag{186}$$
$$\tag{187}$$
$$+ \mathbb{E}[N_2N_3] \tag{188}$$
$$\times \{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2N_3]\} \tag{189}$$
$$\tag{190}$$
$$+ \mathbb{E}[N_2^*N_3] \tag{191}$$
$$\times \{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^*N_3] + 2\mathbb{E}[N_1^*N_2]\mathbb{E}[N_1^*N_3]\} \tag{192}$$
$$\tag{193}$$
$$+ \mathbb{E}[N_1^*N_3] \tag{194}$$
$$\times \{\mathbb{E}[N_1^*N_3]\mathbb{E}[|N_2|^2]\} \tag{195}$$
$$\tag{196}$$
$$+ \mathbb{E}[N_1^*N_2] \tag{197}$$
$$\times \{\mathbb{E}[N_1^*N_2]\mathbb{E}[|N_3|^2]\} \tag{198}$$
$$\tag{199}$$

□

Theorem 13 (Variance)**.** *Let $a_{n,k}$ be the DH coefficients of f and $\tilde{b}$ the SDB of $\tilde{f} = f + \eta$, $\eta \sim \mathcal{N}(0, \sigma^2)$. The coefficients $a_{0,k}$ are real. For a generic coefficient $b_{2,n,k} = a_{1,1}\, a_{n,1}\, a^*_{n+1,k}$, meaning $n \geq 2$, or $n = 0$ and $k \geq 2$,*

$$\begin{aligned}\mathrm{Var}(\tilde{b}_{2,n,k}) = \sigma^2\big(|a_{1,1}|^2|a_{n,1}|^2 + |a_{1,1}|^2|a_{n+1,k}|^2 + |a_{n,1}|^2|a_{n+1,k}|^2\big) \\ + \sigma^4\big(|a_{1,1}|^2 + |a_{n,1}|^2 + |a_{n+1,k}|^2\big) + \sigma^6.\end{aligned} \tag{200}$$

The remaining coefficients, in which a DH coefficient appears twice, have

$$\mathrm{Var}(\tilde{b}_{001}) = 9\sigma^2 a_{0,1}^4 + 36\sigma^4 a_{0,1}^2 + 15\sigma^6, \tag{201}$$
$$\mathrm{Var}(\tilde{b}_{00k}) = \sigma^2\big(a_{0,1}^4 + 4a_{0,1}^2 a_{0,k}^2\big) + \sigma^4\big(6a_{0,1}^2 + 2a_{0,k}^2\big) + 3\sigma^6, \qquad k \geq 2, \tag{202}$$
$$\mathrm{Var}(\tilde{b}_{201}) = \sigma^2\big(|a_{1,1}|^4 + 2a_{0,1}^2|a_{1,1}|^2\big) + \sigma^4\big(4|a_{1,1}|^2 + a_{0,1}^2\big) + 2\sigma^6, \tag{203}$$
$$\mathrm{Var}(\tilde{b}_{21k}) = \sigma^2\big(|a_{1,1}|^4 + 4|a_{1,1}|^2|a_{2,k}|^2\big) + \sigma^4\big(4|a_{1,1}|^2 + 2|a_{2,k}|^2\big) + 2\sigma^6. \tag{204}$$

Proof. As in Th. 10, we use $\tilde{b}_{123} = \tilde{a}_1\tilde{a}_2\tilde{a}_3^*$ to denote $\tilde{b}_{j_1,j_2,k_3} = \tilde{a}_{n_{j_1}k_{j_1}}\tilde{a}_{n_{j_2}k_{j_2}}\tilde{a}^*_{n_{j_1}+n_{j_2}k_3}$. Here, we compute the variance of the bispectrum of a signal f, corrupted with Gaussian noise $\eta \sim \mathcal{N}(0, \sigma^2)$. We denote the bispectrum of the true signal as $b_{123} = a_1a_2a_3^*$.

The variance of the complex random variable $\tilde{b}_{123}$ is given by

$$\mathrm{Var}(\tilde{b}_{123}) = \mathbb{E}[\tilde{b}_{123}\tilde{b}^*_{123}] - \mathbb{E}[\tilde{b}_{123}]\mathbb{E}[\tilde{b}^*_{123}]. \tag{205}$$

We start with the second term. As shown in Th. 10,

$$\mathbb{E}[\tilde{b}_{123}] = a_1a_2a_3^* + a_3^*\mathbb{E}[N_1N_2] \tag{206}$$
$$+ a_2\mathbb{E}[N_1N_3^*] + a_1\mathbb{E}[N_2N_3^*]. \tag{207}$$

Similarly,

$$\mathbb{E}[\tilde{b}_{123}^*] = a_1^* a_2^* a_3 + a_3 \mathbb{E}[N_1^* N_2^*] \tag{208}$$
$$+ a_2^* \mathbb{E}[N_1^* N_3] + a_1^* \mathbb{E}[N_2^* N_3]. \tag{209}$$

Thus,

$$\mathbb{E}[\tilde{b}_{123}]\mathbb{E}[\tilde{b}_{123}^*] \tag{210}$$
$$= |a_1|^2|a_2|^2|a_3|^2 + a_1 a_2 |a_3|^2 \mathbb{E}[N_1^* N_2^*] \tag{211}$$
$$+ a_1 |a_2|^2 a_3^* \mathbb{E}[N_1^* N_3] + |a_1|^2 a_2 a_3^* \mathbb{E}[N_2^* N_3] \tag{212}$$
$$+ a_1^* a_2^* |a_3|^2 \mathbb{E}[N_1 N_2] + |a_3|^2 \mathbb{E}[N_1^* N_2^*]\mathbb{E}[N_1 N_2] \tag{213}$$
$$+ a_3^* a_2^* \mathbb{E}[N_1 N_2]\mathbb{E}[N_1^* N_3] + a_3^* a_1^* \mathbb{E}[N_1 N_2]\mathbb{E}[N_2^* N_3] \tag{214}$$
$$+ a_1^* |a_2|^2 a_3 \mathbb{E}[N_1 N_3^*] + a_2 a_3 \mathbb{E}[N_1 N_3^*]\mathbb{E}[N_1^* N_2^*] \tag{215}$$
$$+ |a_2|^2 \mathbb{E}[N_1 N_3^*]\mathbb{E}[N_1^* N_3] + a_2 a_1^* \mathbb{E}[N_1 N_3^*]\mathbb{E}[N_2^* N_3] \tag{216}$$
$$+ |a_1|^2 a_2^* a_3 \mathbb{E}[N_2 N_3^*] + a_1 a_3 \mathbb{E}[N_2 N_3^*]\mathbb{E}[N_1^* N_2^*] \tag{217}$$
$$+ a_1 a_2^* \mathbb{E}[N_2 N_3^*]\mathbb{E}[N_1^* N_3] + |a_1|^2 \mathbb{E}[N_2 N_3^*]\mathbb{E}[N_2^* N_3]. \tag{218}$$

Recall the complex conjugate relations in Th. 12, which say that

$$\mathbb{E}[N_{nk} N_{n'k'}] = \mathbb{E}[N_{nk}^* N_{n'k'}^*] = \sigma^2 \delta_{n,-n'} \delta_{k',k} \gamma_{n,k'}$$
$$\mathbb{E}[N_{nk} N_{n'k'}^*] = \mathbb{E}[N_{nk}^* N_{n'k'}] = \sigma^2 \delta_{k,k'} \delta_{n,n'}.$$

We substitute these and standardize to $\mathbb{E}[N_{nk} N_{n'k'}]$ and $\mathbb{E}[N_{nk}^* N_{n'k'}]$, resulting in

$$\mathbb{E}[\tilde{b}_{123}]\mathbb{E}[\tilde{b}_{123}^*] \tag{219}$$
$$= |a_1|^2|a_2|^2|a_3|^2 + a_1 a_2 |a_3|^2 \mathbb{E}[N_1 N_2] \tag{220}$$
$$+ a_1 |a_2|^2 a_3^* \mathbb{E}[N_1^* N_3] + |a_1|^2 a_2 a_3^* \mathbb{E}[N_2^* N_3] \tag{221}$$
$$+ a_1^* a_2^* |a_3|^2 \mathbb{E}[N_1 N_2] + |a_3|^2 \mathbb{E}[N_1 N_2]^2 \tag{222}$$
$$+ a_3^* a_2^* \mathbb{E}[N_1 N_2]\mathbb{E}[N_1^* N_3] + a_3^* a_1^* \mathbb{E}[N_1 N_2]\mathbb{E}[N_2^* N_3] \tag{223}$$
$$+ a_1^* |a_2|^2 a_3 \mathbb{E}[N_1^* N_3] + a_2 a_3 \mathbb{E}[N_1^* N_3]\mathbb{E}[N_1 N_2] \tag{224}$$
$$+ |a_2|^2 \mathbb{E}[N_1^* N_3]^2 + a_2 a_1^* \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2^* N_3] \tag{225}$$
$$+ |a_1|^2 a_2^* a_3 \mathbb{E}[N_2^* N_3] + a_1 a_3 \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1 N_2] \tag{226}$$
$$+ a_1 a_2^* \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1^* N_3] + |a_1|^2 \mathbb{E}[N_2^* N_3]^2. \tag{227}$$

Next, we focus on the $\mathbb{E}[\tilde{b}_{123}\tilde{b}_{123}^*]$ term, starting with

$$\tilde{b}_{123}\tilde{b}_{123}^* = \tilde{a}_1 \tilde{a}_2 \tilde{a}_3^* \tilde{a}_1^* \tilde{a}_2^* \tilde{a}_3 \tag{228}$$
$$= (a_1 + N_1)(a_1^* + N_1^*)(a_2 + N_2)(a_2^* + N_2^*)(a_3 + N_3)(a_3^* + N_3^*) \tag{229}$$
$$= (|a_1|^2 + a_1 N_1^* + a_1^* N_1 + |N_1|^2) \tag{230}$$
$$\times (|a_2|^2 + a_2 N_2^* + a_2^* N_2 + |N_2|^2) \tag{231}$$
$$\times (|a_3|^2 + a_3 N_3^* + a_3^* N_3 + |N_3|^2). \tag{232}$$
$$\tag{233}$$

Defining $\xi_i = a_i N_i^* + a_i^* N_i$, we can rewrite as

$$\tilde{b}_{123}\tilde{b}_{123}^* = \tilde{a}_1 \tilde{a}_2 \tilde{a}_3^* \tilde{a}_1^* \tilde{a}_2^* \tilde{a}_3 \tag{234}$$
$$= (|a_1|^2 + \xi_1 + |N_1|^2) \tag{235}$$
$$\times (|a_2|^2 + \xi_2 + |N_2|^2) \tag{236}$$
$$\times (|a_3|^2 + \xi_3 + |N_3|^2). \tag{237}$$
$$\tag{238}$$

and expand as,

$$
\begin{aligned}
&\tilde{b}_{123}\tilde{b}^*_{123} && (239)\\
&\quad = |a_1|^2|a_2|^2|a_3|^2 + |a_1|^2|a_2|^2\xi_3 + |a_1|^2|a_2|^2|N_3|^2 && (240)\\
&\quad + |a_1|^2\xi_2|a_3|^2 + |a_1|^2\xi_2\xi_3 + |a_1|^2\xi_2|N_3|^2 && (241)\\
&\quad + |a_1|^2|N_2|^2|a_3|^2 + |a_1|^2|N_2|^2\xi_3 + |a_1|^2|N_2|^2|N_3|^2 && (242)\\
& && (243)\\
&\quad + \xi_1|a_2|^2|a_3|^2 + \xi_1|a_2|^2\xi_3 + \xi_1|a_2|^2|N_3|^2 && (244)\\
&\quad + \xi_1\xi_2|a_3|^2 + \xi_1\xi_2\xi_3 + \xi_1\xi_2|N_3|^2 && (245)\\
&\quad + \xi_1|N_2|^2|a_3|^2 + \xi_1|N_2|^2\xi_3 + \xi_1|N_2|^2|N_3|^2 && (246)\\
& && (247)\\
&\quad + |N_1|^2|a_2|^2|a_3|^2 + |N_1|^2|a_2|^2\xi_3 + |N_1|^2|a_2|^2|N_3|^2 && (248)\\
&\quad + |N_1|^2\xi_2|a_3|^2 + |N_1|^2\xi_2\xi_3 + |N_1|^2\xi_2|N_3|^2 && (249)\\
&\quad + |N_1|^2|N_2|^2|a_3|^2 + |N_1|^2|N_2|^2\xi_3 + |N_1|^2|N_2|^2|N_3|^2 && (250)\\
& && (251)
\end{aligned}
$$

There are sixty four terms above, recalling that ξ_i represents two terms. Thirty two of the terms will have odd-order moments in N, and will therefore equal zero according to Th. 6, which states we can use Isserlis' Theorem Th. 11 for DH coefficient noise moments. This leaves thirty two even order terms.

$$
\begin{aligned}
&\tilde{b}_{123}\tilde{b}^*_{123} && (252)\\
&\quad = |a_1|^2|a_2|^2|a_3|^2 + |a_1|^2|a_2|^2|N_3|^2 && (253)\\
&\quad + |a_1|^2\xi_2\xi_3 && (254)\\
&\quad + |a_1|^2|N_2|^2|a_3|^2 + |a_1|^2|N_2|^2|N_3|^2 && (255)\\
& && (256)\\
&\quad + \xi_1|a_2|^2\xi_3 && (257)\\
&\quad + \xi_1\xi_2|a_3|^2 + \xi_1\xi_2|N_3|^2 && (258)\\
&\quad + \xi_1|N_2|^2\xi_3 && (259)\\
& && (260)\\
&\quad + |N_1|^2|a_2|^2|a_3|^2 + |N_1|^2|a_2|^2|N_3|^2 && (261)\\
&\quad + |N_1|^2\xi_2\xi_3 && (262)\\
&\quad + |N_1|^2|N_2|^2|a_3|^2 + |N_1|^2|N_2|^2|N_3|^2 && (263)\\
& && (264)
\end{aligned}
$$

Replacing ξ_i and expanding, we find

$$\tilde{b}_{123}\tilde{b}^*_{123} \tag{265}$$
$$= |a_1|^2|a_2|^2|a_3|^2 + |a_1|^2|a_2|^2|N_3|^2 \tag{266}$$
$$+ |a_1|^2(a_2a_3N_2^*N_3^* + a_2a_3^*N_2^*N_3 + a_2^*a_3N_2N_3^* + a_2^*a_3^*N_2N_3) \tag{267}$$
$$+ |a_1|^2|N_2|^2|a_3|^2 + |a_1|^2|N_2|^2|N_3|^2 \tag{268}$$
$$\tag{269}$$
$$+ |a_2|^2(a_1a_3N_1^*N_3^* + a_1a_3^*N_1^*N_3 + a_1^*a_3N_1N_3^* + a_1^*a_3^*N_1N_3) \tag{270}$$
$$+ |a_3|^2(a_1a_2N_1^*N_2^* + a_1a_2^*N_1^*N_2 + a_1^*a_2N_1N_2^* + a_1^*a_2^*N_1N_2) \tag{271}$$
$$+ |N_3|^2(a_1a_2N_1^*N_2^* + a_1a_2^*N_1^*N_2 + a_1^*a_2N_1N_2^* + a_1^*a_2^*N_1N_2) \tag{272}$$
$$+ |N_2|^2(a_1a_3N_1^*N_3^* + a_1a_3^*N_1^*N_3 + a_1^*a_3N_1N_3^* + a_1^*a_3^*N_1N_3) \tag{273}$$
$$\tag{274}$$
$$+ |N_1|^2|a_2|^2|a_3|^2 + |N_1|^2|a_2|^2|N_3|^2 \tag{275}$$
$$+ |N_1|^2(a_2a_3N_2^*N_3^* + a_2a_3^*N_2^*N_3 + a_2^*a_3N_2N_3^* + a_2^*a_3^*N_2N_3) \tag{276}$$
$$+ |N_1|^2|N_2|^2|a_3|^2 + |N_1|^2|N_2|^2|N_3|^2 \tag{277}$$
$$\tag{278}$$
$$\tag{279}$$

Taking the expectation of $\tilde{b}_{123}\tilde{b}^*_{123}$,

$$\mathbb{E}[\tilde{b}_{123}\tilde{b}^*_{123}] \tag{280}$$
$$= |a_1|^2|a_2|^2|a_3|^2 + |a_1|^2|a_2|^2\mathbb{E}[|N_3|^2] \tag{281}$$
$$+ |a_1|^2a_2a_3\mathbb{E}[N_2^*N_3^*] + |a_1|^2a_2a_3^*\mathbb{E}[N_2^*N_3] \tag{282}$$
$$+ |a_1|^2a_2^*a_3\mathbb{E}[N_2N_3^*] + |a_1|^2a_2^*a_3^*\mathbb{E}[N_2N_3] \tag{283}$$
$$+ |a_1|^2|a_3|^2\mathbb{E}[|N_2|^2] + |a_1|^2\mathbb{E}[|N_2|^2|N_3|^2] \tag{284}$$
$$\tag{285}$$
$$+ |a_2|^2a_1a_3\mathbb{E}[N_1^*N_3^*] + |a_2|^2a_1a_3^*\mathbb{E}[N_1^*N_3] \tag{286}$$
$$+ |a_2|^2a_1^*a_3\mathbb{E}[N_1N_3^*] + |a_2|^2a_1^*a_3^*\mathbb{E}[N_1N_3] \tag{287}$$
$$+ |a_3|^2a_1a_2\mathbb{E}[N_1^*N_2^*] + |a_3|^2a_1a_2^*\mathbb{E}[N_1^*N_2] \tag{288}$$
$$+ |a_3|^2a_1^*a_2\mathbb{E}[N_1N_2^*] + |a_3|^2a_1^*a_2^*\mathbb{E}[N_1N_2] \tag{289}$$
$$+ a_1a_2\mathbb{E}[|N_3|^2N_1^*N_2^*] + a_1a_2^*\mathbb{E}[|N_3|^2N_1^*N_2] \tag{290}$$
$$+ a_1^*a_2\mathbb{E}[|N_3|^2N_1N_2^*] + a_1^*a_2^*\mathbb{E}[|N_3|^2N_1N_2] \tag{291}$$
$$+ a_1a_3\mathbb{E}[|N_2|^2N_1^*N_3^*] + a_1a_3^*\mathbb{E}[|N_2|^2N_1^*N_3] \tag{292}$$
$$+ a_1^*a_3\mathbb{E}[|N_2|^2N_1N_3^*] + a_1^*a_3^*\mathbb{E}[|N_2|^2N_1N_3] \tag{293}$$
$$\tag{294}$$
$$+ |a_2|^2|a_3|^2\mathbb{E}[|N_1|^2] + |a_2|^2\mathbb{E}[|N_1|^2|N_3|^2] \tag{295}$$
$$+ a_2a_3\mathbb{E}[|N_1|^2N_2^*N_3^*] + a_2a_3^*\mathbb{E}[|N_1|^2N_2^*N_3] \tag{296}$$
$$+ a_2^*a_3\mathbb{E}[|N_1|^2N_2N_3^*] + a_2^*a_3^*\mathbb{E}[|N_1|^2N_2N_3] \tag{297}$$
$$+ |a_3|^2\mathbb{E}[|N_1|^2|N_2|^2] + \mathbb{E}[|N_1|^2|N_2|^2|N_3|^2]. \tag{298}$$
$$\tag{299}$$

Next, we recall the fourth order moment expansion in Th. 7

$$\mathbb{E}[|N_1|^2|N_2|^2] = \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2] + \mathbb{E}[N_1^*N_2]^2 + \mathbb{E}[N_1N_2]^2 \tag{300}$$
$$\mathbb{E}[|N_1|^2N_2N_3] = \mathbb{E}[|N_1|^2]\mathbb{E}[N_2N_3] + \mathbb{E}[N_1N_3]\mathbb{E}[N_1^*N_2] + \mathbb{E}[N_1N_2]\mathbb{E}[N_1^*N_3]. \tag{301}$$
$$\tag{302}$$

$$\mathbb{E}[\tilde{b}_{123}\tilde{b}^*_{123}] \tag{303}$$
$$= |a_1|^2|a_2|^2|a_3|^2 + |a_1|^2|a_2|^2\mathbb{E}[|N_3|^2] \tag{304}$$
$$+ |a_1|^2 a_2 a_3 \mathbb{E}[N_2^* N_3^*] + |a_1|^2 a_2 a_3^* \mathbb{E}[N_2^* N_3] \tag{305}$$
$$+ |a_1|^2 a_2^* a_3 \mathbb{E}[N_2 N_3^*] + |a_1|^2 a_2^* a_3^* \mathbb{E}[N_2 N_3] \tag{306}$$
$$+ |a_1|^2|a_3|^2\mathbb{E}[|N_2|^2] + |a_1|^2(\mathbb{E}[|N_2|^2]\mathbb{E}[|N_3|^2] + \mathbb{E}[N_2^* N_3]^2 + \mathbb{E}[N_2 N_3]^2) \tag{307}$$
$$\tag{308}$$
$$+ |a_2|^2 a_1 a_3 \mathbb{E}[N_1^* N_3^*] + |a_2|^2 a_1 a_3^* \mathbb{E}[N_1^* N_3] \tag{309}$$
$$+ |a_2|^2 a_1^* a_3 \mathbb{E}[N_1 N_3^*] + |a_2|^2 a_1^* a_3^* \mathbb{E}[N_1 N_3] \tag{310}$$
$$+ |a_3|^2 a_1 a_2 \mathbb{E}[N_1^* N_2^*] + |a_3|^2 a_1 a_2^* \mathbb{E}[N_1^* N_2] \tag{311}$$
$$+ |a_3|^2 a_1^* a_2 \mathbb{E}[N_1 N_2^*] + |a_3|^2 a_1^* a_2^* \mathbb{E}[N_1 N_2] \tag{312}$$
$$+ a_1 a_2(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1^* N_2^*] + \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1^* N_3^*] + \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2^* N_3^*]) \tag{313}$$
$$+ a_1 a_2^*(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1^* N_2] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_1^* N_3^*] + \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2 N_3^*]) \tag{314}$$
$$+ a_1^* a_2(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1 N_2^*] + \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1 N_3^*] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_2^* N_3^*]) \tag{315}$$
$$+ a_1^* a_2^*(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1 N_2] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_1 N_3^*] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_2 N_3^*]) \tag{316}$$
$$+ a_1 a_3(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1^* N_3^*] + \mathbb{E}[N_2 N_3^*]\mathbb{E}[N_2^* N_1^*] + \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2^* N_3^*]) \tag{317}$$
$$+ a_1 a_3^*(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1^* N_3] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_2^* N_1^*] + \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2^* N_3]) \tag{318}$$
$$+ a_1^* a_3(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1 N_3^*] + \mathbb{E}[N_2 N_3^*]\mathbb{E}[N_2^* N_1] + \mathbb{E}[N_2 N_1]\mathbb{E}[N_2^* N_3^*]) \tag{319}$$
$$+ a_1^* a_3^*(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1 N_3] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_2^* N_1] + \mathbb{E}[N_2 N_1]\mathbb{E}[N_2^* N_3]) \tag{320}$$
$$\tag{321}$$
$$+ |a_2|^2|a_3|^2\mathbb{E}[|N_1|^2] + |a_2|^2(\mathbb{E}[|N_1|^2]\mathbb{E}[|N_3|^2] + \mathbb{E}[N_1^* N_3]^2 + \mathbb{E}[N_1 N_3]^2) \tag{322}$$
$$+ a_2 a_3(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^* N_3^*] + \mathbb{E}[N_1 N_3^*]\mathbb{E}[N_1^* N_2^*] + \mathbb{E}[N_1 N_2^*]\mathbb{E}[N_1^* N_3^*]) \tag{323}$$
$$+ a_2 a_3^*(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^* N_3] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_1^* N_2^*] + \mathbb{E}[N_1 N_2^*]\mathbb{E}[N_1^* N_3]) \tag{324}$$
$$+ a_2^* a_3(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2 N_3^*] + \mathbb{E}[N_1 N_3^*]\mathbb{E}[N_1^* N_2] + \mathbb{E}[N_1 N_2]\mathbb{E}[N_1^* N_3^*]) \tag{325}$$
$$+ a_2^* a_3^*(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2 N_3] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_1^* N_2] + \mathbb{E}[N_1 N_2]\mathbb{E}[N_1^* N_3]) \tag{326}$$
$$+ |a_3|^2(\mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2] + \mathbb{E}[N_1^* N_2]^2 + \mathbb{E}[N_1 N_2]^2) + \mathbb{E}[|N_1|^2|N_2|^2|N_3|^2]. \tag{327}$$
$$\tag{328}$$

Plugging complex conjugate relations above,

$$
\begin{aligned}
&\mathbb{E}[\tilde{b}_{123}\tilde{b}^*_{123}] && (329)\\
&= |a_1|^2|a_2|^2|a_3|^2 + |a_1|^2|a_2|^2\mathbb{E}[|N_3|^2] && (330)\\
&+ |a_1|^2 a_2 a_3 \mathbb{E}[N_2 N_3] + |a_1|^2 a_2 a_3^* \mathbb{E}[N_2^* N_3] && (331)\\
&+ |a_1|^2 a_2^* a_3 \mathbb{E}[N_2^* N_3] + |a_1|^2 a_2^* a_3^* \mathbb{E}[N_2 N_3] && (332)\\
&+ |a_1|^2|a_3|^2\mathbb{E}[|N_2|^2] + |a_1|^2(\mathbb{E}[|N_2|^2]\mathbb{E}[|N_3|^2] + \mathbb{E}[N_2^* N_3]^2 + \mathbb{E}[N_2 N_3]^2) && (333)\\
& && (334)\\
&+ |a_2|^2 a_1 a_3 \mathbb{E}[N_1 N_3] + |a_2|^2 a_1 a_3^* \mathbb{E}[N_1^* N_3] && (335)\\
&+ |a_2|^2 a_1^* a_3 \mathbb{E}[N_1^* N_3] + |a_2|^2 a_1^* a_3^* \mathbb{E}[N_1 N_3] && (336)\\
&+ |a_3|^2 a_1 a_2 \mathbb{E}[N_1 N_2] + |a_3|^2 a_1 a_2^* \mathbb{E}[N_1^* N_2] && (337)\\
&+ |a_3|^2 a_1^* a_2 \mathbb{E}[N_1^* N_2] + |a_3|^2 a_1^* a_2^* \mathbb{E}[N_1 N_2] && (338)\\
&+ a_1 a_2(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1 N_2] + \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1 N_3] + \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2 N_3]) && (339)\\
&+ a_1 a_2^*(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1^* N_2] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_1 N_3] + \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2^* N_3]) && (340)\\
&+ a_1^* a_2(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1^* N_2] + \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1^* N_3] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_2 N_3]) && (341)\\
&+ a_1^* a_2^*(\mathbb{E}[|N_3|^2]\mathbb{E}[N_1 N_2] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_1^* N_3] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_2^* N_3]) && (342)\\
&+ a_1 a_3(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1 N_3] + \mathbb{E}[N_2^* N_3]\mathbb{E}[N_2 N_1] + \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2 N_3]) && (343)\\
&+ a_1 a_3^*(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1^* N_3] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_2 N_1] + \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2^* N_3]) && (344)\\
&+ a_1^* a_3(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1^* N_3] + \mathbb{E}[N_2^* N_3]\mathbb{E}[N_2^* N_1] + \mathbb{E}[N_2 N_1]\mathbb{E}[N_2 N_3]) && (345)\\
&+ a_1^* a_3^*(\mathbb{E}[|N_2|^2]\mathbb{E}[N_1 N_3] + \mathbb{E}[N_2 N_3]\mathbb{E}[N_2^* N_1] + \mathbb{E}[N_2 N_1]\mathbb{E}[N_2^* N_3]) && (346)\\
& && (347)\\
&+ |a_2|^2|a_3|^2\mathbb{E}[|N_1|^2] + |a_2|^2(\mathbb{E}[|N_1|^2]\mathbb{E}[|N_3|^2] + \mathbb{E}[N_1^* N_3]^2 + \mathbb{E}[N_1 N_3]^2) && (348)\\
&+ a_2 a_3(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2 N_3] + \mathbb{E}[N_1^* N_3]\mathbb{E}[N_1 N_2] + \mathbb{E}[N_1^* N_2]\mathbb{E}[N_1 N_3]) && (349)\\
&+ a_2 a_3^*(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^* N_3] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_1 N_2] + \mathbb{E}[N_1^* N_2]\mathbb{E}[N_1^* N_3]) && (350)\\
&+ a_2^* a_3(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^* N_3] + \mathbb{E}[N_1^* N_3]\mathbb{E}[N_1^* N_2] + \mathbb{E}[N_1 N_2]\mathbb{E}[N_1 N_3]) && (351)\\
&+ a_2^* a_3^*(\mathbb{E}[|N_1|^2]\mathbb{E}[N_2 N_3] + \mathbb{E}[N_1 N_3]\mathbb{E}[N_1^* N_2] + \mathbb{E}[N_1 N_2]\mathbb{E}[N_1^* N_3]) && (352)\\
&+ |a_3|^2(\mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2] + \mathbb{E}[N_1^* N_2]^2 + \mathbb{E}[N_1 N_2]^2) + \mathbb{E}[|N_1|^2|N_2|^2|N_3|^2]. && (353)\\
& && (354)
\end{aligned}
$$

Recalling $\mathbb{E}[\tilde{b}_{123}]\mathbb{E}[\tilde{b}^*_{123}]$, we compute $\mathrm{Var}(\tilde{b}_{123}) = \mathbb{E}[\tilde{b}_{123}\tilde{b}^*_{123}] - \mathbb{E}[\tilde{b}_{123}]\mathbb{E}[\tilde{b}^*_{123}]$ as

$$\mathrm{Var}(\tilde{b}_{123}) = \text{——} + |a_1|^2|a_2|^2\mathbb{E}[|N_3|^2] \tag{355}$$
$$+ |a_1|^2 a_2 a_3 \mathbb{E}[N_2 N_3] + \text{——} + \text{——} + |a_1|^2 a_2^* a_3^* \mathbb{E}[N_2 N_3] \tag{356}$$
$$+ |a_1|^2|a_3|^2\mathbb{E}[|N_2|^2] + |a_1|^2\mathbb{E}[|N_2|^2]\mathbb{E}[|N_3|^2] + \text{——} + |a_1|^2\mathbb{E}[N_2 N_3]^2 \tag{357}$$
$$\tag{358}$$
$$+ |a_2|^2 a_1 a_3 \mathbb{E}[N_1 N_3] + \text{——} + \text{——} + |a_2|^2 a_1^* a_3^* \mathbb{E}[N_1 N_3] \tag{359}$$
$$+ \text{——} + |a_3|^2 a_1 a_2^* \mathbb{E}[N_1^* N_2] + |a_3|^2 a_1^* a_2 \mathbb{E}[N_1^* N_2] + \text{——} \tag{360}$$
$$+ a_1 a_2 \mathbb{E}[|N_3|^2]\mathbb{E}[N_1 N_2] + a_1 a_2 \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1 N_3] + a_1 a_2 \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2 N_3] \tag{361}$$
$$+ a_1 a_2^* \mathbb{E}[|N_3|^2]\mathbb{E}[N_1^* N_2] + a_1 a_2^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_1 N_3] + \text{——} \tag{362}$$
$$+ a_1^* a_2 \mathbb{E}[|N_3|^2]\mathbb{E}[N_1^* N_2] + \text{——} + a_1^* a_2 \mathbb{E}[N_1 N_3]\mathbb{E}[N_2 N_3] \tag{363}$$
$$+ a_1^* a_2^* \mathbb{E}[|N_3|^2]\mathbb{E}[N_1 N_2] + a_1^* a_2^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_1^* N_3] + a_1^* a_2^* \mathbb{E}[N_1 N_3]\mathbb{E}[N_2^* N_3] \tag{364}$$
$$+ a_1 a_3 \mathbb{E}[|N_2|^2]\mathbb{E}[N_1 N_3] + \text{——} + a_1 a_3 \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2 N_3] \tag{365}$$
$$+ a_1 a_3^* \mathbb{E}[|N_2|^2]\mathbb{E}[N_1^* N_3] + a_1 a_3^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_2 N_1] + a_1 a_3^* \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2^* N_3] \tag{366}$$
$$+ a_1^* a_3 \mathbb{E}[|N_2|^2]\mathbb{E}[N_1^* N_3] + a_1^* a_3 \mathbb{E}[N_2^* N_3]\mathbb{E}[N_2^* N_1] + a_1^* a_3 \mathbb{E}[N_2 N_1]\mathbb{E}[N_2 N_3] \tag{367}$$
$$+ a_1^* a_3^* \mathbb{E}[|N_2|^2]\mathbb{E}[N_1 N_3] + a_1^* a_3^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_2^* N_1] + \text{——} \tag{368}$$
$$\tag{369}$$
$$+ |a_2|^2|a_3|^2\mathbb{E}[|N_1|^2] + |a_2|^2\mathbb{E}[|N_1|^2]\mathbb{E}[|N_3|^2] + \text{——} + |a_2|^2\mathbb{E}[N_1 N_3]^2) \tag{370}$$
$$+ a_2 a_3 \mathbb{E}[|N_1|^2]\mathbb{E}[N_2 N_3] + \text{——} + a_2 a_3 \mathbb{E}[N_1^* N_2]\mathbb{E}[N_1 N_3] \tag{371}$$
$$+ a_2 a_3^* \mathbb{E}[|N_1|^2]\mathbb{E}[N_2^* N_3] + a_2 a_3^* \mathbb{E}[N_1 N_3]\mathbb{E}[N_1 N_2] + a_2 a_3^* \mathbb{E}[N_1^* N_2]\mathbb{E}[N_1^* N_3] \tag{372}$$
$$+ a_2^* a_3 \mathbb{E}[|N_1|^2]\mathbb{E}[N_2^* N_3] + a_2^* a_3 \mathbb{E}[N_1^* N_3]\mathbb{E}[N_1^* N_2] + a_2^* a_3 \mathbb{E}[N_1 N_2]\mathbb{E}[N_1 N_3] \tag{373}$$
$$+ a_2^* a_3^* \mathbb{E}[|N_1|^2]\mathbb{E}[N_2 N_3] + a_2^* a_3^* \mathbb{E}[N_1 N_3]\mathbb{E}[N_1^* N_2] + \text{——} \tag{374}$$
$$+ |a_3|^2\mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2] + |a_3|^2\mathbb{E}[N_1^* N_2]^2 + \text{——} + \mathbb{E}[|N_1|^2|N_2|^2|N_3|^2], \tag{375}$$
$$\tag{376}$$

where —— indicates a term in $\mathbb{E}[\tilde{b}_{123}\tilde{b}^*_{123}]$ that cancels with $\mathbb{E}[\tilde{b}_{123}]\mathbb{E}[\tilde{b}^*_{123}]$. Note that after this cancellation, there are no surviving terms from $\mathbb{E}[\tilde{b}_{123}]\mathbb{E}[\tilde{b}^*_{123}]$. Next, we remove cancellation

placeholders and substitute $\mathbb{E}[N_i N_i^*] = \mathbb{E}[|N_i|^2] = \sigma^2$, finding

$$
\begin{aligned}
\mathrm{Var}(\tilde{b}_{123}) = &+|a_1|^2|a_2|^2\sigma^2 && (377)\\
&+|a_1|^2 a_2 a_3 \mathbb{E}[N_2 N_3] + |a_1|^2 a_2^* a_3^* \mathbb{E}[N_2 N_3] && (378)\\
&+|a_1|^2|a_3|^2\sigma^2 + |a_1|^2\sigma^2\sigma^2 + |a_1|^2\mathbb{E}[N_2 N_3]^2 && (379)\\
& && (380)\\
&+|a_2|^2 a_1 a_3 \mathbb{E}[N_1 N_3] + |a_2|^2 a_1^* a_3^* \mathbb{E}[N_1 N_3] && (381)\\
&+|a_3|^2 a_1 a_2^* \mathbb{E}[N_1^* N_2] + |a_3|^2 a_1^* a_2 \mathbb{E}[N_1^* N_2] && (382)\\
&+a_1 a_2 \sigma^2 \mathbb{E}[N_1 N_2] + a_1 a_2 \mathbb{E}[N_2^* N_3]\mathbb{E}[N_1 N_3] + a_1 a_2 \mathbb{E}[N_1^* N_3]\mathbb{E}[N_2 N_3] && (383)\\
&+a_1 a_2^* \sigma^2 \mathbb{E}[N_1^* N_2] + a_1 a_2^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_1 N_3] && (384)\\
&+a_1^* a_2 \sigma^2 \mathbb{E}[N_1^* N_2] + a_1^* a_2 \mathbb{E}[N_1 N_3]\mathbb{E}[N_2 N_3] && (385)\\
&+a_1^* a_2^* \sigma^2 \mathbb{E}[N_1 N_2] + a_1^* a_2^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_1^* N_3] + a_1^* a_2^* \mathbb{E}[N_1 N_3]\mathbb{E}[N_2^* N_3] && (386)\\
&+a_1 a_3 \sigma^2 \mathbb{E}[N_1 N_3] + a_1 a_3 \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2 N_3] && (387)\\
&+a_1 a_3^* \sigma^2 \mathbb{E}[N_1^* N_3] + a_1 a_3^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_2 N_1] + a_1 a_3^* \mathbb{E}[N_2 N_1^*]\mathbb{E}[N_2^* N_3] && (388)\\
&+a_1^* a_3 \sigma^2 \mathbb{E}[N_1^* N_3] + a_1^* a_3 \mathbb{E}[N_2^* N_3]\mathbb{E}[N_2^* N_1] + a_1^* a_3 \mathbb{E}[N_2 N_1]\mathbb{E}[N_2 N_3] && (389)\\
&+a_1^* a_3^* \sigma^2 \mathbb{E}[N_1 N_3] + a_1^* a_3^* \mathbb{E}[N_2 N_3]\mathbb{E}[N_2^* N_1] && (390)\\
& && (391)\\
&+|a_2|^2|a_3|^2\sigma^2 + |a_2|^2\sigma^2\sigma^2 + |a_2|^2\mathbb{E}[N_1 N_3]^2) && (392)\\
&+a_2 a_3 \sigma^2 \mathbb{E}[N_2 N_3] + a_2 a_3 \mathbb{E}[N_1^* N_2]\mathbb{E}[N_1 N_3] && (393)\\
&+a_2 a_3^* \sigma^2 \mathbb{E}[N_2^* N_3] + a_2 a_3^* \mathbb{E}[N_1 N_3]\mathbb{E}[N_1 N_2] + a_2 a_3^* \mathbb{E}[N_1^* N_2]\mathbb{E}[N_1^* N_3] && (394)\\
&+a_2^* a_3 \sigma^2 \mathbb{E}[N_2^* N_3] + a_2^* a_3 \mathbb{E}[N_1^* N_3]\mathbb{E}[N_1^* N_2] + a_2^* a_3 \mathbb{E}[N_1 N_2]\mathbb{E}[N_1 N_3] && (395)\\
&+a_2^* a_3^* \sigma^2 \mathbb{E}[N_2 N_3] + a_2^* a_3^* \mathbb{E}[N_1 N_3]\mathbb{E}[N_1^* N_2] && (396)\\
&+|a_3|^2\sigma^2\sigma^2 + |a_3|^2\mathbb{E}[N_1^* N_2]^2 + \mathbb{E}[|N_1|^2|N_2|^2|N_3|^2]. && (397)\\
& && (398)
\end{aligned}
$$

Reorganizing,

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{123}) = \sigma^2\{|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2\} && (399)\\
&+ \sigma^4\{|a_1|^2 + |a_2|^2 + |a_3|^2\} && (400)\\
& && (401)\\
&+ \mathbb{E}[N_1N_2]\{\sigma^2(a_1a_2 + a_1^*a_2^*) + \mathbb{E}[N_1N_3](a_2^*a_3 + a_2a_3^*) + \mathbb{E}[N_2N_3](a_1^*a_3 + a_1a_3^*)\} && (402)\\
& && (403)\\
&+ \mathbb{E}[N_1N_3]\{|a_2|^2(a_1a_3 + a_1^*a_3^* + \mathbb{E}[N_1N_3]) + \sigma^2(a_1a_3 + a_1^*a_3^*) && (404)\\
&\quad + \mathbb{E}[N_2N_3](a_1a_2^* + a_1^*a_2) + \mathbb{E}[N_2^*N_3](a_1^*a_2^* + a_1a_2) + \mathbb{E}[N_1^*N_2](a_2a_3 + a_2^*a_3^*)\} && (405)\\
& && (406)\\
&+ \mathbb{E}[N_2N_3]\{|a_1|^2(a_2a_3 + a_2^*a_3^* + \mathbb{E}[N_2N_3]) + \sigma^2(a_2a_3 + a_2^*a_3^*) && (407)\\
&\quad + \mathbb{E}[N_1^*N_3](a_1a_2 + a_1^*a_2^*) + \mathbb{E}[N_1^*N_2](a_1a_3 + a_1^*a_3^*)\} && (408)\\
& && (409)\\
&+ \mathbb{E}[N_2^*N_3]\{\sigma^2(a_2a_3^* + a_2^*a_3) + \mathbb{E}[N_1^*N_2](a_1a_3^* + a_1^*a_3)\} && (410)\\
& && (411)\\
&+ \mathbb{E}[N_1^*N_3]\{\sigma^2(a_1a_3^* + a_1^*a_3) + \mathbb{E}[N_1^*N_2](a_2a_3^* + a_2^*a_3)\} && (412)\\
& && (413)\\
&+ \mathbb{E}[N_1^*N_2]\{|a_3|^2(a_1a_2^* + a_1^*a_2 + \mathbb{E}[N_1^*N_2]) + \sigma^2(a_1a_2^* + a_1^*a_2)\} && (414)\\
& && (415)\\
&+ \mathbb{E}[|N_1|^2|N_2|^2|N_3|^2]. && (416)\\
& && (417)
\end{aligned}
$$

Recall the sixth order moment from Th. 8,

$$
\begin{aligned}
&\mathbb{E}[|N_1|^2|N_2|^2|N_3|^2] = \mathbb{E}[|N_1|^2]\mathbb{E}[|N_2|^2]\mathbb{E}[|N_3|^2] && (418)\\
&\quad + \mathbb{E}[N_1N_2]\{\mathbb{E}[N_1N_2]\mathbb{E}[|N_3|^2] + 2\mathbb{E}[N_1^*N_3]\mathbb{E}[N_2N_3] + 2\mathbb{E}[N_1N_3]\mathbb{E}[N_2^*N_3]\} && (419)\\
&\quad + \mathbb{E}[N_1N_3]\{\mathbb{E}[N_1N_3]\mathbb{E}[|N_2|^2] + 2\mathbb{E}[N_1^*N_2]\mathbb{E}[N_2N_3]\} && (420)\\
&\quad + \mathbb{E}[N_2N_3]\{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2N_3]\} && (421)\\
&\quad + \mathbb{E}[N_2^*N_3]\{\mathbb{E}[|N_1|^2]\mathbb{E}[N_2^*N_3] + 2\mathbb{E}[N_1^*N_2]\mathbb{E}[N_1^*N_3]\} && (422)\\
&\quad + \mathbb{E}[N_1^*N_3]\{\mathbb{E}[N_1^*N_3]\mathbb{E}[|N_2|^2]\} && (423)\\
&\quad + \mathbb{E}[N_1^*N_2]\{\mathbb{E}[N_1^*N_2]\mathbb{E}[|N_3|^2]\} && (424)\\
& && (425)
\end{aligned}
$$

and plug in to find

$$\mathrm{Var}(\tilde{b}_{123}) = \sigma^2\{|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2\} \tag{426}$$
$$+ \sigma^4\{|a_1|^2 + |a_2|^2 + |a_3|^2\} + \sigma^6 \tag{427}$$
$$\tag{428}$$
$$+ \mathbb{E}[N_1N_2]\{\sigma^2(a_1a_2 + a_1^*a_2^* + \mathbb{E}[N_1N_2]) + \mathbb{E}[N_1N_3](a_2^*a_3 + a_2a_3^* + 2\mathbb{E}[N_2^*N_3]) \tag{429}$$
$$+ \mathbb{E}[N_2N_3](a_1^*a_3 + a_1a_3^* + 2\mathbb{E}[N_1^*N_3])\} \tag{430}$$
$$\tag{431}$$
$$+ \mathbb{E}[N_1N_3]\{(|a_2|^2 + \sigma^2)(a_1a_3 + a_1^*a_3^* + \mathbb{E}[N_1N_3]) \tag{432}$$
$$+ \mathbb{E}[N_2N_3](a_1a_2^* + a_1^*a_2 + 2\mathbb{E}[N_1^*N_2]) \tag{433}$$
$$+ \mathbb{E}[N_2^*N_3](a_1^*a_2^* + a_1a_2) + \mathbb{E}[N_1^*N_2](a_2a_3 + a_2^*a_3^*)\} \tag{434}$$
$$\tag{435}$$
$$+ \mathbb{E}[N_2N_3]\{(|a_1|^2 + \sigma^2)(a_2a_3 + a_2^*a_3^* + \mathbb{E}[N_2N_3]) \tag{436}$$
$$+ \mathbb{E}[N_1^*N_3](a_1a_2 + a_1^*a_2^*) + \mathbb{E}[N_1^*N_2](a_1a_3 + a_1^*a_3^*)\} \tag{437}$$
$$\tag{438}$$
$$+ \mathbb{E}[N_2^*N_3]\{\sigma^2(a_2a_3^* + a_2^*a_3 + \mathbb{E}[N_2^*N_3]) + \mathbb{E}[N_1^*N_2](a_1a_3^* + a_1^*a_3 + 2\mathbb{E}[N_1^*N_3])\} \tag{439}$$
$$\tag{440}$$
$$+ \mathbb{E}[N_1^*N_3]\{\sigma^2(a_1a_3^* + a_1^*a_3 + \mathbb{E}[N_1^*N_3]) + \mathbb{E}[N_1^*N_2](a_2a_3^* + a_2^*a_3)\} \tag{441}$$
$$\tag{442}$$
$$+ \mathbb{E}[N_1^*N_2]\{(|a_3|^2 + \sigma^2)(a_1a_2^* + a_1^*a_2 + \mathbb{E}[N_1^*N_2])\}. \tag{443}$$
$$\tag{444}$$

We recall second moment identities from Th. 12 and remember that the SDB only considers terms for $n \geq 0$, so $\delta_{n,-n'} \to \delta_{n,0}\delta_{n',0}$ and $\gamma_{0k'} = 1$. With these conditions,

$$\mathbb{E}[N_{nk}N_{n'k'}] = \sigma^2\delta_{n,0}\delta_{n',0}\delta_{k',k}$$
$$\mathbb{E}[N_{nk}N_{n'k'}^*] = \sigma^2\delta_{n,n'}\delta_{k,k'}.$$

Plugging in second order moments for the SDB,

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{123}) && (445)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (446)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (447)\\
& && (448)\\
&+ \sigma^2\delta_{n_1,0}\delta_{n_2,0}\delta_{k_1,k_2} \times \{\sigma^2(\sigma^2\delta_{n_1,0}\delta_{n_2,0}\delta_{k_1,k_2} + a_1a_2 + a_1^*a_2^*) && (449)\\
&\quad + \sigma^2\delta_{n_1,0}\delta_{n_3,0}\delta_{k_1,k_3}(a_2a_3^* + a_2^*a_3 + 2\sigma^2\delta_{n_2,n_3}\delta_{k_2,k_3}) && (450)\\
&\quad + \sigma^2\delta_{n_2,0}\delta_{n_3,0}\delta_{k_2,k_3}(a_1a_3^* + a_1^*a_3 + 2\sigma^2\delta_{n_1,n_3}\delta_{k_1,k_3})\} && (451)\\
& && (452)\\
&+ \sigma^2\delta_{n_1,0}\delta_{n_3,0}\delta_{k_1,k_3} \times \{(|a_2|^2 + \sigma^2)(\sigma^2\delta_{n_1,0}\delta_{n_3,0}\delta_{k_1,k_3} + a_1a_3 + a_1^*a_3^*) && (453)\\
&\quad + \sigma^2\delta_{n_2,0}\delta_{n_3,0}\delta_{k_2,k_3}(a_1a_2^* + a_1^*a_2 + 2\sigma^2\delta_{n_1,n_2}\delta_{k_1,k_2}) && (454)\\
&\quad + \sigma^2\delta_{n_1,n_2}\delta_{k_1,k_2}(a_2a_3 + a_2^*a_3^*) + \sigma^2\delta_{n_2,n_3}\delta_{k_2,k_3}(a_1a_2 + a_1^*a_2^*)\} && (455)\\
& && (456)\\
&+ \sigma^2\delta_{n_2,0}\delta_{n_3,0}\delta_{k_2,k_3} \times \{(|a_1|^2 + \sigma^2)(\sigma^2\delta_{n_2,0}\delta_{n_3,0}\delta_{k_2,k_3} + a_2a_3 + a_2^*a_3^*) && (457)\\
&\quad + \sigma^2\delta_{n_1,n_2}\delta_{k_1,k_2}(a_1a_3 + a_1^*a_3^*) + \sigma^2\delta_{n_1,n_3}\delta_{k_1,k_3}(a_1a_2 + a_1^*a_2^*)\} && (458)\\
& && (459)\\
&+ \sigma^2\delta_{n_2,n_3}\delta_{k_2,k_3} \times \{\sigma^2(\sigma^2\delta_{n_2,n_3}\delta_{k_2,k_3} + a_2a_3^* + a_2^*a_3) && (460)\\
&\quad + \sigma^2\delta_{n_1,n_2}\delta_{k_1,k_2}(a_1a_3^* + a_1^*a_3 + 2\sigma^2\delta_{n_1,n_3}\delta_{k_1,k_3})\} && (461)\\
& && (462)\\
&+ \sigma^2\delta_{n_1,n_3}\delta_{k_1,k_3} \times \{\sigma^2(\sigma^2\delta_{n_1,n_3}\delta_{k_1,k_3} + a_1a_3^* + a_1^*a_3) && (463)\\
&\quad + \sigma^2\delta_{n_1,n_2}\delta_{k_1,k_2}(a_2a_3^* + a_2^*a_3)\} && (464)\\
& && (465)\\
&+ \sigma^2\delta_{n_1,n_2}\delta_{k_1,k_2} \times \{(|a_3|^2 + \sigma^2)(\sigma^2\delta_{n_1,n_2}\delta_{k_1,k_2} + a_1a_2^* + a_1^*a_2)\}. && (466)\\
& && (467)
\end{aligned}
$$

Next, we remember that our notation $\tilde{b}_{123} = \tilde{a}_1\tilde{a}_2\tilde{a}_3^*$ is shorthand for $\tilde{b}_{j_1,j_2,k_3} = \tilde{a}_{n_{j_1}k_{j_1}}\tilde{a}_{n_{j_2}k_{j_2}}\tilde{a}^*_{n_{j_1}+n_{j_2}k_3}$. Consequently, $\delta_{n_1,0}\delta_{n_3,0}$ and $\delta_{n_2,0}\delta_{n_3,0}$, are equivalent to $\delta_{n_1,0}\delta_{n_2,0}$. The SDB only considers coefficients where $k_{j_1} = k_{j_2} = 1$. Therefore, $\delta_{n_1,0}\delta_{n_2,0}\delta_{k_1,k_2} \to \delta_{n_1,0}\delta_{n_2,0}$, and $\delta_{n_2,0}\delta_{n_3,0}\delta_{k_2,k_3} \to \delta_{n_1,0}\delta_{n_2,0}\delta_{k_3,1}$ for example. Additionally, $\delta_{n_2,n_3} = \delta_{n_2,n_1+n_2} = \delta_{n_1,0}$ and $\delta_{n_1,n_3} = \delta_{n_1,n_1+n_2} = \delta_{n_2,0}$. With this, and simplifying redundant deltas

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{123}) && (468)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (469)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (470)\\
& && (471)\\
&+ \sigma^2\delta_{n_1,0}\delta_{n_2,0} \times \{\sigma^2(\sigma^2 + a_1a_2 + a_1^*a_2^*) && (472)\\
&\quad + \sigma^2\delta_{k_3,1}(a_2a_3^* + a_2^*a_3 + 2\sigma^2) && (473)\\
&\quad + \sigma^2\delta_{k_3,1}(a_1a_3^* + a_1^*a_3 + 2\sigma^2)\} && (474)\\
& && (475)\\
&+ \sigma^2\delta_{n_1,0}\delta_{n_2,0}\delta_{k_3,1} \times \{(|a_2|^2 + \sigma^2)(\sigma^2 + a_1a_3 + a_1^*a_3^*) && (476)\\
&\quad + \sigma^2(a_1a_2^* + a_1^*a_2 + 2\sigma^2) && (477)\\
&\quad + \sigma^2(a_2a_3 + a_2^*a_3^*) + \sigma^2(a_1a_2 + a_1^*a_2^*)\} && (478)\\
& && (479)\\
&+ \sigma^2\delta_{n_1,0}\delta_{n_2,0}\delta_{k_3,1} \times \{(|a_1|^2 + \sigma^2)(\sigma^2 + a_2a_3 + a_2^*a_3^*) && (480)\\
&\quad + \sigma^2(a_1a_3 + a_1^*a_3^*) + \sigma^2(a_1a_2 + a_1^*a_2^*)\} && (481)\\
& && (482)\\
&+ \sigma^2\delta_{n_1,0}\delta_{k_3,1} \times \{\sigma^2(\sigma^2 + a_2a_3^* + a_2^*a_3) && (483)\\
&\quad + \sigma^2\delta_{n_2,0}(a_1a_3^* + a_1^*a_3 + 2\sigma^2)\} && (484)\\
& && (485)\\
&+ \sigma^2\delta_{n_2,0}\delta_{k_3,1} \times \{\sigma^2(\sigma^2 + a_1a_3^* + a_1^*a_3) && (486)\\
&\quad + \sigma^2\delta_{n_1,0}(a_2a_3^* + a_2^*a_3)\} && (487)\\
& && (488)\\
&+ \sigma^2\delta_{n_1,n_2} \times \{(|a_3|^2 + \sigma^2)(\sigma^2 + a_1a_2^* + a_1^*a_2)\}. && (489)\\
& && (490)
\end{aligned}
$$

Next, we combine terms. For the SDB, $n_1 = 0$ only for the first set of terms, where $n_2 = 0$ also. Therefore, $\delta_{n_1,0}\delta_{k_3,1}$ in the third to last term is equivalent to $\delta_{n_1,0}\delta_{n_2,0}\delta_{k_3,1}$, and we will write it as the latter in order to combine it with the terms above. With that, only the last two terms have deltas not equivalent to $\delta_{n_1,0}\delta_{n_2,0}\delta_{k_3,1}$. In the second to last term, $\delta_{n_2,0}$ could happen when $n_1 = 0$ in the first set of SDB coefficients or $n_1 = 1$ in the second set of SDB coefficients. In the last term above, δ_{n_1,n_2} could happen in the first set of coefficients when $n_1 = n_2 = 0$ or in the second set of coefficients when $n_1 = n_2 = 1$. With this in mind, we split the variance cases, temporarily keeping a_1, a_2, a_3 notation for ease of readability,

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{j_1j_2k_3}) \text{ for } \tilde{b}_{j_1j_2k_3} \notin \{\tilde{b}_{001}, \tilde{b}_{00k}, \tilde{b}_{201}, \tilde{b}_{21k}\} && (491)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (492)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (493)\\
& && (494)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{001}) && (495)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (496)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (497)\\
&+ \sigma^2 \times \{\sigma^2(\sigma^2 + a_1a_2 + a_1^*a_2^*) && (498)\\
&\qquad + \sigma^2(a_2a_3^* + a_2^*a_3 + 2\sigma^2) && (499)\\
&\qquad + \sigma^2(a_1a_3^* + a_1^*a_3 + 2\sigma^2)\} && (500)\\
&+ \sigma^2 \times \{(|a_2|^2 + \sigma^2)(\sigma^2 + a_1a_3 + a_1^*a_3^*) && (501)\\
&\qquad + \sigma^2(a_1a_2^* + a_1^*a_2 + 2\sigma^2) && (502)\\
&\qquad + \sigma^2(a_2a_3 + a_2^*a_3^*) + \sigma^2(a_1a_2 + a_1^*a_2^*)\} && (503)\\
&+ \sigma^2 \times \{(|a_1|^2 + \sigma^2)(\sigma^2 + a_2a_3 + a_2^*a_3^*) && (504)\\
&\qquad + \sigma^2(a_1a_3 + a_1^*a_3^*) + \sigma^2(a_1a_2 + a_1^*a_2^*)\} && (505)\\
&+ \sigma^4(\sigma^2 + a_2a_3^* + a_2^*a_3) && (506)\\
&\qquad + \sigma^4(a_1a_3^* + a_1^*a_3 + 2\sigma^2) && (507)\\
&+ \sigma^4(\sigma^2 + a_1a_3^* + a_1^*a_3) && (508)\\
&\qquad + \sigma^4(a_2a_3^* + a_2^*a_3) && (509)\\
&+ \sigma^2(|a_3|^2 + \sigma^2)(\sigma^2 + a_1a_2^* + a_1^*a_2) && (510)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{00k}) && (511)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (512)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (513)\\
&+ \sigma^4(\sigma^2 + a_1a_2 + a_1^*a_2^*) && (514)\\
&+ \sigma^2(|a_3|^2 + \sigma^2)(\sigma^2 + a_1a_2^* + a_1^*a_2) && (515)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{201}) && (516)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (517)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (518)\\
&+ \sigma^4(\sigma^2 + a_1a_3^* + a_1^*a_3) && (519)\\
& && (520)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{21k}) && (521)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (522)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (523)\\
&+ \sigma^2(|a_3|^2 + \sigma^2)(\sigma^2 + a_1a_2^* + a_1^*a_2) && (524)
\end{aligned}
$$

Next, sorting each variance expression into second, fourth, and sixth order noise terms,

$$
\begin{aligned}
&\mathrm{Var}(\tilde{b}_{j_1j_2k_3}) \text{ for } \tilde{b}_{j_1j_2k_3} \notin \{\tilde{b}_{001}, \tilde{b}_{00k}, \tilde{b}_{201}, \tilde{b}_{21k}\} && (525)\\
&= \sigma^6 + \sigma^4(|a_1|^2 + |a_2|^2 + |a_3|^2) && (526)\\
&+ \sigma^2(|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2) && (527)\\
& && (528)
\end{aligned}
$$

$$\begin{aligned}
\mathrm{Var}(\tilde{b}_{001}) &= 15\sigma^6 && (529)\\
&+ \sigma^4\{2|a_1|^2 + 2|a_2|^2 + 2|a_3|^2 && (530)\\
&\quad + 3(a_1a_2 + a_1^*a_2^*) + 2(a_1a_2^* + a_1^*a_2) && (531)\\
&\quad + 3(a_2a_3^* + a_2^*a_3) + 2(a_2a_3 + a_2^*a_3^*) && (532)\\
&\quad + 3(a_1a_3^* + a_1^*a_3) + 2(a_1a_3 + a_1^*a_3^*)\} && (533)\\
&+ \sigma^2\{|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2 && (534)\\
&\quad + |a_2|^2(a_1a_3 + a_1^*a_3^*) && (535)\\
&\quad + |a_1|^2(a_2a_3 + a_2^*a_3^*) && (536)\\
&\quad + |a_3|^2(a_1a_2^* + a_1^*a_2)\} && (537)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\tilde{b}_{00k}) &= 3\sigma^6 && (538)\\
&+ \sigma^4\{|a_1|^2 + |a_2|^2 + 2|a_3|^2 && (539)\\
&\quad + (a_1a_2 + a_1^*a_2^*) + (a_1a_2^* + a_1^*a_2)\} && (540)\\
&+ \sigma^2\{|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2 && (541)\\
&\quad + |a_3|^2(a_1a_2^* + a_1^*a_2)\} && (542)
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Var}(\tilde{b}_{201}) && (543)\\
&= 2\sigma^6 && (544)\\
&+ \sigma^4\{|a_1|^2 + |a_2|^2 + |a_3|^2 + a_1a_3^* + a_1^*a_3\} && (545)\\
&+ \sigma^2\{|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2\} && (546)\\
& && (547)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\tilde{b}_{21k}) &= 2\sigma^6 && (548)\\
&+ \sigma^4\{|a_1|^2 + |a_2|^2 + 2|a_3|^2 + (a_1a_2^* + a_1^*a_2)\} && (549)\\
&+ \sigma^2\{|a_1|^2|a_2|^2 + |a_1|^2|a_3|^2 + |a_2|^2|a_3|^2 && (550)\\
&\quad + |a_3|^2(a_1a_2^* + a_1^*a_2)\} && (551)
\end{aligned}$$

Finally, we reintroduce full notation ($a_1 \to a_{n_{j_1},k_{j_1}}$, $a_2 \to a_{n_{j_2},k_{j_2}}$, and $a_3^* \to a^*_{n_{j_1}+n_{j_2},k_{j_3}}$). We recognize that $x + x^* = 2Re(x)$. For the SDB, we use,

$$\begin{aligned}
j = 0 \to n_0 = 0, k_0 = 1 && (552)\\
j = 2 \to n_2 = 1, k_2 = 1, && (553)
\end{aligned}$$

so that, for several of the terms, we recognize that $a_{n,k} = |a_{n,k}|e^{-in\phi}$, so $a_{01} = |a_{0,1}|e^{-i0\cdot\phi} = |a_{0,1}| = Re(a_{01})$, and $a_{01}^2 = |a_{0,1}|^2e^{-i2\cdot0\cdot\phi} = |a_{0,1}|^2 = Re(a_{01}^2) = Re(a_{0,1})^2$. Rewriting in true SDB notation,

$$\begin{aligned}
&\mathrm{Var}(\tilde{b}_{j_1n_2k_3}) \text{ for } \tilde{b}_{j_1n_2k_3} \notin \{\tilde{b}_{001}, \tilde{b}_{00k}, \tilde{b}_{201}, \tilde{b}_{21k}\} && (554)\\
&= \sigma^6 + \sigma^4(|a_{n_{j_1},k_{j_1}}|^2 + |a_{n_2,1}|^2 + |a^*_{n_{j_1}+n_2,k_{j_3}}|^2) && (555)\\
&+ \sigma^2(|a_{n_{j_1},k_{j_1}}|^2|a_{n_2,1}|^2 + |a_{n_{j_1},k_{j_1}}|^2|a^*_{n_{j_1}+n_2,k_{j_3}}|^2 + |a_{n_2,1}|^2|a^*_{n_{j_1}+n_2,k_{j_3}}|^2) && (556)\\
&\mathrm{Var}(\tilde{b}_{001}) = 15\sigma^6 + 36\sigma^4|a_{0,1}|^2 + 9\sigma^2|a_{0,1}|^4 && (557)\\
&\mathrm{Var}(\tilde{b}_{00k}) = 3\sigma^6 + \sigma^4\{6|a_{0,1}|^2 + 2|a_{0,k}|^2\} + \sigma^2\{|a_{0,1}|^4 + 4|a_{0,1}|^2|a_{0,k}|^2\} && (558)\\
&\mathrm{Var}(\tilde{b}_{201}) = 2\sigma^6 + \sigma^4\{4|a_{1,1}|^2 + |a_{0,1}|^2\} + \sigma^2\{|a_{1,1}|^4 + 2|a_{0,1}|^2|a_{1,1}|^2\} && (559)\\
&\mathrm{Var}(\tilde{b}_{21k}) = 2\sigma^6 + \sigma^4\{4|a_{1,1}|^2 + 2|a_{2,k}|^2\} + \sigma^2\{|a_{1,1}|^4 + 4|a_{1,1}|^2|a_{2,k}|^2\} && (560)
\end{aligned}$$

□

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [N/A].
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for [N/A]).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will also be asked to include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While [Yes] is generally preferable to [No], it is perfectly acceptable to answer [No] provided a proper justification is given (e.g., error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering [No] or [N/A] is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We clearly state that our technical contribution is the disk bispectrum inversion, enabling the selection of a minimal subset of coefficients neccessary for complete image description.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: The limitations of the SDB are the same as the limitations of other bispectra. Bispectra are only invariant to one group action, and the SDB is only invariant to rotation. Additionally, its invariance is global, so it is not well suited for tasks requiring local invariance.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All noise proofs assume i.i.d. Gaussian noise is added at each point. The bispectrum inversion relies on non-zero DH coefficients, which is a canonical assumption in bispectrum inversion literature and can be easily satisfied if some amount of noise is present on the image.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide information on noise, seed, MLP parameter counts and number of hidden layers. The code will be publicly available on Github upon acceptance.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Code will be publicly available on Github upon paper acceptance.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.

- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: In addition to providing adequate details in the core paper, we provide details on MLP architecture and hyperparameters in the appendix. Additionally, code will be made publicly available upon acceptance.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The experiments are run across 5 seeds. Graph markers are the mean, and error bars show standard deviation.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

Justification: We report that our experiments are run on one NVIDIA A100 GPU, and average train times are reported for each feature representation.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We follow the code of ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The proposed method could offer faster training on some tasks, which could reduce energy expenditure.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This does not apply to our paper. we do not release data, pre-trained language models, or image generators.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: To the best of our knowledge, we have cited all appropriate works relevant to our paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We will release the code upon acceptance.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.

- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Our work does not include human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Our work does not include human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: Our paper does not use LLMs for important, original, or non-standard components of the core methods in the research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.

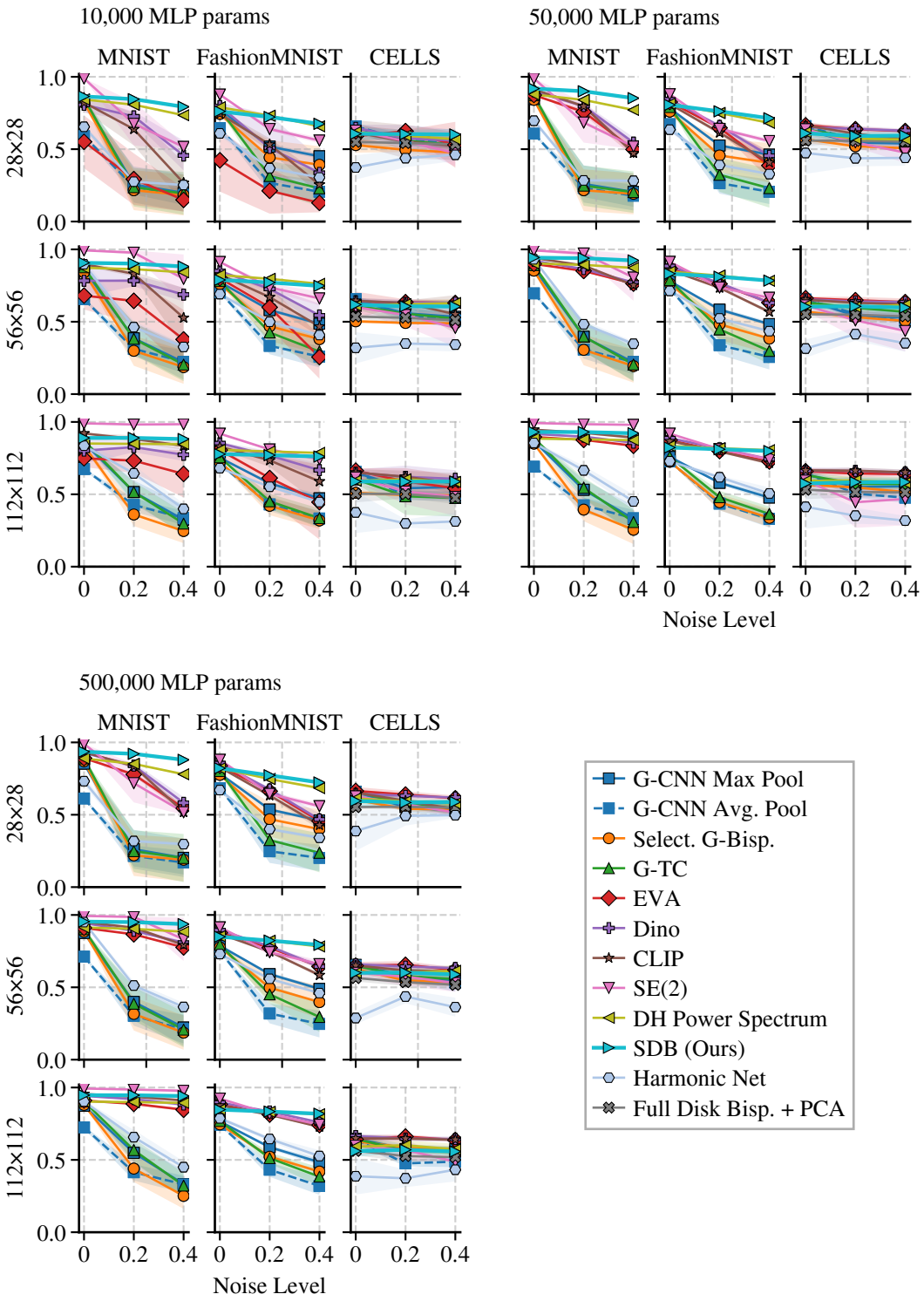

Figure 5: **The SDB's advantage over convolutional descriptors under noise holds at every image size and MLP budget.** Within each block, rows are image sizes $L \in \{28, 56, 112\}$ and columns are datasets (MNIST, FashionMNIST, and osteosarcoma CELLS). The three blocks correspond to MLP budgets of roughly 10,000, 50,000 and 500,000 parameters. Every panel plots test accuracy against the Gaussian noise level $\sigma \in \{0, 0.2, 0.4\}$ on a shared accuracy scale, as mean over seeds with ± 1 standard deviation shaded. The protocol matches Fig. 3. The main-text figure is the $L = 28$, 500,000-parameter slice of this grid.

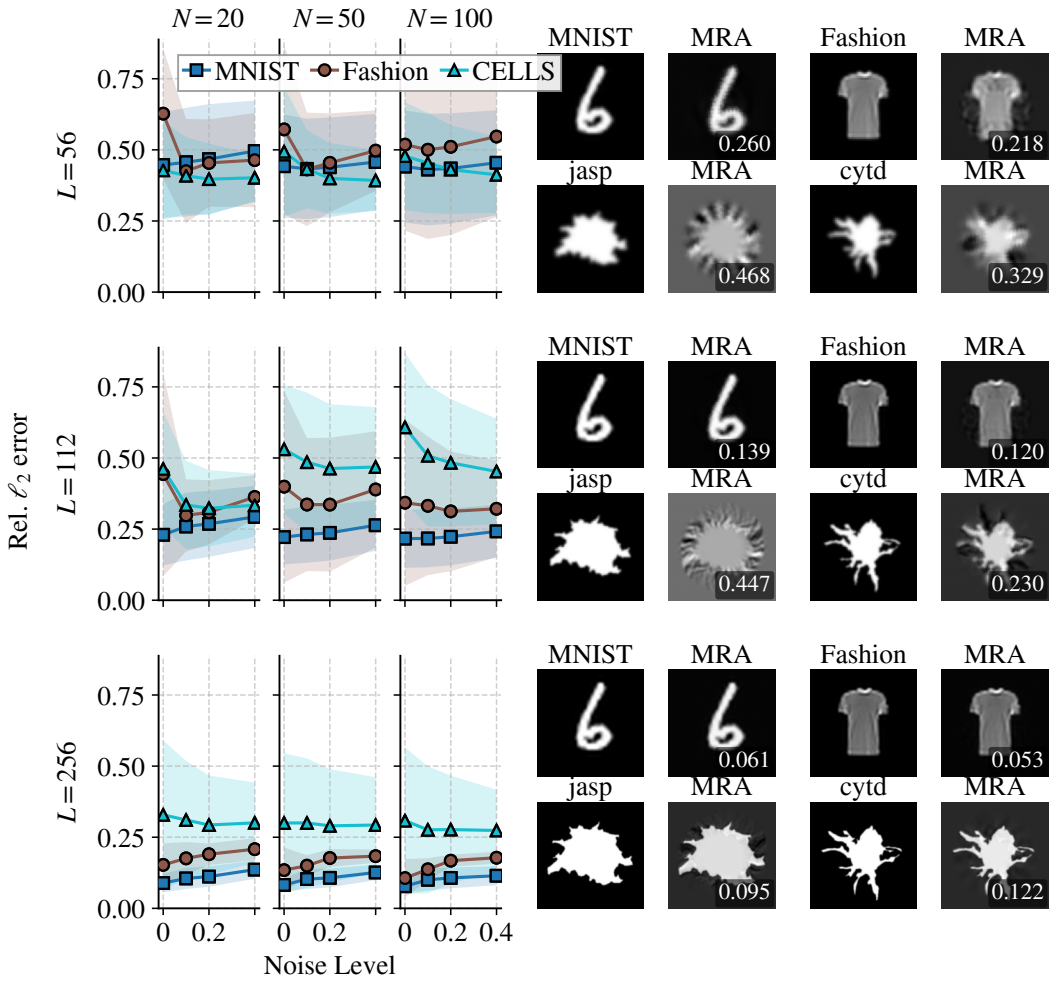

Figure 6: Multi-reference alignment across image sizes. Each row repeats the main-text experiment at a different resolution, $L \in \{56, 112, 256\}$, with bandlimits 44, 84 and 160 respectively. The main-text figure is the $L = 256$ row.

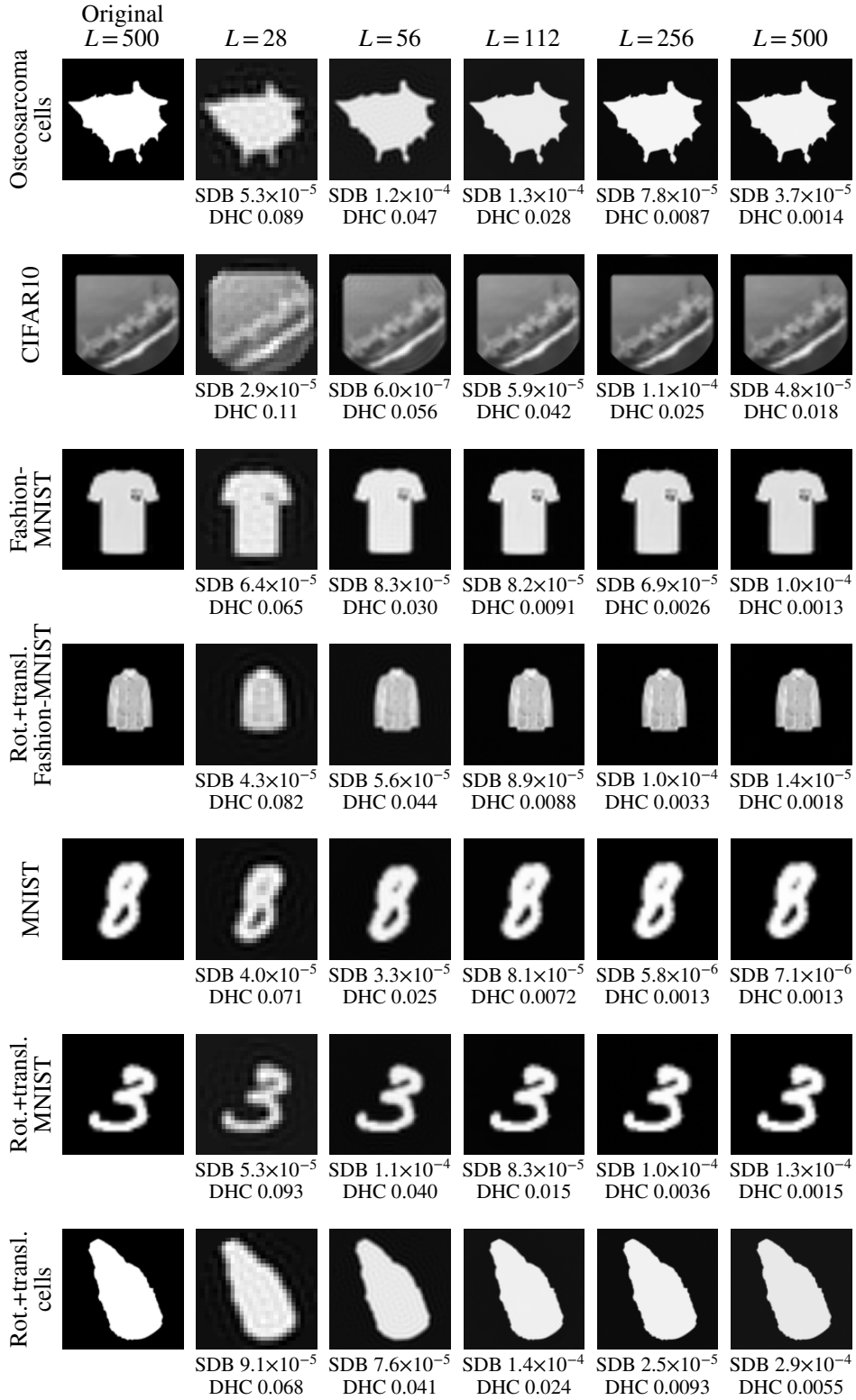

Figure 7: Inversion examples at $L \in \{28, 56, 112, 256, 500\}$, with one row per dataset. The SDB error reports the error from mapping an image's DH coefficients to the SDB and back to DH coefficients. The DH coefficient (DHC) error reports the reconstruction error from mapping an image to its DHCs and back.