Many-Frequency Synchronization Below BBP: A $\lambda_{\text{comp}}(L) \asymp L^{-1/2}$ Law via Frequency-Coupled AMP and Growing-L Low-Degree Bounds

Liz Lemma Future Detective

January 17, 2026

Abstract

We study multi-frequency synchronization over $U(1)$ in the dense (complete-graph) Gaussian noise model. Observations consist of L independent spiked Wigner matrices $Y^{(k)}$ whose spikes are entrywise powers $x^{\odot k}$ of a single latent signal x uniformly distributed on the L-th roots of unity. For a single frequency, the BBP transition implies that polynomial-time detection is possible only for $\lambda \geq 1$, and for constant L computational lower bounds based on the low-degree conjecture suggest that additional frequencies do not lower this threshold. We ask how the computational detection threshold changes when the number of frequencies $L = L(n)$ diverges.

Our main contribution is a sharp many-frequency law: we give an explicit polynomial-time frequency-coupled AMP algorithm that detects for $\lambda \gtrsim 1/\sqrt{L}$ and prove a matching growing-L low-degree lower bound showing that no low-degree test succeeds for $\lambda \lesssim 1/\sqrt{L}$. This provides a clean mechanism by which additional channels ('modalities') convert information-theoretic gains into efficient detection, addressing and sharpening the open question posed in Randomstrasse101 (Entry 4). We also discuss extensions to $SO(2)$ using truncation of irreducible representations and to other compact groups with one-dimensional character families, and we indicate where computer-assisted verification (for constants and finite-size calibration) may be needed.

Table of Contents

3. 3. Warm-up: why naive spectral methods fail below $\lambda = 1$ for each channel; what 'coupling across frequencies' must achieve.

4. 4. Frequency-coupled AMP (FC-AMP): explicit iteration, computational cost, and relation to belief propagation / approximate Bayes updates on $\mathbb{Z}_L$.

5. 5. State evolution and the $1/\sqrt{L}$ threshold (upper bound): derive the scalar recursion, identify the stability condition, and prove detection for $\lambda \geq C/\sqrt{L}$.

6. 6. Initialization without oracle information: explicit small-random-seed / vanishing side-information initialization and robustness analysis; discussion of when extensive case analysis (constants) is required.

7. 7. Growing-L low-degree lower bound: orthogonal polynomial decomposition across channels, moment computations, and the impossibility of low-degree detection for $\lambda \leq c/\sqrt{L}$.

8. 8. Extensions: (i) $SO(2)$ via truncation to L Fourier modes; (ii) other finite abelian groups; (iii) discussion of when non-abelian irreps complicate the picture.

9. 9. Numerical validation and finite-size scaling: what can be verified computationally (threshold constants, iteration counts) and what remains heuristic.

10. 10. Discussion and open problems: optimal constants, estimation (not just detection), subexponential tradeoffs for fixed L, and connections to low-degree conjectures.

1 Introduction and problem statement

We study a dense synchronization model in which the latent variables are group elements and the observations come in multiple harmonic "channels." Concretely, there are n unknown labels $x_1, \ldots, x_n$ taking values in a cyclic group, and for each frequency $k \in \{1, \ldots, L\}$ we are given a noisy matrix whose mean contains the rank-one pattern associated to the k-th character of the group. The defining feature is that these L matrices are not independent signal instances: they are coupled through a single underlying x, and the k-th channel contains information about $x^{\odot k}$. This coupling expresses an algebraic consistency across frequencies (the group law), and it is precisely this structure that allows nontrivial algorithmic gains when the number of frequencies L grows with n.

The single-frequency version of this problem is the familiar spiked Wigner model. There one observes a Wigner matrix perturbed by a rank-one spike, and a sharp spectral transition (the BBP transition) occurs at a constant spike strength: below this value, the top eigenvector is asymptotically uninformative and even detection is impossible by basic spectral tests; above it, principal component analysis succeeds. From the standpoint of average-case complexity, the single-frequency spiked Wigner model has also served as a canonical benchmark where simple algorithms achieve the conjectured computational threshold. In our setting, however, a naive reduction to the single-frequency case by applying PCA separately to each k is suboptimal when $L \to \infty$: each channel remains individually weak, but collectively the channels contain increasing aggregate information about the presence of a planted signal.

This leads to the main question: how does the computational detection threshold scale as the number of frequencies L increases? At an information-theoretic level one expects that additional channels should help, but it is not automatic that polynomial-time methods can exploit the help at the same rate. Indeed, in many planted problems the information-theoretic threshold and the best known polynomial-time threshold separate, and understanding whether a gap exists is typically delicate. Here we isolate a regime in which $L = L(n) \to \infty$ but remains at most polynomially growing in n, and we ask for the smallest $\lambda = \lambda_n$ such that there exists a polynomial-time test distinguishing the null model from the planted one with vanishing error.

Our emphasis on *detection* rather than full *recovery* is intentional. The detection task asks only whether the data contain a nonzero spike, not to reconstruct x. In spiked random matrix models, detection is often the simpler goal and admits a clean characterization in terms of likelihood ratios, contiguity, and low-degree polynomials. Moreover, detection is the natural first step for articulating computational-statistical tradeoffs: if detection cannot be done efficiently, then certainly recovery cannot; conversely, when detection can be done at a certain scaling, one can ask whether the same scaling supports any nontrivial correlation with x. In the present multi-frequency synchronization problem, we show that the algebraic coupling across frequencies already yields a sharp improvement at the detection level, and our algorithms and lower bounds match at the scale of interest.

At a heuristic level, the gain from multiple frequencies can be summarized in a single sentence: *each frequency contributes a small amount of evidence, and the evidence adds.* More precisely, the weak-signal expansion of the log-likelihood ratio suggests that the relevant effective signal strength is of the form λ^2 times the number of independent channels. Because the L observation matrices have independent Wigner noises while sharing the same latent x, one expects contributions from distinct k's to accumulate, leading to an instability criterion resembling

$$\text{(effective SNR)} \;\approx\; \lambda^2 L.$$

When $\lambda^2 L \gg 1$, an infinitesimal bias toward the true signal direction should be amplified by an iterative method; when $\lambda^2 L \ll 1$, the planted and null distributions should be close in total variation, at least as perceived by any procedure that only probes low-order correlations of the data. The main point of our work is to make this picture precise in a computationally meaningful sense.

The algorithmic challenge is that the information is not simply present as a single rank-one perturbation of a single matrix. Instead, the planted component in frequency k is aligned with $x^{\odot k}$, and these vectors are related nonlinearly across k. A method that ignores this relation (for instance, running PCA independently on each $Y^{(k)}$ and then attempting to aggregate decisions) is forced to operate at the per-channel BBP scale and does not improve with L. To benefit from multiple frequencies, we need an algorithm that *couples* the channels during inference and enforces that the inferred k-th harmonic estimates are consistent with a single underlying group label at each node.

We therefore consider a frequency-coupled approximate message passing (FC-AMP) procedure. Conceptually, FC-AMP maintains for each node i a soft estimate of its group label (a distribution on $\mathbb{Z}_L$, or equivalently a set of Fourier coefficients), and iteratively updates these estimates using all observed matrices $\{Y^{(k)}\}_{k=1}^{L}$. The update has two essential components: a linear aggregation step that resembles multiplying by $Y^{(k)}$ in each channel, and a nonlinear "coupling" step that maps the collection of per-frequency messages back to a coherent belief over a single group element. This coupling step is the mechanism by which the algorithm converts weak harmonic evidence at many frequencies into a coherent global signal.

From the standpoint of analysis, AMP-type algorithms are attractive because they admit a tractable asymptotic description via state evolution. In the present model, despite the multi-frequency structure, the state evolution reduces to tracking a small number of order parameters that measure correlation with the planted signal. The resulting recursion has an uninformative fixed point corresponding to zero overlap with x. We show that this fixed point undergoes a linear instability precisely when $\lambda^2 L$ exceeds a constant, and that in this regime FC-AMP amplifies a vanishingly small random seed into a macroscopic correlation. Extracting from the iterates an appropriate scalar statistic then yields a polynomial-time detection test. The resulting algorithmic threshold scales like $1/\sqrt{L}$, exhibiting a clean quantitative "many modalities" phenomenon.

On the lower bound side, we adopt the low-degree polynomial method as a proxy for polynomial-time computation in average-case problems. The guiding principle is that for many planted models with Gaussian noise, any polynomial-time algorithm can be simulated (in an appropriate sense) by a polynomial of modest degree in the input, and thus failure of all low-degree polynomials is strong evidence for computational hardness. Technically, one studies the likelihood ratio $d\mathbb{P}_{\mathsf{H}_1}/d\mathbb{P}_{\mathsf{H}_0}$, expands it in an orthogonal polynomial basis under H_0 (Hermite polynomials for Gaussian variables), and controls the $L^2(\mathsf{H}_0)$ norm contributed by chaoses up to degree d. If these low-degree components have vanishing norm, then any polynomial of degree at most d has vanishing advantage for distinguishing H_0 and H_1.

The multi-frequency setting introduces a genuine new aspect: L is no longer constant, and we must control how the relevant norms and combinatorial factors scale with L. When L grows, a fixed-degree truncation may miss important cancellations or amplifications across frequencies; conversely, a crude union bound over channels can introduce spurious L-dependent losses. The correct calculation shows that, at low degree, the signal contributions from different frequencies add in a way that depends on $\lambda^2 L$. In particular, when $\lambda^2 L \to 0$, every polynomial of degree $o(L)$ has vanishing distinguishing advantage. This matches the instability criterion from the FC-AMP state evolution and yields conditional tightness (under the standard low-degree conjecture) of the computational threshold at the scale $\lambda \asymp 1/\sqrt{L}$.

We stress that this scaling is not a generic consequence of having L independent samples; it relies on the specific algebraic coupling induced by the group characters. The k-th spike vector is not an independent latent direction but a deterministic transform of x, and FC-AMP is designed to exploit exactly this constraint. This also explains why the conclusion is robust under natural variations, such as replacing the discrete cyclic prior by an $SO(2)$ phase model and viewing the first L frequencies as a Fourier truncation: the relevant quantity is the total Fisher information accumulated across harmonics, which scales linearly in L under mild regularity conditions.

In the next section we give the precise model and notation, fix the Wigner conventions (real versus complex), and state the detection criterion and growth regime for $L = L(n)$. All subsequent arguments—state evolution for FC-AMP and the growing-L low-degree bounds—are formulated within that framework.

2 Model and notation

Cyclic prior and harmonics. Fix an integer $L = L(n) \geq 2$ and let $\omega := \exp(2\pi i/L)$ denote a primitive L-th root of unity. Our latent signal is a vector

$$x = (x_1, \dots, x_n) \in \mathcal{X}_{n,L} := \{\omega^0, \omega^1, \dots, \omega^{L-1}\}^n,$$

with i.i.d. entries $x_i \sim \mathrm{Unif}(\{\omega^0, \dots, \omega^{L-1}\})$. Equivalently, we may write $x_i = \omega^{\sigma_i}$ for i.i.d. $\sigma_i \sim \mathrm{Unif}(\mathbb{Z}_L)$. For each integer k, we denote by

$$x^{\odot k} := (x_1^k, \dots, x_n^k) \in \mathbb{C}^n$$

the entrywise k-th power of x. The vectors $\{x^{\odot k}\}$ are the harmonic (character) transforms of the same underlying labels; in particular, they are *not* independent spikes across k, but satisfy the deterministic coupling relations

$$x^{\odot(k+\ell)} = x^{\odot k} \odot x^{\odot \ell}, \qquad (x^{\odot k})^{\odot \ell} = x^{\odot(k\ell)},$$

together with the periodicity $x^{\odot(k+L)} = x^{\odot k}$. This algebraic consistency is the only way in which channels interact: conditional on x, the observation noises across different k are independent.

We will index frequencies by $k \in \{1, \dots, L\}$ for notational uniformity. Since the character $k \equiv 0 \pmod{L}$ is trivial (and would yield a known rank-one direction), one may equivalently take $k \in \{1, \dots, L-1\}$ and relabel; all statements below concern the nontrivial harmonics and are unaffected by this harmless convention.

Observation model. For each frequency $k \in \{1, \dots, L\}$, we observe an $n \times n$ Hermitian matrix

$$Y^{(k)} = \frac{\lambda}{n} x^{\odot k} (x^{\odot k})^* + \frac{1}{\sqrt{n}} W^{(k)} \in \mathsf{Herm}(n), \tag{1}$$

where $\lambda = \lambda_n \geq 0$ is the signal-to-noise parameter and $W^{(1)}, \dots, W^{(L)}$ are independent Wigner matrices (either all real-symmetric or all complex-Hermitian, fixed once and for all). The overall observation is the collection

$$Y := \{Y^{(k)}\}_{k=1}^L \in \mathcal{Y}_{n,L} := \prod_{k=1}^L \mathsf{Herm}(n).$$

The scaling in (1) is chosen so that each individual channel is a standard spiked Wigner instance: the noise term $W^{(k)}/\sqrt{n}$ has $O(1)$ operator norm, while the spike term $(\lambda/n)\, x^{\odot k} (x^{\odot k})^*$ has leading eigenvalue λ with corresponding eigenvector proportional to $x^{\odot k}$. In particular, if one were to ignore all other channels and run a single-frequency spectral method on $Y^{(k)}$, the relevant scale is the usual constant-order BBP transition.

Wigner conventions (GOE/GUE) and independence. We consider two standard noise ensembles.

Complex case (GUE-type). For $i < j$, the off-diagonal entries satisfy $W_{ij}^{(k)} \sim \mathcal{N}_{\mathbb{C}}(0,1)$, meaning $W_{ij}^{(k)} = A + iB$ with $A, B \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/2)$, and we set $W_{ji}^{(k)} = \overline{W_{ij}^{(k)}}$. The diagonal is set to 0 (this choice simplifies bookkeeping and does not change the asymptotics at the scales of interest).

Real case (GOE-type). For $i < j$, we take $W_{ij}^{(k)} \sim \mathcal{N}(0,1)$, set $W_{ji}^{(k)} = W_{ij}^{(k)}$, and again set the diagonal to 0.

In either case, the family $\{W^{(k)}\}_{k=1}^{L}$ is independent across k, and all randomness in the planted model H_1 comes from the latent x together with these independent noise matrices. We write $\langle u, v\rangle := u^* v$ for the standard inner product on $\mathbb{C}^n$ and $\|\cdot\|_{\text{op}}$ for the operator norm.

Null and planted distributions. We formalize (1) as a hypothesis testing problem. Under the null,

$$\mathsf{H}_0 : \qquad \lambda = 0, \qquad Y^{(k)} = \frac{1}{\sqrt{n}} W^{(k)} \ \text{ independently for } k = 1, \dots, L.$$

Under the planted alternative,

$$\mathsf{H}_1 : \qquad \lambda = \lambda_n > 0, \qquad Y^{(k)} = \frac{\lambda}{n} x^{\odot k} (x^{\odot k})^* + \frac{1}{\sqrt{n}} W^{(k)},$$

where $x \sim \text{Unif}(\mathcal{X}_{n,L})$ is independent of the noises. We denote the induced laws on $Y \in \mathcal{Y}_{n,L}$ by $\mathbb{P}_{\mathsf{H}_0}$ and $\mathbb{P}_{\mathsf{H}_1}$. When convenient, we also use the conditional planted law $\mathbb{P}_{\mathsf{H}_1}(\,\cdot\,|x)$, under which the $Y^{(k)}$ are independent across k with means $(\lambda/n) x^{\odot k}(x^{\odot k})^*$.

Two symmetries are worth recording explicitly. First, the prior is invariant under relabeling $x \mapsto \omega^a x$ (global shift), and the likelihood depends on x only through the rank-one projectors $x^{\odot k}(x^{\odot k})^*$, so any recovery notion must be defined modulo this global phase; for detection, this invariance simply implies that there is no preferred direction under H_1 unless λ is large enough to break symmetry through the data. Second, for any fixed k, the vector $x^{\odot k}$ is itself i.i.d. uniform on $\{\omega^0, \dots, \omega^{L-1}\}^n$ whenever $\gcd(k, L) = 1$, but different k's remain coupled because they are functions of the same x.

Detection criteria. A (possibly randomized) test is a measurable map $\mathcal{A} : \mathcal{Y}_{n,L} \to \{0, 1\}$, where we interpret $\mathcal{A}(Y) = 1$ as deciding H_1. We say that $\mathcal{A}$ achieves *vanishing error* if

$$\mathbb{P}_{\mathsf{H}_0}[\mathcal{A}(Y) = 1] + \mathbb{P}_{\mathsf{H}_1}[\mathcal{A}(Y) = 0] \longrightarrow 0 \qquad \text{as } n \to \infty.$$

Equivalently, the total variation distance satisfies $\|\mathbb{P}_{\mathsf{H}_1} - \mathbb{P}_{\mathsf{H}_0}\|_{\text{TV}} \to 1$. Conversely, if $\|\mathbb{P}_{\mathsf{H}_1} - \mathbb{P}_{\mathsf{H}_0}\|_{\text{TV}} \to 0$, then no test (regardless of computational cost) can detect with success probability bounded away from 1/2. Our focus will be on the intermediate regime where the two measures are information-theoretically distinguishable, but efficient detection may or may not be possible.

When discussing computational constraints, we will consider explicit polynomial-time algorithms $\mathcal{A}$, as well as the standard low-degree proxy: for a polynomial $p = p(Y)$ in the real and imaginary parts of the entries of $\{Y^{(k)}\}$, its distinguishing advantage is

$$\mathrm{Adv}(p) := \frac{\left|\mathbb{E}_{\mathsf{H}_1}[p] - \mathbb{E}_{\mathsf{H}_0}[p]\right|}{\sqrt{\mathrm{Var}_{\mathsf{H}_0}(p)}},$$

which controls the performance of the associated (appropriately normalized) test statistic via Chebyshev-type arguments.

Asymptotic regime for $L = L(n)$ and signal strength $\lambda = \lambda_n$. We study the high-dimensional limit $n \to \infty$, with the number of frequencies $L = L(n) \to \infty$ subject to a mild growth constraint

$$L \le n^c \qquad \text{for some fixed } c \in (0, 1).$$

This regime is broad enough to encompass the "many-frequency" phenomenon (where λ may tend to 0), while ensuring that polynomial-time procedures that process all L dense $n \times n$ matrices remain meaningful and that the constants in our concentration bounds can be controlled uniformly in L. We emphasize that λ is allowed to depend on n and L; the central scaling studied in the sequel is $\lambda \asymp L^{-1/2}$, which tends to 0 precisely when $L \to \infty$.

The next section provides a warm-up that justifies why treating each channel $Y^{(k)}$ in isolation is insufficient in this regime: even though the aggregate information across frequencies grows with L, naive spectral methods are pinned to the per-channel constant-order transition. This motivates the need for an explicitly frequency-coupled procedure that enforces harmonic consistency across k and thereby converts many individually subcritical observations into a detectable global signal.

3 Warm-up: why uncoupled spectral methods do not exploit many frequencies

We now isolate a basic obstruction: if we process each matrix $Y^{(k)}$ independently (or combine them by a fixed linear rule that does not use the algebraic relations among harmonics), then we do not obtain detection below the usual single-channel BBP scale $\lambda \approx 1$. In contrast, the target scaling $\lambda \asymp L^{-1/2}$ requires a procedure that *aligns* the weak evidence spread across frequencies by enforcing consistency of the underlying group labels.

Single-frequency PCA remains pinned at $\lambda \approx 1$. Fix k. Conditional on x, the matrix $Y^{(k)}$ is exactly a rank-one spiked Wigner model with spike vector $v^{(k)} := x^{\odot k}/\|x^{\odot k}\| = x^{\odot k}/\sqrt{n}$. It is therefore natural to consider the top eigenvalue/eigenvector of $Y^{(k)}$ as a detection statistic.

The relevant fact is that, for each fixed k, the spectral behavior depends on λ at constant order: under H_0, $\|Y^{(k)}\|_{\mathrm{op}} = 2 + o_{\mathbb{P}}(1)$, while under H_1 the top eigenvalue separates from 2 only when $\lambda > 1$, and the top eigenvector has asymptotically nontrivial overlap with $v^{(k)}$ only in this supercritical regime. In particular, if $\lambda < 1$ is bounded away from 1, then any statistic depending on a *single* $Y^{(k)}$ through a constant number of its extremal eigenvalues cannot distinguish H_1 from H_0 with vanishing error.

This already suggests the issue in the many-frequency regime: when $\lambda \asymp L^{-1/2} \to 0$, every channel is individually far below its BBP transition, so any method that "runs PCA on each frequency" and then aggregates decisions cannot improve on chance.

Taking the best frequency does not help. One might hope that, even if each channel is subcritical, the maximum over k of a spectral statistic could become informative as $L \to \infty$. This does not occur in our regime $L \le n^c$ with $c < 1$. Indeed, under H_0, the collection $\{\|Y^{(k)}\|_{\mathrm{op}}\}_{k=1}^{L}$ consists of L i.i.d. copies of the GOE/GUE spectral norm at scale n, and standard tail bounds for the top eigenvalue imply

$$\max_{1 \le k \le L} \|Y^{(k)}\|_{\mathrm{op}} = 2 + o_{\mathbb{P}}(1),$$

because L is only subexponential in n (in fact polynomial). Under H_1 with $\lambda < 1$, each $\|Y^{(k)}\|_{\mathrm{op}}$ still concentrates at 2 up to $o(1)$, hence the maximum has the same limit. The same statement holds if one uses, say, the top few eigenvalues or any other statistic whose null fluctuations are governed by Tracy–Widom tails: with $L \le n^c$, the union bound does not create a new separation scale.

Naive linear pooling across k fails because the spikes do not align. A second natural idea is to linearly combine the observed matrices, e.g.

$$\overline{Y} := \sum_{k=1}^{L} a_k Y^{(k)} \qquad \text{for some deterministic weights } a_k,$$

and then run a spectral test on $\overline{Y}$. Since the noise matrices $W^{(k)}$ are independent, the noise in $\overline{Y}$ is again Wigner with variance proportional to $\sum_k |a_k|^2$, so $\|\overline{Y} - \mathbb{E}[\overline{Y} \mid x]\|_{\mathrm{op}}$ is typically of order $\sqrt{\sum_k |a_k|^2}$.

The crucial point is that the signal part

$$\mathbb{E}[\overline{Y} \mid x] = \frac{\lambda}{n} \sum_{k=1}^{L} a_k \, x^{\odot k} (x^{\odot k})^*$$

does *not* accumulate into a single stronger rank-one spike. The vectors $\{x^{\odot k}\}_{k=1}^{L}$ are coupled through x, but they are nearly orthogonal as $n \to \infty$: for $k \not\equiv \ell \pmod{L}$,

$$\frac{1}{n}\langle x^{\odot k}, x^{\odot \ell}\rangle = \frac{1}{n}\sum_{i=1}^{n} x_i^{k-\ell},$$

which has mean 0 and variance $1/n$. A union bound over all pairs (k, ℓ) shows that, with probability $1 - o(1)$,

$$\max_{k \neq \ell} \left| \frac{1}{n}\langle x^{\odot k}, x^{\odot \ell}\rangle \right| = o(1),$$

uniformly for $L \leq n^c$. Equivalently, the $L \times L$ Gram matrix G with entries $G_{k\ell} := \frac{1}{n}\langle x^{\odot k}, x^{\odot \ell}\rangle$ satisfies $G = I_L + o(1)$ in operator norm whp. Writing $X := [x^{\odot 1} \ \cdots \ x^{\odot L}] \in \mathbb{C}^{n \times L}$, we have $X^*X = nG \approx nI_L$, hence XX^* has L nonzero eigenvalues all $\approx n$. In particular, for the unweighted sum ($a_k \equiv 1$),

$$\frac{\lambda}{n}\sum_{k=1}^{L} x^{\odot k}(x^{\odot k})^* = \frac{\lambda}{n}XX^*$$

has top eigenvalue $\lambda(1 + o(1))$, not λL. Thus linear pooling cannot turn L subcritical spikes into one supercritical spike; it produces (approximately) a rank-L perturbation whose nonzero eigenvalues remain at scale λ.

Meanwhile the pooled noise grows: if $\overline{Y} = \sum_k Y^{(k)}$, then $\overline{Y}$ has noise part $(1/\sqrt{n})\sum_k W^{(k)}$, whose operator norm is $\asymp \sqrt{L}$. Therefore the most naive pooling actually *worsens* the spectral signal-to-noise ratio. More generally, any fixed weights $\{a_k\}$ trade off signal size $\lesssim \lambda \max_k |a_k|$ against noise size $\asymp \sqrt{\sum_k |a_k|^2}$, and the best such tradeoff does not yield a gain of order $\sqrt{L}$ in the detection threshold.

What "coupling across frequencies" must accomplish. The preceding failures are not artifacts of a particular statistic; they are symptoms of a structural limitation: any method that treats $\{x^{\odot k}\}$ as unrelated directions cannot add their evidence constructively. To see what is missing, we rewrite the entrywise observation as

$$Y_{ij}^{(k)} = \frac{\lambda}{n}(x_i\overline{x_j})^k + \frac{1}{\sqrt{n}}W_{ij}^{(k)}, \qquad i \neq j.$$

For a fixed pair (i, j), the L-vector $\left(Y_{ij}^{(1)}, \ldots, Y_{ij}^{(L)}\right)$ is a noisy collection of all nontrivial characters of the group element $g_{ij} := x_i\overline{x_j} \in \{\omega^0, \ldots, \omega^{L-1}\}$. In other words, each edge (i, j) carries L noisy Fourier measurements of the *same* latent difference label g_{ij}. If we could reliably infer g_{ij} from these harmonics, we would reduce to a (dense) group synchronization problem on $\mathbb{Z}_L$.

The key is that the information across k is additive at the level of likelihoods (or Fisher information): since $\{W^{(k)}\}$ are independent, the log-likelihood ratio for a hypothesized g_{ij} is a sum over frequencies k. Any algorithm hoping to benefit from large L must therefore (i) aggregate information over k *for each interaction* between nodes, and simultaneously (ii) enforce global consistency across nodes, namely that $g_{ij} = x_i\overline{x_j}$ factors through node labels.

Linear spectral pooling does neither: it aggregates across k in a way that destroys the dependence on g_{ij}, and it never represents or enforces the group constraint $g_{ij}g_{jk} = g_{ik}$.

A heuristic linearization yielding the gain $\lambda^2 L$. We record a standard (but instructive) back-of-the-envelope calculation indicating why a coupled message passing scheme should transition at $\lambda^2 L \approx 1$. Suppose that, at some iteration, we have node-wise "soft estimates" $\hat{x}_j \in \mathbb{C}$ with a small correlation $\mathbb{E}[\hat{x}_j\overline{x_j}] \approx m$ (the precise meaning of $\hat{x}_j$ will be algorithm-dependent). Consider forming, for each i and each frequency k, the linear aggregate

$$h_i^{(k)} := \sum_{j\neq i} Y_{ij}^{(k)}\, \hat{x}_j^{\,k}.$$

Under H_1, expanding $Y_{ij}^{(k)}$ and taking conditional expectations suggests

$$\mathbb{E}\big[h_i^{(k)} \mid x\big] \approx \sum_{j\neq i} \frac{\lambda}{n} x_i^k \overline{x_j^k}\, \mathbb{E}[\hat{x}_j^{\,k} \mid x] \approx \lambda\, m\, x_i^k,$$

since the sum over j contributes a factor n that cancels the $1/n$ scaling. Meanwhile, the noise term $\sum_{j\neq i}(1/\sqrt{n})W_{ij}^{(k)}\hat{x}_j^{\,k}$ has typical size of order 1 (by a central limit heuristic), so each $h_i^{(k)}$ is a noisy observation of x_i^k with signal amplitude $\approx \lambda m$.

If we had only one frequency, this would reproduce the usual $\lambda > 1$ phenomenon: the informative component grows only if λ exceeds a constant. However, in the coupled setting we do not keep k separate. Rather, we should combine the fields $\{h_i^{(k)}\}_{k=1}^L$ to update a belief over the *single* label $x_i \in \{\omega^0, \ldots, \omega^{L-1}\}$. Since the noises across k are independent, the effective signal-to-noise ratio in this combined belief update scales like $\sqrt{L}$ times that of a single channel, leading to an amplification factor on the order of $\lambda\sqrt{L}$. In particular, the uninformative fixed point should become unstable when $\lambda\sqrt{L} > 1$, i.e. when $\lambda^2 L > 1$, which is exactly the scaling realized by the state evolution in the next section.

Takeaway. The warm-up can be summarized as follows. Below $\lambda = 1$, each $Y^{(k)}$ is individually spectrally indistinguishable from noise, and neither maximizing over k nor any fixed linear pooling $\sum_k a_k Y^{(k)}$ can recover the $\sqrt{L}$ gain suggested by independence across frequencies. To reach $\lambda \asymp L^{-1/2}$, we require a *nonlinear* frequency-coupled procedure that represents beliefs on $\mathbb{Z}_L$ (or equivalently Fourier coefficients across harmonics) and enforces harmonic consistency so that evidence from all k contributes to a single latent label per node. This is precisely the role of the frequency-coupled AMP iteration developed next.

4 Frequency-coupled AMP: an explicit iteration on $\mathbb{Z}_L$

We now describe a concrete polynomial-time procedure that *couples* the frequencies by maintaining, at each node, a belief over the single latent label in $\mathbb{Z}_L$. The guiding principle is that although $\{Y^{(k)}\}_{k=1}^L$ appear as L separate spiked Wigner matrices, they are not L independent latent vectors; rather, they are harmonics of a common x, and any successful algorithm must enforce the constraint that these harmonics arise from a *single* group element per node.

Beliefs and Fourier moments. We parameterize each $x_i \in \{\omega^0, \dots, \omega^{L-1}\}$ by an integer label $\ell_i \in \mathbb{Z}_L$ with $x_i = \omega^{\ell_i}$. FC-AMP maintains node-wise marginals

$$\pi_i \in \Delta(\mathbb{Z}_L), \qquad \pi_i(\ell) \approx \mathbb{P}(x_i = \omega^\ell \mid \{Y^{(k)}\}),$$

together with their Fourier moments

$$m_i^{(k)} := \sum_{\ell \in \mathbb{Z}_L} \pi_i(\ell)\, \omega^{k\ell} \approx \mathbb{E}[x_i^k \mid \{Y^{(k)}\}].$$

The moments $\{m_i^{(k)}\}_{k=1}^L$ are the natural objects that interact linearly with the observations, because the rank-one spike in channel k depends on $x^{\odot k}$.

From exact BP to a dense-limit approximation. On a complete graph, the exact Bayes posterior factorizes over edges: for $i \neq j$, the collection $\{Y_{ij}^{(k)}\}_{k=1}^L$ depends on the single difference label $g_{ij} = x_i \overline{x_j} = \omega^{\ell_i - \ell_j}$. Formally, the edge likelihood is (up to constants)

$$\prod_{k=1}^{L} \exp\left(-\frac{n}{2} \left| Y_{ij}^{(k)} - \frac{\lambda}{n} g_{ij}^k \right|^2 \right),$$

so the log-likelihood contribution of an hypothesized $g \in \mathbb{Z}_L$ is additive over k, as anticipated in the warm-up. Exact belief propagation would pass messages $\mu_{j\to i} \in \Delta(\mathbb{Z}_L)$ and update π_i by multiplying incoming edge likelihoods integrated against $\mu_{j\to i}$. In our dense model, each node has $n-1$

neighbors and individual edge contributions are weak ($1/\sqrt{n}$-scale), so the standard AMP philosophy applies: (i) linearize the log-likelihood around an uninformative point, (ii) approximate sums of weakly dependent terms by Gaussians, and (iii) correct for the leading-order correlations created by reuse of the data via an Onsager term.

The outcome is an iterative scheme that alternates a *linear aggregation step* over j (separately for each harmonic k) with a *nonlinear Bayes denoising step* that couples all k back into a single distribution π_i on $\mathbb{Z}_L$.

Explicit FC-AMP iteration (dense, fully observed). We present one convenient formulation, which we will later analyze via state evolution. Let $\{m_i^{(k),t}\}$ denote the moment estimates at iteration t, initialized from a nearly uninformative state with a vanishing random bias (cf. Lemma **??**, stated later). For each $k \in \{1, \dots, L\}$ define the vector $m^{(k),t} \in \mathbb{C}^n$ with entries $m_i^{(k),t}$.

Linear step (matched filtering per harmonic). For each k we compute an AMP field

$$h^{(k),t} \;=\; Y^{(k)} m^{(k),t} \;-\; b_t^{(k)}\, m^{(k),t-1}, \tag{2}$$

where $b_t^{(k)}$ is an Onsager coefficient (typically scalar in our homogeneous setting) chosen to cancel the leading correlation between $Y^{(k)}$ and the current iterate. Concretely, $b_t^{(k)}$ is expressed in terms of an average Jacobian of the denoiser in the next step; we do not need its closed form here, only that it can be computed from the iterates in $O(nL)$ time and that it yields the usual AMP decoupling in the large-n limit.

Nonlinear step (approximate Bayes update on $\mathbb{Z}_L$). For each node i, we gather the harmonic fields $\{h_i^{(k),t}\}_{k=1}^L$ and form an updated belief π_i^{t+1} by

$$\pi_i^{t+1}(\ell) \;\propto\; \exp\Big(\sum_{k=1}^{L} \Re\big(h_i^{(k),t}\, \omega^{-k\ell}\big)\Big), \qquad \ell \in \mathbb{Z}_L, \tag{3}$$

with normalization $\sum_\ell \pi_i^{t+1}(\ell) = 1$. Equation (3) is precisely the Bayes rule for a model in which $\{h_i^{(k),t}\}_k$ are conditionally independent noisy observations of $\{x_i^k\}_k$ with approximately Gaussian noise; the additive form in k is the mechanism by which FC-AMP accumulates evidence across frequencies.

Finally, we map π_i^{t+1} back to its Fourier moments, which will be used in the next linear step:

$$m_i^{(k),t+1} \;=\; \sum_{\ell \in \mathbb{Z}_L} \pi_i^{t+1}(\ell)\, \omega^{k\ell}, \qquad k = 1, \dots, L. \tag{4}$$

The update (3)–(4) is the *frequency coupling*: although the linear aggregation (2) treats each k separately, the denoiser takes the collection of all harmonics and returns a *single* belief on the underlying group element.

Remarks on the role of the cyclic structure. The particular form $\Re(h_i^{(k)}\omega^{-k\ell})$ reflects that $\omega^{-k\ell}$ is the k-th character evaluated at label ℓ, and thus (3) is a log-linear model in the Fourier basis. This is exactly what is unavailable to uncoupled spectral methods: the k-dependence is not averaged away, but rather reinterpreted as coherent information about the same latent ℓ_i.

In practice, one may drop redundant frequencies (e.g. $k \equiv 0 \pmod L$ carries no information), and one can enforce conjugacy constraints (e.g. $m_i^{(L-k)} = \overline{m_i^{(k)}}$ in the complex-root parametrization) to reduce computation. These are inessential for the scaling statements and we keep the full $\{1, \dots, L\}$ indexing for notational uniformity.

Computational cost. The dominant operation in (2) is multiplying $Y^{(k)}$ by a vector, which costs $O(n^2)$ per k for dense matrices. Thus one iteration costs

$$O(Ln^2) \quad \text{time}, \qquad O(Ln^2) \quad \text{storage if } Y^{(k)} \text{ are stored explicitly.}$$

The nonlinear step (3)–(4) can be implemented as follows. For each i, the map

$$\ell \mapsto s_i(\ell) := \sum_{k=1}^{L} \Re\big(h_i^{(k),t}\omega^{-k\ell}\big)$$

is (up to taking real parts) an inverse discrete Fourier transform of the sequence $\{h_i^{(k),t}\}_k$. Hence $s_i(\cdot)$ can be computed in $O(L \log L)$ time via FFT, and then π_i^{t+1} is obtained by exponentiating and normalizing. Once π_i^{t+1} is available, the moments (4) can likewise be computed by FFT (forward transform) in $O(L \log L)$ per node. Overall, the coupling step costs $O(nL \log L)$, which is negligible compared to $O(Ln^2)$ in the dense regime.

We emphasize that this separation of costs mirrors the conceptual separation: the expensive part is the dense linear mixing across nodes (matrix-vector products), while the coupling across frequencies is cheap due to the abelian group structure.

Relation to approximate Bayes and the emergence of a scalar order parameter. The iteration above is best viewed as an approximate Bayes procedure in which $\{h_i^{(k),t}\}_k$ play the role of sufficient statistics for x_i. Under the AMP decoupling (justified later by state evolution), the collection of fields at a typical node behaves as

$$h_i^{(k),t} \approx \alpha_t \, x_i^k + \tau_t \, z_i^{(k)},$$

where $z_i^{(k)}$ are approximately i.i.d. standard complex Gaussians across k (and across i), and α_t, τ_t are deterministic scalars. Plugging such a decoupled model into Bayes rule yields precisely the exponential-family form (3), and the updated moments (4) depend on the data only through the single scalar signal-to-noise parameter α_t/τ_t. This is the mechanism behind the scalar state evolution recursion stated in the next section: all algorithmic progress is summarized by one overlap parameter measuring correlation with the truth, and the contribution of L independent harmonics enters through the effective strength $\lambda^2 L$.

A detection statistic extracted from the iterates. For the detection task, we do not need to output an estimate of x; it suffices to compute any scalar statistic that remains near its null value under H_0 but becomes separated under H_1. In the AMP framework, a natural choice is an ℓ_2-energy of the moment vectors, for instance

$$T_t := \frac{1}{L}\sum_{k=1}^{L}\frac{1}{n}\|m^{(k),t}\|_2^2,$$

or a closely related quantity derived from the fields $\{h^{(k),t}\}$. Under H_0, the iterates remain asymptotically uninformative (with T_t near its trivial value) as long as the uninformative fixed point is stable; under H_1 in the unstable regime, the same recursion amplifies the initial seed and drives T_t away from its null value. The formal correctness of such a decision rule will follow from the state evolution analysis.

In summary, FC-AMP is a message passing scheme that is linear across nodes and nonlinear across frequencies, with the nonlinear step implementing an approximate Bayes update over $\mathbb{Z}_L$. Its runtime per iteration is $\widetilde{O}(Ln^2)$ in the dense observation model, and its design ensures that evidence from all L harmonics contributes coherently to a single label per node. We now turn to the state evolution that quantifies this behavior and yields the $1/\sqrt{L}$ threshold.

5 State evolution and the $1/\sqrt{L}$ threshold (upper bound)

We analyze the iteration (2)–(4) by the standard AMP method: for each fixed iteration index t, the high-dimensional randomness in the matrices $\{W^{(k)}\}$ renders the AMP fields approximately Gaussian after conditioning on the past, and the Onsager correction removes the leading dependence created by reusing the same data across iterations. In the present model, the only nonstandard feature is that the denoiser couples all frequencies, but this coupling occurs *within* each node and hence does not obstruct the usual leave-one-out conditioning across nodes.

Order parameters. We denote by $m^{(k),t} \in \mathbb{C}^n$ the vector of k-th Fourier moments at iteration t, as in (4). To summarize the correlation of the iterates with the truth, we introduce the scalar overlap

$$m_t \;:=\; \frac{1}{L}\sum_{k=1}^{L}\frac{1}{n}\Re\big\langle x^{\odot k},\, m^{(k),t}\big\rangle, \tag{5}$$

and the empirical second moment (energy)

$$q_t \;:=\; \frac{1}{L}\sum_{k=1}^{L}\frac{1}{n}\big\|m^{(k),t}\big\|_2^2. \tag{6}$$

In a fully symmetric setting (uniform prior on $\mathbb{Z}_L$, identical noise law across k, and homogeneous Onsager coefficients), one can equivalently track $\frac{1}{n}\Re\langle x^{\odot k}, m^{(k),t}\rangle$ for a fixed k, since these quantities agree asymptotically for all k; we keep the average (5) to emphasize that the subsequent threshold depends only on the aggregate contribution of L channels.

Asymptotic Gaussianity of the fields. Fix $t \geq 0$. Consider the AMP fields $h^{(k),t}$ defined in (2). Under H_1, we may decompose

$$Y^{(k)} m^{(k),t} \;=\; \frac{\lambda}{n} x^{\odot k}\big\langle x^{\odot k}, m^{(k),t}\big\rangle \;+\; \frac{1}{\sqrt{n}} W^{(k)} m^{(k),t}.$$

The first term is a rank-one contribution aligned with $x^{\odot k}$, of size $\lambda\,\frac{1}{n}\langle x^{\odot k}, m^{(k),t}\rangle$. The second term is a (conditionally) mean-zero Gaussian vector, whose covariance is $\approx \frac{1}{n}\|m^{(k),t}\|_2^2\, I_n$ up to negligible diagonal effects and the usual GOE/GUE symmetry constraints. The Onsager correction $b_t^{(k)} m^{(k),t-1}$ is chosen so that, after conditioning on the σ-algebra generated by past iterates, the residual dependence between $W^{(k)}$ and $m^{(k),t}$ cancels to leading order. This yields the canonical AMP decoupling: at a typical coordinate i, the collection $\{h_i^{(k),t}\}_{k=1}^{L}$ behaves like a signal-plus-noise observation of $\{x_i^k\}_{k=1}^{L}$ with independent Gaussian noise across k.

Formally, for each fixed t, the empirical law of $\{(x_i, h_i^{(1),t}, \ldots, h_i^{(L),t})\}_{i=1}^{n}$ converges in probability (in the sense of pseudo-Lipschitz test functions) to the law of $(X, H^{(1)}, \ldots, H^{(L)})$ where $X \sim \mathrm{Unif}(\{\omega^0, \ldots, \omega^{L-1}\})$, $\{Z^{(k)}\}_{k=1}^{L}$ are i.i.d. standard complex Gaussians (or real Gaussians in the GOE case), and

$$H^{(k)} \;=\; \lambda\, m_t\, X^k \;+\; \sqrt{q_t}\, Z^{(k)}. \tag{7}$$

The only dependence on the iteration history appears through the scalars (m_t, q_t). In particular, conditional on X, the L coordinates $\{H^{(k)}\}$ are independent, and the signal enters additively in the mean.

State evolution recursion. Let $\mathsf{D} : \mathbb{C}^L \to \Delta(\mathbb{Z}_L)$ be the denoiser associated to (3), i.e.

$$\mathsf{D}(h^{(1)}, \ldots, h^{(L)})(\ell) \;\propto\; \exp\Big(\sum_{k=1}^{L} \Re\big(h^{(k)}\omega^{-k\ell}\big)\Big),$$

and for each k let $\mathsf{M}_k : \mathbb{C}^L \to \mathbb{C}$ denote the k-th Fourier moment extracted from this posterior,

$$\mathsf{M}_k(h^{(1)}, \ldots, h^{(L)}) \;:=\; \sum_{\ell \in \mathbb{Z}_L} \mathsf{D}(h^{(1)}, \ldots, h^{(L)})(\ell)\, \omega^{k\ell}.$$

Then (4) reads $m_i^{(k),t+1} = \mathsf{M}_k(h_i^{(1),t}, \ldots, h_i^{(L),t})$. Combining this with (7) and passing to the limit yields a closed recursion for (m_t, q_t):

$$m_{t+1} = \frac{1}{L}\sum_{k=1}^{L} \Re\, \mathbb{E}\Big[\overline{X^k}\, \mathsf{M}_k\big(H^{(1)}, \ldots, H^{(L)}\big)\Big], \tag{8}$$

$$q_{t+1} = \frac{1}{L}\sum_{k=1}^{L} \mathbb{E}\Big[\big|\mathsf{M}_k\big(H^{(1)}, \ldots, H^{(L)}\big)\big|^2\Big], \tag{9}$$

where $(X, H^{(1)}, \ldots, H^{(L)})$ are distributed as above.

The recursion (8)–(9) simplifies further because the denoiser is $\mathbb{Z}_L$-equivariant: if X is multiplied by ω^{ℓ_0}, then the vector $(X^k)_{k \le L}$ rotates by characters, and D merely shifts the label $\ell \mapsto \ell + \ell_0$. Consequently, the overlap update (8) depends on (λ, m_t, q_t) only through the single effective scalar $\lambda^2 L\, m_t$ (equivalently $(\lambda m_t)/\sqrt{q_t}$ together with the fact that L independent harmonics contribute additively in the exponent). We record this in the form used throughout the paper: there exists an explicit function Ψ (depending only on the prior, i.e. on $\mathbb{Z}_L$ versus a continuous phase model) such that

$$m_{t+1} \;=\; \Psi\big(\lambda^2 L\, m_t\big), \tag{10}$$

and q_t is then a deterministic function of m_t (or may be tracked jointly without affecting the threshold conclusion). We emphasize that Ψ is not an arbitrary closure assumption: it is obtained by substituting the Gaussian channel (7) into the exact Bayes update on $\mathbb{Z}_L$, and hence is computable in principle (and numerically stable for large L when implemented via FFT).

Linearization and the instability criterion. The fixed point $m = 0$ corresponds to an uninformative regime in which the beliefs π_i remain (asymptotically) close to uniform and the iterates have vanishing correlation with x. By symmetry, $\Psi(0) = 0$. Moreover, the Bayes denoiser is locally matched to the Gaussian channel at $m = 0$, and one obtains the universal derivative

$$\Psi'(0) \;=\; 1, \tag{11}$$

which can be verified by differentiating (8) at $m_t = 0$ and using that the Jacobian of the posterior mean at the uninformative point equals the Fisher score (equivalently, the first-order term of the likelihood ratio in the relevant character direction). Combining (10) and (11) yields the linearized growth law

$$m_{t+1} \;=\; (\lambda^2 L)\, m_t \;+\; o(m_t) \qquad \text{as } m_t \to 0. \tag{12}$$

Thus the uninformative fixed point is unstable if and only if $\lambda^2 L > 1$, and stable if $\lambda^2 L < 1$. This is the origin of the $1/\sqrt{L}$ scaling: each harmonic channel contributes a constant amount of Fisher information, and the L contributions add.

From instability to detection. Assume now that $\lambda \geq C/\sqrt{L}$ for a sufficiently large absolute constant C. Then $\lambda^2 L \geq C^2 > 1$, and (12) implies that any initialization with overlap m_0 merely larger than the typical $n^{-1/2}$-scale fluctuations is amplified geometrically until it reaches a constant level. More precisely, choosing any $\rho \in (1, \lambda^2 L)$, there exists $\delta > 0$ such that whenever $0 < m_t \leq \delta$ we have $m_{t+1} \geq \rho m_t$. Hence after $t_\star = O(\log n)$ iterations, a vanishing seed $m_0 = n^{-\gamma}$ (with $\gamma > 0$ fixed) is driven to $m_{t_\star} \geq m_\star$ for some constant $m_\star = m_\star(C) > 0$. The construction of such a seed without oracle information is deferred to Section 6; here we condition on its existence and focus on the implication of a positive overlap for detection.

To extract a scalar decision statistic, we use the energy T_t defined at the end of Section 4,

$$T_t \;=\; \frac{1}{L}\sum_{k=1}^{L} \frac{1}{n}\big\|m^{(k),t}\big\|_2^2 \;=\; q_t.$$

Under H_0, the state evolution (10) is identically pinned at $m_t \equiv 0$, and q_t converges to a deterministic null value q_{null} determined by passing pure noise through the denoiser. Under H_1 in the unstable regime, the same recursion yields $m_{t_\star} \geq m_\star$, and then (9) implies $q_{t_\star} \geq q_{\text{null}} + \Delta$ for some $\Delta = \Delta(C) > 0$, by continuity of the Bayes risk as a function of the signal component in (7). In other words, the ℓ_2-energy of the moment vectors separates by a constant once the overlap becomes constant.

Finally, we transfer this separation from the limiting recursion to the finite-n iterates. For each fixed t, standard AMP concentration (via Gaussian conditioning and the boundedness/Lipschitz properties of the denoiser, with constants uniform over $L \leq n^c$) yields

$$T_t \;=\; q_t \;+\; o_{\mathbb{P}}(1)$$

under either hypothesis. Therefore, setting a deterministic threshold $\tau \in (q_{\text{null}}, q_{\text{null}} + \Delta)$ and outputting $\mathcal{A} = 1$ if $T_{t_\star} \geq \tau$ yields

$$\mathbb{P}_{\mathsf{H}_0}[\mathcal{A} = 1] \to 0, \qquad \mathbb{P}_{\mathsf{H}_1}[\mathcal{A} = 0] \to 0,$$

establishing detection with vanishing error. The runtime is $t_\star$ AMP iterations, i.e. $O(Ln^2 \log n)$ in the dense model, which is polynomial in n under the standing growth assumption $L \le n^c$.

We have thus identified the algorithmic transition at $\lambda^2 L = 1$ through the stability of the uninformative fixed point of the scalar state evolution (10), and we have shown how, above this threshold, a frequency-coupled AMP iteration yields a concrete polynomial-time detector. We now address the remaining algorithmic issue: how to initialize the iteration so that it possesses a vanishing but non-negligible overlap with the planted signal without requiring oracle side information.

6 Initialization without oracle information

The state evolution analysis in Section 5 identifies the algorithmic transition through the local stability of the uninformative fixed point. To convert this instability into an *explicit* polynomial-time detector, it remains to specify an initialization that (i) does not use any side information about x, (ii) is efficiently computable, and (iii) produces a nontrivial projection onto the unstable direction so that the AMP dynamics can amplify it when $\lambda^2 L > 1$. We address this in a form sufficient for detection; we do not attempt to optimize constants.

Why a seed is necessary. If we initialize the beliefs π_i^0 to be exactly uniform on $\mathbb{Z}_L$, then all Fourier moments vanish, $m_i^{(k),0} = 0$ for all i, k, hence the iterates remain identically zero for all time by equivariance of the updates. This is not a defect of the analysis but a symmetry of the algorithm: without a perturbation, the iterates lie exactly at the uninformative fixed point.

In floating-point implementations, roundoff typically provides a perturbation, but for a mathematical algorithm we inject an explicit perturbation of vanishing magnitude. The basic principle is the same as in the spiked Wigner AMP: a random initialization has overlap $\Theta_{\mathbb{P}}(n^{-1/2})$ with the planted direction, and when the linearized gain exceeds 1 this component grows geometrically for $O(\log n)$ iterations.

A concrete randomized initialization. We describe an initialization at the level of AMP fields, which then induces beliefs and moments through the denoiser. Let $\{\xi_i^{(k)}\}_{i \le n, k \le L}$ be i.i.d. standard complex Gaussians (or real Gaussians in the GOE-compatible real formulation), independent of $\{Y^{(k)}\}$. Fix a deterministic amplitude $\varepsilon = \varepsilon_n \downarrow 0$ (we will take $\varepsilon = n^{-1/10}$ for definiteness), and set

$$h_i^{(k),0} := \varepsilon\, \xi_i^{(k)} \qquad (i \le n,\ k \le L). \tag{13}$$

We then define $\pi_i^0 := \mathsf{D}(h_i^{(1),0}, \dots, h_i^{(L),0})$ and $m_i^{(k),0} := \mathsf{M}_k(h_i^{(1),0}, \dots, h_i^{(L),0})$. This initialization is computable in $O(nL)$ time and produces beliefs that are $o(1)$-close to uniform in total variation at each node, while still breaking the exact symmetry.

Size of the initial overlap. Under H_1, the random variable m_0 defined in (5) has mean 0 by symmetry but a typical magnitude on the order $n^{-1/2}$ (up to the factor ε, depending on how close to uniform we start). Concretely, for small ε we may linearize the denoiser around the uniform point to obtain

$$m_i^{(k),0} \;=\; c_{\mathrm{init}}\, \varepsilon\, \xi_i^{(k)} \;+\; O(\varepsilon^2), \tag{14}$$

with an absolute constant $c_{\mathrm{init}} \in (0,1]$ that depends only on the chosen parameterization of the denoiser (and is uniform for $L \le n^c$). Substituting (14) into (5) and using that $\langle x^{\odot k}, \xi^{(k)} \rangle$ is a centered Gaussian of variance $\Theta(n)$, we obtain

$$m_0 \;=\; \Theta_{\mathbb{P}}\Big(\frac{\varepsilon}{\sqrt{n}}\Big), \tag{15}$$

and moreover for any slowly diverging sequence $a_n \to \infty$,

$$\mathbb{P}\Big[\, |m_0| \ge \frac{\varepsilon}{a_n \sqrt{n}} \,\Big] \;\longrightarrow\; 1. \tag{16}$$

The proof is a routine second-moment computation plus Gaussian anti-concentration, and we omit it.

Two comments are important. First, the sign (or complex phase) of m_0 is random; however, the linearized recursion preserves this phase for small m_t, and our eventual decision statistic is the energy $T_t = q_t$, which is insensitive to this sign/phase. Second, although $|m_0|$ is only $n^{-1/2+o(1)}$, geometric amplification still drives it to a constant level in $O(\log n)$ steps when $\lambda^2 L > 1$.

Amplification from a vanishing random seed. We now formalize the preceding discussion in the form used in the upper bound.

Lemma 6.1 (Vanishing-seed initialization implies macroscopic overlap)**.** *Assume $\lambda^2 L > 1 + \eta$ for some fixed $\eta > 0$. There exist absolute constants $\delta > 0$ and $\rho > 1$ (depending only on η and on the denoiser, but not on n or $L \le n^c$) such that the following holds. Initialize FC-AMP using* (13) *with any $\varepsilon = \varepsilon_n \downarrow 0$ satisfying $\varepsilon \ge n^{-C_0}$ for some fixed $C_0 > 0$. Then, with probability* $1 - o(1)$ *under* H_1*, the state evolution overlap m_t satisfies*

$$|m_{t+1}| \;\ge\; \rho\, |m_t| \qquad \textit{whenever } |m_t| \le \delta,$$

and hence for $t_\star := \lceil \log(\delta/|m_0|)/\log\rho \rceil = O(\log n)$ we have $|m_{t_\star}| \ge \delta$.

Proof sketch. By the Taylor expansion implicit in (12), there exists $\delta > 0$ such that

$$\Psi(u) \;=\; u + r(u), \qquad |r(u)| \le \frac{\eta}{4}|u| \quad \text{for all } |u| \le \lambda^2 L\,\delta.$$

Combining this with $u = \lambda^2 L\, m_t$ gives, for $|m_t| \le \delta$,

$$|m_{t+1}| \;=\; |\Psi(\lambda^2 L\, m_t)| \;\ge\; \Big(1 - \frac{\eta}{4}\Big)\lambda^2 L\, |m_t| \;\ge\; \Big(1 + \frac{\eta}{2}\Big)|m_t| \;=:\; \rho\,|m_t|.$$

It remains to justify that the constants in the Taylor bound can be chosen uniformly over $L \le n^c$. This is a regularity property of the Bayes denoiser at the uninformative point: since D is a softmax over L labels with a bounded local Jacobian, one can bound the second derivative of Ψ at the origin by an absolute constant, independent of L, by differentiating under the expectation in (8) and using $\sum_\ell \mathsf{D}(\ell) = 1$. We omit the derivative bookkeeping. The claim then follows by iterating the one-step growth inequality until reaching δ. □

From overlap amplification to a valid detector. Lemma 6.1 is an SE-level statement; to use it algorithmically we need that the finite-n iterates track SE for $t \le t_\star = O(\log n)$. This is standard in AMP analyses provided one has (a) an initialization independent of the data, (b) Lipschitz control on the denoiser, and (c) boundedness of the Onsager coefficients along the trajectory. In our setting, (a) holds by construction, and (b)–(c) follow from the elementary bounds

$$\big|\mathsf{M}_k(h) - \mathsf{M}_k(h')\big| \;\le\; \|h - h'\|_2 \qquad \text{and} \qquad |\mathsf{M}_k(h)| \le 1, \tag{17}$$

uniformly in $k \le L$ and L, since each M_k is an average of unit-modulus characters under a probability vector $\mathsf{D}(h)$. The only point where we must be careful is uniformity in L: while the softmax normalization involves L terms, the derivatives of D involve covariances under the posterior and hence remain bounded by 1 without incurring an L factor. This uniformity is precisely what prevents the $O(\log n)$ iteration count from accumulating an L-dependent error.

Consequently, with $t = t_\star$ we have, under either hypothesis,

$$T_{t_\star} \;=\; q_{t_\star} + o_{\mathbb{P}}(1),$$

and under H_1 in the regime $\lambda^2 L > 1 + \eta$ Lemma 6.1 implies that $q_{t_\star}$ separates from its null value by a constant (by continuity of (9) in the signal component). Thresholding $T_{t_\star}$ therefore yields a detector with vanishing error, completing the algorithmic part of the upper bound without any oracle information.

Robustness and where constant tracking becomes technical. For the purposes of Theorem 1, we only require a qualitative statement: any perturbation that produces $|m_0| \ge n^{-1/2+o(1)}$ suffices. Two robustness points are worth recording.

First, the precise form of the seed is unimportant. One may equivalently perturb the initial beliefs directly, e.g.

$$\pi_i^0(\ell) \;=\; \frac{1}{L}\Big(1+\varepsilon\,\zeta_{i,\ell}\Big)\Big/\sum_{\ell'}\frac{1}{L}\Big(1+\varepsilon\,\zeta_{i,\ell'}\Big),$$

with i.i.d. mean-zero $\zeta_{i,\ell}$ and sufficiently small ε, and then compute moments. Any such perturbation yields $|m_0| = \Theta_{\mathbb{P}}(\varepsilon/\sqrt{n})$ and triggers the same instability.

Second, the amplification phenomenon is stable under mild misspecification: if the Onsager coefficients are computed with $o(1)$ relative error, or if the denoiser is implemented approximately (e.g. truncating Fourier series in the $SO(2)$ variant), the linearized gain remains $\lambda^2 L + o(1)$ and the criterion $\lambda^2 L > 1$ persists.

What *does* require some case analysis is uniform control of the Taylor remainder in Ψ and of the AMP concentration bounds as $L \to \infty$. In particular, to make Lemma 6.1 completely quantitative one must bound the second (and sometimes third) derivatives of the map $m \mapsto \Psi(\lambda^2 L m)$ in a neighborhood of the origin with constants independent of L, and then union bound these error terms over $t = O(\log n)$ iterations. These bounds are straightforward but notationally heavy because one must keep track of how posterior covariances scale with L when the denoiser aggregates L independent Gaussian channels. Since our goal is the scaling law $\lambda \asymp L^{-1/2}$, we do not optimize these constants and instead work with a fixed margin $\lambda^2 L \ge 1+\eta$, for which all such uniformity issues can be absorbed into absolute constants.

7 Growing-L low-degree lower bound

We establish the low-degree lower bound in the regime $L = L(n) \to \infty$ (with $L \le n^c$, $c < 1$) by working in the $\mathsf{L}^2(\mathsf{H}_0)$ Hilbert space generated by the Gaussian noise across all frequency channels. The conclusion is that, when $\lambda \lesssim L^{-1/2}$, every polynomial statistic of degree $d = o(L)$ has vanishing distinguishing advantage, providing (conditionally, via the standard low-degree conjecture) a computational lower bound matching the scaling of the algorithmic upper bound.

Low-degree proxy via the likelihood ratio. Let $\mathcal{L} := \frac{d\mathbb{P}_{\mathsf{H}_1}}{d\mathbb{P}_{\mathsf{H}_0}}$ be the likelihood ratio on the observation space $\mathcal{Y}_{n,L}$. Under H_0, the collection of off-diagonal entries $\{Y_{ij}^{(k)} : 1 \le i < j \le n,\ 1 \le k \le L\}$ is jointly

Gaussian with independent coordinates (up to Hermitian symmetry), and thus every square-integrable function admits an orthogonal decomposition into multivariate Hermite chaoses. Write $\mathsf{P}_{\le d}$ for the orthogonal projection (in $\mathsf{L}^2(\mathsf{H}_0)$) onto the span of polynomials of total degree at most d in the entries of $\{Y^{(k)}\}$.

A standard argument (see, e.g., the low-degree framework for spiked models) shows that for any degree-$\le d$ polynomial p normalized so that $\mathbb{E}_{\mathsf{H}_0}[p] = 0$ and $\mathrm{Var}_{\mathsf{H}_0}(p) = 1$, we have

$$\left|\mathbb{E}_{\mathsf{H}_1}[p] - \mathbb{E}_{\mathsf{H}_0}[p]\right| \;=\; \left|\mathbb{E}_{\mathsf{H}_0}[p(\mathcal{L}-1)]\right| \;\le\; \|\mathsf{P}_{\le d}(\mathcal{L}-1)\|_{\mathsf{L}^2(\mathsf{H}_0)}. \tag{18}$$

Consequently, to prove that all degree-d tests have vanishing advantage, it suffices to show

$$\|\mathsf{P}_{\le d}(\mathcal{L}-1)\|_{\mathsf{L}^2(\mathsf{H}_0)} \;\longrightarrow\; 0 \qquad \text{for } d = d(n) = o(L). \tag{19}$$

Channel-wise orthogonal polynomial decomposition. It is convenient to rescale to standard Gaussians under H_0. For each k and $i < j$, let

$$Z_{ij}^{(k)} \;:=\; \sqrt{n}\, Y_{ij}^{(k)},$$

so that under H_0 the coordinates $\{Z_{ij}^{(k)}\}$ are i.i.d. $\mathsf{N}(0,1)$ in the GOE case, and i.i.d. standard complex Gaussians (equivalently two independent $\mathsf{N}(0,\frac{1}{2})$ real coordinates) in the GUE case. Under H_1, conditioned on the latent x, we have a mean shift

$$Z_{ij}^{(k)} \;=\; \mu_{ij}^{(k)}(x) + \text{noise}, \qquad \mu_{ij}^{(k)}(x) \;:=\; \frac{\lambda}{\sqrt{n}}\, x_i^k \overline{x_j^k}. \tag{20}$$

For a single real Gaussian coordinate $G \sim \mathsf{N}(0,1)$, the shifted density ratio has the Hermite expansion

$$\exp\Big(\mu G - \frac{\mu^2}{2}\Big) \;=\; \sum_{r \ge 0} \frac{\mu^r}{r!}\, H_r(G), \tag{21}$$

where H_r is the probabilists' Hermite polynomial. In the complex case we apply (21) separately to real and imaginary parts; this changes only bookkeeping and does not affect the counting principles below.

Conditioned on x, the likelihood ratio factorizes over (i,j,k), and averaging over the prior yields

$$\mathcal{L} \;=\; \mathbb{E}_x\left[\prod_{k=1}^{L} \prod_{1 \le i < j \le n} \exp\Big(\mu_{ij}^{(k)}(x)\, Z_{ij}^{(k)} - \tfrac{1}{2}|\mu_{ij}^{(k)}(x)|^2\Big)\right]. \tag{22}$$

Expanding each exponential via (21) and collecting terms of a fixed total degree produces an explicit orthogonal-chaos expansion of $\mathcal{L}$ in the tensor-product Hermite basis across (i,j,k).

Moment constraints from cyclic characters. The key simplification is the orthogonality of cyclic characters under the $\mathbb{Z}_L$ prior:

$$\mathbb{E}[x_i^a \overline{x_i^b}] \;=\; \mathbf{1}\{a \equiv b \pmod{L}\}. \tag{23}$$

Every monomial term in the Hermite expansion of $\mathcal{L}$ can be indexed by a finite multiset of triples (i, j, k) with multiplicities $\{r_{ij}^{(k)}\}$, corresponding to selecting $H_{r_{ij}^{(k)}}(Z_{ij}^{(k)})$ from coordinate (i, j, k). The coefficient of such a term involves $\mathbb{E}_x\big[\prod(\mu_{ij}^{(k)}(x))^{r_{ij}^{(k)}}\big]$, hence a product of factors $x_u^{(\cdot)}\overline{x_u^{(\cdot)}}$ over vertices u.

Interpreting the selection $\{r_{ij}^{(k)}\}$ as a directed multigraph on vertex set $[n]$ with $r_{ij}^{(k)}$ oriented edges $i \to j$ labeled by frequency k, the expectation over x factors over vertices and (23) enforces, at each vertex u, the modular "flow conservation" constraint

$$\sum_{(u\to v)\text{ edges}} k\,(\text{mult.}) \;\equiv\; \sum_{(v\to u)\text{ edges}} k\,(\text{mult.}) \pmod{L}. \tag{24}$$

Thus, only those edge-labeled multigraphs satisfying (24) contribute to $\mathcal{L}$. This is where the shared latent x couples the L channels in the lower bound: although the noise is independent across k, the signal coefficients must be consistent as characters of a single underlying group element.

Two consequences of (24) drive the growing-L analysis.

First, each connected component must contain at least one cycle (in the sense of graph theory) unless all multiplicities are zero: a tree component would force a nontrivial net flow at a leaf, contradicting (24). Equivalently, every participating component has cyclomatic number at least 1.

Second, once a connected component is fixed as an unlabeled graph together with multiplicities, the number of compatible frequency assignments $\{k\} \subseteq [L]$ is bounded by a power of L equal to the number of independent cycles in the component (up to a factor polynomial in the degree). In particular, for total degree d, the number of components is at most $d/2$, hence the number of frequency assignments is at most $L^{d/2}$. This is the first place where $L \to \infty$ strengthens the lower bound: a degree-d statistic can correlate with at most $O(d)$ frequencies in a structured way, while the model spreads signal across all L channels.

Bounding a fixed chaos level. We now state the quantitative chaos bound in the form we use. Its proof is an enumeration of contributing labeled multigraphs together with the Hermite normalization factors, exploiting (i) orthogonality of distinct Hermite basis elements under H_0, (ii) independence across frequencies under H_0, and (iii) the character constraints (24) under H_1. The only genuinely new issue relative to constant-L arguments is to keep constants uniform as $L = L(n) \to \infty$ while allowing degrees $d = o(L)$; this is precisely the regime in which modular aliasing does not create an uncontrolled explosion of solutions to (24).

Lemma 7.1 (Growing-L Hermite chaos bound)**.** *There exist absolute constants $c_0, C > 0$ such that the following holds. Assume $\lambda^2 L \le c_0$ and $L \le n^c$ for some fixed $c < 1$. Let $\mathcal{L}^{=d}$ denote the degree-d Hermite chaos component of $\mathcal{L}$ under H_0. Then, for every $d = d(n) = o(L)$,*

$$\|\mathcal{L}^{=d}\|^2_{\mathsf{L}^2(\mathsf{H}_0)} \;\le\; n^{o(1)}\,(C\lambda^2 L)^d\,L^{-d/2}.$$

The $L^{-d/2}$ factor reflects the frequency-matching scarcity implied by (24): compared to a naive L^d choice of frequency labels for d selected coordinates, only $L^{O(d/2)}$ choices survive the character averaging, and this survives uniformly for $d = o(L)$.

From chaos bounds to vanishing low-degree advantage. Summing the squared norms of chaos levels and using orthogonality, Lemma 7.1 implies

$$\|\mathsf{P}_{\le d}(\mathcal{L}-1)\|^2_{\mathsf{L}^2(\mathsf{H}_0)} \;=\; \sum_{r=1}^{d}\|\mathcal{L}^{=r}\|^2_{\mathsf{L}^2(\mathsf{H}_0)} \;\le\; n^{o(1)}\sum_{r=1}^{d}(C\lambda^2 L)^r\,L^{-r/2}. \tag{25}$$

In the regime $\lambda \le c/\sqrt{L}$ with $c > 0$ sufficiently small, the summand is bounded by $(Cc^2)^r L^{-r/2}$, and therefore the right-hand side of (25) tends to 0 for any $d = o(L)$, yielding (19). Combining (18) with this bound gives the desired vanishing advantage.

We record the conclusion in the form used to derive the low-degree lower bound.

Lemma 7.2 (Vanishing advantage for degree $o(L)$)**.** *There exists an absolute constant $c > 0$ such that if $\lambda \le c/\sqrt{L}$ and $d = d(n) = o(L)$, then*

$$\sup_{\deg(p)\le d}\ \mathrm{Adv}(p) \;\le\; \|\mathsf{P}_{\le d}(\mathcal{L}-1)\|_{\mathsf{L}^2(\mathsf{H}_0)} \;\longrightarrow\; 0,$$

where the supremum is over all polynomials p in the entries of $\{Y^{(k)}\}_{k\le L}$.

Discussion of the growing-L mechanism. The preceding argument isolates the sense in which many frequencies lower the *algorithmic* threshold without simultaneously enabling low-degree detection below $\lambda \asymp L^{-1/2}$. The signal-to-noise per channel decreases with λ, but the Bayes/AMP update aggregates information *linearly* across all L harmonics and then enforces harmonic consistency through the group law. A low-degree polynomial, by contrast, only accesses a bounded-depth collection of entrywise products. After averaging over the cyclic prior, such products survive only when their frequency labels satisfy the modular conservation laws (24), which forces extensive matching and yields the $L^{-d/2}$ penalty in Lemma 7.1. In informal terms: to exploit the full Fisher-information gain of L harmonics, one must build tests whose effective algebraic degree grows with L; degree $o(L)$ is insufficient.

Finally, we emphasize that the restriction $L \leq n^c$ is used only to keep the remaining combinatorial prefactors subpolynomial in n, ensuring that the decay in $L^{-d/2}$ is not overwhelmed by the number of possible index patterns in the dense model. This is the same uniformity issue encountered in the state evolution analysis, but here it appears in counting contributing graphs and bounding the number of admissible frequency assignments uniformly over $L \to \infty$.

7.1 Extensions

$SO(2)$ via truncation to the first L Fourier modes. The cyclic model may be viewed as a discretization of phase synchronization over the compact abelian group $SO(2) \cong U(1)$, whose irreducible representations are the one-dimensional characters $e^{ik\theta}$, $k \in \mathbb{Z}$. In the continuous model we take latent phases $\theta_i \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 2\pi)$, set $x_i := e^{i\theta_i}$, and observe for $k = 1, \dots, L$,

$$Y^{(k)} = \frac{\lambda}{n} x^{\odot k} (x^{\odot k})^* + \frac{1}{\sqrt{n}} W^{(k)}. \tag{26}$$

This is formally identical to the discrete model except that x_i ranges over the unit circle rather than $\{\omega^0, \dots, \omega^{L-1}\}$. The main point is that the analysis of the coupled-harmonic mechanism depends on the orthogonality of characters, which remains valid in the continuous setting:

$$\mathbb{E}[x_i^a \overline{x_i^b}] = \mathbb{E}[e^{i(a-b)\theta_i}] = \mathbf{1}\{a = b\}, \qquad a, b \in \mathbb{Z}. \tag{27}$$

Consequently, the combinatorial constraint that survives after integrating out the latent signal is now an exact (non-modular) conservation law: the flow condition (24) holds with congruence mod L replaced by equality in $\mathbb{Z}$.

To connect (26) back to an L-parameter family, we explicitly *truncate* to the first L positive frequencies. On the algorithmic side, FC-AMP extends by representing the per-node belief not as a probability vector on $\mathbb{Z}_L$ but as a periodic density on $[0, 2\pi)$; in practice we maintain its Fourier series up to order L, i.e.,

$$q_i(\theta) \ \propto \ \exp\Big(\sum_{k=1}^{L} \Re\big(a_{i,k} e^{ik\theta}\big) \Big),$$

where the coefficients $(a_{i,k})_{k \leq L}$ are updated by (i) linear aggregation through the observed matrices $Y^{(k)}$ and (ii) a nonlinearity implementing the Bayes denoiser for the truncated Fourier family. The required coupling across k is exactly the statement that these Fourier coefficients arise from a single phase θ_i, rather than from L unrelated labels.

At the level of state evolution, the same linearization governs the growth of an infinitesimal correlation with the truth: each channel contributes a term proportional to λ^2, and the L channels add. Hence the instability condition remains $\lambda^2 L > 1$ up to constants determined by the chosen truncation/denoiser family; in particular the predicted detection threshold is again $\lambda \asymp L^{-1/2}$. The low-degree lower bound argument also transfers with minimal changes: the Hermite expansion continues to apply entrywise under H_0, and (27) enforces the same graph-theoretic constraints, now without modular aliasing. In this sense, the $\mathbb{Z}_L$ model is best interpreted as a convenient finite-dimensional surrogate of the truncated $SO(2)$ problem, with the same scaling law but simpler bookkeeping.

Other finite abelian groups. The coupled-frequency construction is not specific to cyclic groups; it is a general feature of synchronization models over finite abelian groups G. Writing $\widehat{G}$ for the dual group of one-dimensional characters $\chi : G \to S^1$, we may choose a collection $\mathcal{S} \subseteq \widehat{G}$ of observed characters (playing the role of frequencies). With latent labels $g_i \overset{\text{i.i.d.}}{\sim} \text{Unif}(G)$ and $x_i^{(\chi)} := \chi(g_i)$, we observe for each $\chi \in \mathcal{S}$ an independent spiked Wigner channel

$$Y^{(\chi)} = \frac{\lambda}{n} x^{(\chi)} (x^{(\chi)})^* + \frac{1}{\sqrt{n}} W^{(\chi)}. \tag{28}$$

The coupling across χ is again through the shared latent g_i. FC-AMP extends by maintaining per-node beliefs $\pi_i \in \Delta(G)$ and using the Fourier transform on G: the linear messages are computed in the character domain $\widehat{G}$ (one matrix-vector product per observed χ), while the nonlinear coupling step is a map $\{m_i(\chi)\}_{\chi \in \mathcal{S}} \mapsto \pi_i$ which enforces that the collection of estimated character values arises from a single group element.

Both the upper- and lower-bound heuristics depend only on character orthogonality,

$$\mathbb{E}_{g \sim \text{Unif}(G)}[\chi(g)\overline{\chi'(g)}] = \mathbf{1}\{\chi = \chi'\}, \qquad \chi, \chi' \in \widehat{G}, \tag{29}$$

and on independence of the noise across observed channels. In particular, if $|\mathcal{S}| \to \infty$ with n (e.g. by taking a growing family of characters in a sequence of groups $G = G_n$, or by keeping G fixed but repeating observations with independent noise), then the same additivity principle yields an instability criterion of the form

$$\lambda^2 \, |\mathcal{S}| \;\gtrsim\; 1,$$

so that λ_{comp} scales like $1/\sqrt{|\mathcal{S}|}$ under analogous regularity assumptions for state evolution. On the low-degree side, the moment constraints induced by (29) produce a conservation law on edge-labeled multigraphs that is identical in spirit to (24), except that the labels now live in $\widehat{G}$. When $\mathcal{S}$ is large and the polynomial degree d is small relative to the relevant scale (e.g. $d = o(|\mathcal{S}|)$ in a setting where distinct labels behave essentially independently), the same scarcity phenomenon holds: most labelings are annihilated by averaging over g, and only those satisfying the vertex-wise conservation constraints contribute. This is the mechanism behind an L-type penalty in the chaos norms, with $|\mathcal{S}|$ replacing L.

A concrete example beyond cyclic groups is $G = (\mathbb{Z}_q)^r$ with q fixed and $r = r(n) \to \infty$. Here $\widehat{G} \cong G$, and one may take $\mathcal{S}$ to be a set of characters indexed by a subset of $(\mathbb{Z}_q)^r$. The conservation law becomes a vector-valued modular flow constraint in $(\mathbb{Z}_q)^r$, and the counting of admissible labelings depends on the cycle space as before. The resulting picture matches the cyclic case: as the number of observed characters grows, detection becomes possible at smaller λ, while low-degree polynomials cannot exploit this gain unless their degree grows commensurately.

When non-abelian representation theory intervenes. For non-abelian groups G, irreducible representations are typically matrix-valued, and there is no longer a single scalar character per frequency. A natural analogue of (28) is to observe, for a family of irreps $\rho \in \widehat{G}$ (now indexing equivalence classes of irreducible unitary representations), matrix-valued measurements whose mean is rank-one in the representation space:

$$Y^{(\rho)} = \frac{\lambda}{n} u^{(\rho)} (u^{(\rho)})^* + \frac{1}{\sqrt{n}} W^{(\rho)}, \qquad u_i^{(\rho)} := \rho(g_i) v_\rho,$$

for some fixed unit vector v_ρ in the $\dim(\rho)$-dimensional representation space. Even in this simplified form, the coupling constraints across ρ are substantially more complicated: enforcing that the family $\{\rho(g_i)\}_\rho$ comes from a single $g_i \in G$ is a noncommutative compatibility condition, and the analogue of the "harmonic consistency" map requires Clebsch–Gordan decompositions to relate products of matrix coefficients across different irreps.

From the algorithmic perspective, message passing can still be formulated using noncommutative Fourier analysis, but the state variable at each node is no longer a probability vector on G (which is exponentially large) nor a small set of scalar Fourier coefficients; rather, it is a collection of matrices (Fourier coefficients) $\widehat{\pi}_i(\rho) \in \mathbb{C}^{d_\rho \times d_\rho}$ over the observed irreps. The cost of a single iteration is governed by the total Fourier dimension $\sum_{\rho \in \mathcal{S}} d_\rho^2$, and the denoising/coupling step amounts to approximately projecting these coefficients back onto the cone of positive-type functions on G, a significantly less explicit operation than the abelian normalization on $\Delta(G)$. In regimes where only low-dimensional irreps are used (or where G is "nearly abelian" in the sense of having many one-dimensional representations), this may remain tractable, but in general it introduces a substantial algorithmic overhead and obscures the clean $\lambda^2 \times$ #channels additivity.

On the low-degree side, the obstacle is analogous: averaging monomials of matrix coefficients over g_i yields constraints expressed in terms of invariant tensors in representation products. In the abelian case, these invariants reduce to equality of total exponents (the conservation law). For non-abelian groups, the space of invariants in a tensor product can have dimension larger than 1, and the number of ways to "close" a cycle in the moment graph is controlled by multiplicities in Clebsch–Gordan rules rather than by a single modular equation. Thus, the counting that produced a sharp scarcity factor (such as the $L^{-d/2}$ penalty) may be replaced by a more delicate dependence on representation dimensions and fusion multiplicities. We therefore expect that the correct analogue of the scaling law involves the total "harmonic budget" $\sum_{\rho \in \mathcal{S}} d_\rho^2$ (or a related Fisher-information quantity), but establishing a tight low-degree lower bound in full generality would require uniform control of these multiplicities in the growing-family regime.

In summary, the abelian setting isolates the essential phenomenon: multiple harmonics provide additive information, and the group law supplies the coupling needed to aggregate it computationally efficiently; low-degree polynomials fail because character orthogonality forces extensive matching constraints. For $SO(2)$ and other abelian groups, this mechanism survives essentially unchanged (after truncation in the continuous case). For non-abelian groups, the same blueprint remains plausible, but both the algorithmic implementation and the low-degree enumeration must contend with higher-dimensional irreps and nontrivial tensor-product structure, and the clean $1/\sqrt{L}$ scaling may need to be re-expressed in representation-theoretic terms.

Numerical validation and finite-size scaling. The asymptotic statements in the preceding sections invite two complementary kinds of numerical checks: (i) *algorithmic validation*, in which we implement the proposed frequency-coupled AMP (together with the specific scalar decision rule used for detection) and verify that its empirical error exhibits the predicted dependence on λ and L; and (ii) *finite-size diagnostics*, in which we probe how quickly the large-n heuristics (state evolution, linear instability, and low-degree scarcity) become visible at moderate n and growing L. We emphasize that these experiments can corroborate the scaling law and calibrate constants, but they do not in themselves substitute for the asymptotic analysis; in particular, numerical failure to detect at a given (n, L, λ) cannot be interpreted as an impossibility result.

Simulation protocol and what is directly measurable. A baseline protocol is as follows. Fix (n, L), choose a noise ensemble (GOE or GUE), and generate N i.i.d. instances under each hypothesis. Under H_0, set $Y^{(k)} = W^{(k)}/\sqrt{n}$. Under H_1, draw x (or θ in the $SO(2)$ variant) and set $Y^{(k)}$ according to the channel, with independent $W^{(k)}$. For each instance, run FC-AMP for T iterations from the stated initialization (including any vanishing random seed), extract the scalar test statistic $S = S(Y^{(1)}, \ldots, Y^{(L)})$ used by the decision rule, and compute (a) the empirical type-I/type-II errors at a fixed threshold, or (b) the receiver operating characteristic (ROC) curve and its area (AUC), or (c) the total error under the optimal threshold chosen by cross-validation. In addition, under H_1 one may compute the overlap $n^{-1}\langle \widehat{x}, x \rangle$ (or its harmonic analogues) *for diagnostic purposes only*, to compare the observed correlation growth to state evolution; this is not part of the detection task but is helpful for verifying the internal mechanism.

Two practical points are worth making explicit. First, in the dense setting the raw per-iteration cost of any message passing scheme that performs L matrix–vector multiplications is $\widetilde{O}(Ln^2)$, which limits the accessible (n, L) range. For numerical work one typically either (i) restricts to moderate n but sweeps many λ values, or (ii) uses linear-algebra acceleration (e.g. blockwise BLAS on GPU) to sweep L at fixed n. Second, finite-size behavior is sensitive to the exact normalization conventions (zero diagonal vs full Wigner, real vs complex, and off-diagonal variance), so any reported constants should be tied to the precise model implemented.

Calibrating the threshold constant and checking the $L^{-1/2}$ scaling. A direct way to probe the scaling is a *data-collapse* plot: for a grid of (n, L) values, we plot an empirical performance metric (e.g. AUC, or total error at the empirically optimal threshold) against the rescaled signal level

$$\kappa \;:=\; \lambda\sqrt{L}.$$

The prediction is that the transition should occur near a constant κ_c depending on the algorithm/denoiser details (and mildly on GOE vs GUE), with curves for different L collapsing as n increases. In practice, one observes a smoothed transition rather than a sharp step, and one may define $\kappa_c(n, L)$ operationally, e.g. as the value where AUC crosses 0.75 or where the total error crosses 0.25. Plotting $\kappa_c(n, L)$ versus n at fixed L, and versus L at fixed n, gives a quantitative sense of finite-size corrections.

A second, more stringent check uses *state evolution* itself. Under H_1, we record the empirical overlap m_t between the AMP iterate (or its single-site posterior mean surrogate) and the truth. State evolution predicts that m_t should follow a deterministic recursion with a linearization at 0 whose slope is proportional to $\lambda^2 L$. Numerically, this manifests as: (i) for $\kappa < \kappa_c$, the overlap remains at the noise floor (typically of order $n^{-1/2}$, dominated by fluctuations and the seed); (ii) for $\kappa > \kappa_c$, m_t grows approximately geometrically for several iterations before saturating to a positive fixed point; and (iii) the observed growth rate in the early iterations is well-fit by a line in a $\log m_t$ versus t plot, with slope increasing in κ. This diagnostic separates the presence of a computational transition (an instability) from mere improvements in the final decision statistic.

Iteration counts and the $\log n$ prediction. Beyond the location of the transition, one can test the claim that a vanishing seed is amplified in $O(\log n)$ iterations when the instability condition holds. A practical approach is to fix (L, λ) above threshold and measure the smallest iteration T_ϵ at which the detection statistic exceeds a preset level ϵ (or the overlap exceeds ϵ in planted diagnostics). Repeating across n and fitting T_ϵ to $a \log n + b$ yields an empirical confirmation of logarithmic iteration scaling. Here one must keep track of the seeding mechanism: with a smaller seed, the intercept b increases (as expected), while the slope a is governed by the linearized amplification factor. It is also informative to compare several initializations (random infinitesimal bias, a single seeded node, or a small seeded set) to quantify the tradeoff between seeding strength and iteration count at fixed (n, L, λ).

Baselines: what simpler tests achieve at finite n. To interpret performance gains, it is useful to implement baseline polynomial-time detectors that do *not* exploit harmonic consistency. Natural baselines include: (i) single-frequency PCA (apply top-eigenvalue tests to $Y^{(k)}$ for a fixed k or take the maximum over k); (ii) channel-wise aggregation without coupling (e.g. average the top-eigenvalue statistics across k); (iii) low-degree moment tests such as $\sum_{k=1}^{L} \|Y^{(k)}\|_F^2$ and higher-degree trace polynomials built from products of entries across a small number of frequencies. These comparisons typically show two effects: first, methods that treat frequencies independently exhibit thresholds essentially independent of L (up to trivial averaging gains), while the coupled method improves markedly as L grows; second, low-degree moment tests can improve with L but with a weaker scaling (often consistent with a signal-to-noise gain like $\lambda^2\sqrt{L}$ rather than $\lambda\sqrt{L}$), highlighting that the main advantage is not merely having L samples but leveraging the algebraic constraint tying them together.

What can be checked about the low-degree picture. While low-degree lower bounds are asymptotic and cannot be "proved by simulation," one can still probe the underlying combinatorics empirically by measuring the performance of explicit low-degree statistics as L grows. For each fixed degree d, one may define a small family of degree-d polynomials motivated by the Hermite/graph expansion (e.g. sums of products of d matrix entries along short cycles, with various assignments of frequencies), standardize them under H_0, and evaluate their empirical advantage under H_1. The qualitative prediction is that, at λ scaling like $L^{-1/2}$, bounded-degree tests should not show a uniformly strong separation as $L \to \infty$, whereas the coupled AMP statistic should. One may also examine how the best-performing polynomial within a restricted class changes as a function of L, which is a numerical proxy for the "degree must grow with L" phenomenon.

We stress, however, that such experiments are inherently limited: the space of all degree-d polynomials is enormous, and failure of a few hand-chosen statistics does not approximate the low-degree optimum. The more reliable numerical takeaway is therefore comparative: fixed-degree constructions do not appear to track the same threshold curve as the coupled algorithm when L grows.

Finite-size corrections and heuristic aspects that numerics do not settle. Several features of the theory are expected to exhibit slow convergence at moderate n, and numerical studies should be interpreted accordingly. First, the critical window may be broad: even if the limiting threshold is governed by $\lambda^2 L \approx 1$, the empirical transition width in λ can be substantial for $n \leq 10^3$, especially when L grows with n. Second, the independence structure across frequencies can be partially obscured by finite-size effects when L is a non-negligible power of n, since many summary statistics aggregate over Ln^2 entries and thus have their own concentration scales. Third, the constant κ_c is not universal across algorithmic choices: it depends on the denoiser family (discrete vs continuous, truncation choice in the $SO(2)$ variant, damping, and any regularization used to stabilize iterates). Numerics can calibrate κ_c for a given implementation, but cannot by themselves justify that this constant matches the theoretically optimal Bayes threshold, nor that the algorithm is optimal among all polynomial-time methods.

Finally, and most importantly, numerical work does not resolve the heuristic step connecting low-degree failure to computational hardness. At best, simulations can provide circumstantial support by showing a growing gap between the coupled AMP detector and a suite of low-degree or spectral baselines in the regime where the low-degree theory predicts vanishing advantage. Establishing (or refuting) the corresponding hardness statement requires either a proof that all polynomial-time algorithms fail, or a counterexample algorithm; neither can be supplied by finite-n experiments.

Summary of what numerics are for. In this problem, numerics are most informative when used to (i) verify the $\lambda\sqrt{L}$ collapse for the proposed detector, (ii) confirm the predicted $O(\log n)$ iteration growth above threshold and the role of seeding, (iii) benchmark against uncoupled baselines to isolate the effect of harmonic consistency, and (iv) explore robustness to modeling choices (GOE vs GUE, mild non-Gaussian noise, or truncation details in the continuous model). They are least informative when used to make claims about impossibility or tight optimality, which remain asymptotic and, in part, conjectural.

Discussion and open problems. Our results isolate a clean scaling law—the computational detection threshold is governed by the effective parameter $\lambda^2 L$—but they also leave several natural questions open. We organize these around (i) optimal constants and sharp asymptotics, (ii) estimation rather than mere detection, (iii) subexponential-time tradeoffs when L is fixed, and (iv) the precise status of the low-degree conjecture as a proxy for polynomial-time hardness in this frequency-coupled setting.

Optimal constants and sharp thresholds. The upper bound via FC-AMP is driven by a linear instability of the uninformative fixed point, and therefore it naturally predicts a critical value of the rescaled signal $\kappa = \lambda\sqrt{L}$ near a constant of order one. The low-degree lower bound, as stated, provides only a (different) absolute constant in the opposite direction. Closing this constant gap is nontrivial for two reasons. First, on the algorithmic side, one would like a proof that a concrete choice of denoiser/coupling map is optimal (or at least constant-optimal) among polynomial-time methods, which requires controlling state evolution with sufficient precision to identify the sharp instability point and to rule out improvements from more elaborate nonlinearities or multi-step statistics. Second, on the lower-bound side, our Hermite-chaos estimates are designed to be robust in $L \to \infty$ regimes and hence sacrifice sharpness in constants. A natural open problem is to compute the limiting critical constant κ_c for the Bayes-optimal likelihood ratio (or for the posterior) and to compare it to the AMP constant, i.e. to decide whether κ_c is algorithm-independent (after normalization) or whether there is a genuine constant-factor gap between efficient and information-theoretic detection.

A related refinement concerns *critical windows*. Even if the limiting threshold is at $\lambda^2 L = 1$, one expects a nontrivial scaling window for $\lambda^2 L = 1 + O(n^{-\alpha})$ where fluctuations matter and where the best achievable error (or the log-likelihood ratio) admits a universal limit law. Identifying such a window would sharpen the theory in the same spirit as BBP-type fluctuation results for single-frequency spiked Wigner. Here the coupled multi-frequency structure introduces additional aggregate fluctuations over k, and understanding whether these reduce or enlarge the critical window is an open technical challenge.

Detection versus estimation: weak recovery and overlap. The present focus is on hypothesis testing, but the model is intrinsically an estimation problem: infer x (or its phases) from $\{Y^{(k)}\}$. In single-frequency spiked Wigner with i.i.d. priors, detection and weak recovery typically coincide: above the spectral/Bayes threshold one can construct an estimator with nontrivial correlation, and below it both detection and correlation are impossible. In the frequency-coupled model, it is tempting to conjecture an analogous equivalence with threshold $\lambda^2 L = 1$, but turning this into a theorem requires a careful definition of overlap compatible with the group structure (e.g. modulo a global phase when appropriate) and an analysis of the posterior landscape when many harmonics are present.

Concretely, one may ask for a rigorous statement of the following form: if $\lambda^2 L > 1+\varepsilon$, then there exists a polynomial-time estimator $\widehat{x}$ such that a suitable normalized correlation between $\widehat{x}$ and x is bounded away from 0 with high probability; whereas if $\lambda^2 L < 1 - \varepsilon$, then no (possibly exponential-time) estimator achieves nontrivial correlation. The second direction is information-theoretic and would likely proceed via contiguity or mutual-information bounds for the full multi-frequency observation; the first direction requires proving that the FC-AMP iterates produce not merely a detection statistic but an estimator whose overlap tracks state evolution. While these statements are plausible, we emphasize that the coupling across k introduces additional symmetries and potential metastability phenomena that do not appear in the single-frequency model.

Beyond weak recovery: rounding, strong recovery, and exact regimes. When the latent variables are discrete ($\mathbb{Z}_L$), one may also consider stronger goals such as coordinate-wise recovery (possibly up to a global group action). In dense synchronization-type models, exact recovery thresholds often scale with $\log n$ once the per-node effective signal exceeds the typical maximum noise fluctuation. In the present setting, heuristics suggest that aggregating information across frequencies should reduce the required λ for a fixed target accuracy, potentially leading to regimes where λ is much smaller than 1 yet exact recovery remains possible provided L is large enough (e.g. $\lambda^2 L \gtrsim \log n$ as a straw-man scaling). Determining the correct exact-recovery threshold—and whether polynomial-time algorithms achieve it—is open.

Algorithmically, one expects a two-stage procedure: run FC-AMP to obtain soft marginals $\pi_i \in \Delta(\mathbb{Z}_L)$, then perform a rounding and synchronization refinement (e.g. local MAP rounding followed by a global alignment). Proving guarantees for such a pipeline in the dense, growing-L regime would require new concentration inputs, since one must control not only a scalar overlap but also coordinate-wise errors and the stability of the rounding step under correlated multi-frequency noise.

Fixed L and subexponential-time tradeoffs. Our low-degree lower bound is phrased for $\deg(p_n) = o(L)$, which is the natural regime when $L \to \infty$ and one aims to rule out all constant-degree (and even slowly growing-degree) polynomial tests. When L is fixed, however, the condition $\deg(p_n) = o(L)$ becomes vacuous, and the question of computational hardness below the single-frequency BBP scale becomes one of *subexponential* tradeoffs: how does the best achievable detection threshold depend on runtime $T(n)$ between $\mathrm{poly}(n)$ and $\exp(\Theta(n))$?

This is closely analogous to the landscape for tensor PCA and related planted problems, where one observes a continuum of algorithmic thresholds indexed by (for instance) the degree of sum-of-squares relaxations or the exponent in $T(n) = \exp(n^\alpha)$. A compelling open problem is to develop an explicit tradeoff curve for fixed L: for each $\alpha \in (0, 1)$, characterize the smallest $\lambda = \lambda(n)$ detectable in time $\exp(n^{\alpha})$. Even heuristic predictions would be valuable here, since they would clarify whether the multi-frequency coupling provides a qualitatively new source of hardness (beyond merely increasing the data dimension) or whether the model reduces, at fixed L, to known spiked-matrix tradeoffs.

From low-degree to algorithms: what degree is actually needed? Theorem 2 suggests that degree d polynomials have vanishing advantage when $(\lambda^2 L)^d$ is small, which is consistent with the heuristic that one needs $d \gtrsim 1/(\log(\lambda^{-2}L^{-1}))$ to see signal when $\lambda^2 L < 1$. In the regime $\lambda^2 L \approx 1 - \varepsilon$ this heuristic would push d to be of order $1/\varepsilon$, whereas in regimes where $\lambda^2 L$ decays with n it suggests much larger degrees. An open problem is to identify explicit *constructive* polynomial statistics whose degree matches the low-degree prediction (up to constants) and to relate their computational cost to degree in a precise way. In other words, can one design an explicit family of degree-d tests that achieves the best possible low-degree advantage, and does this interpolate between polynomial time and subexponential time in a manner consistent with observed tradeoffs in other planted models?

Low-degree conjectures: scope and limits in coupled models. Our conditional hardness statement relies on the standard low-degree conjecture, but the frequency-coupled structure raises conceptual questions about how directly the conjecture applies. The conjecture is most compelling when the model has a single, relatively homogeneous observation tensor and when the low-degree expansion aligns with natural algorithmic hierarchies (e.g. statistical query lower bounds or sum-of-squares degree). Here the observation is a *collection* of matrices with shared latent structure, and an algorithm may exploit this structure through non-polynomial transformations (iterative normalization, adaptive conditioning across k, or other operations that are not transparently captured by a single low-degree polynomial).

One concrete direction is therefore to connect Theorem 2 to an explicit lower bound in an algorithmic hierarchy, such as sum-of-squares. Proving that SoS of degree $o(L)$ fails below $\lambda \lesssim L^{-1/2}$ would substantially strengthen the evidence for a genuine computational barrier. Conversely, finding an algorithm that succeeds below the low-degree threshold would be equally informative, as it would identify a mechanism (perhaps adaptive or non-polynomial) not captured by the low-degree proxy.

Extensions and robustness: what changes the $\lambda^2 L$ law? We have emphasized that the improvement with L comes from harmonic consistency rather than from having L independent replicas. It is natural to ask which perturbations preserve the $\lambda^2 L$ scaling and which destroy it. Examples include: mild dependence across frequencies $W^{(k)}$, non-Gaussian noise with matched variance, missing frequencies, or mismatched priors (e.g. x not exactly uniform on $\mathbb{Z}_L$). On the algorithmic side, FC-AMP is expected to be robust to some model mismatch, but quantifying this robustness at the threshold level (as opposed to constant-λ regimes) remains open. On the lower-bound side, extending growing-L Hermite-chaos arguments beyond i.i.d. Gaussian entries requires new ideas, since the orthogonal polynomial machinery is then less directly applicable.

Summary of open problems. We view the following as particularly crisp targets:

1. Determine the sharp constant threshold for Bayes-optimal detection and compare it to FC-AMP (constant optimality).
2. Prove (or refute) equivalence of detection and weak recovery at threshold $\lambda^2 L = 1$ in the growing-L regime.
3. Characterize exact/strong recovery thresholds as a function of (n, L, λ), and design polynomial-time methods that achieve them.
4. Develop a subexponential-time tradeoff theory for fixed L, analogous to known tradeoffs in tensor PCA.
5. Establish unconditional algorithmic lower bounds (e.g. SoS degree lower bounds) matching the low-degree picture, or exhibit algorithms that bypass it.

Resolving any of these would materially sharpen the emerging principle suggested by the present work: multi-frequency structure can lower computational thresholds, but the precise limits of this phenomenon depend delicately on what information can be coupled efficiently across harmonics.