This is the accepted manuscript made available via CHORUS. The article has been published as:

Exactly solvable model of randomly coupled twisted superconducting bilayers

Andrew C. Yuan

Phys. Rev. B **108**, 184515 — Published 28 November 2023

DOI: 10.1103/PhysRevB.108.184515

Exactly Solvable Model of Randomly Coupled Twisted Superconducting Bilayers

Andrew C. Yuan[1]

[1] *Department of Physics, Stanford University, Stanford, CA 94305, USA*

(Dated: October 16, 2023)

Motivated by recent experiments on twisted junctions of cuprate superconductors (SC), it was proposed [1] that at zero temperature, a random first order Josephson coupling $J_1(\mathbf{r})\cos\phi$ generates an "effective" global second order coupling, $J_2\cos(2\phi)$, with a sign that favors $\phi = \pm\pi/2$, i.e., spontaneous breaking of time reversal symmetry (TRS). To obtain a more controlled understanding of the suggested "disorder-induced-order" mechanism, we construct an exactly solvable lattice mean field model and prove that when the disorder-average $\bar{J}_1 = 0$, the model exhibits a TRS breaking phase for all temperatures below the SC transition, i.e., $T_c = T_{\mathrm{TRSB}}$, regardless of the specific form of disorder. In the presence of nonzero $\bar{J}_1 \neq 0$, we show that the two transitions split linearly for small $\bar{J}_1 \ll \kappa$ (where κ is the in-plane SC stiffness), and that T_{TRSB} vanishes for $\bar{J}_1 > J_c$ where $J_c = \overline{J_1^2}/\kappa$ in the weak disorder limit.

I. INTRODUCTION

Twisting and stacking 2D quantum materials have proven to be powerful tools for creating new materials [2–4] and probing their properties [5, 6]. In particular, the Josephson coupling between two d-wave superconductors has been studied as a function of the twist angle θ in an attempt to reveal their underlying pairing symmetry. Early experiments in cuprate superconductors did not show the expected angle dependence of the Josephson coupling [7]. However, in a beautiful recent set of experiments [8], an observed $|\cos(2\theta)|$ dependence of the Josephson coupling has been observed, potentially resolving the discrepancy.

Though still more recent experiments [9] have reproduced earlier results (with the Josephson coupling having no angle dependence), we conjecture that this discrepancy has an extrinsic origin and focus on the experiment showing the expected angle dependence. In this case, at $\theta = 45°$, the lowest-order Josephson coupling J_1 vanishes due to d-wave symmetry, but a substantial second-order coupling J_2 has been observed. Notably, such a phenomenon was predicted and studied by proposing an intrinsic 2nd order Joesphson coupling as the underlying mechanism [10]. Possible implications such as time-reversal symmetry breaking (TRSB), chiral superconductivity and gapped topological behavior near $\theta = 45°$ have aroused attention within the community [10–12].

Experimentally, the precise microscopic mechanism of the 2nd order coupling and the sign of J_2 are still under investigation. The extreme anisotropy of BSCCO (cuprates used in the experiments) implies that the expected magnitude of the intrinsic J_2 would be too small to explain observations [1]. Therefore, an alternative mechanism involving locally symmetry-breaking (nematic) inhomogeneities was proposed [1]. The local nematicity induces a spatially varying 1st order Josephson coupling, which can generate an effective J_2 with the required sign to favor TRSB, i.e., phase difference $\phi = \pm\pi/2$ between the two twisted superconductors. The magnitude of the effective J_2 is enhanced by a factor of ξ/ξ_{sc} where ξ is the disorder correlation length and ξ_{sc} is the superconducting coherence length, and thus for reasonably large ξ, the proposed mechanism is sufficiently large to explain the experimental observations.

It is worth mentioning that the proposal in Ref. [1] has a few restrictions due to the nature of the implemented perturbative approach. (1). It focuses on the zero temperature scenario where the in-plane nearest neighbor interaction can be approximated by the standard nonlinear sigma model (NLSM). Therefore, whether the TRSB behavior can be extended to finite temperatures, and more importantly, up to temperatures comparable to the SC transition T_c (to achieve high $T_{\mathrm{TRSB}} \approx T_c$ topological superconductivity), remains an open question. (2). The approach is only valid in the limit of weak (Gaussian) disorder. While this aligns with the physical conditions typically encountered in cases like twisted cuprates, it doesn't address the validity of similar conclusions in scenarios involving arbitrarily large (possibly non-Gaussian) disorder. Notably, the connection between the strength of disorder couplings and the magnitude of the SC gap in the TRSB phase suggests that robust topological superconductivity might necessitate substantial disorder interactions. (3). In the extreme limit in which each twisted superconductor is a single 2D layer, the perturbative approach also encounters difficulties due to a logarithmic divergence in the computation of the "effective" J_2[1]. Whether this affects the claim of disorder induced TRSB behavior also is yet to be determined.

In this paper, we will provide an alternative approach, and solve exactly a model with all-to-all interactions [13] for which mean-field theory is exact. The constructed model permits us to circumvent (ill-controlled) perturbative techniques and answer many of the open questions discussed previously, albeit we pay the price of a less physical model. Nevertheless, we prove that at twist

[1] This infrared divergence was treated using a variation mean field approach, in which an emergent length scale served as a physical cut-off.

angle $\theta_c = 45°$ so that $\bar{J}_1 = 0$, the model exhibits TRSB behavior for all temperatures below the superconducting transition, i.e., $T_c = T_{\text{TRSB}}$[2], regardless of the specific form and coupling strength of the disorder[3]. Near twist angle $\theta \approx 45°$ so that the disorder-average $\bar{J}_1 \neq 0$, we show that the two transitions split linearly for small $\bar{J}_1 \ll \kappa$ (where κ is the in-plance SC stiffness), with the TRSB phase eventually disappearing when $\bar{J}_1$ reaches $\bar{J}_1 = \overline{J_1^2}/\kappa$ in the limit of weak disorder $\overline{J_1^2} \to 0$. In the case of twisted BSCCO, due to the large anisotropy, this implies that topological superconductivity can only be achieved in a narrow region around the critical twist angle $\theta_c = 45°$.

The paper is organized as follows. Sec. (II) introduces the classical statistical all-to-all model under investigation as well as the motivation of its construction and its relation to the standard nearest-neighbor XY model. Sec. (III) computes the exact free energy of the all-to-all model and disucsses its inherent symmetries. The correspondence between spontaneous symmetry breaking (SSB) and the respective transition temperatures is made explicit. Sec. (IV) illustrates the perturbative results, which are well-controlled in given limits and independent of the specific form of the disorder. They are then compared with the corresponding numerical results shown in in Fig. IV. Finally, Sec. (V) proves the main statement, i.e., Theorem (V.2). The paper will only outline and cite the necessary theorems, while leaving the proof in the Appendix.

II. ALL-TO-ALL MODEL: MOTIVATION AND SETUP

We will represent the twist junction at twist-angle θ by a model Hamiltonian consisting of two classical XY ferromagnets representing the phase degrees of freedom of the individual superconducting planes with random-in sign local couplings between the two:

$$\mathcal{H} = -\kappa \sum_{\langle \mathbf{r}\mathbf{r}' \rangle, l} \sigma_l(\mathbf{r}) \cdot \sigma_l(\mathbf{r}') - \sum_{\mathbf{r}} J(\mathbf{r}) \sigma_1(\mathbf{r}) \cdot \sigma_2(\mathbf{r}) \quad (1)$$

where the system is defined on a square lattice, $\sigma_l(\mathbf{r}) \in \mathbb{S}^1$ corresponds to the superconducting phase $\phi_l(\mathbf{r})$ at site $\mathbf{r}$ and layer $l = \pm$, and the last term involving $\sigma_1(\mathbf{r}) \cdot \sigma_2(\mathbf{r}) = \cos\phi(\mathbf{r})$ (where $\phi = \phi_1 - \phi_2$) represents the first order Josephson coupling between the 2 layers. The disorder realization $\boldsymbol{J} = (J(\mathbf{r}) : \mathbf{r} \in \mathbb{Z}^2)$ satisfy $J(\mathbf{r}) = \bar{J}_\theta + \delta J(\mathbf{r})$, where $\bar{J}_\theta = \bar{J}_0 \cos(2\theta)$ and θ corresponds to the twist angle, and $\delta\boldsymbol{J} = (\delta J(\mathbf{r}) : \mathbf{r} \in \mathbb{Z}^2)$ are independently identically distributed (iid) *even*[4] random variables[5] such that the disorder average (denoted by $\overline{\cdots}$) satisfies $\overline{\delta J(\mathbf{r})} = 0$ and $\overline{(\delta J(\mathbf{r}))^2} = \varepsilon^2$. Note that the twist angle θ enters entirely through $\bar{J}_\theta$ (and thus will often be omitted).

Physically, the lattice spacing of our model corresponds to the spatial correlation length ξ of the disordered inter-layer Josephson coupling, whose length scale can be estimated from experimental data [1]. κ is the SC phase stiffness of an individual CuO_2 plane. Since the cuprate superconductors are highly anisotropic, both $\bar{J}$ and ε are much smaller than κ, with typical magnitudes discussed in Ref. [1], though we will not make this restriction in our analysis unless otherwise stated. This model resembles a random field XY model that was studied by Cardy and Ostlund [14], however it is distinct in the fact that the type of disorder we account for preserves time-reversal symmetry (TRS).

Due to the difficulty of treating the model introduced in Eq. (1), we shall instead consider a modified version of the model with all-to-all interactions in each plane for which many aspects of the system can be computed exactly from a mean-field approach. We thus define the "mean-field Hamiltonian,"

$$\mathcal{H}_{\text{MF}} = -\frac{d\kappa}{V} \sum_{\mathbf{r},\mathbf{r}',l} \sigma_l(\mathbf{r}) \cdot \sigma_l(\mathbf{r}') - \sum_{\mathbf{r}} J(\mathbf{r}) \sigma_1(\mathbf{r}) \cdot \sigma_2(\mathbf{r}) \quad (2)$$

where, in common with the model introduced in Ref. [13], the nearest-neighbor in-plane interaction in Eq. (1) is replaced by an infinite range coupling between in-plane spins, d is the dimension of the in-plane system[6], and V is the volume (number of lattice sites in a single layer). It should be emphasized, however, that the inter-layer interaction is kept local in space.

A. Relation of the Two Models

Consider the nearest-neighbor model in Eq. (1) and notice that the in-plane nearest-neighbor interaction between spins $\sigma \in \mathbb{S}^1$ can be rewritten as

$$\kappa \sum_{\langle \mathbf{r}\mathbf{r}' \rangle} \sigma_l(\mathbf{r}) \cdot \sigma_l(\mathbf{r}') = \frac{\kappa}{2} \sum_{\mathbf{r}} \sigma_l(\mathbf{r}) \cdot \sum_{\mathbf{r}' : \langle \mathbf{r}\mathbf{r}' \rangle} \sigma_l(\mathbf{r}') \quad (3)$$

$$= \kappa d \sum_{\mathbf{r}} \sigma_l(\mathbf{r}) \cdot \left[\frac{1}{2d} \sum_{\mathbf{r}' : \langle \mathbf{r}\mathbf{r}' \rangle} \sigma_l(\mathbf{r}') \right] \quad (4)$$

[2] The absence of vestigial TRSB above T_c will turn out to be trivial within the context of mean-field theory

[3] Similar exact statements can be made to the non-disordered inherent J_2 model with all-to-all interactions. In fact, the proof will be much simpler and we suggest the reader understand the non-disordered problem before moving on to the disordered problem. See Appendix (F) (or more specifically, Theorem (F.5)) for details.

[4] $\delta J(\mathbf{r})$ and $-\delta J(\mathbf{r})$ have the same distribution

[5] It should be emphasized that the disorder does not necessarily have to be of Gaussian form.

[6] In experiment, $d = 2$, but we leave it as a general parameter since our mean-field calculations hold for all d.

For every lattice site $\mathbf{r}$, its $2d$ nearest-neighbor $\mathbf{r}'$ spins generates an effective magnetic field which interacts with $\sigma(\mathbf{r})$. The mean field approximation we consider here is then replacing the nearest-neighbor field with a system-average magnetic field, i.e.,

$$\frac{1}{2d}\sum_{\mathbf{r}':\langle \mathbf{r}\mathbf{r}'\rangle}\sigma_l(\mathbf{r}') \mapsto \frac{1}{V}\sum_{\mathbf{r}'}\sigma_l(\mathbf{r}') \tag{5}$$

Where V is the system size, i.e., number of lattice sites. We thus obtain $\mathcal{H}_{\mathrm{MF}}$ in Eq. (2).

It should be emphasized that the two models (Eq. (1) and (2)) are only expected to coincide asymptotically when the effective number of neighboring spins is large as occurs in long-range interactions or in high dimensions ($d \to \infty$)[7], though, similar to any other mean-field theory, it is generally believed that the two share the same physics in $d \geq 4$ dimensions [13]. In lower dimensions $d = 2, 3$, fluctuations due to short-range interactions of Eq. (1) may be important and we comment more on this at the end of this paper.

III. EXACT FREE ENERGY IN THERMODYNAMIC LIMIT

Using a Hubbard-Stratonovich transformation (see Appendix (A)), we can obtain the free energy density of $\mathcal{H}_{\mathrm{MF}}$ as

$$\mathcal{F}_{\mathrm{MF}} = \lim_{V\to\infty}\overline{\frac{-1}{\beta V}\log \mathcal{Z}_{V,\boldsymbol{J}}} = \min_{w_\pm\in\mathbb{R}^2} F(w_+, w_-) \tag{6}$$

Where the minimizers $w_l^\star$ (the star * is to distinguish from arbitrary values w_l) correspond to the magnetizations[8] of each layer $l = \pm$, and

$$F = \frac{1}{4d\kappa}(w_+^2 + w_-^2) - \frac{1}{\beta}\overline{G_J} \tag{7}$$

$$G_J = \ln \iint d\sigma_+ d\sigma_- \; e^{-\beta H_J}, \tag{8}$$

$$-H_J = w_+\cdot\sigma_+ + w_-\cdot\sigma_- + J\sigma_+\cdot\sigma_- \tag{9}$$

Where H_J denotes a two-site Hamiltonian[9] with spins $\sigma_\pm \in \mathbb{S}^1$. To simplify notation, let us define φ as the phase difference between $w_\pm \in \mathbb{R}^2$ and

$$a^2 = \frac{1}{4}(w_+^2 + w_-^2), \quad \eta = \frac{2|w_+||w_-|}{w_+^2 + w_-^2} \tag{10}$$

Where $0 \leq \eta \leq 1$ is dimensionless and denotes the relative magnitudes between the two layer[10]. Then (see Appendix (B))

$$F = \frac{1}{d\kappa}a^2 - \frac{1}{\beta}\overline{G_J} \tag{11}$$

$$G_J = \ln\int I_0(2\beta a\sqrt{1+\eta\cos\phi})e^{\beta J\cos(\phi+\varphi)}d\phi \tag{12}$$

Where I_0 denotes the modified Bessel function.

A. Symmetries

Notice that F is independent of the average phase of $w_\pm \in \mathbb{R}^2$ and invariant under $\varphi \mapsto -\varphi$. Hence, F has an innate $U(1)\times\mathbb{Z}_2$ symmetry. A nonzero minimizer $a^\star > 0$ denotes a spontaneous symmetry breaking (SSB) in $U(1)$ and corresponds to the SC transition $T_{U(1)} \equiv T_c$. Similarly, a nonzero $\varphi^\star \neq 0$ denotes SSB in $\mathbb{Z}_2$ and corresponds to the TRSB transition, i.e., $T_{\mathbb{Z}_2} \equiv T_{\mathrm{TRSB}}$. Since the phase difference φ is only well-defined when $a > 0$, it follows that the critical temperature corresponding to each SSB satisfies $T_{U(1)} \geq T_{\mathbb{Z}_2}$ within mean-field theory. It is also worth mentioning that F has an additional $\mathbb{Z}_2$ symmetry regarding exchanging the layers $w_1 \leftrightarrow w_2$, and if broken, would correspond to a minimizer with $\eta^\star \neq 1$. However, we will show that $\eta^\star = 1$ always holds (mathematically, for $\bar{J} = 0$ and numerically for $\bar{J} \neq 0$). Hence, there is no confusion as to which order parameter $T_{\mathbb{Z}_2}$ refers to.

IV. PERTURBATIVE RESULTS

A. Near $T_{U(1)}$

Using a as a small parameter, F can be expanded near $T_{U(1)}$ so that

$$F = -c_2 a^2 + c_4\frac{1}{4}a^4 + O(a^6) \tag{13}$$

$$c_2 = \beta\left(1 + \eta\overline{r_1}\cos\varphi\right) - \frac{1}{d\kappa} \tag{14}$$

$$c_4 = \beta^3\Bigg[1 - \frac{1}{2}\eta^2 + \eta^2\overline{r_1^2} + 2\eta\overline{r_1}\cos\varphi + \eta^2\left(\overline{r_1^2} - \frac{\overline{r_2}}{2}\right)\cos 2\varphi\Bigg] \tag{15}$$

Where $r_\nu = I_\nu(\beta J)/I_0(\beta J)$ and I_ν are the modified Bessel functions.

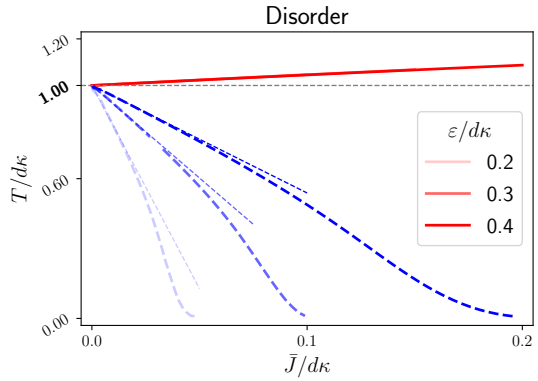
(i) If $\bar{J} > 0$ so that $\overline{r_1} > 0$, then the quadratic term is first term in the series expansion to contain dependence on φ, η so that F is minimized by

[7] A similar mean-field model (called the Curie-Weiss model) was constructed for the standard Ising model, and it was shown that the free energy density of the two models coincides asymptotically in the limit $d \to \infty$ [15].

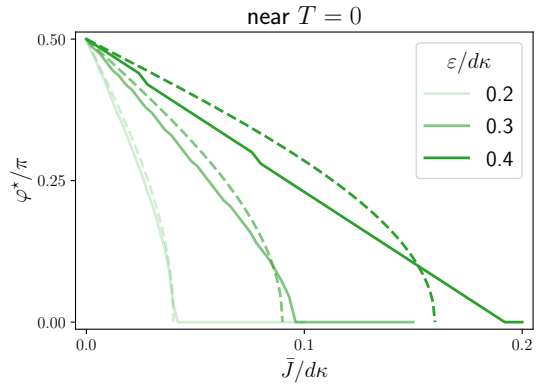
[8] The actual magnetizations of each layer have a 1-1 correspondence with the minimizers $w_l^\star$ (see Appendix (A)). A nontrivial magnetization in layer l implies $w_l^\star \neq 0$ and vice-versa.

[9] Notice that F can be regarded as the disordered-average free energy of $H_J + (w_+^2 + w_-^2)/4d\kappa$

[10] $\eta^\star = 1$ implies that the two layers have the same magnitude in magnetization $|w_+^\star| = |w_-^\star|$, while $\eta^\star = 0$ implies that only one of the layers is magnetized, e.g., $|w_+^\star| > |w_-^\star| = 0$.



(a)



(b)

Figure 1. Mean-Field. The numerical results assumes that the disorder δJ is of Gaussian distribution. This is the only time we assume an explicit distribution of δJ. (a). The solid/dashed lines represent the BKT/Ising transition temperatures $T_{U(1)}, T_{\mathbb{Z}_2}$. The shorter dashed lines denote the linear approximation of $T_{\mathbb{Z}_2}$ with respect to $\bar{J} = \bar{J}_\theta$. (b). The solid line denotes the minimizing $\varphi^\star$ computed at $\beta d\kappa = 100$, while the dashed lines denote the asymptotic limit as $\varepsilon \to 0^+$ given by Eq. (20).

$\eta^\star = 1, \varphi^\star = 0$ near $T_{U(1)}$. This indicates a split transition, i.e., $T_{U(1)} > T_{\mathbb{Z}_2}$.

(ii) Conversely, if $\bar{J} = 0$ so that $\overline{r_1} = 0$, the quartic terms must be considered so that F is minimized[11] by $\eta^\star = 1, \varphi^\star = \pm\pi/2$ near $T_{U(1)}$. Physically speaking, the disorder generates an "effective" $\cos 2\varphi$ term so that $T_{U(1)} = T_{\mathbb{Z}_2}$[12].

[11] Notice that $I_1^2 \geq I_0 I_2$ [16] and thus $r_1^2 > r_2/2$

[12] The series expansion indeed proves that TRSB behavior onsets precisely at the superconducting transition, provided that the transition is of second order, i.e., the minimzer $a^\star \to 0$ as $T \to T_{U(1)}$. However, it does not answer whether the TRSB behavior sustains at lower temperatures. Indeed, in general short-range lattice models (e.g., standard XY or Ising model), the Ginibre inequality [17] guarantees that (ferromagnetic) ordering increases as the temperature decreases. However, due to the long-range interactions of the current model, other formal techniques must be developed, which are postponed to Theorem. (V.2).

Similar to Ginzburg-Landau, the $U(1)$ transition occurs exactly when the quadratic coefficient c_2 transitions from a negative to a positive value, i.e., $c_2 = 0$, and thus $T_{U(1)}$ is determined self-consistently via finding the fixed point of

$$\hat{T} = 1 + \overline{r_1} \tag{16}$$

Where $\cdot\hat{\cdot}\cdot = \cdots/d\kappa$ is dimensionless[13].

Numerical calculations also suggest that $\eta^\star = 1$ for general $\bar{J} > 0$ and temperatures T, albeit we do not have a rigorous proof. Assuming $\eta^\star = 1$, the 2nd transition $T_{\mathbb{Z}_2}$ can then be determined exactly via the self-consistently equations (see Appendix (C)), i.e.,

$$|\hat{w}_\pm| = 2\overline{\langle \cos\phi_1 \rangle_J}, \quad \hat{T} = \overline{\langle (\sin\phi_+ - \sin\phi_-)^2 \rangle_J} \tag{17}$$

Where $\langle \cdots \rangle_J$ denotes the thermal average with respect to the two-site Hamiltonian H_J with $\varphi = 0$. Fig. 1a shows the 2 transitions (solid and dashed lines) $T_{U(1)}, T_{\mathbb{Z}_2}$ as a function of the dimensionless quantities $\hat{\bar{J}}$ and $\hat{\varepsilon}$. The smaller dashed lines show the linear approximation of the transition temperatures near $\hat{T} \approx 1$ and with $\bar{J}/\varepsilon \to 0$ (see Appendix (D)), that is,

$$\hat{T}_{U(1)} = 1 + \frac{1}{2}\hat{\bar{J}}\left[1 - \frac{\hat{\varepsilon}^2}{2^2} + O(\hat{\varepsilon}^4)\right] + O(\hat{\bar{J}}^2) \tag{18}$$

$$\hat{T}_{\mathbb{Z}_2} = 1 - \hat{\bar{J}}\left[\frac{2}{3}\frac{1}{\hat{\varepsilon}^2} + O(1)\right] + O(\hat{\bar{J}}^2) \tag{19}$$

Notice that the slope of $\hat{T}_{\mathbb{Z}_2}$ with respect to $\bar{J}$ is $\propto 1/\hat{\varepsilon}^2$. This implies that as the disorder strength $\hat{\varepsilon} \to 0$, the region of $\mathbb{Z}_2$ SSB becomes extremely narrow in $\bar{J}$, despite $T_{\mathbb{Z}_2} = T_{U(1)}$ when $\bar{J} = 0$.

B. Near $T = 0$ and small disorder $\varepsilon \to 0^+$

On a first glance of the two-site Hamiltonian H_J in Eq. (9), one would naively attempt to obtain near $T = 0$ results by using the nonlinear sigma model (NLSM)[14] with respect to spins $\sigma_\pm \in \mathbb{S}^1$. However, due to disorder J, the expansion near $T = 0$ will diverge as $a \to 0^+$ and thus cannot be readily minimized (see Appendix (E)). However, in the limit of small disorder $\varepsilon \to 0^+$, we can safely assume that the minimizer $a^\star(\varepsilon)$ is sufficiently close to the non-disordered minimizer $a^\star(\varepsilon = 0) \neq 0$, and thus the divergence of F near $a \to 0^+$ does not affect us. Therefore, we find that (see Appendix (E)) in the limit $\bar{J}/\varepsilon \to 0$,

$$\cos\varphi^\star = \hat{\bar{J}}/\hat{\varepsilon}^2 \tag{20}$$

[13] Notice that in the absence of disorder, $T_{U(1)} = d\kappa$ and thus warrants the definition of dimensionless quantities

[14] Approximating $\cos\theta \approx 1 - \theta^2/2$ in the exponent to obtain a Gaussian measure, which in our case, is only valid when $\beta a \gg 1$.

In particular, if $\hat{\bar{J}} \leq \hat{\varepsilon}^2$, then the system exhibits $\mathbb{Z}_2$ SSB near $T = 0$, where $\varphi^\star = \pm\pi/2$ when $\bar{J} = 0$. Indeed, Fig. 1b compares the asymptotic limit (dashed lines) with the numerical calculations (solid) for different disorder strength ε.

V. GENERAL T

We emphasize that beyond special limits, non-perturbative techniques are essential in obtaining exact result and deeper insights to strong interactions. Indeed, for general temperatures T (and disorder strength ε), one can prove rigorously that[15]

Theorem V.1 (Orientation, see (F.1))**.** *If $\bar{J} \geq 0$, then F is minimized when $|\varphi^\star| \leq \pi/2$. Equivalently, F is minimized when $\bar{J}\cos\varphi^\star \geq 0$.*

Theorem V.2 (TRSB, see (F.2))**.** *If $\bar{J} = 0$, then F is minimized via $\varphi^\star = \pm\pi/2$ and $\eta^\star = 1$*

Where the proof only relies on the probability distribution of δJ being even (see Appendix (F) for details). Due to the definitions of a, φ, η, it should be noted that φ, η are only well-defined when $a > 0$, and thus the previous theorems implicitly assume that $a > 0$.

Theorem (V.1) is the formal statement which confirms the physical expectation that the system wishes to minimize the Josephson coupling in Eq. (2). More specifically, let us naively replace the disorder term in Eq. (2) with its average value $J(\mathbf{r}) \mapsto \bar{J}$ so that the average value generates a 1st order Josephson $J_1 \sim \bar{J}$ coupling. On the other hand, the disorder $\delta J(\mathbf{r})$ is expected to generate an "effective" 2nd order Josephson coupling term J_2 with a sign that favors a phase difference $\phi = \pm\pi/2$ (as described in previous work [1] near zero temperature). In this naive setup, the interaction between the two layers is described by a competition between the 1st and 2nd order Josephson coupling, schematically written as $-\bar{J}\cos\phi + J_2\cos 2\phi$. If $J_1 > 0$ (< 0, respectively), then we would expect the ground state phase difference to be $|\phi| \leq \pi/2$ ($\geq \pi/2$).

More importantly, theorem (V.2) proves the existence of a TRSB phase generated by disorder, indicating that the subtle logarithm divergence seen in Ref. [1] may not affect the overall conclusion. In fact, the TRSB phase extends up to the superconducting transition, which suggests the possibility of disorder induced high T_c topological superconductivity.

[15] We note that the formal techniques developed here can also be used to show that the analagous all-to-all model with an inherent non-disordered $J_2\cos 2\phi$ inter-layer interaction has the same properties. In fact, the proof is much simpler than the disordered case and we suggest that the reader understand the non-disordered problem before moving on to the disordered problem. See Appendix (F) (or more specifically, Theorem (F.5)) for details.

VI. SUMMARY AND DISCUSSION

The paper presents a model that is exactly solvable for all-to-all interactions. Notably, it introduces rigorous methods to demonstrate the presence of a disorder-induced phase with broken time-reversal symmetry (TRSB)[16]. When $\bar{J} = 0$, we prove that the TRSB phase persists up to the superconducting transition temperature ($T_{\text{TRSB}} = T_c$), independent of the specific disorder characteristics. The same techniques are applied to analyze the non-disordered (inherent J_2) problem with vanishing 1st order Josephson coupling $J_1 = 0$ [10], yielding analogous conclusions (see Theorem (F.5)). Consequently, the developed framework could serve as a formal approach to validate results anticipated by Ginzburg-Landau theory in closely related scenarios, such as those referenced in [18, 19], and possibly in a more general setting as well.

We also note that while twisted bilayer superconductors [9] were the motivation for the proposed exactly solvable model, similar constructions can potentially be applied in a boarder context, since the model is not concerned with the microscopic details. Indeed, effective Josephson couplings incuded by frustration from three (or more) coupled bands [18–23], inter-band scattering within a 2-band system [24] or boundary effects [25, 26] have been previously proposed as possible mechanisms of generating TRSB behavior. Consequently, one would anticipate that disorder can induce similar phenomena within such proposals, as demonstrated in this study focusing on the bilayer system.

It should be noted that the mean-field model is only quantitatively reliable when the effective number of neighboring spins is large as occurs when there are long-range interactions or in high dimensions ($d \to \infty$) [15]. In low dimensions ($d = 2, 3$), fluctuations due to short-range interactions may induce distinct statistical behavior, e.g., such as the possibility of a *vestigial* TRSB phase, i.e., $T_{\text{TRSB}} > T_c$ [27]. Indeed, a similar possibility was suggested for the non-disorder inherent J_2 problem by applying an RG scheme [28], though further research (both numerical [29] and mathematical [30]) seems to suggest otherwise.

[16] The physical interpretation of this result, as discussed previously in section (V), can be seen in the following manner. The disordered average $\bar{J}$ generates a 1st order Josephson coupling $J_1 \sim \bar{J}$, while the disorder term $\delta J(\mathbf{r})$ is expected to generate an "effective" 2nd order Josephson coupling $J_2 \sim \varepsilon^2$ with a sign that favors a phase difference $\phi = \pm\pi/2$ (as described in previous work [1] near zero temperature). Taken together, the interlayer interaction is schematically related to the non-disorder problem, i.e., $-J_1\cos\phi + J_2\cos 2\phi$. Note that within the context of Ginzburg Landau theory, where $\psi_\pm \in \mathbb{C}$ are the Landau order parameters for each layer $l = \pm$ [10], only the quadratic terms determine the critical transition temperature. Since the J_1 term is quadratic (schematically written as $\psi_+^\dagger\psi_- + \text{h.c.}$) and the J_2 term is quartic (i.e., $(\psi_+^\dagger\psi_-)^2 + \text{h.c.}$), we see that the two transitions must coincide at $J_1 = 0$, and split when $J_1 \neq 0$.

Another worry would be due to the Imry-Ma argument [31], which precludes long-range ordering of a continuous symmetry in the presence of disorder in low dimensions. However, the Imry-Ma argument considers the scenario where the disorder has the same degree of freedom as the spins (e.g., if $\sigma \in \mathbb{S}^1$, then so is the disorder field), while the disorder considered in this paper is "effectively" within a hyper-plane of the continuous $U(1)$ symmetry, i.e., the disorder term $J(\mathbf{r})$ is coupled to the phase difference ϕ, and acts along the x-axis via the interaction $J(\mathbf{r})\cos\phi$, while ϕ is in $\mathbb{S}^1 \cong U(1)$. The difference can result in distinct statistical behavior as discussed in Ref. [32, 33], in which it was proved mathematically that the system exhibits long-range order for sufficiently low temperatures and weak disorder. This provides confidence that the certain aspects of the mean-field phenomena will extend to low dimensions (e.g., the existence of TRSB), circumventing the standard Imry-Ma argument.

VII. ACKNOWLEDEGEMENTS

I am grateful for Steve A. Kivelson's support and generosity during this project and also for providing extensive comments and suggestions on the draft. This work was supported, in part, by NSF Grant No. DMR-2000987 at Stanford University.

[1] A. C. Yuan, Y. Vituri, E. Berg, B. Spivak, and S. A. Kivelson, Inhomogeneity-induced time-reversal symmetry breaking in cuprate twist junctions, Phys. Rev. B **108**, L100505 (2023).

[2] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, *et al.*, Correlated insulator behaviour at half-filling in magic-angle graphene superlattices, Nature **556**, 80 (2018).

[3] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Unconventional superconductivity in magic-angle graphene superlattices, Nature **556**, 43 (2018).

[4] M. Yankowitz, S. Chen, H. Polshyn, Y. Zhang, K. Watanabe, T. Taniguchi, D. Graf, A. F. Young, and C. R. Dean, Tuning superconductivity in twisted bilayer graphene, Science **363**, 1059 (2019).

[5] H. Polshyn, M. Yankowitz, S. Chen, Y. Zhang, K. Watanabe, T. Taniguchi, C. R. Dean, and A. F. Young, Large linear-in-temperature resistivity in twisted bilayer graphene, Nature Physics **15**, 1011 (2019).

[6] A. Inbar, J. Birkbeck, J. Xiao, T. Taniguchi, K. Watanabe, B. Yan, Y. Oreg, A. Stern, E. Berg, and S. Ilani, The quantum twisting microscope, Nature **614**, 682 (2023).

[7] Q. Li, Y. Tsay, M. Suenaga, R. Klemm, G. Gu, and N. Koshizuka, $Bi_2Sr_2CaCu_2O_8 + \delta$ bicrystal c-axis twist Josephson junctions: a new phase-sensitive test of order parameter symmetry, Physical review letters **83**, 4160 (1999).

[8] S. Zhao, N. Poccia, X. Cui, P. Volkov, H. Yoo, R. Engelke, Y. Ronen, R. Zhong, G. Gu, S. Plugge, *et al.*, Emergent interfacial superconductivity between twisted cuprate superconductors, in *THE 12TH INTERNATIONAL CONFERENCE ON INTRINSIC JOSEPHSON EFFECT AND HORIZONS OF SUPERCONDUCTING SPINTRONICS ABSTRACT BOOK* (2021) p. 73.

[9] Y. Zhu, H. Wang, Z. Wang, S. Hu, G. Gu, J. Zhu, D. Zhang, and Q.-K. Xue, Persistent Josephson tunneling between $Bi_2Sr_2CaCu_2O_{8+x}$ flakes twisted by 45° across the superconducting dome, arXiv e-prints , arXiv:2301.03838 (2023), arXiv:2301.03838 [cond-mat.supr-con].

[10] O. Can, T. Tummuru, R. P. Day, I. Elfimov, A. Damascelli, and M. Franz, High-temperature topological superconductivity in twisted double-layer copper oxides, Nature Physics **17**, 519 (2021).

[11] T. Tummuru, S. Plugge, and M. Franz, Josephson effects in twisted cuprate bilayers, Physical Review B **105**, 064501 (2022).

[12] A. Mercado, S. Sahoo, and M. Franz, High-temperature majorana zero modes, Physical Review Letters **128**, 137002 (2022).

[13] S. Friedli and Y. Velenik, *Statistical mechanics of lattice systems: a concrete mathematical introduction* (Cambridge University Press, 2017).

[14] J. L. Cardy and S. Ostlund, Random symmetry-breaking fields and the XY model, Phys. Rev. B **25**, 6899 (1982).

[15] C. J. Thompson, Ising model in the high density limit, Communications in Mathematical Physics **36**, 255 (1974).

[16] V. Thiruvenkatachar and T. Nanjundiah, Inequalities concerning Bessel functions and orthogonal polynomials, in *Proceedings of the indian academy of sciences-section a*, Vol. 33 (Springer India New Delhi, 1951) p. 373.

[17] J. Ginibre, General formulation of griffiths' inequalities, Communications in mathematical physics **16**, 310 (1970).

[18] T. A. Bojesen, E. Babaev, and A. Sudbø, Phase transitions and anomalous normal state in superconductors with broken time-reversal symmetry, Physical Review B **89**, 104509 (2014).

[19] I. Maccari and E. Babaev, Effects of intercomponent couplings on the appearance of time-reversal symmetry breaking fermion-quadrupling states in two-component london models, Physical Review B **105**, 214520 (2022).

[20] W.-C. Lee, S.-C. Zhang, and C. Wu, Pairing state with a time-reversal symmetry breaking in feas-based superconductors, Phys. Rev. Lett. **102**, 217002 (2009).

[21] D. F. Agterberg, V. Barzykin, and L. P. Gor'kov, Conventional mechanisms for exotic superconductivity, Phys. Rev. B **60**, 14868 (1999).

[22] T. K. Ng and N. Nagaosa, Broken time-reversal symmetry in josephson junction involving two-band superconductors, Europhysics Letters **87**, 17003 (2009).

[23] V. Stanev and Z. Tešanović, Three-band superconductivity and the order parameter that breaks time-reversal symmetry, Phys. Rev. B **81**, 134522 (2010).

[24] A. Bobkov and I. Bobkova, Time-reversal symmetry breaking state near the surface of an s±superconductor, Physical Review B **84**, 134527 (2011).
[25] S. R. Bahcall, Boundary effects and the order parameter symmetry of high-t c superconductors, Physical review letters **76**, 3634 (1996).
[26] D. Rainer, H. Burkhardt, M. Fogelström, and J. Sauls, Andreev bound states, surfaces and subdominant pairing in high tc superconductors, Journal of Physics and Chemistry of Solids **59**, 2040 (1998).
[27] A.C. Yuan, Y. Vituri, E. Berg, B. Spivak and S. Kivelson, Vestigial Time-reversal Symmetry Breaking in Randomly Coupled Superconducting Bilayers. To appear.
[28] M. Zeng, L.-H. Hu, H.-Y. Hu, Y.-Z. You, and C. Wu, Phase-fluctuation induced time-reversal symmetry breaking normal state, arXiv preprint arXiv:2102.06158 (2021).
[29] F.-F. Song and G.-M. Zhang, Phase coherence of pairs of cooper pairs as quasi-long-range order of half-vortex pairs in a two-dimensional bilayer system, Physical Review Letters **128**, 195301 (2022).
[30] A. C. Yuan, Absence of Vestigial Time-reversal Symmetry Breaking above T_c in the Strong-coupling Limit of a Twisted Bilayer Superconductor., arXiv preprint arXiv:2308.06988 10.48550/arXiv.2308.06988 (2023).
[31] Y. Imry and S.-k. Ma, Random-field instability of the ordered state of continuous symmetry, Physical Review Letters **35**, 1399 (1975).
[32] N. Crawford, On random field induced ordering in the classical xy model, Journal of Statistical Physics **142**, 11 (2011).
[33] N. Crawford, Random field induced order in low dimension, Europhysics Letters **102**, 36003 (2013).

Appendix A: Hubbard-Stratonovich Transform and the Free Energy

In our main text, we assumed that $J(\mathbf{r})$ are iid (independent identically distributed) even random variables with mean $\bar{J}$ and variance $\overline{(\delta J)^2}$. However, since our results will not depend on the specific form of the distribution, in this section, it's instructive to consider the case where δJ is δ-distributed, that is, $\delta J = \pm\varepsilon$[17] where $\varepsilon > 0$ and the $\pm$ sign is chosen randomly at each lattice site $\mathbf{r}$. The generalization to even probability distributions $\mathbb{P}$ (i.e., $\mathbb{P}[\delta J] = \mathbb{P}[-\delta J]$ for all values of δJ) is straightforward and we will comment on the extension in end of this section, i.e., Appendix (A 1). To simplify notation, we will also use $\mathbb{E}_{\delta J}[\cdots]$ to denote averaging over the disorder δJ.

The Hubbard-Stratonovich (HS) transform is an exact transformation using the following fact [13]

$$\exp\left(ax^2\right) = \frac{1}{\sqrt{\pi a}} \int_{\mathbb{R}} \exp\left(-\frac{y^2}{a} + 2xy\right) dy \tag{A1}$$

In particular, we find that

$$\exp\left(\frac{\beta d\kappa}{V}\left(\sum_{\mathbf{r}} \sigma_l(\mathbf{r})\right)^2\right) = \frac{\beta V}{4\pi d\kappa} \int_{\mathbb{R}^2} dw_l \exp\left[-\frac{\beta V}{4d\kappa} w_l^2 + \beta w_l \cdot \sum_{\mathbf{r}} \sigma_l(\mathbf{r})\right] \tag{A2}$$

Notice that by using the HS transform, we obtain a term linear in $\sigma_l(\mathbf{r})$, which implies that the exponential term can be decomposed into the product of local functions $\prod_{\mathbf{r}} f_l(\sigma_l(\mathbf{r}))$, each depending on the spin configuration $\boldsymbol{\sigma}_l \equiv (\sigma_l(\mathbf{r}) : \mathbf{r} \in V)$ implicitly through $\sigma_l(\mathbf{r})$ at lattice site $\mathbf{r}$. Since we intend to integrate over all spin configurations, i.e., $\sigma_l(\mathbf{r}) \in \mathbb{S}^1$ for all $\mathbf{r}$, we can interchange the order of product and integration, i.e., schematically,

$$\int_{(\mathbb{S}^1)^V} d\boldsymbol{\sigma}_l \prod_{\mathbf{r}} f_l(\sigma_l(\mathbf{r})) = \left[\int_{\mathbb{S}^1} d\sigma_l f(\sigma_l)\right]^V, \quad d\boldsymbol{\sigma}_l \equiv \prod_{\mathbf{r}} d\sigma_l(\mathbf{r}) \tag{A3}$$

More specifically, for a given disorder realization $\boldsymbol{J} \equiv (J(\mathbf{r}) : \mathbf{r} \in V)$,

$$Z_{V,\boldsymbol{J}} = \left(\frac{\beta V}{4\pi d\kappa}\right)^2 \iint_{\mathbb{R}^2\times\mathbb{R}^2} dw_\pm \exp\left[-\frac{\beta V}{4d\kappa}(w_+^2 + w_-^2)\right] \prod_{\mathbf{r}} \iint_{\mathbb{S}^1\times\mathbb{S}^1} d\sigma_\pm \exp\left[\beta \sum_l w_l \cdot \sigma_l + J(\mathbf{r})\sigma_+ \cdot \sigma_-\right] \tag{A4}$$

By using the HS transform, the lattice site $\mathbf{r}$ dependence of the spins σ_l inside the integral disappears and thus what remains is the possible $\mathbf{r}$ dependence of the inter-layer interaction $J(\mathbf{r})$. Let $V_\pm$ denote the number of lattice sites $\mathbf{r}$

[17] Notice that in the main text, we used ε^2 to denote the disorder strength $\overline{(\delta J)^2}$. Therefore, if we wish to be consistent, we would require that $\delta J = \pm\varepsilon/\sqrt{2}$. However, since this is a constant factor which will not affect our analysis, we shall adopt the simpler notation and omit the $1/\sqrt{2}$.

with $\delta J(\mathbf{r}) = \pm\varepsilon$ so that $V_+ + V_- = V$ and let $\rho(\delta \boldsymbol{J}) = V_+/V$ so that

$$Z_{V,\boldsymbol{J}} = \left(\frac{\beta V}{4\pi d\kappa}\right)^2 \iint_{\mathbb{R}^2\times\mathbb{R}^2} dw_\pm \exp\left[-\frac{\beta V}{4d\kappa}(w_+^2 + w_-^2) + V_+ G_{\bar{J}+\varepsilon}(w_+, w_-) + V_- G_{\bar{J}-\varepsilon}(w_+, w_-))\right] \tag{A5}$$

$$= \left(\frac{\beta V}{4\pi d\kappa}\right)^2 \iint_{\mathbb{R}^2\times\mathbb{R}^2} dw_\pm \exp\left[-\beta V \psi_{\bar{J},\rho(\delta\boldsymbol{J})}(w_+, w_-)\right] \tag{A6}$$

Where $G_{J=\bar{J}\pm\varepsilon}$ is defined using the two-site Hamiltonian in Eq. (9) and

$$\psi_{\bar{J},\rho(\delta\boldsymbol{J})}(w_+, w_-) = \frac{1}{4d\kappa}(w_+^2 + w_-^2) - \frac{1}{\beta}\left(\rho G_{\bar{J}+\varepsilon} + (1-\rho)G_{\bar{J}-\varepsilon}\right) \tag{A7}$$

Therefore, we see that the partition function $Z_{V,\boldsymbol{J}}$ only depends on $\boldsymbol{J} = (J(\mathbf{r}) : \mathbf{r} \in V)$ implicitly via the average value $\bar{J}$ and $\rho(\delta\boldsymbol{J})$ and thus warrants the notation $Z_{V,\boldsymbol{J}} = Z_{V,\bar{J},\rho(\delta\boldsymbol{J})}$. Notice that $\rho(\delta\boldsymbol{J})$ can be rewritten as

$$\rho(\delta\boldsymbol{J}) = \frac{1}{V}\sum_{\mathbf{r}} \frac{\delta J(\mathbf{r}) + \varepsilon}{2\varepsilon} \tag{A8}$$

By the central limit theorem, $\rho(\delta J) \to \mathbb{E}_{\delta\boldsymbol{J}}[\rho(\delta\boldsymbol{J})] = 1/2$ as $V \to \infty$. Therefore, we will consider the partition function $Z_{V,\bar{J},\rho=1/2}$ defined with $\psi_{\bar{J},\rho=1/2}$, and show that the free energies of the 2 partition functions must converge in the thermodynamic limit $V \to \infty$. More specifically, we shall prove the following.

Theorem A.1 (Free Energy)**.**

$$\lim_{V\to\infty} \mathbb{E}_{\delta\boldsymbol{J}} \frac{-1}{\beta V} \log Z_{V,\bar{J},\rho(\delta\boldsymbol{J})} = \lim_{V\to\infty} \frac{-1}{\beta V} \log Z_{V,\bar{J},\rho=1/2} = \min_{w_\pm\in\mathbb{R}^2} \psi_{\bar{J},\rho=1/2}(w_+, w_-) \tag{A9}$$

We leave the proof of this main statement to the end of this section. Instead, let us try to understand why the statement "should" be true and its physical implications. Indeed, when considering $Z_{V,\bar{J},\rho=1/2}$, due to the exponential weight $\propto e^{-\beta V \psi_{\bar{J},\rho=1/2}}$, at least schematically, only the minimizers $w_\pm^\star$ of $\psi_{\bar{J},\rho=1/2}$ should contribute to the free energy in the thermodynamic limit $V \to \infty$. A similar logic applies when computing the average magnetization of each layer $l = \pm$, i.e., by taking the derivative of the weight $e^{-\beta V \psi_{V,\bar{J},\rho(\delta\boldsymbol{J})}}$ with respect to w_l, we obtain

$$\left\langle \frac{1}{V}\sum_{\mathbf{r}} \sigma_l(\mathbf{r}) \right\rangle_{V,\boldsymbol{J}} = \frac{1}{2d\kappa}\langle w_l \rangle_{V\psi_{\bar{J},\rho(\delta\boldsymbol{J})}} \tag{A10}$$

Where the right-hand-side the average with respect to the measure $\propto e^{-\beta V\psi}$. In the thermodynamic limit $V \to \infty$, only the minimizer $w_\pm^\star$ will contribute to the average on the right-hand-side, and thus at least schematically,

$$\lim_{V\to\infty} \mathbb{E}_{\boldsymbol{J}} \left\langle \frac{1}{V}\sum_{\mathbf{r}} \sigma_l(\mathbf{r}) \right\rangle_{V,\boldsymbol{J}} = \frac{1}{2d\kappa} w_l^\star \tag{A11}$$

Technically, due to exact $U(1)$ symmetry of the mean-field model, the average magnetization $\langle m \rangle = 0$. Rather, we should compute the average squared magnetization $\langle m^2 \rangle$ to account for $U(1)$ symmetry. However, the reasoning is the same and thus we see that there is a 1-1 correspondence between the minimizers $w_l^\star$ and the magnetization of each layer $l = \pm$, as we claimed in the main text.

The rest of this subsection will be contributed to proving Theorem (A.1), where the proof involves 2 steps. The first step is that the free energy corresponding to $Z_{V,\bar{J},\rho=1/2}$ converges and has the limit given on the right-hand-side (RHS). The second step is to show the first equality, i.e., the two partition functions have the same limit. We will also need the following lemma, obtained from Exercise 2.6 in Ref. [13], which we repeat here (albeit slightly modified for our case) for completeness.

Lemma A.2 (Exercise 2.6 in Ref. [13])**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be analytic and bounded below by $O(x^2)$ for sufficiently large x. Then*

$$\lim_{N\to\infty} \sqrt{N} \int_{-\infty}^{\infty} e^{-N(f(x) - \min f(x))dx} > 0 \tag{A12}$$

Proof. Since this was left as an exercise in the reference, we shall provide a short proof here for completeness. Indeed, since $f(x)$ is analytic and nontrivial, we see that its minimums cannot have a limit. Since $f \to \infty$ at large x, we see that it can only have a finite number of minima. Therefore, without loss of generality, we shall assume that f has a unique minimum at $x = 0$ and take $f(0) = 0$. Without loss of generality, we shall also assume that $f''(0) > 0$, since if not, we can always take the lowest nonzero coefficient (which must be of even order of derivation since it is a min) in the Taylor series. By Taylor series, we see that for sufficiently small $\epsilon > 0$, there exists $c(\epsilon), C(\epsilon) > 0$ such that

$$c(\epsilon) \leq \frac{f(x)}{x^2} \leq C(\epsilon), \quad |x| \leq \epsilon \tag{A13}$$

Notice that we can take $c(\epsilon), C(\epsilon) \to f''(0)/2$ as $\epsilon \to 0$. There also exists R and M such that if $|x| \geq R$, then $f(x) \geq Mx^2$. Hence, we see that

$$\sqrt{N} \int e^{-Nf(x)} dx = \int e^{-Nf(x/\sqrt{N})} dx \tag{A14}$$

$$= \int_{|x| \leq \epsilon\sqrt{N}} e^{-Nf(x/\sqrt{N})} + \int_{\epsilon\sqrt{N} < |x| \leq R\sqrt{N}} e^{-Nf(x/\sqrt{N})} + \int_{|x| > R\sqrt{N}} e^{-Nf(x/\sqrt{N})} \tag{A15}$$

otice that

$$\int_{|x| > R\sqrt{N}} e^{-Nf(x/\sqrt{N})} dx \leq \int_{|x| > R\sqrt{N}} e^{-Mx^2} \to 0, \quad N \to \infty \tag{A16}$$

Since f has a unique min valued $\min f = 0$ at $x = 0$, we see that there exists $g(\epsilon)$ such that $f(x) \geq g(\epsilon) > 0$ for $|x| > \epsilon$ and thus

$$\int_{\epsilon\sqrt{N} < |x| \leq R\sqrt{N}} e^{-Nf(x/\sqrt{N})} dx \leq e^{-Ng(\epsilon)} \sqrt{N} \times (R - \epsilon) \to 0, \quad N \to \infty \tag{A17}$$

Hence, only the first term will contribute to the limit. Indeed, notice that

$$\int_{|x| \leq \epsilon\sqrt{N}} e^{-Nf(x/\sqrt{N})} \leq \int_{|x| \leq \epsilon\sqrt{N}} e^{-c(\epsilon)x^2} dx \leq \sqrt{\frac{\pi}{c(\epsilon)}} \tag{A18}$$

And similarly,

$$\int_{|x| \leq \epsilon\sqrt{N}} e^{-Nf(x/\sqrt{N})} \geq \int_{|x| \leq \epsilon\sqrt{N}} e^{-C(\epsilon)x^2} dx \tag{A19}$$

$$\liminf_{N \to \infty} \int_{|x| \leq \epsilon\sqrt{N}} e^{-Nf(x/\sqrt{N})} \geq \sqrt{\frac{\pi}{C(\epsilon)}} \tag{A20}$$

Therefore, we have

$$\sqrt{\frac{\pi}{C(\epsilon)}} \leq \liminf_{N \to \infty} \int e^{-Nf(x/\sqrt{N})} \leq \limsup_{N \to \infty} \int e^{-Nf(x/\sqrt{N})} \leq \sqrt{\frac{\pi}{c(\epsilon)}} \tag{A21}$$

Take $\epsilon \to 0$ and we see that the statement follows. □

Proof of Theorem (A.1). Notice that the first step is nothing but a multi-dimensional application of the previous Lemma (A.2). Indeed, F is analytic with respect to $|w|^2, \varphi, \eta$ and thus

$$\lim_{V \to \infty} \frac{-1}{\beta V} \log Z_{V,\bar{J},\rho=1/2} = \min_{w_\pm \in \mathbb{R}^2} \psi_{\bar{J},\rho=1/2}(w_+, w_-) \tag{A22}$$

Let us now attempt to show the 1st equality. Indeed, notice that

$$Z_{V,\bar{J},\rho(\delta J)} = Z_{V,\bar{J},\rho=1/2} \left\langle e^{-\beta V(\psi_{\bar{J},\rho(\delta J)} - \psi_{\bar{J},\rho=1/2})} \right\rangle_{V,\bar{J},\rho=1/2} \tag{A23}$$

Where the $\langle \cdots \rangle$ is the average with respect to the weight $\propto e^{-\beta V \psi_{\bar{J},\rho=1/2}(w_+,w_-)}$, integrated over $w_\pm$, and normalized via $Z_{V,\bar{J},\rho=1/2}$. Therefore,

$$\frac{1}{V}|\log Z_{V,\bar{J},\rho(\delta J)} - \log Z_{V,\bar{J},\rho=1/2}| \leq \frac{1}{V}\left|\log \left\langle e^{-\beta V(\psi_{\bar{J},\rho(\delta J)} - \psi_{\bar{J},\rho=1/2})}\right\rangle_{V,\bar{J},\rho=1/2}\right| \tag{A24}$$

Notice that there exists some constant $C > 0$ such that

$$|\psi_{\bar{J},\rho(\delta J)} - \psi_{\bar{J},\rho=1/2}| \leq \frac{C}{\beta}\left|\rho(\delta J) - \frac{1}{2}\right| \tag{A25}$$

Hence,

$$\frac{1}{V}|\log Z_{V,\bar{J},\rho(\delta J)} - \log Z_{V,\bar{J},\rho=1/2}| \leq C\left|\rho(\delta J) - \frac{1}{2}\right| \tag{A26}$$

Therefore, we see that after disorder averaging, we have

$$\mathbb{E}_{\delta J}\frac{1}{V}|\log Z_{V,\bar{J},\rho(\delta J)} - \log Z_{V,\bar{J},\rho=1/2}| \leq C\mathbb{E}_{\delta J}\left|\rho(\delta J) - \frac{1}{2}\right| \to 0, \quad V \to \infty \tag{A27}$$

Where the limit follows from the central limit theorem. Therefore, we see that the statement follows. □

1. Extension to Arbitrary (Even) Distributions of δJ

We comment that when δJ is δ-distributed, i.e., $\delta J = \pm\varepsilon$, we see that $\psi_{\bar{J},\rho=1/2}$ in Theorem (A.1) is equal to F given in Eq. (7). In fact, F is the natural extension to arbitrary even probability distributions of δJ. More specifically, in the case of an arbitrary even probability distribution $\mathbb{P}[\delta J] = \mathbb{P}[-\delta J]$, notice that $V_+ G_{\bar{J}+\varepsilon} + V_- G_{\bar{J}-\varepsilon}$ in Eq. (A5) is replaced schematically[18] with

$$\sum_{\delta J} V_{\delta J} G_{\bar{J}+\delta J} \tag{A28}$$

where the summation is over all possible values of δJ, and $V_{\delta J}$ is the number of lattice sites with disorder value δJ. In the limit $V \to \infty$, for "well-behaved" probability distributions, we expect $V_{\delta J}/V \to \mathbb{P}[\delta J]$ and thus the quantity can be replaced by

$$\sum_{\delta J} V_{\delta J} G_{\bar{J}+\delta J} \sim V \times \mathbb{E}_{\delta J} G_{J+\delta J} \tag{A29}$$

And the remainder of the proof is similar. Admittedly, what "well-behaved" means rigorously is a small subtlety that we do not delve into, since it is not essential to the techniques developed in the proof.

Appendix B: Rewriting the Free Energy Density

In the main text, we claimed that F can be rewritten as Eq. (11) using the new variables a, η, φ. This was proven using the integral representation of the modified Bessel function, i.e.,

$$I_0(|z|) = \int_{\mathbb{S}^1} \exp \Re(ze^{-i\phi})\frac{d\phi}{2\pi}, \quad z \in \mathbb{C} \cong \mathbb{R}^2 \tag{B1}$$

[18] In the general continuum case, one would actually sum over partitions of $\mathbb{R}$, i.e., $V_{\delta J}$ is the number of lattice sites with disorder values in the range $(\delta J - \Delta/2, \delta J + \Delta/2)$ where $\delta J = n \times \Delta$ and $n \in \mathbb{Z}$, and the summation is over all $n \in \mathbb{Z}$. Ultimately, one would take $V \to \infty$ and $\Delta \to 0$ so that the summation approximates an integral.

More specifically, if we use the natural isomorphism and view $w_\pm \in \mathbb{C} \cong \mathbb{R}^2$, then

$$G_J = \int_{\phi_+=-\pi}^{\pi} \int_{\phi_-=-\pi}^{\pi} \frac{d\phi_\pm}{(2\pi)^2} \exp\left[\beta \Re\left(w_+ e^{-i\phi_+} + w_- e^{-i\phi_-}\right)\right] e^{\beta J \cos(\phi_+ - \phi_-)} \tag{B2}$$

$$= \int_{\phi=-\pi}^{\pi} \frac{d\phi}{2\pi} e^{\beta J \cos(\phi_+ - \phi_-)} \int_{\phi_-=-\pi}^{\pi} \frac{d\phi_-}{2\pi} \exp\left[\beta \Re\left((w_+ e^{-i\phi} + w_-) e^{-i\phi_-}\right)\right], \quad \phi = \phi_+ - \phi_- \tag{B3}$$

$$= \int_{\phi=-\pi}^{\pi} \frac{d\phi}{2\pi} e^{\beta J \cos(\phi_+ - \phi_-)} I_0\left(\beta |w_+ e^{-i\phi} + w_-|\right) \tag{B4}$$

A change of variables using $w_\pm \mapsto a, \eta, \varphi$ will result in Eq. (11).

Appendix C: Self-Consistency Equations

In this section, we obtain the self-consistency equations for $T_{\mathbb{Z}_2}$ assuming that $\eta^\star = 1$ (or equivalently, we only need to consider $F(w_+, w_-)$ with $|w_+| = |w_-| = |w|$). Indeed, similar to the $U(1)$ transition with respect to a, the $\mathbb{Z}_2$ transition occurs exactly when the 2nd order derivative with respect to φ has a sign transition. More specifically, the critical temperature $T_{\mathbb{Z}_2}$ must satisfy

$$\left.\frac{\partial^2}{\partial \varphi^2}\right|_{\varphi=0} F = 0, \quad \left.\frac{\partial}{\partial |w|}\right|_{\varphi=0} F = 0 \tag{C1}$$

Where the 2nd equality is to determine the minimizing $|w^\star|$. Notice that

$$K_J \equiv e^{G_J} = \iint d\phi_\pm \exp[\beta |w| (\cos(\phi_+ - \varphi/2) + \cos(\phi_- + \varphi/2) - \beta J \cos(\phi_+ - \phi_-)] \tag{C2}$$

Where we have set $w_\pm = |w| e^{\pm i\varphi/2}$ (possible since K_J is independent of the average phase). It's then easy to check that the self-consistency equations reduce to Eq. (17) in the main text.

Appendix D: Critical Temperatures

In this section, we derive the series expansion of the critical temperatures $T_{U(1)}, T_{\mathbb{Z}_2}$ with respect to $\bar{J} \ll \varepsilon$. For simplicity, we shall set $d\kappa = 1$. Indeed, $T_{U(1)}$ is determined by the self-consistency equation

$$T = 1 + \mathbb{E}_J r_1(\beta J) \tag{D1}$$

where $r_\nu \equiv I_\nu / I_0$ and I_ν are the modified Bessel functions. When $\bar{J} = 0$, it's clear that $T_{U(1)} = 1$ and thus to obtain the linear approximation, we set $T = 1 + c_1 \bar{J} + \cdots_{\bar{J}}$ where $\cdots_{\bar{J}}$ denotes higher order terms in $\bar{J}$. Therefore, keeping only linear terms, we find that

$$c_1 \bar{J} + \cdots = \mathbb{E}_{\delta J} r_1((\bar{J} + \delta J)(1 - c_1 \bar{J} + \cdots_{\bar{J}})) \tag{D2}$$

$$= \mathbb{E}_{\delta J} r_1(\delta J + (1 - c_1 \delta J)\bar{J} + \cdots_{\bar{J}}) \tag{D3}$$

$$= \mathbb{E}_{\delta J} r_1(\delta J) + \bar{J} \times \mathbb{E}_{\delta J}[r_1'(\delta J)(1 - c_1 \delta J)] + \cdots_{\bar{J}} \tag{D4}$$

$$= \bar{J} \times \mathbb{E}_{\delta J} r_1'(\delta J) + \cdots_{\bar{J}} \tag{D5}$$

Where the last equality uses the fact that $\delta J \mapsto r_1(\delta J)$ is odd. Therefore,

$$T = 1 + \bar{J} \times \overline{r_1'(\delta J)} + \cdots_{\bar{J}} \tag{D6}$$

$$= 1 + \bar{J} \times \left[\frac{1}{2} - \frac{1}{2^3}\overline{(\delta J)^2} + \cdots\right] + \cdots \bar{J} \tag{D7}$$

Where the last equality uses the Taylor series expansion of the derivative r_1'. The 2nd transition $T_{\mathbb{Z}_2}$ can be determined similarly (along with the corresponding magnetization $w^\star$ at $T_{\mathbb{Z}_2}$). More specifically, we can set

$$T = 1 + s_1 \bar{J} + \cdots_{\bar{J}}, \quad |w|^2 = t_1 \bar{J} + \cdots_{\bar{J}} \tag{D8}$$

Where used the fact that $T_{U(1)} = T_{\mathbb{Z}_2}$ when $\bar{J} = 0$ and that F is analytic with respect to $|w|^2$ rather than $|w|$ (the singularity in derivative with respect to $|w|$ occurs at $|w| = 0$). Notice that

$$\frac{1}{I_0(\beta J)} K_J = \frac{1}{I_0(\beta J)} \int I_0(2\beta|w|\cos(\phi/2)) e^{\beta J \cos\phi} d\phi \tag{D9}$$

$$= 1 + \frac{1}{2}(\beta|w|)^2(1 + r_1) + \cdots_{\beta|w|} \tag{D10}$$

Where $\cdots_{\beta|w|}$ denotes higher order terms in $\beta|w|$ and $r_\nu = r_\nu(\beta J)$ as in the main text. Similarly, notice that

$$\frac{1}{I_0(\beta J)} K_J \times 2\langle \cos\phi_+ \rangle_J = \frac{1}{I_0(\beta J)} \int 2\cos(\phi/2) I_1(2\beta|w|\cos(\phi/2)) e^{\beta J \cos\phi} d\phi \tag{D11}$$

$$= \beta|w|(1 + r_1) + \frac{1}{2^2}(\beta|w|)^3 \left(\frac{3}{2} + 2r_1 + \frac{1}{2} r_2 \right) + \cdots_{\beta|w|} \tag{D12}$$

And that

$$\frac{1}{I_0(\beta J)} K_J \times \langle (\sin\phi_+ - \sin\phi_-)^2 \rangle_J = \frac{1}{I_0(\beta J)} \int (1 - \cos\phi)(I_0(2\beta|w|\cos(\phi/2)) + I_2(2\beta|w|\cos(\phi/2)) e^{\beta J \cos\phi} d\phi \tag{D13}$$

$$= (1 - r_1) + \frac{3}{2^3}(\beta|w|)^2(1 - r_2) + \cdots_{\beta|w|} \tag{D14}$$

Therefore,

$$2\langle \cos\phi_+ \rangle_J = \beta|w|(1 + r_1) - \frac{1}{2^3}(\beta|w|)^3((1 + 2r_1)^2 - r_2) + \cdots_{\beta|w|} \tag{D15}$$

$$\langle (\sin\phi_+ - \sin\phi_-)^2 \rangle_J = (1 - r_1) - \frac{1}{2^3}(\beta|w|)^2(1 + 3r_2 - 4r_1^2) + \cdots_{\beta|w|} \tag{D16}$$

Using the self-consistency equations in Eq. (17), we find that up to linear terms in $\bar{J}$, we have the following independent equations

$$s_1 = \mathbb{E}_{\delta J} \left[+r_1'(\delta J) - \frac{1}{2^3} t_1 (1 + 4r_1^2(\delta J) - r_2(\delta J) \right] \tag{D17}$$

$$s_1 = \mathbb{E}_{\delta J} \left[-r_1'(\delta J) - \frac{1}{2^3} t_1 (1 - 4r_1^2(\delta J) + 3r_2(\delta J) \right] \tag{D18}$$

Solving the system of equations, we find

$$t_1 = 2 \times \frac{\overline{r_1'(\delta J)}}{\overline{r_1(\delta J)^2} - \overline{r_2(\delta J)}/2} \tag{D19}$$

$$s_1 = -\frac{1}{2^2} \frac{\overline{r_1'(\delta J)}(1 + \overline{r_2(\delta J)}}{\overline{r_1(\delta J)^2} - \overline{r_2(\delta J)}/2} \tag{D20}$$

In particular, we find that in the limit of $\varepsilon \to 0^+$,

$$s_1 = -\frac{2}{3} \frac{1}{\varepsilon^2} + O(1) \tag{D21}$$

Appendix E: The nonlinear sigma model near $T = 0$ and small disorder $\varepsilon \to 0^+$

1. Why it Breaks Down at Large ε

As discussed in the main text, one may naively attempt to obtain near $T = 0$ results by using the nonlinear sigma model. More specifically, we would like to compute

$$e^{G_J} = \iint_{[-\pi,\pi]^2} d\phi_\pm \exp\left[\beta \sum_l |w_l| \cos\phi_l + \beta J \cos(\varphi + \phi_+ - \phi_-) \right] \tag{E1}$$

by replacing $\cos\phi_l \approx 1 - \phi_l^2/2$ and extending the integration limits to $[-\pi,\pi]^2 \mapsto \mathbb{R}^2$ since only small values of ϕ_l are expected to contribute to the integral in the limit where $\beta|w_l| \to \infty$ (as an example, the asymptotic expansion of the modified Bessel function $I_0(x)$ where $x \to \infty$ utilizes this expansion). The limit $\beta|w_l^\star| \to \infty$ is expected to be true since near $T = 0$ since $\beta \to \infty$ and the minimizers $w_l^\star$ are related to the magnetiztations of each layer, which are expected to be nonzero for sufficiently low temperatures. However, we will find that the expansion is ill-regulated when $|w_l|$ is small and thus we cannot readily find the minimizers $w_l^\star$ using the expansion.

Since ultimately the expansion fails for general disorder strength ε, it is instructive to consider the algebraically simple case $|w_+| = |w_-| = |w|$ and $\varphi = \pi/2$. Notice that these restrictions are sufficient since we will rigorously prove in Appendix (F), that if $\bar{J} = 0$, then the minimizer must satisfy $\eta^\star = 1, \varphi^\star = \pm\pi/2$ for all temperatures. In this case, up to 2nd order in ϕ_l, we have

$$e^{G_J} \approx e^{2\beta|w|} \iint_{\mathbb{R}^2} d\phi_\pm \exp\left[-\frac{1}{2}\beta|w|(\phi_+^2 + \phi_-^2) - \beta J(\phi_+ - \phi_-)\right] \tag{E2}$$

$$= e^{2\beta|w|} e^{\beta J^2/|w|} \frac{\pi}{\beta|w|} \tag{E3}$$

$$-\frac{1}{\beta}\mathbb{E}_J G_J \approx -\frac{\mathbb{E}_J J^2}{|w|} - 2|w| + \frac{1}{\beta}\log\frac{\beta|w|}{\pi} \tag{E4}$$

$$F \approx -\frac{\mathbb{E}_J J^2}{|w|} - 2|w| + \frac{1}{2d\kappa}|w|^2 + \frac{1}{\beta}\log\frac{\beta|w|}{\pi} \tag{E5}$$

On the right hand side, there are two terms which $\to -\infty$ in the limit $|w| \to 0^+$, i.e., the first and last terms. In the case where $\beta \to \infty$, one can argue that the last term $\to 0$ and should not be considered. However, the first term $\propto -1/|w|$ is independent of temperature and thus cannot be excluded. It's then clear that we cannot readily minimize F using the nonlinear sigma expansion.

2. Why it Works at Small ε

As discussed in the main text, in the limit of small disorder $\varepsilon \to 0^+$, the minimizers $a^*(\varepsilon), \eta^*(\varepsilon)$ are expected to be sufficiently close to their non-disordered counterparts $a^*(\varepsilon = 0), \eta^*(\varepsilon = 0)$, and thus the divergence at $a \to 0^+$ does not affect us. In this case, we can assume that $|w_+| = |w_-| = |w| = 2d\kappa$ and find that

$$e^{G_J} \approx e^{2\beta|w|} e^{\beta J\cos\varphi} \iint_{\mathbb{R}^2} d\phi_\pm \exp\left[-\frac{1}{2}\beta\phi^T A\phi - \beta\phi^T\psi\right] \tag{E6}$$

Where $\phi = (\phi_+, \phi_-)$, $\psi = J\sin\varphi(1,-1)$, and

$$A = \begin{bmatrix} |w| + J\cos\varphi & -J\cos\varphi \\ -J\cos\varphi & |w| + J\cos\varphi \end{bmatrix} \tag{E7}$$

Therefore,

$$e^{G_J} \approx e^{2\beta|w|} e^{\beta J\cos\varphi} e^{\frac{1}{2}\psi^T A\psi} \frac{\pi}{\beta\sqrt{\det A}} \tag{E8}$$

$$-\frac{1}{\beta}\mathbb{E}_J G_J \approx -\sin^2\varphi\,\mathbb{E}_J\left[\frac{J^2}{|w|}\left(1 + \frac{2J}{|w|}\cos\varphi\right)^{-1}\right] - \bar{J}\cos\varphi - 2|w| + \frac{1}{\beta}\log\frac{\beta\sqrt{\det A}}{\pi} \tag{E9}$$

Therefore, in the limit $\beta \to \infty$ and $\bar{J} \ll \varepsilon \to 0^+$, we see that

$$F \approx -\sin^2\varphi\frac{\overline{J^2}}{|w|} + \bar{J}\cos\varphi + \cdots \tag{E10}$$

$$= -\frac{\overline{J^2}}{|w|}\left[1 + \left(\frac{\bar{J}|w|}{2\overline{J^2}}\right)^2 - \left(\cos\varphi - \left(\frac{\bar{J}|w|}{2\overline{J^2}}\right)\right)^2\right] \tag{E11}$$

Where $\cdots$ contains higher order terms or terms that are independent of φ. Therefore, we find that the minimizer $\varphi^\star$ satisfies

$$\cos\varphi^* = \frac{\bar{J}|w|}{2\overline{J^2}} \sim \frac{\bar{J}d\kappa}{(\delta J)^2}, \quad \bar{J}/\varepsilon \to 0 \tag{E12}$$

3. Subtleties of The Previous Solution

Notice that there is a small subtlety regarding the first term in Eq. (E9). Indeed, for the nonlinear sigma approximation to converge, we require that $\det A > 0$ and thus J cannot be "too" negative, as indicated by the denominator of the first term. However, if we were to naively take the disorder average for a distribution which is not compactly supported (i.e., nonzero probability for large negative J), the first term would diverge. Therefore, for large negative values of J, one would need to use the asymptotically exact formula instead of the nonlinear sigma approximation, and the first term would actually be the disorder average over J which is not "too negative", e.g..,

$$-\sin^2\varphi \mathbb{E}_J\left[1\left\{1+\frac{2J}{|w|}\cos\varphi \geq \frac{1}{2}\right\}\frac{J^2}{|w|}\left(1+\frac{2J}{|w|}\cos\varphi\right)^{-1}\right] \tag{E13}$$

Where $1\{A\}$ is the indicator function, i.e., $=1$ if A is true, and $=0$ otherwise, and the cutoff $1/2$ was arbitrarily chosen. In the $\varepsilon \to 0^+$ limit, the contributions of large negative J would $\to 0$ and thus doesn't affect our argument (hence, the abuse of notation).

Appendix F: Rigorous results for general temperatures T: TRSB behavior

In this section, we shall prove the following two statements.

Theorem F.1 (Orientation)**.** *Let the probability distribution of δJ be even and let $a > 0$ where a is defined in Eq. (10). If $\bar{J} \geq 0$, then F is minimized when $|\varphi^\star| \leq \pi/2$. Equivalently, F is minimized when $\bar{J}\cos\varphi^\star \geq 0$.*

Theorem F.2 (TRSB)**.** *Let the probability distribution of δJ be even and let $a > 0$ where a is defined in Eq. (10). If $\bar{J} = 0$, then F is minimized when $|\varphi^\star| = \pi/2$ and $\eta^\star = 1$ where φ, η are defined in Eq. (10).*

Indeed, the disordered average $\bar{J}$, at least schematically, corresponds to an interaction of the form $-\bar{J}\cos\phi(\mathbf{r})$ in the Hamiltonian, and thus one would expect that the $\mathbb{Z}_2$ order parameter $\varphi^\star$ would be such that $\bar{J}\cos\varphi^\star \geq 0$ so the Hamiltonian can be minimized. Theorem (F.1) formalizes this concept within the context of mean field theory. Similarly, it was previously argued (based on physical reasoning) in Ref. [1] that a disordered system would exhibit time-reversal-symmetry breaking (TRSB) behavior near $\bar{J} = 0$. Theorem (F.2) shows that this is indeed true in context of mean-field theory, and that the TRSB behavior is independent of the specific form of the disorder (as long as it is even with respect to δJ).

To prove the given statements, let us introduce some useful lemmas as follows.

Lemma F.3. *Let $f : \mathbb{S}^1 \to \mathbb{C}$ denote an "well-behaved" function (say continuous). Then*

$$\int_{\theta=-\pi}^{\pi} f(\cos\theta, \sin\theta)d\theta = \sum_{\xi,\eta=\pm1}\int_{\theta=0}^{\pi/2} f(\xi\cos\theta, \eta\sin\theta)d\theta \tag{F1}$$

$$= \frac{1}{4}\sum_{\xi,\eta=\pm1}\int_{\theta=-\pi}^{\pi} f(\xi|\cos\theta|, \eta|\sin\theta|)d\theta \tag{F2}$$

Proof. Let $\sigma = (\cos\theta, \sin\theta) \in \mathbb{S}^1$ and let $\xi, \eta = \pm1$ denote the sign of the x, y components of σ, i.e., that of $\cos\theta, \sin\theta$, respectively, so that $\cos\theta = \xi|\cos\theta|$ and $\sin\theta = \eta|\sin\theta|$. The statement is then clear after restricting the integration limits to $\theta \in (0, \pi/2)$ so that $\cos\theta, \sin\theta \geq 0$. □

Let us also introduce the concept of *conic combinations*, that is, a function $f(x, y, z)$ is a *conic combination* of $x, y, z \in \mathbb{R}$ if f can be written as a summation of terms of the form $x^k y^l z^m$ (with nonnegative coefficients) as $k, l, m \in \mathbb{N}$, i.e.,

$$f(x, y, z) = \sum_{k,l,m\geq0} c_{klm}x^k y^l z^m, \quad c_{klm} \geq 0 \tag{F3}$$

And we call f a *conic function* of x, y, z.

For example, the modified Bessel function $I_0(x)$ is a conic function of x (when viewing it as a Taylor series). Since the explicit values of the nonnegative coefficients c_{klm} will not matter, we will use $\{x, y, z\}$ to denote the conic function $f(x, y, z)$. If $f(x, y, z)$ is even/odd under parity $x, y, z \mapsto -x, -y, -z$, we will denote it by $\{\cdots\}^\pm$, respectively. Notice

that the product of 2 conic functions is also conic, i.e., if $f(x,y)$ and $g(z)$ is conic, then so is $f(x,y)g(z)$ and thus we can write

$$\{x,y\}\{z\} = \{x,y,z\} \tag{F4}$$

We now claim the following.

Lemma F.4.

$$\sum_{x,y,z,...=\pm 1} \{x,y,z,...\} \geq 0 \tag{F5}$$

Proof. Notice that

$$\sum_{x,y,z,...=\pm 1} x^k y^l z^m \cdots = \left[\sum_{x=\pm 1} x^k\right]\left[\sum_{y=\pm 1} y^l\right]\left[\sum_{z=\pm 1} z^m\right]\cdots \tag{F6}$$

$$\geq 0 \tag{F7}$$

Where we used the fact that $\sum_{x=\pm 1} x^k = 2$ if k is even and $= 0$ if k is odd. Therefore, the lemma follows. □

1. A Digression to the Non-disorder J_2 Problem

With the previous lemmas in mind, we are now ready to prove the given statements in Theorem (F.1) and (F.2). However, since the proof is relatively convoluted for the disordered problem, we shall first demonstrate its application on the analogous all-to-all model of the non-disordered (inherent) $J_2 \cos 2\phi$ inter-layer interaction, which turns out to be substantially simpler. More specifically, we shall prove the following.

Theorem F.5. *Consider the mean-field Hamiltonian with inherent $J_2 > 0$ inter-layer coupling, i.e.,*

$$\mathcal{H}_{\text{MF}} = -\frac{d\kappa}{V}\sum_{\mathbf{r},\mathbf{r}',l} \sigma_l(\mathbf{r})\cdot\sigma_l(\mathbf{r}') + \mathcal{H}_{\text{int}} \tag{F8}$$

$$\mathcal{H}_{\text{int}} = \sum_{\mathbf{r}}(-J_1\cos\phi(\mathbf{r}) + J_2\cos 2\phi(\mathbf{r})) \tag{F9}$$

$$= \sum_{\mathbf{r}}(-J_1\sigma_+(\mathbf{r})\cdot\sigma_-(\mathbf{r}) + J_2(2(\sigma_+(\mathbf{r})\cdot\sigma_-(\mathbf{r}))^2 - 1)) \tag{F10}$$

Then the free energy density of $\mathcal{H}_{\text{MF}}$ is

$$\mathcal{F}_{\text{MF}} \equiv \lim_{V\to\infty}\frac{-1}{\beta V}\log \mathcal{Z}_{V,J_1,J_2} = \min_{w_\pm\in\mathbb{R}^2} F(w_+,w_-) \tag{F11}$$

Where

$$F = \frac{1}{4d\kappa}(w_+^2 + w_-^2) - \frac{1}{\beta}G_{J_1,J_2} \tag{F12}$$

$$G_{J_1,J_2} = \ln\iint d\sigma_+ d\sigma_- e^{-\beta H_{J_1,J_2}} \tag{F13}$$

$$-H_{J_1,J_2} = w_+\cdot\sigma_+ + w_-\cdot\sigma_- + J_1\sigma_+\cdot\sigma_- - J_2(2(\sigma_+\cdot\sigma_-)^2 - 1) \tag{F14}$$

Furthermore, define φ, a, η as in the main text, i.e. Eq. (10), for the disordered problem with respect to $w_\pm \in \mathbb{R}^2$.

1. *(Orientation) If $J_1 \geq 0$, then F is minimized when $|\varphi^\star| \leq \pi/2$ for all $\beta, a, \eta > 0$. Equivalently, F is minimized when $J_1\cos\varphi^\star \geq 0$.*
2. *(TRSB) If $J_1 = 0$, then the free energy F is minimized by $|\varphi^\star| = \pi/2$ and $\eta^\star = 1$ for all $\beta, a > 0$.*

Proof. The proof of the free energy density of $\mathcal{H}_{\text{MF}}$ follows a similar argument as that of Theorem (A.1). Therefore, let us focus on the final 2 statement. Indeed, since the statement holds for all $\beta, a > 0$, we shall simplify notation in our proof by setting $\beta = 2a = 1$ (One can of course repeat the proof with general values of $\beta, a > 0$). We shall also write $K = e^G$ as the partition function of the two-site Hamiltonian H (where we have omitted the subscripts J_1, J_2).

1. Let us consider that case where $J_1 \geq 0$ and compare the possibilities of φ and $\pi - \varphi$, i.e., the sign of

$$K(\varphi) - K(\pi - \varphi) \tag{F15}$$

Notice that

$$K_{J_1,J_2}(\varphi) = \int \frac{d\theta}{2\pi} I_0(\sqrt{1+\eta\cos\theta}) e^{J_1\cos(\theta+\varphi) - J_2\cos 2(\theta+\varphi)} \tag{F16}$$

Therefore,

$$K_{J_1,J_2}(\pi - \varphi) = K_{-J_1,J_2}(\varphi) \tag{F17}$$

And thus we shall consider

$$K_{J_1,J_2}(\varphi) - K_{-J_1,J_2}(\varphi) = 2\int \frac{d\theta}{2\pi} I_0(\sqrt{1+\eta\cos\theta}) e^{-J_2\cos 2(\theta+\varphi)} \sinh(J_1\cos(\theta+\varphi)) \tag{F18}$$

$$= 2\int \frac{d\theta}{2\pi} I_0(\sqrt{1+\eta\cos(\theta-\varphi)}) e^{-J_2\cos 2\theta} \sinh(J_1\cos\theta) \tag{F19}$$

Notice that

$$I_0(x) = \sum_{n=0}^{\infty} \frac{1}{n!^2}\left(\frac{x}{2}\right)^{2n} \tag{F20}$$

Therefore, $I_0(\sqrt{1+\eta\cos(\theta-\varphi)})$ is a conic combination of $1+\eta\cos(\theta-\varphi)$. Furthermore, let $\xi_\theta, \eta_\theta = \pm 1$ denote the sign of the x, y (real, imaginary) components of $e^{i\theta}$, and similarly for ξ_ϕ, η_ϕ. Then we have

$$1+\eta\cos(\theta-\varphi) = 1 + \xi_\theta\xi_\phi \times \eta|\cos\theta\cos\varphi| + \eta_\theta\eta_\varphi \times \eta|\sin\theta\sin\varphi| \tag{F21}$$

In particular, we see that $I_0(\sqrt{1+\eta\cos(\theta-\varphi)})$ is a conic combination of $\xi_\theta\xi_\varphi$ and $\eta_\theta\eta_\varphi$, i.e.,

$$I_0(\sqrt{1+\eta\cos(\theta-\varphi)}) = \{\xi_\theta\xi_\varphi, \eta_\theta\eta_\varphi\} \tag{F22}$$

Also notice that

$$e^{-J_2\cos 2\theta}\sinh(J_1\cos\theta) = \xi_\theta \times e^{-J_2(2\cos^2\theta - 1)}\sinh(J_1|\cos\theta|) \tag{F23}$$

$$= \{\xi_\theta\}^- \tag{F24}$$

Therefore, we have

$$K_{J_1,J_2}(\varphi) - K_{-J_1,J_2}(\varphi) = 2\int \frac{d\theta}{2\pi}\{\xi_\theta\xi_\varphi, \eta_\theta\eta_\varphi\}\{\xi_\theta\eta_\theta\}^- \tag{F25}$$

$$= \frac{1}{4}\sum_{\xi_\theta,\eta_\theta=\pm1}\int \frac{d\theta}{2\pi}\{\xi_\theta\xi_\varphi, \eta_\theta\eta_\varphi\}\{\xi_\theta\}^- \tag{F26}$$

$$= \frac{1}{4}\sum_{\xi_\theta,\eta_\theta=\pm1}\int \frac{d\theta}{2\pi}\{\xi_\theta, \eta_\theta\}\{\xi_\theta\xi_\varphi\}^- \tag{F27}$$

$$= \xi_\varphi \times \underbrace{\frac{1}{4}\sum_{\xi_\theta,\eta_\theta=\pm1}\int \frac{d\theta}{2\pi}\{\xi_\theta, \eta_\theta\}\{\xi_\theta\}^-}_{\geq 0} \tag{F28}$$

Where the 2nd equality uses Lemma (F.3), the 3rd equality uses the fact that $\xi_\theta \mapsto \xi_\theta\xi_\varphi$ is a bijective mapping from $\{\pm1\} \to \{\pm1\}$ (and similarly $\eta_\theta \mapsto \eta_\theta\eta_\varphi$), and the 4th equality uses Lemma (F.4). Therefore, $K(\varphi) - K(\pi-\varphi)$ has the same sign as ξ_φ, and thus the statement regarding orientation follows.

2. Setting $J_1 = 0$, we find that

$$K_{J_2}(\varphi) = \int \frac{d\theta}{2\pi} I_0(\sqrt{1+\eta\cos\theta}) e^{-J_2\cos 2(\theta+\varphi)} \tag{F29}$$

Therefore, we see that

$$\frac{\partial}{\partial\varphi}K_{J_2}(\varphi) = 2J_2\int\frac{d\theta}{2\pi}I_0(\sqrt{1+\eta\cos\theta})e^{-J_2\cos 2(\theta+\varphi)}\sin 2(\theta+\varphi) \tag{F30}$$

$$= 2J_2\int\frac{d\theta}{2\pi}I_0(\sqrt{1+\eta\cos(\theta-\varphi)})e^{-J_2\cos 2\theta}\sin(2\theta) \tag{F31}$$

As before, we have

$$I_0(\sqrt{1+\eta\cos(\theta-\varphi)}) = \{\xi_\theta\xi_\varphi, \eta_\theta\eta_\varphi\} \tag{F32}$$

Also notice that

$$e^{-J_2\cos 2\theta}\sin(2\theta) = 2e^{-J_2(2\cos^2\theta-1)}\sin\theta\cos\theta \tag{F33}$$

$$= \{\xi_\theta\eta_\theta\}^- \tag{F34}$$

Therefore, we have

$$\frac{\partial}{\partial\varphi}K_{J_2}(\varphi) = 2J_2\int\frac{d\theta}{2\pi}\{\xi_\theta\xi_\varphi, \eta_\theta\eta_\varphi\}\{\xi_\theta\eta_\theta\}^- \tag{F35}$$

$$= \frac{J_2}{2}\sum_{\xi_\theta,\eta_\theta=\pm1}\int\frac{d\theta}{2\pi}\{\xi_\theta\xi_\varphi, \eta_\theta\eta_\varphi\}\{\xi_\theta\eta_\theta\}^- \tag{F36}$$

$$= \frac{J_2}{2}\sum_{\xi_\theta,\eta_\theta=\pm1}\int\frac{d\theta}{2\pi}\{\xi_\theta, \eta_\theta\}\{\xi_\theta\xi_\varphi\eta_\theta\eta_\varphi\}^- \tag{F37}$$

$$= \xi_\varphi\eta_\varphi \times \underbrace{\frac{J_2}{2}\sum_{\xi_\theta,\eta_\theta=\pm1}\int\frac{d\theta}{2\pi}\{\xi_\theta, \eta_\theta\}\{\xi_\theta\eta_\theta\}^-}_{\geq 0} \tag{F38}$$

Where the 2nd equality uses Lemma (F.3), the 3rd equality uses the fact that $\xi_\theta \mapsto \xi_\theta\xi_\varphi$ is a bijective mapping from $\{\pm1\} \to \{\pm1\}$ (and similarly $\eta_\theta \mapsto \eta_\theta\eta_\varphi$), and the 4th equality uses Lemma (F.4). Therefore, $\partial K_{J_2}(\varphi)/\partial\varphi$ has the same sign as $\xi_\varphi\eta_\varphi$. In particular, this implies that $K_{J_2}(\varphi)$ is increasing when $\varphi = 0 \to \pm\pi/2$ and decreases as $\varphi = \pm\pi/2 \to \pi$, and thus K_{J_2} is maximized (indicating F is minimized) when $|\varphi^\star| = \pi/2$.

Let us now optimize η. Indeed, if we take $\varphi = \pm\pi/2$, we find that

$$K_{J_2}(\eta;\varphi=\pm\pi/2) = \int\frac{d\theta}{2\pi}I_0(\sqrt{1+\eta\cos\theta})e^{J_2\cos 2\theta} \tag{F39}$$

Taking the derivative with respect to η and we obtain

$$\frac{\partial}{\partial\eta}K_{J_2}(\eta) = \int\frac{d\theta}{2\pi}I_1(\sqrt{1+\eta\cos\theta})\frac{1}{2\sqrt{1+\eta\cos\theta}} \times \cos\theta e^{J_2\cos 2\theta} \tag{F40}$$

Notice that

$$\frac{I_1(x)}{x} = \sum_{n=0}^{\infty}\frac{1}{n!(n+1)!}\left(\frac{x}{2}\right)^{2n} \tag{F41}$$

Therefore, we see that $I_1(\sqrt{1+\eta\cos\theta})/\sqrt{1+\eta\cos\theta}$ is a conic combination of $1+\eta\cos\theta$ (similar to the case of I_0), and thus we have

$$I_1(\sqrt{1+\eta\cos\theta})\frac{1}{2\sqrt{1+\eta\cos\theta}} = \{\xi_\theta\} \tag{F42}$$

Also notice that

$$\cos\theta e^{J_2\cos 2\theta} = \{\xi_\theta\}^- \tag{F43}$$

Therefore, we have

$$\frac{\partial}{\partial\eta}K_{J_2}(\eta) = \int\frac{d\theta}{2\pi}\{\xi_\theta\}\{\xi_\theta\}^- \geq 0 \tag{F44}$$

Where we applied Lemma (F.3) and (F.4) as before. Therefore, we see that F is minimized when $|\varphi^\star| = \pi/2$ and $\eta^\star = 1$.

□

2. Returning to the Disorder Problem

Let us now prove the statements for the disordered problem. Since the statements will hold for all temperatures $\beta > 0$ and $a > 0$, we shall simplify notation in our proof by setting $\beta = 2a = 1$ (One can of course repeat the proof with general values of $\beta, a > 0$). We shall also write $K_J = e^{G_J}$ as the partition function of the two-site Hamiltonian H_J in Eq. (9).

Proof of Theorem (F.1). Notice that minimizing F with respect to φ is the equivalent to maximizing $\mathbb{E}_J G_J$ with respect to φ and thus we will prove the latter. It's clear that $K_J(\varphi) = K_J(-\varphi)$. Therefore, we shall fix $|\varphi| \leq \pi/2$ and compare φ with $\pi - \varphi$. Notice that

$$K_J(\pi - \varphi) = K_{-J}(\varphi) \tag{F45}$$

Also notice that

$$\mathbb{E}_J G_J = \int_{\delta J > 0} (G_{\bar{J}+\delta J} + G_{\bar{J}-\delta J}) \mathbb{P}_{\delta J} \tag{F46}$$

Therefore,

$$\mathbb{E}_J G_J(\varphi) - \mathbb{E}_J G_J(\pi - \varphi) = \int_{\delta J > 0} (G_{\bar{J}+\delta J}(\varphi) + G_{\bar{J}-\delta J}(\varphi) - G_{-\bar{J}-\delta J}(\varphi) - G_{-\bar{J}+\delta J}(\varphi)) \mathbb{P}_{\delta J} \tag{F47}$$

Therefore, it is sufficient to show that

$$G_{\bar{J}+\delta J}(\varphi) + G_{\bar{J}-\delta J}(\varphi) - G_{-\bar{J}-\delta J}(\varphi) - G_{-\bar{J}+\delta J}(\varphi) \geq 0 \tag{F48}$$

For simplicity, we will suppress the φ notation in $G_J(\varphi)$ since each term is understood to be at φ. Notice that there are two cases which we will treat independently; (1) $\bar{J} \geq \delta J \geq 0$, and (2) $\delta J \geq \bar{J} \geq 0$.

(1) In case (1), it is sufficient to show that

$$G_p \geq G_{-p}, \quad \forall p \geq 0 \tag{F49}$$

Or equivalently, we need to show that

$$K_p - K_{-p} = \int \frac{d\theta}{2\pi} I_0(\sqrt{1 + \eta \cos\theta}) \sinh(p \cos(\theta - \varphi)) \geq 0 \tag{F50}$$

Let ξ_θ, η_θ be the signs of the x, y components of $e^{i\theta}$ (i.e., $\cos\theta, \sin\theta$, respectively) as in Lemma. (F.3), and similarly, $\xi_\varphi, \eta_\varphi$ for φ. By using the Taylor expansion of the modified Bessel function I_0, we see that I_0 is a conic function of ξ_θ, and thus

$$K_p - K_{-p} = \sum_{\xi_\theta, \eta_\theta = \pm 1} \int d\theta \{\xi_\theta\} \{\xi_\theta \xi_\varphi, \eta_\theta \eta_\varphi\}^- \tag{F51}$$

$$= \xi_\varphi \sum_{\xi_\theta, \eta_\theta = \pm 1} \int d\theta \{\xi_\theta\} \{\xi_\theta, \eta_\theta \xi_\varphi \eta_\varphi\}^- \tag{F52}$$

Since we are summing over all $\eta_\theta = \pm 1$, we can perform the bijective transformation $\eta_\theta \mapsto \eta_\theta \xi_\varphi \eta_\varphi$, to find that

$$K_p - K_{-p} = \xi_\varphi \sum_{\xi_\theta, \eta_\theta = \pm 1} \int d\theta \{\xi_\theta\} \{\xi_\theta, \eta_\theta\}^- \tag{F53}$$

Since the integrand does not depend on $\xi_\varphi, \eta_\varphi$ (which are not summed over ± 1), by Lemma (F.4), we see that $K_p - K_{-p}$ has the same sign as ξ_φ. Since we assumed that $|\varphi| \leq \pi/2$, we see that

$$K_p - K_{-p} \geq 0 \tag{F54}$$

(2) In case (2), it is sufficient to show that for all $p \geq q \geq 0$, we have

$$G_p + G_{-q} \geq G_{-p} + G_q \tag{F55}$$

Equivalently, we need to show that

$$K_p K_{-q} \geq K_{-p} K_q \tag{F56}$$
$$(K_p - K_{-p})(K_q + K_{-q}) \geq (K_p + K_{-p})(K_q - K_{-q}) \tag{F57}$$
$$\frac{K_p - K_{-p}}{K_p + K_{-p}} \geq \frac{K_q - K_{-q}}{K_q + K_{-q}} \tag{F58}$$

Therefore, it's sufficient to show that the following function is increasing for $p \geq 0$.

$$p \mapsto \frac{K_p - K_{-p}}{K_p + K_{-p}} \tag{F59}$$

Alternatively, it's sufficient to show that

$$\frac{\partial}{\partial p}\left(\frac{K_p - K_{-p}}{K_p + K_{-p}}\right) \geq 0 \tag{F60}$$
$$\frac{2}{(K_p + K_{-p})^2}\left(\dot{K}_p K_{-p} + K_p \dot{K}_{-p}\right) \geq 0 \tag{F61}$$

Where $\dot{K}$ is short for the dertivative of $p \mapsto K_p$ with respect to p. Notice that

$$\dot{K}_p K_{-p} + K_p \dot{K}_{-p} = \iint I_0(g_1) I_0(g_2) e^{p[\cos(\theta_1+\varphi) - \cos(\theta_1+\varphi)]}[\cos(\theta_1+\varphi) + \cos(\theta_1+\varphi)] \tag{F62}$$

Where $g_i = \sqrt{1 + \eta \cos\theta_i}$ and the double integral is over $\theta_1, \theta_2 \in (-\pi, \pi)$. Therefore, the **rest of this proof** will be to show that the $\dot{K}_p K_{-p} + K_p \dot{K}_{-p} \geq 0$ for all $p \geq 0$. Notice that by interchanging $\theta_1 \leftrightarrow \theta_2$, we find that

$$\dot{K}_p K_{-p} + K_p \dot{K}_{-p} = \iint \mu(\theta - \varphi) f(\theta) \tag{F63}$$

where $\theta = (\theta_1, \theta_2)$ and $\theta - \varphi = (\theta_1 - \varphi, \theta_2 - \varphi)$ and that

$$f(\theta) = [\cos\theta_1 + \cos\theta_2] \cosh\left(p(\cos\theta_1 - \cos\theta_2)\right) \tag{F64}$$
$$\mu(\theta) = I_0(g_1) I_0(g_2) \geq 0 \tag{F65}$$

Let us introduce the coupled variables (change of variables),

$$\chi = \frac{\theta_1 + \theta_2}{2}, \quad \delta = \frac{\theta_1 - \theta_2}{2} \tag{F66}$$

And let ξ_χ, η_χ be the sign of the x, y components of $e^{i\chi}$ (i.e., the signs of $\cos\chi, \sin\chi$, respectively), and similarly for ξ_δ, η_δ, as in Lemma. (F.3). Then we have

$$f(\theta) = \{\xi_\chi \xi_\delta\}^- \{\eta_\chi \eta_\delta\}^+ \tag{F67}$$

And that

$$\mu(\theta - \varphi) = \sum_{\xi_\psi, \eta_\psi = \pm 1} \int \{\xi_\delta \xi_\chi \xi_\varphi, \xi_\delta \eta_\chi \eta_\varphi, \xi_\chi \eta_\chi \xi_\varphi \eta_\varphi\} d\psi \tag{F68}$$

Where $\xi_\varphi, \eta_\varphi$ are those corresponding to the phase φ (we leave the proof of this claim to Lemma (F.6)). Combining them together, we find that

$$\dot{K}_p K_{-p} + K_p \dot{K}_{-p} = \sum_{\xi_\chi, \eta_\chi, \xi_\delta, \eta_\delta \xi_\psi, \eta_\psi = \pm 1} \iiint d\chi d\delta d\psi \{\xi_\chi \xi_\delta\}^- \{\eta_\chi \eta_\delta\}^+ \{\xi_\delta \xi_\chi \xi_\varphi, \xi_\delta \eta_\chi \eta_\varphi, \xi_\chi \eta_\chi \xi_\varphi \eta_\varphi\} \tag{F69}$$

Since we are summing over all $\xi_\chi, \eta_\chi = \pm 1$, we can perform the bijective transformation $\xi_\chi, \eta_\chi \mapsto \xi_\chi \xi_\varphi, \eta_\chi \eta_\varphi$, to find that

$$\dot{K}_p K_{-p} + K_p \dot{K}_{-p} = \xi_\varphi \sum_{\xi_\chi, \eta_\chi, \xi_\delta, \eta_\delta \xi_\psi, \eta_\psi = \pm 1} \iiint d\chi d\delta d\psi \{\xi_\chi \xi_\delta\}^- \{\eta_\chi \eta_\delta\}^+ \{\xi_\delta \xi_\chi, \xi_\delta \eta_\chi, \xi_\chi \eta_\chi\} \tag{F70}$$

Notice that the integrand inside the summation is now independent of $\xi_\varphi, \eta_\varphi$ (which is not summed over ± 1). Therefore, (using the fact that the product of conic functions is again conic) by Lemma (F.4), we see that $\dot{K}_p K_{-p} + K_p \dot{K}_{-p}$ has the same sign as ξ_φ. Since we assumed that $|\varphi| \leq \pi/2$, we see that

$$\dot{K}_p K_{-p} + K_p \dot{K}_{-p} \geq 0 \tag{F71}$$

□

Proof of Theorem (F.2). Notice that minimizing F with respect to φ, η is the equivalent to maximizing $\mathbb{E}_J G_J$ with respect to φ, η and thus we will prove the latter. More specifically,

1. We will first show that for any $\eta \in [0,1]$, if $\bar{J} = 0$, then the mapping $\varphi \mapsto \mathbb{E}_J G_J$ is an increasing function when $0 \leq \varphi \leq \pi/2$ and a decreasing function when $\pi/2 \leq \varphi \leq \pi$, so that $\varphi = \pi/2$ maximizes $\mathbb{E}_J G_J$. Since $G_J(\varphi) = G_J(-\varphi)$, we see that $\varphi = \pm\pi/2$ maximizes $\mathbb{E}_J G_J$.

2. We will then show that when $\bar{J} = 0$ and $\varphi = \pm\pi/2$, the mapping $\eta \mapsto \mathbb{E}_J G_J$ is an increasing function and thus $\eta = 1$ maximizes $\mathbb{E}_J G_J$.

1. Notice that

$$\mathbb{E}_{\delta J} G_{\delta J} = \int_{\delta J > 0} (G_{\delta J} + G_{-\delta J}) \mathbb{P}_{\delta J} \tag{F72}$$

Therefore, it's sufficent to show that $\varphi \mapsto G_{\delta J} + G_{-\delta J} = \log K_{\delta J} K_{-\delta J}$ is increasing when $0 \leq \varphi \leq \pi/2$ and a decreasing function when $\pi/2 \leq \varphi \leq \pi$. Indeed, let us consider

$$\frac{\partial}{\partial \varphi} K_{\delta J} K_{-\delta J} = (-1) \iint d\theta_1 d\theta_2 \mu(\theta - \varphi)(\sin\theta_1 - \sin\theta_2) e^{\delta J(\cos\theta_1 - \cos\theta_2)} \tag{F73}$$

Where $g_i = \sqrt{1 + \eta \cos\theta_i}$, the double integral is over $\theta_1, \theta_2 \in (-\pi, \pi)$, $\theta = (\theta_1, \theta_2)$ and $\theta - \varphi = (\theta_1 - \varphi, \theta_2 - \varphi)$ and

$$\mu(\theta) = I_0(g_1) I_0(g_2) \tag{F74}$$

By the symmetry regarding interchanging $\theta_1 \leftrightarrow \theta_2$, we see that

$$\frac{\partial}{\partial \varphi} K_{\delta J} K_{-\delta J} = \iint d\theta_1 d\theta_2 \mu(\theta - \varphi) f(\theta) \tag{F75}$$

Where

$$f(\theta) = (-1)(\sin\theta_1 - \sin\theta_2) \sinh(\delta J(\cos\theta_1 - \cos\theta_2)) \tag{F76}$$

Let us use the coupled integration variables, i.e.,

$$\chi = \frac{\theta_1 + \theta_2}{2}, \quad \delta = \frac{\theta_1 - \theta_2}{2} \tag{F77}$$

And let ξ_χ, η_χ be the sign of the x, y components of $e^{i\chi}$ (i.e., the signs of $\cos\chi, \sin\chi$, respectively), and similarly for ξ_δ, η_δ, as in Lemma. (F.3). Then we have

$$f(\theta) = \{\eta_\delta \xi_\chi\}^- \{\eta_\delta \eta_\chi\}^- = \{\xi_\chi\}^- \{\eta_\chi\}^- \tag{F78}$$

And by Lemma (F.6), we have

$$\mu(\theta - \varphi) = \sum_{\xi_\psi, \eta_\psi = \pm 1} \int_{\psi = -\pi}^{\pi} \{\xi_\delta \xi_\chi \xi_\varphi, \xi_\delta \eta_\chi \eta_\varphi, \xi_\chi \eta_\chi \xi_\varphi \eta_\varphi\} d\psi d\psi \tag{F79}$$

Combining them together, we find that

$$\frac{\partial}{\partial\varphi}K_{\delta J}K_{-\delta J} = \sum_{\xi_\chi,\eta_\chi,\xi_\delta,\eta_\delta,\xi_\psi,\eta_\psi=\pm1}\iiint d\chi d\delta d\psi\{\xi_\chi\}^-\{\eta_\chi\}^-\{\xi_\delta\xi_\chi\xi_\varphi,\xi_\delta\eta_\chi\eta_\varphi,\xi_\chi\eta_\chi\xi_\varphi\eta_\varphi\} \tag{F80}$$

Since we are summing over all $\xi_\chi,\eta_\chi=\pm1$, we can perform the bijective transformation $\xi_\chi,\eta_\chi\mapsto\xi_\chi\xi_\varphi,\eta_\chi\eta_\varphi$, to find that

$$\frac{\partial}{\partial\varphi}K_{\delta J}K_{-\delta J} = \xi_\varphi\eta_\varphi\sum_{\xi_\chi,\eta_\chi,\xi_\delta,\eta_\delta,\xi_\psi,\eta_\psi=\pm1}\iiint d\chi d\delta d\psi\{\xi_\chi\}^-\{\eta_\chi\}^-\{\xi_\delta\xi_\chi,\xi_\delta\eta_\chi,\xi_\chi\eta_\chi\}$$

Where we used the fact that $\{x\}^-$ is odd under $x\mapsto -x$. Notice that the integrand inside the summation is now independent of ξ_φ,η_φ (which is not summed over ±1). Therefore, (using the fact that the product of conic functions is again conic) by Lemma (F.4), we see that $\partial K_{\delta J}K_{-\delta J}/\partial\varphi$ has the same sign as $\xi_\varphi\eta_\varphi$. Therefore, $K_{\delta J}K_{-\delta J}$ is increasing when $0\le\varphi\le\pi/2$ and decreasing when $\pi/2\le\varphi\le\pi$.

2. Let us now set $\varphi=\pi/2$ since we now know that it must maximize $\mathbb{E}_{\delta J}G_{\delta J}$ for any $\eta\in[0,1]$, and consider

$$\frac{\partial}{\partial\eta}K_{\delta J}K_{-\delta J} = \iint d\theta_1 d\theta_2\nu(\theta)e^{\delta J(\sin\theta_1-\sin\theta_2)} \tag{F81}$$

Where $\theta=(\theta_1,\theta_2)$ and

$$\nu(\theta) = \frac{I_1(g_1)}{g_1}\cos\theta_1 I_0(g_2) + I_0(g_1)\frac{I_1(g_2)}{g_2}\cos\theta_2 \tag{F82}$$

Where $g_i=\sqrt{1+\eta\cos\theta_i}$ as before. By using the Taylor expansion of the modified Bessel function, it's clear that

$$\frac{I_1(g_1)}{g_1} = \{\xi_1\},\quad I_0(g_2)=\{\xi_2\} \tag{F83}$$

Where ξ_i,η_i are the signs of $\cos\theta_i,\sin\theta_i$. Hence,

$$\nu(\theta) = \{\xi_1,\xi_2\} \tag{F84}$$

By symmetry regarding interchanging $\theta_1\leftrightarrow\theta_2$, we see that

$$\frac{\partial}{\partial\eta}K_{\delta J}K_{-\delta J} = \iint\nu(\theta)\cosh(\delta J(\sin\theta_1-\sin\theta_2)) \tag{F85}$$

$$= \sum_{\xi_1,\eta_1,\xi_2,\eta_2=\pm1}\iint\{\xi_1,\xi_2\}\{\eta_1,-\eta_2\}^+ \tag{F86}$$

Notice that we can perform the bijective transform $\eta_2\mapsto-\eta_2$ and the statement then follows from Lemma (F.4).

□

Lemma F.6.

$$\mu(\theta-\varphi) = \sum_{\xi_\psi,\eta_\psi=\pm1}\int_{\psi=-\pi}^{\pi}\{\xi_\delta\xi_\chi\xi_\varphi,\xi_\delta\eta_\chi\eta_\varphi,\xi_\chi\eta_\chi\xi_\varphi\eta_\varphi\}d\psi \tag{F87}$$

Where $\theta-\varphi=(\theta_1-\varphi,\theta_2-\varphi)$, and ξ_s,η_s are the signs of $\cos s,\sin s$ for $s=\chi,\delta,\psi,\varphi$, and χ,δ are the coupled variables defined from θ_1,θ_2 as done in Eq. (F66) and/or (F77).

Proof. Since we wish to compute $\mu(\theta-\varphi)$, let us use the notation $h_i=\sqrt{1+\eta\cos(\theta_i-\varphi)}$. Then we have

$$\mu(\theta-\varphi) = I_0(h_1)I_0(h_2) \tag{F88}$$

$$= \iint e^{h_1\cos\psi_1+h_2\cos\psi_2}\frac{d\psi_1}{2\pi}\frac{d\psi_2}{2\pi} \tag{F89}$$

$$= \iint\exp\left[\Re(h_1+h_2e^{i\psi})e^{i\psi_1}\right]\frac{d\psi_1}{2\pi}\frac{d\psi}{2\pi},\quad \psi=\psi_2-\psi_1 \tag{F90}$$

$$= \int I_0(|h_1+h_2e^{i\psi}|)\frac{d\psi}{2\pi} \tag{F91}$$

$$= \frac{1}{2}\int\left[I_0(|h_1+h_2e^{i\psi}|)+I_0(|h_1-h_2e^{i\psi}|)\right]\frac{d\psi}{2\pi} \tag{F92}$$

Notice that

$$I_0(x) = \sum_{n\geq 0} \frac{1}{(n!)^2} \left(\frac{x}{2}\right)^{2n} \tag{F93}$$

And that

$$|h_1 \pm h_2 e^{i\psi}|^2 = (h_1^2 + h_2^2) \pm 2h_1 h_2 \cos\psi \tag{F94}$$

Therefore, we see that

$$\sum_{s=\pm 1} I_0(|h_1 + sh_2 e^{i\psi}|) = \sum_{n\geq 0} \frac{1}{2^{2n}(n!)^2} \sum_{s=\pm 1} |h_1 + sh_2 e^{i\psi}|^{2n} \tag{F95}$$

For any $n \geq 0$, we see that

$$\begin{aligned}
\sum_{s=\pm 1} |h_1 + sh_2 e^{i\psi}|^{2n} &= \sum_{s=\pm 1} \left[(h_1^2 + h_2^2) + s(2h_1 h_2 \cos\psi)\right]^n && \text{(F96)}\\
&= \sum_{m\leq n} \binom{n}{m} (h_1^2 + h_2^2)^{n-m} (2h_1 h_2 \cos\psi)^m \sum_{s=\pm 1} s^m && \text{(F97)}\\
&= \sum_{m=2k\leq n} \underbrace{\binom{n}{m} 2^{2k+1} \cos^{2k}\psi}_{\geq 0} \times (h_1^2 + h_2^2)^{n-2k} (h_1 h_2)^{2k} && \text{(F98)}
\end{aligned}$$

Notice that the coefficient in front of $(h_1^2 + h_2^2)^{n-2k}(h_1 h_2)^{2k}$ is nonnegative and thus we only need to consider powers of $h_1^2 + h_2^2$ and $(h_1 h_2)^2$. Indeed, notice that

$$\begin{aligned}
h_1^2 + h_2^2 &= 2(1 + \eta\cos(\chi - \varphi)\cos\delta) && \text{(F99)}\\
&= 2(1 + \eta\cos\delta(\cos\chi\cos\varphi + \sin\chi\sin\varphi)) && \text{(F100)}\\
&= 2(1 + \eta|\cos\delta|\xi_\delta(|\cos\chi||\cos\varphi|\xi_\chi\xi_\varphi + |\sin\chi||\sin\varphi|\eta_\chi\eta_\varphi)) && \text{(F101)}\\
&= 2 + \underbrace{2\eta|\cos\delta\cos\chi\cos\varphi|}_{\geq 0} \times \xi_\delta\xi_\chi\xi_\varphi + \underbrace{2\eta|\cos\delta\sin\chi\sin\varphi|}_{\geq 0} \times \xi_\delta\eta_\chi\eta_\varphi && \text{(F102)}\\
&= \{\xi_\delta\xi_\chi\xi_\varphi, \xi_\delta\eta_\chi\eta_\varphi\} && \text{(F103)}
\end{aligned}$$

And that

$$\begin{aligned}
(h_1 h_2)^2 &= 1 - \eta^2 + \eta^2(\cos^2(\chi - \varphi) + \cos^2\delta) + 2\eta\cos(\chi - \varphi)\cos\delta && \text{(F104)}\\
&= 1 - \eta^2 + \eta^2(\cos^2\chi\cos^2\varphi + \sin^2\chi\sin^2\varphi + \cos^2\delta) + \frac{1}{2}\eta^2\sin 2\chi\sin 2\varphi && \text{(F105)}\\
&\qquad + 2\eta\cos\delta(\cos\chi\cos\varphi + \sin\chi\sin\varphi) && \text{(F106)}\\
&= \underbrace{1 - \eta^2 + \eta^2(\cos^2\chi\cos^2\varphi + \sin^2\chi\sin^2\varphi + \cos^2\delta)}_{\geq 0} + \underbrace{2\eta^2|\sin\chi\cos\chi\sin\varphi\cos\varphi|}_{\geq 0} \times \xi_\chi\eta_\chi\xi_\varphi\eta_\varphi && \text{(F107)}\\
&\qquad + \underbrace{2\eta|\cos\delta\cos\chi\cos\varphi|}_{\geq 0} \times \xi_\delta\xi_\chi\xi_\varphi + \underbrace{2\eta|\cos\delta\sin\chi\sin\varphi|}_{\geq 0} \times \xi_\delta\eta_\chi\eta_\varphi && \text{(F108)}\\
&= \{\xi_\delta\xi_\chi\xi_\varphi, \xi_\delta\eta_\chi\eta_\varphi, \xi_\chi\eta_\chi\xi_\varphi\eta_\varphi\} && \text{(F109)}
\end{aligned}$$

Combining this together and using the fact that the product of conic functions is again conic, we see that

$$\frac{1}{2}\sum_{s=\pm 1} I_0(|h_1 + sh_2 e^{i\psi}|) = \{\xi_\delta\xi_\chi\xi_\varphi, \xi_\delta\eta_\chi\eta_\varphi, \xi_\chi\eta_\chi\xi_\varphi\eta_\varphi\} \tag{F110}$$

And thus the statement follows. □