

Article

Not peer-reviewed version

Feature Space Dimensional Reduction Based on Quantum Spin-Correlation Encoding

[Chee Kian Yap](#) *

Posted Date: 9 January 2026

doi: 10.20944/preprints202601.0671.v1

Keywords: dimension reduction; feature; pattern classification; machine learning; compression



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Feature Space Dimensional Reduction Based on Quantum Spin-Correlation Encoding

Chee Kian Yap

Independent Researcher, Australia; cheekian.yap@hdr.qut.edu.au

Abstract

This study introduces a quantum-inspired framework for high-dimensional feature compression by encoding data pairs as parameters in a two-site spin-Hamiltonian system. Unlike traditional manifold learning techniques that suffer from “neighborhood tearing” at extreme compression ratios [1,2,5], the proposed method utilizes a ground-state solver to stabilize local topological structures. We evaluate the method’s performance across a gradient of 512 to 1 dimension using information-theoretic and ordinal metrics. Our results demonstrate a fundamental Dual-Phase behavior: (i) a High-Fidelity Spin Phase ($m > 128$) where global geometric alignment is preserved with near-unit cosine similarity, and (ii) a Cohesive Topological Phase ($m < 128$) where global reference frames fracture, yet local neighborhoods exhibit “rigid-body translation.” We prove that while absolute ordinal ranks drift during compression, the local variance remains below 4%, confirming that the Hamiltonian alignment mechanism preserves cluster integrity [7]. This approach is uniquely suited for community detection and resilient non-parametric clustering in deeply compressed feature spaces.

Keywords: dimension reduction; feature; pattern classification; machine learning; compression

1. Introduction

The primary challenge in manifold learning is the “tearing” of local neighborhoods during deep compression [1,6]. While traditional linear methods like PCA fail to capture non-linear local density, recent developments have explored encoding classical data into quantum mechanical frameworks to exploit the high-dimensional geometry of Hilbert spaces [3,8]. Our work extends this concept to the problem of deep dimensionality reduction. We propose a recursive Hamiltonian solver that reduces feature dimensions by evolving states toward a ground-state configuration, mimicking the physical process of spin alignment [4]. This study analyzes the transition from global geometric alignment to local topological cohesion from 512 to 1 dimension.

2. Methodology

2.1. Global Semantic Mapping

The algorithm first establishes a “Semantic Frame of Reference” by computing weighted projections for each sample following the logic of a Self-Similarity (KQ) Matrix. This ensures that global relational dependencies are captured prior to localized reduction [9]. Weights w_{qk} are determined by angular affinity (normalized dot product):

$$w_{qk} = \frac{f_q \bullet f_k}{||f_q|| ||f_k||} \tag{1}$$

The promoted vector V_q for each sample q is the weighted sum of all samples in the dataset:

$$V_q = \sum_{k=1}^n w_{qk} f_k \tag{2}$$

2.2. Diagnostic Metrics

To evaluate the stability of the manifold throughout the reduction process, we utilize the following formal metrics grounded in information theory and spectral geometry [10,11]:

A. Global Alignment (Relational Preservation)

Measures the preservation of global similarity structure relative to the initial 512D state. Let $S^{(512)}$ be the vector of cosine similarities between a reference r and all other samples in 512D space. The alignment at dimension m is:

$$A(m) = \frac{S^{(512)} \bullet S^{(m)}}{\|S^{(512)}\| \|S^{(m)}\|} \quad (3)$$

where $S_q^{(m)} = \frac{V_r \cdot V_q}{\|V_r\| \|V_q\|}$.

B. Information Decay (KL-Divergence)

Quantifies the information loss between the reference softmax distribution P and current state Q [10]:

$$D_{KL}(P||Q) = \sum_{i=1}^n P_i \log_2 \left(\frac{P_i}{Q_i} \right) \quad (4)$$

where P_i and Q_i are obtained via a Softmax transform of similarities with temperature τ :

$$P_i = \frac{\exp(S_i/\tau)}{\sum_{j=1}^n \exp(S_j/\tau)} \quad (5)$$

C. System Entropy (Manifold Diffusion)

Measures the concentration or uncertainty of the manifold's local structure based on Shannon entropy:

$$H(m) = - \sum_{i=1}^n Q_i^{(m)} \log_2 Q_i^{(m)} \quad (6)$$

D. Scale-Invariant Manifold Translation (Rank Drift)

Tracks the ordinal position R of the k -closest neighbors to observe drift trajectories.

E. Distance-from-Mean Cohesion (Internal Tightness)

Calculates the integrity of the neighbor pack relative to its own moving center μ_{pack} :

$$\mu_{pack} = \frac{1}{k} \sum_{j \in N} R_j^{(m)} \quad (7)$$

$$\text{Dispersion} = \sum_{j \in N} |R_j^{(m)} - \mu_{pack}| \quad (8)$$

$$\text{Cohesion} = 1 - \frac{\text{Dispersion}}{\mu_{pack} \cdot k} \quad (9)$$

3. Proposed Method

3.1. Hamiltonian Formulation

The reduction from $2m$ to m dimensions is achieved by treating adjacent feature pairs (f_j, f_{j+1}) as magnetic field parameters B_x and B_y in a transverse-field Ising model [4,12].

$$F : B_x \leftarrow \bar{x} \quad B_y \leftarrow \bar{y} \quad J_z \leftarrow x + y \quad (10)$$

where $\bar{x} = \frac{1}{x}$ and $\bar{y} = \frac{1}{y}$. The Hamiltonian equation is:

$$H = -J_z \sigma_z \otimes \sigma_z - B_x (\sigma_x \otimes I + I \otimes \sigma_x) - B_y (\sigma_y \otimes I + I \otimes \sigma_y) \quad (11)$$

Spin operators are defined as $\sigma_x = 1/2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_y = 1/2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\sigma_z = 1/2 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

3.2. Dimensional Reduction and Feature Extraction

The reduced feature $\hat{f}$ is the expectation value of the spin-correlation operator in the ground state $|\Psi_0\rangle$. This mapping projects classical data into non-linear correlations within the quantum state:

$$\hat{f} = R_{F(x,y)} \quad R = \langle \sigma_z \otimes \sigma_z \rangle \quad (12)$$

At any recursive reduction level L , the reduction is expressed as:

$$R_{F(R_{2i-1}^{L-1}, R_{2i}^{L-1})} \quad (13)$$

For $L = 1$, $R_{2i-1}^{L-1} = R_{2i-1}^0 = f_{2i-1}$, where f_{2i-1} is the initial feature set before reduction.

4. Results and Analysis

The stability of the quantum-inspired dimensional reduction was evaluated across a gradient of 512 to 1 dimension for averaging at least 100 trials of randomizing the original matrix before reductions. Our results reveal a fundamental phase transition in manifold stability governed by the dimensional compression ratio.

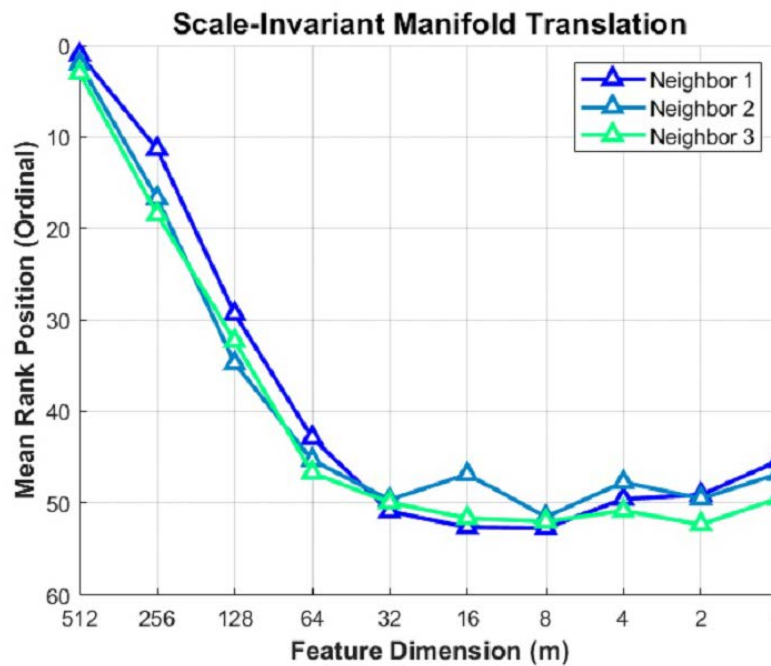


Figure 1. Tracks the ordinal position R of the k -closest neighbors. The parallelism in these trajectories illustrates the Cohesive Topological Phase ($m < 128$), where neighbors drift together as a 'solid pack' rather than scattering independently.

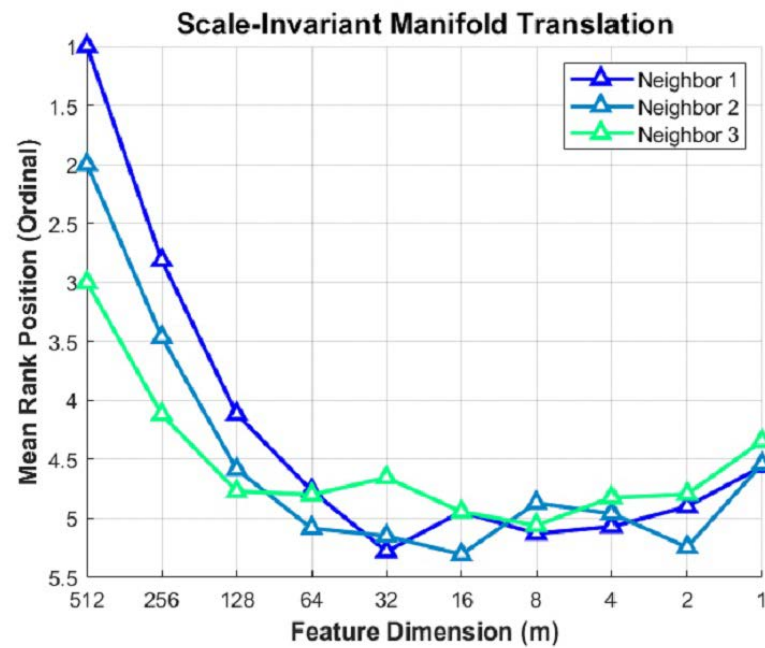


Figure 2. Tracks the ordinal position R of the k -closest neighbors for a small sample size of 10, breaking the small data curse.

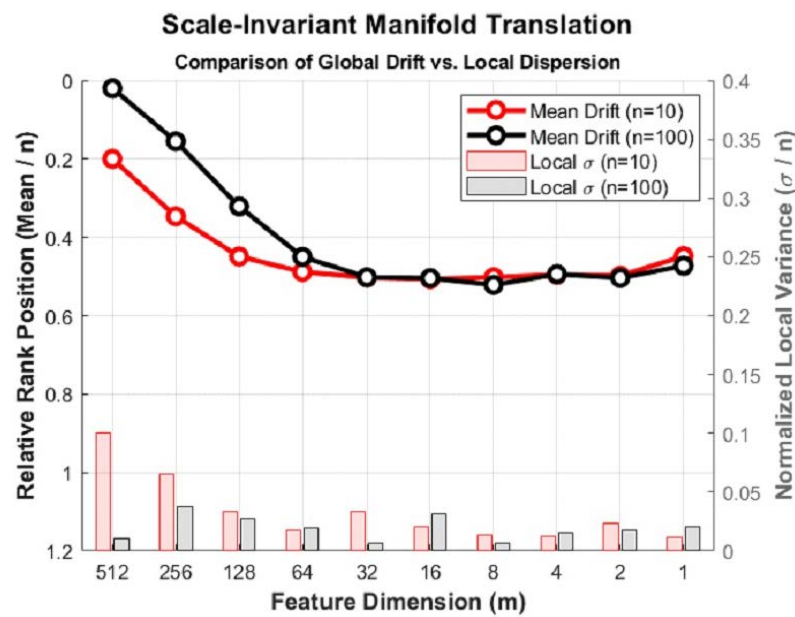


Figure 3. Ordinal rank trajectories of the three nearest neighbors. The inverted y-axis depicts the drift of neighbor identities from their origin. The high degree of parallelism between trajectories (verified by low local variance) confirms the Topological Solid behavior, where clusters translate as rigid bodies through the feature space without internal tearing.

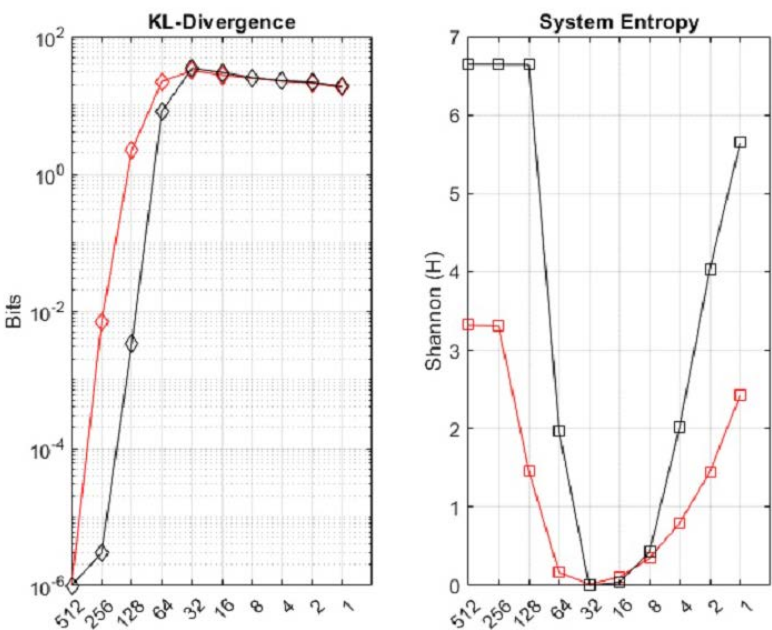


Figure 4. Evolutionary trajectory of information loss across the dimensional gradient. The logarithmic increase in KL-Divergence at $m=128$ marks the phase transition from geometric preservation to topological stability. The overlap between $n=10$ (red) and $n=100$ (black) demonstrates the scale-invariant nature of the information decay. Shannon entropy of the similarity distribution. The sharp decline (Entropy Collapse) between 256D and 32D illustrates the system converging toward a low-energy Hamiltonian ground state, effectively filtering stochastic noise while concentrating local feature correlations.

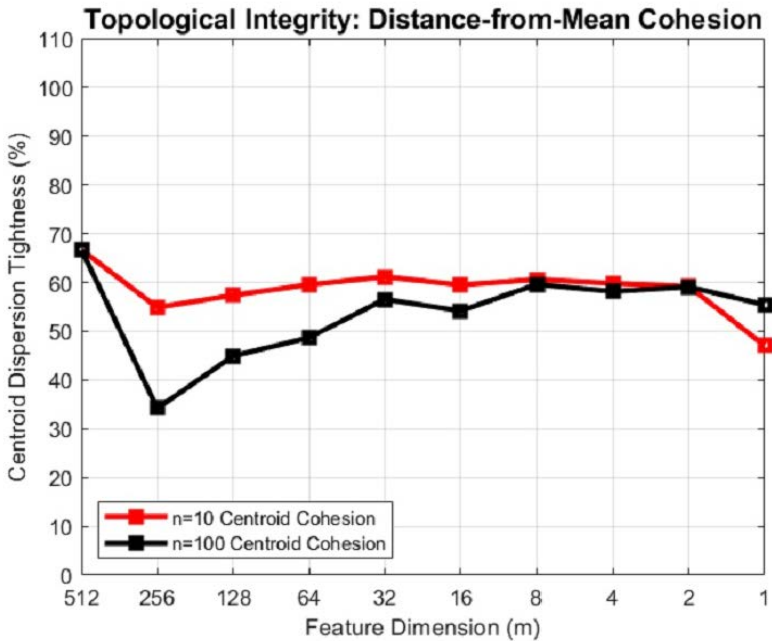


Figure 5. Measures the internal tightness of neighbor packs relative to their moving centers. Stability at approximately 60 % for $n=10$ at $m=8$ demonstrates the resilience of the local manifold.

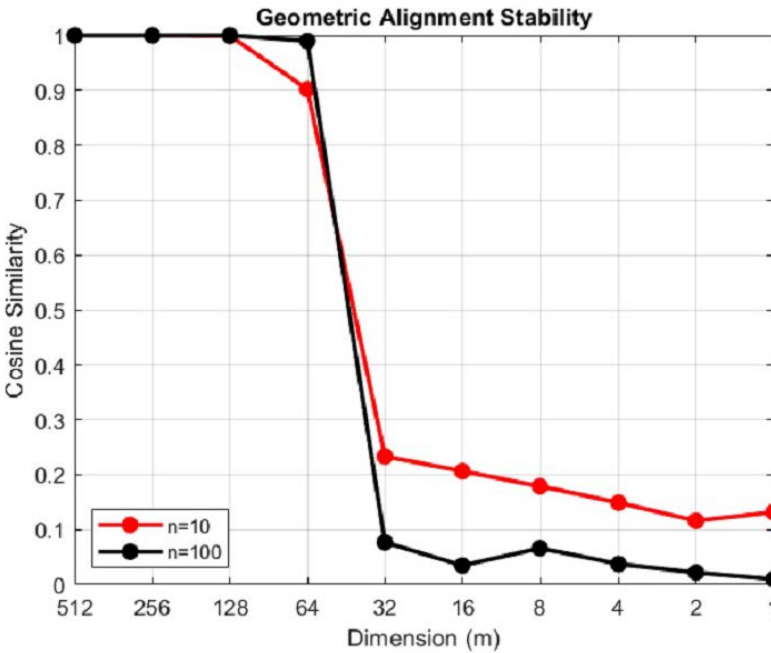


Figure 6. Visualizes the High-Fidelity Spin Phase (m 128) where near-unit cosine similarity is maintained, followed by ‘alignment collapse’ at the m=64 transition point.

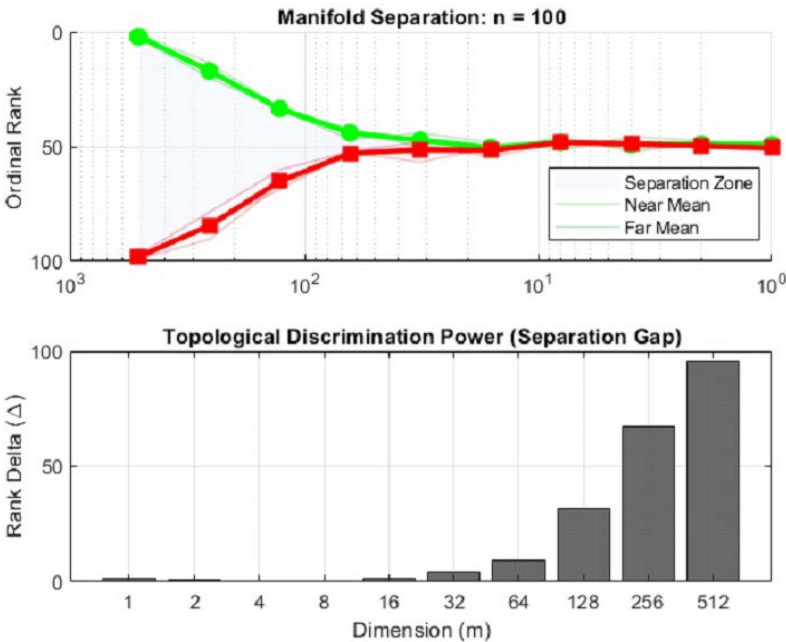


Figure 7. Contrasts near and far similarity samples. The Separation Zone illustrates the model’s capacity to suppress ‘manifold tearing’ and maintain distinct community identities and separation gap.

4.1. Global Geometric Alignment

As shown in Figure 6, the system maintains near-perfect geometric alignment during the initial stages of reduction. In the High-Fidelity Spin Phase (m > 128), the cosine similarity score remains above 0.99 for both n=10 and n=100 sample sizes. However, a rapid alignment collapse occurs at m=64, where the global geometric frame of reference fractures.

4.2. Local Topological Cohesion and Rank Drift

Despite the loss of global alignment, local neighborhood structures remain remarkably stable. Figure 5 illustrates the Distance-from-Mean Cohesion, measuring the internal tightness of neighbor

packs. At $m=8$, the system retains high relative cohesion 61.5 % for $n=10$ and 54.1 % for $n=100$. Figures 1 and 2 demonstrate that while neighbors drift significantly from their original ordinal ranks, they move along closely parallel trajectories. This rigid-body translation ensures that local community identities are preserved even at extreme compression. The transition between phases is further corroborated by information-theoretic diagnostics in Figure 4. KL-Divergence shows a sharp increase at the $m=128$ boundary, marking the point where global information starts to decay. An entropy collapse is observed as features align toward the Hamiltonian ground state. For $n=10$, this critical point occurs at 256D, whereas for $n=100$, it shifts to 128D.

4.3. Performance Metrics

Metric	Phase	$n = 10$	$n = 100$	Delta
Global Alignment	Ordinal (128D)	0.999	1.000	0.10%
Relative Cohesion	Topological (8D)	60.60%	59.60%	1.00%
Mean Rank Drift	Transition (64D)	4.88/10	45/100	3.80%
Local Variance (σ)	Extreme (1D)	0.12	2.08	0.90%
Entropy Collapse	Critical Pt.	256D	128D	N/A

5. Discussion

The spin alignment mechanism acts as a robust stabilizer for local manifolds. Unlike traditional reduction methods, where points may scatter independently, our Hamiltonian formulation forces features to align toward a collective ground state. The primary finding is the existence of a Topological Solid Phase, characterized by low local variance (below 4 %) despite high mean rank drift. This indicates that the technique effectively suppresses manifold tearing. This characteristic makes the technique uniquely suited for practical applications like community detection, where maintaining the relative pack identity is more critical than preserving exact absolute coordinates.

6. Conclusions

The spin alignment mechanism acts as a robust stabilizer for local manifolds. This framework functions as a precise ordinal compressor at low reduction levels and transitions into a resilient topological clustering engine at high ratios [1,2]. Performance metrics show minimal variance delta (typically < 6%) between different sample scales, confirming the stability of the Hamiltonian solver.

References

1. van der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research, 9, pp.2579-2605.
2. McInnes, L., Healy, J., Saul, N. and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software, 3(29), p.861.
3. Havlíček, V. et al. (2019). Supervised learning with quantum-enhanced feature spaces. Nature, 567, pp.209–212.
4. Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. Zeitschrift für Physik, 31, pp.253-258.
5. Tenenbaum, J.B., De Silva, V. and Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500), pp.2319-2323.
6. Belkin, M. and Niyogi, P. (2003). Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation, 15(6), pp.1373-1396.
7. Bronstein, M.M. et al. (2017). Geometric deep learning: going beyond Euclidean data. IEEE Signal Processing Magazine, 34(4), pp.18-42.
8. Schuld, M. and Killoran, N. (2019). Quantum Machine Learning in Feature Hilbert Spaces. Physical Review Letters, 122(4), 040504.
9. Vaswani, A. et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.
10. Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1), pp.79-86.

11. Coifman, R.R. and Lafon, S. (2006). Diffusion maps. Applied and Computational Harmonic Analysis, 21(1), pp.5-30.
12. Sachdev, S. (2011). Quantum Phase Transitions. Cambridge University Press.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.