

Étude de systèmes de spins fortement corrélés et désordonnés au moyen d'états quantiques neuronaux de fondation

Rapport final de PSC

Max Anglade, Nikola Audit, Rami Chagnaud, Nathan Dupuy, Adel Mana

Supervisé par Prof. Filippo Vicentini
avec l'aide d'Antoine Misery

24 avril 2026

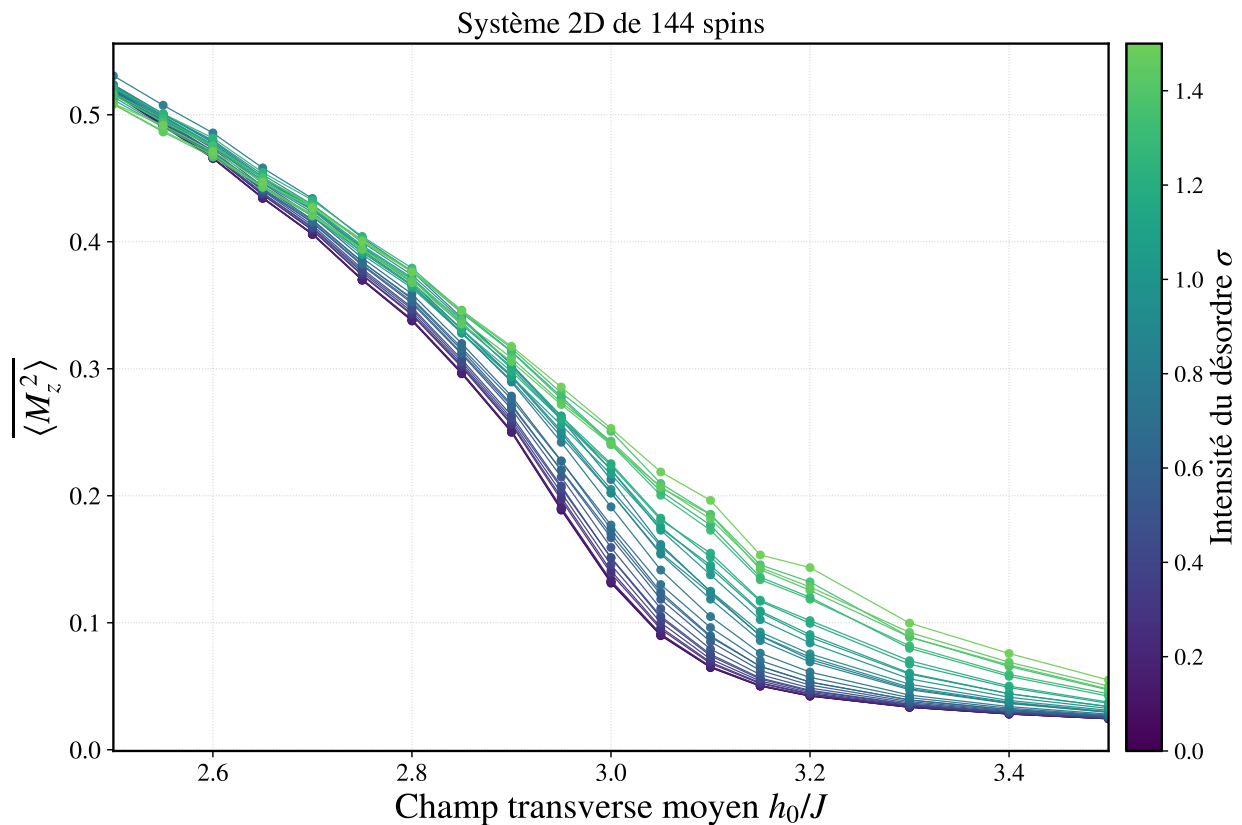


FIGURE 1 – Magnétisation au carré $\langle M_z^2 \rangle$ moyennée sur 80 réalisations du désordre en fonction du champ transverse moyen h_0 , pour différentes intensités du désordre σ , dans le modèle d'Ising en champ transverse désordonné gaussien 2D pour 144 spins.

Réaliser cette figure revient à approximer l'état fondamental de 100 000 hamiltoniens différents, chacun associé à un espace de Hilbert de dimension 2^{144} . Les méthodes développées dans notre PSC nous ont permis de réaliser cette tâche en 15 heures de calcul, en entraînant puis en évaluant un unique modèle de fondation, implémenté par un réseau de neurones de type *Transformer* à 200 000 paramètres.

Table des matières

1	Introduction	3
1.1	Motivations	3
1.2	Problème physique étudié	4
1.3	État de l’art	4
1.4	Méthodes utilisées	5
1.5	Objectifs	5
1.6	Chronologie du projet	5
1.7	Méthodes de travail et répartition des tâches	6
1.8	Plan du rapport	7
2	Objets et méthodes théoriques	7
2.1	Compression de l’état quantique	7
2.2	Monte Carlo Variationnel	8
2.3	Descente de gradient	8
2.4	Descente de gradient naturel	10
2.5	Ansatz variationnels	11
2.5.1	Ansatz de Jastrow	12
2.5.2	Machine de Boltzmann	12
2.5.3	Réseau de Neurones Convolutionnel (CNN)	13
2.6	Validation de la convergence : Analyse du V-score	13
2.7	Conditions aux bords	14
3	Transition de phase ferromagnétique-paramagnétique dans le modèle d’Ising en champ transverse sans désordre	14
3.1	Le Modèle d’Ising	15
3.2	La transition de phase ferromagnétique-paramagnétique	15
3.3	Protocole d’étude	16
3.3.1	Choix des ansatz	16
3.3.2	Entraînement des NQS	17
3.4	Résultats physiques concernant la transition de phase sans désordre	17
3.4.1	Magnétisation et transition de phase en dimension 1	17
3.4.2	Magnétisation et transition de phase en dimension 2	18
3.4.3	Énergie fondamentale : extensivité et linéarité	19
3.5	Coût computationnel	19
4	Méthodologie pour étudier le modèle d’Ising en champ transverse désordonné	20
4.1	Modélisation physique de la notion de désordre	20
4.2	Une nouvelle méthode computationnelle d’étude de la magnétisation du matériau	21
4.2.1	Le Modèle de Fondation	21
4.2.2	Un nouvel ansatz : Le Transformer	22
4.2.3	Une première validation : retrouver la transition de phase sans désordre	24
4.2.4	Application du modèle de fondation à l’étude du désordre	24
4.2.5	Vérifier la performance du réseau entraîné	24
4.2.6	Évaluer $\langle M_z^2 \rangle$ de façon numériquement efficace et rigoureuse : l’échantillonnage préférentiel (Importance Sampling)	26
4.2.7	Choix pratiques pour l’évaluation de la magnétisation	27
4.2.8	Sources d’incertitude et propagation	28
4.3	Problématiques rencontrées lors de l’élaboration de cette méthode	29
4.3.1	Mise en place de techniques d’échantillonnage adaptées	29
4.3.2	Choix des hyperparamètres du Transformer	31
4.3.3	Difficultés de convergence aux grandes tailles	32
4.3.4	Ressources computationnelles	33
5	Résultats finaux concernant l’influence du désordre sur la transition de phase ferromagnétique-paramagnétique dans le modèle d’Ising en champ transverse	34
5.1	Mise en évidence des transitions de phase	34

5.2	Validation des entraînements du réseau	35
5.3	Effet qualitatif du désordre sur le comportement du système	36
5.4	Effet quantitatif du désordre sur le comportement du système	38
5.5	Prolongement potentiel de notre étude	40
5.5.1	Entraînement du réseau	40
5.5.2	Evaluation du réseau	40
5.6	Potentiel d'exploitation et d'innovation	40
6	Évènements et ressources produites en lien avec notre PSC	41
6.1	Pour reproduire nos résultats	41
6.2	Construction d'un outil pédagogique et de visualisation	41
6.3	Participation à la publication de la bibliothèque <i>Netket Foundation</i> et de notebooks d'enseignement	41
6.4	Participation à un Hackathon d'informatique quantique à l'EPFL	41
7	Conclusion	42
8	Remerciements	43

1 Introduction

1.1 Motivations

Depuis son avènement au début du XXe siècle, **la physique quantique a permis une avancée technologique majeure**, qualifiée de première révolution quantique. Celle-ci repose sur la compréhension et le contrôle collectif des particules (électrons, photons), donnant naissance à des inventions majeures comme le transistor (base de toute l'électronique moderne) et le laser.

Aujourd'hui, nous assistons à une seconde révolution quantique. Contrairement à la première, elle vise à contrôler des systèmes quantiques individuels (atomes uniques, photons, spins) pour exploiter directement les principes de superposition et d'intrication. Cette révolution repose sur quatre piliers principaux : la communication (cryptologie quantique), les ordinateurs quantiques, les capteurs et les simulateurs quantiques. Les grandes puissances scientifiques déploient ainsi d'ambitieux projets contribuant à cette seconde révolution (par exemple le Plan Quantique en France).

Dans ces avancées, les systèmes étudiés sont complexes et composés de plusieurs sous-systèmes ou corps distincts. **L'étude d'un système de plusieurs corps et de leurs interactions est très complexe et en général insoluble analytiquement. Développer des méthodes numériques appropriées pour résoudre ce problème est donc nécessaire.** En mécanique quantique, toutes les propriétés d'un système sont encodées dans le spectre de son Hamiltonien, qui, s'il pouvait être diagonalisé, permettrait une prédiction relativement simple des propriétés statiques et dynamiques des systèmes quantiques, notamment à travers l'état fondamental. Une idée serait alors de diagonaliser exactement l'Hamiltonien par des méthodes numériques. Il suffirait alors de calculer les grandeurs physiques d'intérêt. Cependant, **la complexité des algorithmes exacts croît exponentiellement avec la taille des systèmes, les rendant inutilisables en pratique.**

En considérant par exemple un système de spins, nous constatons que la dimension de l'espace de Hilbert double à chaque particule ajoutée. Le plus grand supercalculateur actuel, Frontier, dispose de 9 PB (9×10^{15} octets) de mémoire. Chaque nombre complexe nécessitant 16 octets pour être stocké, Frontier peut contenir environ 5×10^{14} nombres complexes, soit un vecteur d'état complet pour seulement 49 particules de spin $\frac{1}{2}$. De plus, cette capacité ne permet que le stockage de l'état quantique, et non les calculs associés.

Ainsi, afin de tirer parti des opportunités offertes par la seconde révolution quantique, **il devient nécessaire de développer de nouvelles méthodes numériques permettant de surmonter l'explosion combinatoire de la dimension de l'espace des états.**

L'objectif de notre PSC est donc double : d'abord comprendre et développer une nouvelle méthode computationnelle pour traiter certains de ces systèmes complexes, puis l'appliquer à l'étude d'un système quantique désordonné.

1.2 Problème physique étudié

Au sein de ce vaste domaine qu'est l'étude du problème à N corps, **nous avons choisi de nous intéresser aux propriétés magnétiques de certains matériaux particuliers, et notamment à l'existence de transitions de phase ferromagnétique-paramagnétique.** Il s'agit d'un comportement potentiellement exploitable, car il signifie dans notre cas qu'une faible modification du champ magnétique auquel on soumet le matériau modifie complètement ses propriétés magnétiques. La démarche face à ce type de problème est de proposer des modèles essayant de rendre compte du phénomène étudié, puis de chercher s'ils s'approchent d'observations expérimentales. Trouver un modèle pertinent permet ensuite de prédire l'influence de la modification de certains paramètres expérimentaux.

Cependant, les systèmes physiques réels, à l'instar des matériaux macroscopiques, sont toujours sujets à des perturbations, par exemple en raison de leurs impuretés. Nous souhaitons donc **étudier dans quelle mesure l'introduction de désordre perturbe les comportements de transition de phase ferromagnétique-paramagnétique prédits par les modèles sans désordre.** Dans ce cadre, le *modèle d'Ising en champ transverse* est le plus simple pour tenter de décrire le magnétisme, et a déjà été largement étudié. Il était donc le candidat naturel pour notre première introduction du désordre dans un modèle bien connu.

Notre objectif scientifique sera donc d'étudier l'existence, ou non, d'une transition de phase ferromagnétique-paramagnétique dans le modèle d'Ising en champ transverse désordonné pour une chaîne en 1D et un réseau carré périodique en 2D. Néanmoins, afin d'atteindre cet objectif, nous allons d'abord devoir effectuer des développements méthodologiques, comme présenté ci-dessous.

1.3 État de l'art

Sur le plan méthodologique, plusieurs approches ont été développées pour surmonter le coût computationnel exponentiel et parvenir à approximer les grandeurs physiques d'intérêt des systèmes quantiques à N corps avec $N \gg 1$. Les méthodes Monte Carlo quantique (années 1950-60) [1] utilisent l'échantillonnage stochastique pour estimer les observables, elles ne permettent pas l'accès à l'état fondamental même. Le *Density Matrix Renormalization Group* (DMRG)[2] (années 1990 - 2000) a révolutionné l'étude des systèmes 1D grâce aux états produits de matrices (MPS), avec des extensions en 2D (PEPS) [3] aux coûts cependant exponentiels. Plus récemment, les états quantiques neuronaux [4, 5] exploitent les réseaux de neurones, offrant une expressivité élevée sans les limitations géométriques du DMRG. C'est cette approche que nous adoptons dans ce travail. Les *Foundation Neural-Network Quantum States (FNQS)* [6] permettent de résoudre le problème simultanément pour plusieurs hamiltoniens. Développés en 2024-2025, ils représentent la dernière avancée dans le domaine et seront centraux dans notre travail. Il a été démontré que les FNQS permettent d'étudier des systèmes désordonnés en 1D, mais les papiers sur le sujet [6, 7] sont récents et difficilement reproductibles, car ils n'explicitent pas une partie importante des détails techniques de mise en œuvre de la méthode. Comprendre tous ces détails pour reproduire puis essayer d'appliquer ces méthodes à des systèmes physiques plus complexes nécessite donc un long travail.

Concernant le problème physique étudié, notre modèle appartient à la classe des modèles d'Ising désordonnés en champ transverse, communément appelée *Random Transverse Field Ising Model* (RTFIM). La version 1D de cette famille de modèles a été largement caractérisée, notamment par les résultats asymptotiquement exacts de Fisher [8] obtenus par renormalisation à désordre fort. La situation en 2D est plus délicate. Plusieurs approches numériques ont été développées pour étudier des variantes de ce modèle avec désordre : méthodes de renormalisation à désordre fort [9], études *quantum Monte Carlo* (QMC) [10, 11]. Ces approches présentent toutefois chacune leurs limitations : les méthodes de renormalisation fournissent des propriétés critiques asymptotiques mais pas directement l'état fondamental à taille finie ; les méthodes QMC opèrent toujours à température finie et peuvent être mises en défaut pour des modèles présentant de la frustration ; les méthodes de réseaux de tenseurs restent coûteuses en 2D avec désordre. De plus, à notre connaissance, la version spécifique du RTFIM ferromagnétique avec désordre gaussien n'a pas été explorée dans la littérature.

Notre approche par états quantiques neuronaux de fondation constitue donc à la fois une alternative méthodologique aux approches existantes, et l'occasion d'explorer un modèle non traité auparavant.

1.4 Méthodes utilisées

Les méthodes modernes basées sur le Monte Carlo Variationnel (VMC) et les États Quantiques Neuronaux (NQS) fournissent des résultats prometteurs pour des systèmes de spins fortement corrélés de grande taille, avec ou sans désordre. Pour cette raison, nous les avons choisies pour notre étude, qui permettra donc aussi de tester, valider et analyser les performances de cette approche sur un problème non trivial.

Les techniques de **Monte Carlo Variationnel** contournent la dimensionnalité exponentielle de l'espace de Hilbert en décrivant l'état quantique par une fonction d'onde paramétrée $|\psi_\theta\rangle$ (un *ansatz*). L'estimation des observables, notamment l'énergie, repose alors sur une approche stochastique : les sommes exactes sont remplacées par des moyennes empiriques calculées sur un échantillonnage de configurations. En vertu du principe variationnel, l'état fondamental est approché en optimisant itérativement les paramètres θ afin de minimiser l'énergie du système.

Dans ce cadre, les **États Quantiques Neuronaux** constituent une classe d'ansatz particulièrement puissante. En encodant la fonction d'onde via un réseau de neurones, ils offrent une représentation compacte et très expressive, capable de capturer efficacement les corrélations complexes des systèmes physiques étudiés.

Pour être capables d'approcher numériquement le comportement de systèmes désordonnés, nous avons mis en œuvre des **Modèles de fondation** [6], qui consistent à entraîner un même réseau de neurones capable de prendre en argument des paramètres physiques variables (dans notre cas la valeur du champ transverse h en chaque site). En pratique, après avoir été entraîné sur des configurations de désordre données, ces modèles sont capables d'*interpolation*, c'est-à-dire d'approcher l'état fondamental de nouvelles configurations de désordre non fournies lors de l'entraînement. Pour disposer de l'expressivité nécessaire à la réalisation d'une tâche si complexe, nous avons employé une classe particulière de réseaux de neurones, appelés **Transformers**[12, 13], à l'état de l'art dans de nombreux domaines d'application de l'apprentissage profond, particulièrement en traitement du langage naturel et vision par ordinateur.

1.5 Objectifs

Notre objectif est double. Nous avons d'abord pour **objectif méthodologique de reproduire en pratique la méthode introduite en 2025 par Rende et al. [6] fondée sur les *Foundation NQS* pour étudier les systèmes quantiques désordonnés, identifier des paramètres satisfaisants pour atteindre une bonne précision à un coût computationnel raisonnable, et fournir un code utilisable implémentant ces outils**. Ce faisant, nous avons été amenés à **étendre la méthode de Rende et al. de deux manières**. Premièrement, nous combinons dans un même entraînement la paramétrisation continue en champ moyen h_0 (utilisée par Rende et al. sur le cas sans désordre) et la paramétrisation par réalisation du désordre (utilisée sur le cas désordonné à h_0 fixé). Deuxièmement, nous mettons en évidence et exploitons une propriété de généralisation hors-distribution sur l'intensité σ du désordre gaussien.

Atteindre ce premier objectif méthodologique permet ensuite de travailler sur notre **objectif scientifique : étudier les conséquences de l'introduction d'un désordre gaussien sur la transition de phase ferromagnétique-paramagnétique du modèle d'Ising en champ transverse en 1D et 2D**. Pour ce faire, nous avons dû adapter plusieurs aspects de la méthode présentée dans la littérature, et développer des solutions à certains problèmes spécifiques rencontrés.

1.6 Chronologie du projet

Les trois premiers mois de notre PSC nous ont permis d'atteindre plusieurs objectifs cruciaux. Nous avons d'abord compris les enjeux liés à la recherche de l'état fondamental de systèmes quantiques complexes, et l'apport considérable de l'approche utilisant des états neuronaux quantiques. Ensuite, **nous nous sommes consacrés à la compréhension des outils théoriques nécessaires à notre étude**, tels que le *Monte Carlo Variationnel* et la *Descente de Gradient Naturel*. En parallèle, nous avons exploré en profondeur différents ansatz et affiné notre compréhension des compromis intrinsèques à notre méthode. **L'implémentation et les simulations numériques réalisées sur des systèmes simples ont joué un rôle essentiel**, non seulement pour valider nos choix méthodologiques, mais aussi pour nous familiariser en pratique avec chaque outil. Elles nous ont aussi permis de justifier certains choix empiriques, et de vérifier la concordance entre nos travaux et la réalité physique que nous cherchons à modéliser.

Cette première période nous a donc permis d'acquérir le socle de connaissances théoriques et d'expérience pratique absolument nécessaires pour la suite.

Une fois toutes ces méthodes apprivoisées, nous avons pu passer à **une étude physique** du modèle de *Ising en champ transverse sans désordre* ce qui nous a pris environ un mois. Ce système demeure théoriquement assez simple, et ses propriétés sont bien connues, ce qui en fait un bon sujet d'étude préliminaire. Par conséquent, **nous avons remis en évidence une transition de phase ferromagnétique-paramagnétique pour ce modèle, à la fois en dimension 1 et 2**. Pour ce faire, nous avons dû recourir aux ordinateurs de l'X, car nos simulations devenaient trop coûteuses en puissance de calcul.

Cette première étude en profondeur d'un modèle physique à l'aide des méthodes de VMC et NQS constituait le jalon intermédiaire naturel que nous devons atteindre avant de prétendre être en mesure d'introduire du désordre. Cependant, s'agissant encore d'une étape d'apprentissage et de maîtrise des outils, nous n'avons pas cherché à la rendre aussi complète que celle du modèle désordonné, qui contient le modèle sans désordre avec un formalisme plus rigoureux.

Durant les trois derniers mois du projet, **nous avons finalement mis au point une méthode expérimentale adaptée à l'étude que nous souhaitons mener**, qui tient compte de toutes les spécificités des modèles de fondation et de la nature physique de la transition de phase que nous souhaitons étudier. Ce faisant, nous avons rencontré plusieurs problèmes de taille, auxquels nous avons dû apporter des solutions pratiques.

Une fois cette méthode établie, nous **l'avons finalement appliquée au modèle d'Ising en champ transverse désordonné**, ce qui constitue l'objectif scientifique de notre PSC. Nous avons ainsi constaté dans quelle mesure le désordre modifie la nature de la transition de phase ferromagnétique-paramagnétique de ce modèle. Il s'agit de l'étape la plus poussée du point de vue des concepts et des ressources computationnelles mobilisées.

Par ailleurs, **nous avons participé à des projets annexes grâce à notre PSC** : la création d'un outil de visualisation de l'entraînement d'un NQS, la **participation à la publication d'une bibliothèque numérique conçue pour implémenter des modèles de fondation à l'aide de NQS dans le Journal of Open Source Software**, et la participation à un hackathon d'informatique quantique à l'EPFL fin février.

Tout au long du projet, **notre avancée a été ralentie par une difficulté majeure, inhérente à toute méthode computationnelle : l'accès à des ressources computationnelles importantes**. En effet, le centre de calcul de l'X est en panne depuis plus d'un an, et en particulier inaccessible durant toute la durée de notre PSC. Nous avons donc dû nous contenter des salles informatiques de l'École, dont les performances sont très limitées en comparaison avec celle d'un centre de calcul dédié à la recherche professionnelle. Nous avons donc constamment dû trouver un compromis entre la volonté d'avoir des résultats aussi précis que possible, ce qui nécessite une grande puissance de calcul, et la réalité des moyens qui nous étaient accessibles. Nous avons parfois trouvé des moyens détournés pour contourner ce problème. Néanmoins, **pour lancer les simulations finales nous permettant d'atteindre notre objectif scientifique, nous avons payé pour avoir un accès momentané à un centre de calcul cloud** : runpods.io.

1.7 Méthodes de travail et répartition des tâches

Notre travail sur ce PSC a été rythmé par des réunions hebdomadaires d'environ **trois heures avec notre tuteur**, Filippo Vicentini, et Antoine Misery, respectivement enseignant-chercheur et doctorant au *Centre de Physique Théorique de l'X (CPHT)*. Elles consistaient en des cours sur les méthodes et outils théoriques que nous devons prendre en main, des présentations de nos travaux et résultats intermédiaires, et des discussions au sujet de nos incompréhensions ou des points délicats du sujet, ainsi que des pistes que nous souhaitons explorer. Entre ces réunions, nous réalisons des simulations, implémentions des méthodes numériques, faisons des recherches et lisons des articles sur de nouveaux sujets.

Afin d'acquérir une vision globale de tous les aspects des méthodes que nous utilisons, nous avons consacré les trois premiers mois à une phase d'apprentissage théorique. Nous sommes tous restés généralistes ; en effet, à ce stade, il aurait été prématuré de nous spécialiser, et aurait laissé de trop grosses zones d'ombre sur notre démarche pour chacun des membres du groupe.

Cependant, **le problème de la répartition des tâches s'est très vite imposé**. Par la suite, nous avons donc commencé à nous répartir plus systématiquement le travail, selon nos appétences

personnelles. Lors de la phase initiale, nous avons ainsi commencé par nous répartir les hyperparamètres à étudier. Ensuite, lors de l'étude du système sans désordre, nous avons séparé le travail à plus grande échelle : un groupe s'est spécialisé dans l'utilisation de la RBM et son application à la 1D, un autre a pris en main les CNN et les a appliqués à la 2D. Lors de l'élaboration de notre protocole d'étude du système désordonné, nous avons réparti les blocs principaux entre les membres du groupe, selon nos appétences personnelles : la compréhension théorique du Transformer et l'étude de ses hyperparamètres, la résolution de nos problématiques d'échantillonnage, la vérification des performances de nos modèles, l'implémentation de l'Échantillonnage Préférentiel (*Importance Sampling*) et la résolution des problèmes rencontrés en grande taille.

Nous avons toujours veillé à expliquer nos conclusions à tous les membres du groupe, pour que chacun ait toutes les clés de compréhension en main. De plus, notre répartition du travail est restée souple : une personne allait régulièrement en aider une autre si un sujet se révélait délicat, ou si elle avait un éclairage utile à apporter.

1.8 Plan du rapport

Nous présentons dans **la partie 2** les outils théoriques qui constituent les briques essentielles à la compréhension et à la mise en œuvre de nos méthodes VMC et FNQS. Dans la **partie 3**, nous retrouvons les résultats connus sur le modèle d'Ising en champ transverse sans désordre, en utilisant des NQS classiques. Ensuite, nous exposons la **partie 4** la méthodologie que nous avons développée pour appliquer les états quantiques neuronaux de fondation à l'étude du modèle d'Ising en champ transverse désordonné, et revenons sur les points durs que nous avons rencontrés en chemin ainsi que sur les solutions développées pour les résoudre. En **partie 5**, nous présentons nos résultats physiques mettant en évidence dans quelle mesure le désordre influe sur la transition de phase du modèle d'Ising en champ transverse. Enfin, nous évoquons en **partie 6** les projets annexes et événements auxquels nous avons pris part grâce à notre PSC.

2 Objets et méthodes théoriques

Notre PSC explore une approche prometteuse pour pallier les limitations de stockage et de calcul rencontrées lors de l'étude de systèmes quantiques complexes (systèmes de N spins, atomes de Rydberg...). Cette approche repose sur l'utilisation de l'apprentissage machine appliqué aux problèmes de physique quantique.

Cette méthode s'appuie sur trois piliers :

- la paramétrisation de l'espace de Hilbert de dimension 2^N par un espace réduit, via une compression non linéaire. Les états ainsi représentés sont appelés *états quantiques neuronaux* (NQS) [4, 5].
- l'approximation de grandeurs physiques comme l'énergie d'une fonction d'onde par une méthode d'échantillonnage de Monte Carlo utilisant l'algorithme de Metropolis-Hastings [4].
- la minimisation de l'énergie par une descente de gradient intégrée dans la méthode de Monte Carlo variationnelle (VMC) [4].

2.1 Compression de l'état quantique

Tout d'abord, afin de régler le problème de la croissance exponentielle de l'espace des états, nous allons compresser l'information. Soit W l'espace complexe des paramètres, on définit alors la fonction :

$$\Psi : W \rightarrow \mathcal{H}, \quad \theta \mapsto |\psi_\theta\rangle$$

On nomme ces fonctions les *ansatz variationnels* (ou *ansatz*).

Si $\dim[W] \sim \text{Poly}(N) \ll \dim[\mathcal{H}]$, alors la complexité n'est plus que polynomiale et le problème de complexité est réglé. Cela nécessite cependant de choisir une forme d'ansatz qui permet de représenter le plus d'états quantiques possible.

D'autre part, nous savons qu'en dimension finie, notre vecteur d'état peut être décomposé de manière unique sur la base des états qui sont dans notre système de spins 2^N . (Les x_i représentant les spins soit en haut ou en bas).

À partir de θ , pour pouvoir utiliser efficacement les ansatz, on doit pouvoir connaître la décomposition de l'état $|\psi_\theta\rangle$ dans la base. C'est pourquoi les ansatz sont généralement donnés sous la

forme :

$$\Psi_\theta(x_1, \dots, x_N) = \langle x_1, \dots, x_N | \psi_\theta \rangle$$

Nos ansatz codent donc une fonction des états de base vers \mathbb{C} . À chaque θ est associé un nouvel état car c'est alors une nouvelle décomposition dans la base des états de notre espace de Hilbert.

2.2 Monte Carlo Variationnel

Notre but principal est de calculer l'énergie fondamentale du système. Nous devons donc trouver un moyen de calculer cette énergie seulement avec les ansatz variationnels dépendants de $|\psi_\theta\rangle$.

Or, on peut montrer, en décomposant l'état $|\psi_\theta\rangle$ dans la base des états, qu'on a :

$$\langle E \rangle_\theta = \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \sum_x \frac{|\psi_\theta(x)|^2}{\langle \psi_\theta | \psi_\theta \rangle} \frac{\langle x | \hat{H} | \psi_\theta \rangle}{\langle x | \psi_\theta \rangle} = \sum_x p_\theta(x) H_{\text{loc}}^\theta(x) = \mathbb{E}_{p_\theta(x)} [H_{\text{loc}}^\theta(x)]$$

Avec

$$H_{\text{loc}}^\theta(x) = \frac{\langle x | \hat{H} | \psi_\theta \rangle}{\langle x | \psi_\theta \rangle}, \quad p_\theta(x) = \frac{|\psi_\theta(x)|^2}{\langle \psi_\theta | \psi_\theta \rangle}, \quad \psi_\theta(x) = \langle x | \psi_\theta \rangle$$

Ainsi $\psi_\theta(x)$ représente physiquement l'amplitude de probabilité associée à l'état x pour le paramètre θ .

De plus, nous allons utiliser les méthodes d'approximation de Monte Carlo :

$$\langle E \rangle_\theta = \mathbb{E}_{p_\theta(x)} [H_{\text{loc}}^\theta(x)] \approx \frac{1}{N_s} \sum_{x \in \{x\}} H_{\text{loc}}^\theta(x)$$

où N_s est le nombre d'échantillons et où x suit la loi $p_\theta(x)$.

Le calcul ci-dessus reste inchangé pour toute observable quantique arbitraire \hat{O} .

Si la variance de l'estimateur croît polynomialement avec la taille du système $\text{Var}[H_{\text{loc}}] \sim \text{Poly}(N)$, alors un nombre polynomial d'échantillons suffira pour toutes les tailles de système, et nous disposerons de méthodes de calcul dont le coût évolue de manière polynomiale.

La combinaison de la compression qui résout le problème de stockage et l'utilisation des méthodes Monte Carlo qui permet de diminuer la complexité de calcul se nomme *le Monte Carlo Variationnel*.

Afin de calculer la moyenne de l'énergie dans un état donné grâce à l'approximation de Monte Carlo, nous devons échantillonner des variables aléatoires x selon la loi $p_\theta(x)$. Or, d'après la formule trouvée pour $p_\theta(x)$, échantillonner des variables aléatoires selon cette loi nécessite à priori de calculer sa constante de normalisation $\langle \psi_\theta | \psi_\theta \rangle$.

Pour ne pas avoir à calculer cette constante de normalisation trop coûteuse numériquement, nous allons échantillonner la loi $p_\theta(x)$ avec l'algorithme de Metropolis-Hastings.

Cet algorithme nous permet de construire une chaîne de Markov qui possède une mesure invariante pré-déterminée, ici $p_\theta(x)$. En notant $P(x, y)$ la matrice de transition de cette chaîne de Markov et $Q(x, y)$ une matrice symétrique positive au choix, on a :

$$P(x, y) = Q(x, y) \min\left(\frac{|\phi_\theta(y)|^2}{|\phi_\theta(x)|^2}, 1\right)$$

En ajoutant pour chaque état x de la chaîne de Markov, le terme $H_{\text{loc}}^\theta(x)$ à la somme définissant $\langle E \rangle_\theta$, on aura, (même si les premiers états ne seront pas distribués selon $p_\theta(x)$) convergence vers $\langle E \rangle_\theta$.

2.3 Descente de gradient

La méthode de Monte Carlo variationnelle nous permet de calculer et minimiser l'énergie d'une fonction d'onde variationnelle $|\psi_\theta\rangle$. Notre objectif est d'approcher l'état fondamental du système, et en mécanique quantique, nous savons que $\forall \theta, E_\theta \geq E_0$.

Notre but est alors d'atteindre des valeurs d'énergie toujours plus basses car nous savons qu'elles ne seront que des bornes supérieures de l'énergie de l'état fondamental.

Nous réalisons donc une **descente de gradient** pour trouver le minimum global de l'énergie E en fonction de θ , fournissant ainsi une approximation de l'état fondamental, que l'on pourra noter $|\psi_\theta\rangle$.

À partir de l'itération n , nous savons alors quelle doit être la valeur de θ pour progresser à l'étape $n + 1$:

$$\theta^{(n+1)} = \theta^{(n)} - \eta \nabla_{\theta^*} E(\theta)$$

où η est notre pas d'apprentissage, ou *learning rate*.

De nouveau, en décomposant dans notre base d'états, il est possible de montrer que

$$\partial_{\theta_k^*} E = \frac{1}{N_s} J^\dagger \mathbf{E}_{\text{loc}}$$

avec $J_{ij} = \partial_{\theta_j} \log \psi_\theta(x_i)$ la matrice jacobienne et $\mathbf{E}_{\text{loc}} = H_{\text{loc}}(\sigma) - \langle E \rangle$ le vecteur d'énergies locales recentrées (de dimension N_s le nombre d'échantillons).

On peut également montrer que $V_\theta(H_{\text{loc}}(x_i)) = V(\hat{H})$. Cette formule est intéressante car elle nous dit que l'estimation de l'énergie devient de plus en plus précise à mesure que l'on converge vers l'état fondamental, fournissant ainsi un indicateur de convergence fondé sur la variance de l'estimateur (2.6) ainsi qu'une garantie de précision, nécessaire à la convergence de l'algorithme de minimisation

Étude préliminaire du learning rate

Nous avons alors écrit le code de la descente de gradient afin de nous familiariser avec la syntaxe de Netket et pour ne pas utiliser de suite les options déjà toutes prêtes de cette bibliothèque. Un paramètre critique de notre optimisation par descente de gradient est le learning rate. En effet, un learning rate trop petit, bien que permettant une optimisation plus précise, rend impossible une convergence vers l'état fondamental en un nombre raisonnable d'itérations (de l'ordre de 500). À l'inverse, bien qu'un learning rate trop grand puisse donner une convergence plus rapide loin de l'état fondamental, il limite la précision du résultat obtenu, et peut même empêcher la convergence (voir Fig. 2). Il y a donc un juste milieu qui ne peut être déterminé qu'empiriquement, afin d'utiliser le learning rate optimal pour la suite de notre étude.

Nos résultats

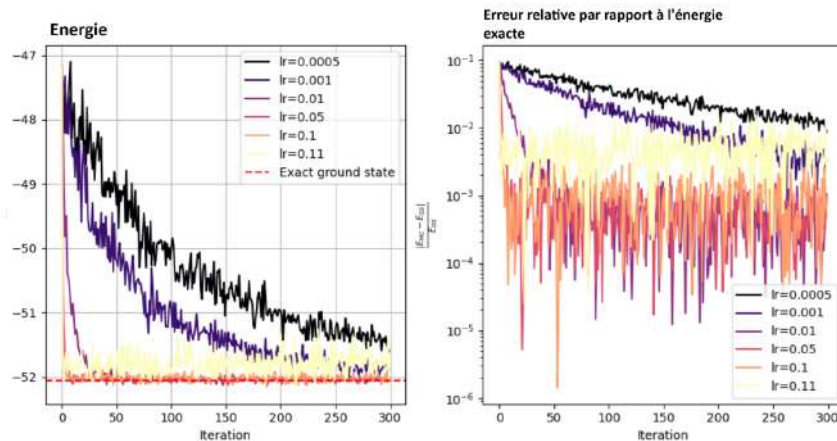


FIGURE 2 – Étude de l'influence du learning rate. Un learning rate trop grand donne une convergence imprécise, tandis qu'un learning rate conduit à une convergence trop lente. Pour la machine de Boltzmann appliquée à un hamiltonien de Ising, la valeur optimale se trouve vers 5×10^{-3} .

La Fig. 2 montre la vitesse de convergence d'une machine de Boltzmann (voir 2.5.2) vers l'état fondamental du système pour différents learning rates. L'erreur est ici calculée par rapport à la valeur exacte de l'énergie du fondamental, le système étudié étant suffisamment petit pour permettre la diagonalisation du Hamiltonien. On y voit l'existence d'un learning rate optimal aux alentours de 5×10^{-3} .

2.4 Descente de gradient naturel

La descente de gradient classique minimise l'énergie dans l'espace des paramètres de l'ansatz, qui n'ont pas de signification physique directe mais sont une simple représentation arbitraire d'un même état quantique. Ainsi, différentes paramétrisations d'un même état peuvent générer des gradients et trajectoires d'optimisation différentes. Dans un paysage énergétique non convexe et bruité par l'échantillonnage Monte Carlo, ces variations peuvent entraîner des vitesses de convergence et comportements numériques disparates, posant un problème physique et pratique.

Pour résoudre ces problèmes, nous avons recours à la descente de gradient naturel [14], qui opère un changement de métrique : au lieu de considérer une variation infinitésimale de θ qui minimise l'énergie, on considère une variation infinitésimale de l'état quantique sous la métrique de Fubini-Study (voir Fig. 3). Contrairement à la descente de gradient classique, qui ajuste les paramètres selon le gradient de l'énergie, ici l'objectif est de maximiser la baisse d'énergie par une modification infinitésimale de l'état quantique. Les paramètres sont alors mis à jour de manière à induire des variations contrôlées et physiquement significatives de l'état, indépendamment de leur représentation. **Cette approche garantit une optimisation plus stable et pertinente sur l'espace des états quantiques.** De plus, elle a l'avantage d'être invariante par reparamétrisation.

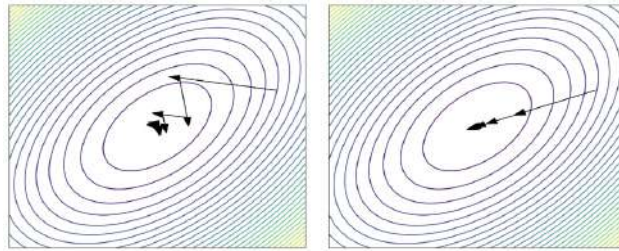


FIGURE 3 – Illustration de la descente de gradient classique à gauche, naturel à droite

Nous pouvons montrer que le changement des paramètres qui permet la modification souhaitée de l'état quantique à l'étape n de l'optimisation est :

$$\theta^{(n+1)} = \theta^{(n)} + \eta \dot{\theta}$$

avec η notre learning rate et $\dot{\theta}$ donné par la formule $\dot{\theta} = S^{-1} \nabla_{\theta} E$ où

$$S_{ij} = \left\langle \frac{\partial \ln \psi_{\theta}}{\partial \theta_i} \middle| \frac{\partial \ln \psi_{\theta}}{\partial \theta_j} \right\rangle, \quad \text{donc} \quad S = J^{\dagger} J, \quad \text{avec} \quad J_{ij} = \partial_{\theta_j} \log \psi_{\theta}(x_i)$$

Nous avons réexprimé S à l'aide de la matrice jacobienne dont nous avons déjà implémenté le calcul dans l'algorithme de descente de gradient.

Cependant, un problème peut apparaître avec l'équation qui définit $\dot{\theta}$. En effet, si les valeurs propres de S sont trop petites, celles de son inverse seront très grandes ce qui peut conduire à des valeurs trop grandes pour $\dot{\theta}$ et donc non convenables.

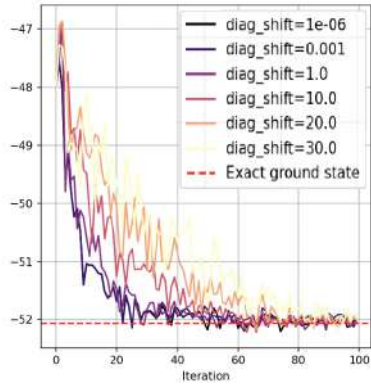
Étude préliminaire de l'impact du diagonal shift Pour remédier à ce problème, nous introduisons un paramètre : le *diagonal shift* λ lève la singularité de S . Pour avoir $\dot{\theta}$, nous résolvons :

$$\dot{\theta} = (S + \lambda I)^{-1} \nabla_{\theta} E$$

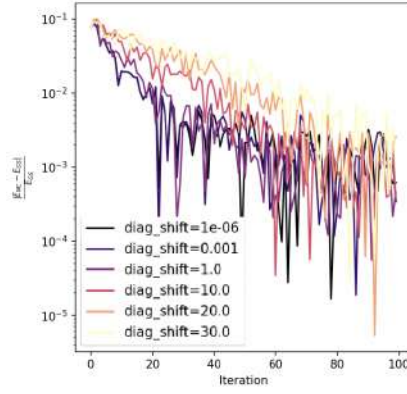
Nous savons calculer S et $\nabla_{\theta} E$ à chaque itération, il suffit donc d'inverser la matrice, par exemple avec la décomposition de Cholesky (X1895) qui décompose la matrice définie positive $(S + \lambda I)^{-1}$ en un produit de matrices triangulaires supérieures ce qui permet d'obtenir l'inverse après résolution de deux systèmes triangulaires, facilement implémentable numériquement.

Nos Résultats

La Figure 4 montre que plus le diagonal shift est gros, plus il ralentit la convergence de l'énergie (car la matrice diagonale λI efface alors l'influence de la matrice S), et qu'un faible diagonal shift



(A) Energie en fonction du nombre d'itérations pour plusieurs diagonal shift



(B) Evolution de l'erreur relative en fonction du nombre d'itérations (échelle log)

FIGURE 4 – Étude de l'influence du diagonal shift dans la minimisation de l'énergie par descente de gradient naturel, l'ansatz utilisé est ici Jastrow

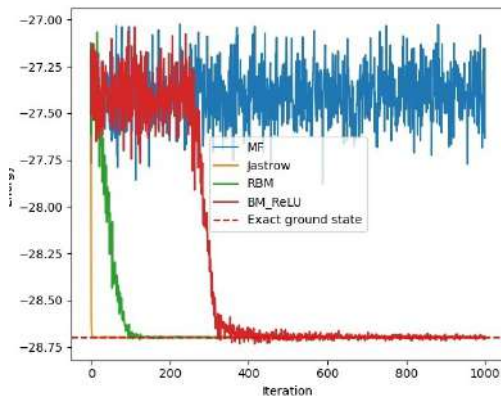
n'empêche pas la convergence, et évite les problèmes de valeurs propres presque nulles. Il nous faut donc, là encore, trouver le bon compromis.

Cependant, avant d'appliquer cette descente de gradient naturel, nous avons dû soigneusement choisir l'ansatz, car ils comportent chacun des hypothèses sur les états quantiques et ne peuvent donc pas décrire le même ensemble d'états.

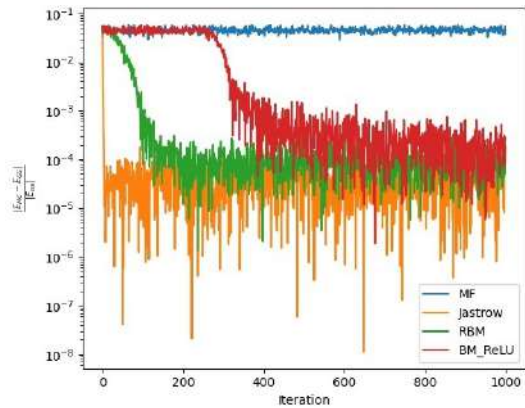
Notre objectif est donc de trouver l'ansatz qui permet de décrire le plus grand ensemble d'états et qui s'approche autant que possible de l'état du fondamental.

2.5 Ansatz variationnels

L'architecture et la complexité d'un ansatz conditionne sa capacité à approcher l'état fondamental d'un hamiltonien donné, et du système physique qui lui est associé. A titre d'exemple, la figure 5A montre les différentes vitesses de convergence de différents ansatz qui seront développés ci-dessous. L'étude est faite sur le hamiltonien de Ising (voir la partie 3.1), pour un petit système 1D puis 2D. L'ansatz de champ moyen (MF) étant trop peu performant et inintéressant pour notre étude, il ne sera pas expliqué par la suite. On y voit clairement que la vitesse et la qualité de convergence sont très différentes d'un ansatz à l'autre, ce qui justifie l'attention que nous apporterons par la suite au choix de notre ansatz.



(A) Evolution de l'énergie en fonction du nombre d'itérations



(B) Evolution de l'erreur relative en fonction du nombre d'itérations (échelle log)

FIGURE 5 – Comparaison des performances de différents ansatz sur le hamiltonien de Ising. Les ansatz sont le champ moyen (MF), l'ansatz de Jastrow (Jastrow), la machine de Boltzmann (RBM), une machine de Boltzmann à deux couches (BM-ReLU). Il est ici clair que Jastrow performe le mieux. Cependant, cela s'explique par la petite taille du système étudié et donc la simplicité des interactions en jeu. Jastrow perd rapidement en performance avec la taille du système.

2.5.1 Ansatz de Jastrow

Tout d'abord, le premier ansatz que nous avons manipulé est celui de Jastrow :

$$\phi_J = \exp\left(X^\top J X\right) = \exp\left(\sum_{i=1}^n \sum_{j=1}^n X_i J_{ij} X_j\right)$$

$X \in \mathbb{R}^n$ est un vecteur colonne qui représente les spins haut ou bas et $J \in \mathbb{R}^{n \times n}$ est une matrice symétrique. Cette fonction nous donne la valeur du coefficient du vecteur de la base $X \in \mathbb{R}^n$ pour un J fixé.

Le paramètre θ est donc ici la matrice J .

Physiquement, cet ansatz est symétrique par échange de deux particules. Les termes non diagonaux J_{ij} reflètent les interactions entre spins.

Enfin, nous pouvons observer que **les interactions ne s'effectuent que 2 par 2 car les termes sont en $X_i J_{ij} X_j$ liant le spin sur le site i et le spin sur le site j** . Nous pouvons donc encore améliorer notre ansatz pour qu'il représente mieux les interactions entre spins.

2.5.2 Machine de Boltzmann

Ensuite, afin de dépasser les limitations de Jastrow qui ne peut pas représenter une interaction à plus de 2 particules et donc inutilisable pour un mouvement collectif des spins, nous avons utilisé une machine de Boltzmann.

$$\log \Psi_{\text{RBM}}(X) = \sum_{k=1}^{\alpha N} \log \cosh\left(b_k + \sum_{i=1}^N W_{ik} x_i\right)$$

avec N le nombre de spins, α le ratio définissant le nombre de neurones cachés $M = \alpha N$, $b \in \mathbb{R}^{\alpha N}$ le vecteur des biais cachés, et $W \in \mathbb{R}^{\alpha N \times N}$ la matrice des poids.

La machine de Boltzmann généralise l'ansatz de Jastrow en n'imposant pas que les interactions s'effectuent 2 par 2.

Une machine de Boltzmann est en réalité un réseau de neurones avec une couche visible et une couche cachée.

Le paramètre θ que nous optimisons au cours de la descente de gradient est ici les poids des liaisons entre les neurones contenus dans la matrice W ainsi que les biais visibles et cachés contenus dans les vecteurs a et b .

Nous avons alors entrepris d'étudier l'influence de α sur l'optimisation, constituant le seul hyperparamètre de notre RBM. Un hyperparamètre est un paramètre fixé avant l'optimisation, il n'est pas appris, il configure simplement l'architecture globale de l'ansatz.

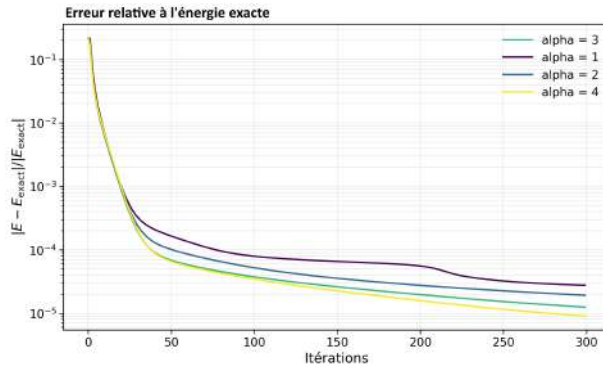


FIGURE 6 – Étude de l'influence de α dans la RBM, descente de gradient exacte sans échantillonnage pour isoler l'influence de α . Augmenter α diminue le nombre d'itérations nécessaires pour atteindre la convergence, mais il est cependant important de souligner que cela se fait au prix d'itérations plus coûteuses. Il faut donc faire un compromis entre nombre d'itérations et coût des itérations

La Fig. 6 montre qu'augmenter α permet une convergence plus précise, cependant il est important de souligner que le temps de calcul pour effectuer une itération augmente polynomialement avec α .

2.5.3 Réseau de Neurons Convolutionnel (CNN)

Bien que la machine de Boltzmann puisse représenter des interactions plus complexes que l'ansatz de Jastrow, elle n'a que deux couches de neurones et ne peut pas représenter pleinement la complexité des interactions entre spins en deux dimensions.

Augmenter le nombre de couches et de neurones garantit, en vertu du théorème d'approximation par des réseaux de neurones [15] [16], des capacités d'approximation croissantes. L'information sera plus subtile et donc on saisira une plus grande partie de la physique du problème. Cependant, le nombre de paramètres explose et les temps de calcul deviennent trop lourds pour présenter un réel intérêt.

Alors que la machine de Boltzmann envisage toutes les interactions entre spins en traitant le système dans sa globalité, le réseau de neurones convolutionnel tire parti du fait que les interactions sont essentiellement locales.

Au lieu d'envisager toutes les corrélations à longue distance dès la première couche, le CNN utilise des filtres locaux appelés noyaux de convolution (*kernels*).

- **Les Kernels** : Ce sont de petites matrices de poids (de taille ajustable) qui parcourent la configuration d'entrée pour détecter des motifs spécifiques (corrélations locales, alignements de spins, etc.). Il agit comme un détecteur de motifs locaux : il met en évidence certaines structures physiques, telles que des corrélations ou motifs d'interaction entre sites voisins.
- **Feature Maps et Channels** : Chaque kernel génère une "carte de caractéristiques" (*feature map*). En multipliant le nombre de filtres via différents *channels*, le réseau peut apprendre simultanément une grande variété de motifs physiques.

Le paramètre θ que nous optimisons au cours de la descente de gradient est ici l'ensemble des paramètres de kernel. Contrairement à la RBM où le nombre de connexions explose avec la taille du système, le nombre de paramètres d'un CNN dépend de la taille des kernels et non de celle de la grille. Ce partage de poids (*weight sharing*) réduit drastiquement le coût computationnel et facilite l'optimisation par descente de gradient.

En empilant plusieurs couches de convolution, le réseau agrège progressivement les informations locales pour construire des représentations de plus en plus abstraites et globales. Cette structure profonde permet de saisir des phénomènes physiques complexes que l'architecture peu profonde de la RBM ne peut atteindre. Enfin, les CNN sont équivariants par translation : en translatant le système d'entrée d'un vecteur, le système de sortie sera translaté de ce même vecteur. Dans le cas de réseau symétrique et invariant par translation, pour une taille infinie, on comprend donc que les CNN s'adaptent correctement à la géométrie de notre système.

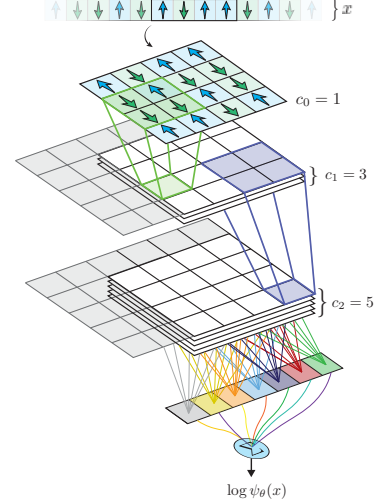


FIGURE 7 – Architecture d'un CNN, schéma issu de [17]

2.6 Validation de la convergence : Analyse du V-score

Pour quantifier la fiabilité de nos simulations et valider la précision de l'Ansatz variationnel, nous utilisons le *V-score* [18]. Il s'agit d'un indicateur de référence défini comme la variance de l'énergie normalisée par son carré :

$$V\text{-score} = \frac{\text{Var}(E)}{\langle \hat{H} \rangle^2} = \frac{\langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2}{\langle \hat{H} \rangle^2}$$

L'utilisation de la variance comme critère strict de convergence repose sur une propriété fondamentale de la mécanique quantique. En décomposant l'état $|\Psi\rangle$ sur la base des états propres $|n\rangle$ d'énergies E_n (soit $|\Psi\rangle = \sum_n c_n |n\rangle$), la variance de l'Hamiltonien s'écrit :

$$\sigma_H^2 = \langle \Psi | \hat{H}^2 | \Psi \rangle - \langle \Psi | \hat{H} | \Psi \rangle^2 = \sum_n |c_n|^2 (E_n - \langle \hat{H} \rangle)^2$$

Cette égalité montre que $\sigma_H^2 = 0$ si et seulement si $|\Psi\rangle$ est un état propre pur du système. Dans

le cadre de la méthode variationnelle, minimiser l'énergie tout en réduisant le V-score garantit que la fonction d'onde optimisée ne se contente pas de trouver un minimum local de l'espérance, mais converge structurellement vers le véritable état fondamental. Une simulation est considérée comme convergée dès lors que $V\text{-score} \leq 10^{-3}$, et d'autant plus précise que cette inégalité est largement respectée.

2.7 Conditions aux bords

Intéressons-nous au choix des conditions aux bords pour nos simulations. Comme évoqué plus haut, nos méthodes permettent de modéliser des systèmes physiques de taille finie, de l'ordre d'une centaine de spins maximum en temps raisonnable. Néanmoins, **les comportements que nous cherchons à mettre en lumière et à étudier sont ceux de matériaux réels à la limite thermodynamique, ce qui correspond à une taille infinie $L \rightarrow +\infty$.**

Une façon de faire consiste à **choisir des conditions aux bords périodiques** pour notre réseau de spins, ce qui revient à juxtaposer une infinité de fois un même système physique d'un nombre fini de spins, et à étudier son comportement. Cette interprétation laisse penser que le comportement de systèmes finis avec des conditions aux bords périodiques s'approchera mieux de celui de systèmes infinis qu'avec des conditions aux bords ouvertes. Cette propriété des conditions aux bords périodiques est confirmée par la Figure 8.

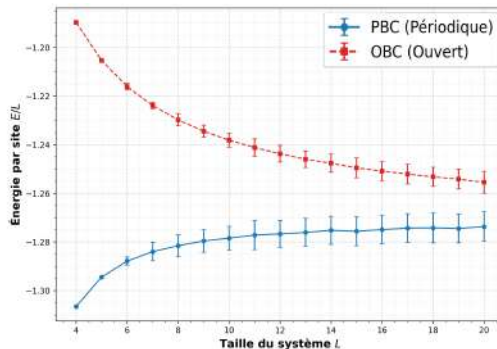


FIGURE 8 – Comparaison de l'énergie de l'état fondamental du système en fonction de la taille L pour les conditions aux bords périodiques et ouvertes.

Il est clair que les deux courbes doivent converger vers la même valeur (car les conditions de bords n'ont pas d'importance pour un système infini), et on constate sur la Figure 8 que le modèle périodique atteint plus rapidement son asymptote. Les résultats obtenus par ce modèle approcheront donc mieux des systèmes de grande échelle.

Ces observations justifient que nous utiliserons toujours des conditions aux bords périodiques par la suite.

3 Transition de phase ferromagnétique-paramagnétique dans le modèle d'Ising en champ transverse sans désordre

Fort de ces connaissances théoriques et de la compréhension des difficultés et enjeux pratiques de nos méthodes, nous pouvons désormais étudier le modèle d'Ising en champ transverse, avec des conditions aux bords périodiques, qui décrit l'interaction de N spins par une constante de couplage J entre plus proches voisins, le système étant soumis à un champ magnétique transverse d'amplitude h . Les résultats produits dans cette partie sont déjà bien connus, mais cette première étude nous a permis d'apprendre à mettre en œuvre nos méthodes dans la pratique et de valider leur principe.

L'hamiltonien de ce modèle est donné par

$$\hat{H} = J \sum_{\langle i,j \rangle} \hat{S}_i^z \hat{S}_j^z + h \sum_{i=1}^N \hat{S}_i^x$$

où le symbole $\langle \dots \rangle$ désigne les paires de plus proches voisins. Nous étudions le cas d'un système ferromagnétique, c'est-à-dire $J < 0$. Par souci de simplification, dans la suite, nous prendrons toujours $J = -1$, car le rapport h/J suffit à caractériser le comportement du système.

3.1 Le Modèle d'Ising

Comme expliqué dans l'introduction, nous souhaitons décrire le comportement réel de matériaux à l'aide du modèle d'Ising. Analysons donc plus en profondeur les hypothèses et la modélisation physique derrière ce modèle.

Dans l'Hamiltonien d'Ising, nous observons la présence de 2 termes distincts. Le premier terme représente l'interaction entre deux spins orientés selon l'axe z , l'énergie étant minimale lorsque les deux spins sont orientés de la même manière. Le deuxième terme correspond à l'interaction d'un spin avec un champ magnétique transverse donnée par $H = -S \cdot h$ où S est le vecteur de spins (dans notre hamiltonien, nous avons pris un signe positif devant le champ transverse mais cela revient à renverser l'axe des z).

Bien qu'il puisse paraître à première vue trop simpliste, le modèle décrit remarquablement bien le comportement de certains matériaux réels. Par exemple, le composé de niobate de cobalt (CoNb_2O_6) s'organise sous la forme de chaînes de spins quasi-1D. Lorsqu'on applique un champ magnétique perpendiculaire (transverse) à ces chaînes, on observe expérimentalement la transition de phase quantique prédite par le modèle de base, prouvant que cette théorie s'applique bien au monde physique expérimental [19].

Cependant, ce modèle présente des limites :

- Tout d'abord, comme l'hamiltonien d'interaction entre 2 spins est donné par : $H = -J S_A \cdot S_B$, l'expression du premier terme montre que ce modèle ne prend que la composante selon z de cette interaction. C'est donc un modèle à forte anisotropie.
- Le modèle ne prend en compte que les interactions avec le premier voisin, ce qui est une hypothèse justifiable en première approximation étant donné la distance caractéristique d'action des forces microscopiques.
- Finalement, le champ magnétique n'est appliqué que selon un axe ce qui est possible en pratique mais manque de généralité. En effet, si un matériau que nous prenons par exemple cristallin, présente plusieurs orientations structurales différentes, plusieurs zones d'orientation similaire des spins, le champ magnétique extérieur ne pourra pas être transverse à toutes ces phases.

Un modèle plus complet serait donc le modèle d'Heisenberg qui s'écrit :

$$H = \sum_{i,j} J_{ij} \vec{S}_i \cdot \vec{S}_j + \sum_i \vec{h} \cdot \vec{S}_i$$

Cependant, ce modèle et le modèle d'Ising font également l'hypothèse que les spins ne peuvent pas se déplacer et qu'ils sont un par site. Or, nous savons par le principe de Pauli que si nous considérons les fermions d'un métal, ils peuvent être deux sur le même site avec des nombres quantiques distincts. Nous pourrions alors considérer pour enrichir notre modèle que les sites soient occupés par une, deux ou aucune particule.

En somme, le modèle d'Ising, malgré ses limites, s'avère être une bonne modélisation de certains matériaux magnétiques. **En outre, il constitue un modèle classique de matériaux, qui a été très étudié ce qui nous permettra de tester les méthodes numériques présentées précédemment ainsi que plusieurs ansatz différents.** Etant plus simple, nous avons donc choisi de commencer par étudier ce modèle afin de nous familiariser avec les méthodes et objets numériques et étudier la transition de phase sans désordre.

3.2 La transition de phase ferromagnétique-paramagnétique

Revenons tout d'abord brièvement sur la notion de transition de phase et son application à notre modèle.

Une transition de phase correspond à un changement abrupt des propriétés macroscopiques d'un système thermodynamique, induit par une variation continue d'un paramètre de contrôle (température, pression, champ magnétique...). Cette transition est associée à une non-analyticité des fonctions thermodynamiques, en un paramètre critique. En ce paramètre critique, la longueur de corrélation entre les sites peut diverger également vers l'infini, dans le cas des transitions de phase d'ordre 2, comme expliqué ci-dessous.

Nous pouvons définir pour chaque transition de phase un paramètre d'ordre qui vérifie trois propriétés :

- il est nul dans la phase "non ordonnée"
- il est non nul dans la phase "ordonnée"
- il change de comportement à la transition soit de manière continue dans le cas d'une transition de phase d'ordre 2 soit de manière discontinue si la transition de phase est d'ordre 1.

Pour autant, la présence d'un paramètre qui vérifie ces conditions à N fini n'est pas suffisant pour statuer sur la présence d'une transition de phase car cela peut être seulement un changement de régime du système. **Il nous faudra donc, dans l'hypothèse de l'existence d'une transition de phase, tester des tailles de plus en plus grandes pour voir si la transition devient de plus en plus marquée ou "stagne"**. Si elle est de plus en plus marquée, cela signifie que la transition est bien une transition de phase s'accompagnant d'une non-analyticité de grandeurs thermodynamiques au paramètre critique.

Afin de comprendre plus en détail quelle observable ou paramètre d'ordre étudier et avoir une intuition sur ce que nous pourrions avoir comme résultat, commençons par comparer le modèle d'ising classique et quantique.

Transition de phase dans le cas classique : Dans le cas classique, le modèle a été très étudié et l'Hamiltonien est donné par :

$$H = J \sum_{\langle i,j \rangle} S_i S_j$$

Dans le cas classique sans champ magnétique, la transition de phase n'existe que pour des dimensions $d > 1$ et est par rapport à la température. Le paramètre d'ordre à étudier est alors $m = \langle M_z \rangle$.

Retour au cas quantique Dans notre cas, c'est le champ transverse qui est responsable de la transition de phase. La transition de phase se produit entre une phase ferromagnétique à petit h , régie par les interactions entre spins et une phase paramagnétique régie par l'interaction des spins avec le champ transverse.

Nous constatons également que $\langle M_z \rangle$ ne peut plus être le paramètre d'ordre puisque par symétrie du problème selon l'axe z , la magnétisation moyenne est toujours nulle : nous ne pouvons donc pas l'utiliser pour distinguer les deux phases. A la place, nous étudions $m = \langle M_z^2 \rangle$. Nous nous attendons à ce qu'il soit strictement positif pour des valeurs faibles de champ magnétique transverse et égal à 0 sinon puisque les états seront, en analogie avec le cas classique, alignés avec le champ extérieur selon x . **En somme, nous sommes ramenés à étudier l'évolution de M_z^2 en fonction du champ transverse, pour étudier l'éventuelle présence d'un changement de son comportement.**

Finalement, la notion d'exposant critique quantifie le comportement du paramètre d'ordre proche du paramètre critique, dans notre cas le champ transverse. Dans un système de taille macroscopique, pour le nombre de particules $N \rightarrow \infty$:

$$m \sim |h - h_c|^\beta$$

avec β l'exposant critique lié au paramètre critique.

3.3 Protocole d'étude

3.3.1 Choix des ansatz

Pour cette étude, en 1D nous avons utilisé l'ansatz *Restricted Boltzmann Machine*, qui fonctionne bien pour ce problème et reste assez simple.

Contrairement au cas unidimensionnel, le modèle d'Ising en champ transverse sur un réseau bidimensionnel ne dispose pas de solution analytique exacte simple. La complexité du problème augmente considérablement en raison de la croissance du nombre de degrés de liberté et de la nature des corrélations quantiques. Pour étudier cette grille de $L \times L$ spins, nous utilisons cette fois un réseau de neurones convolutionnel (CNN). Ce choix s'explique par le fait que la CNN s'adapte bien physiquement à une grille 2D : elle exploite ses symétries comme l'invariance par translation du réseau et est particulièrement efficace pour capturer les corrélations spatiales locales.

3.3.2 Entraînement des NQS

Notre protocole de simulation part d’une remarque importante : les cas limites $h = 0$ et $J \ll h$ sont plus simples, car le hamiltonien est alors régi par un terme dominant. En revanche, pour h et J du même ordre, les effets des deux termes sont en compétition, ce qui engendre plus de complexité dans le comportement du système. En conséquence, notre réseau de neurones aura plus de mal à apprendre l’état fondamental du système dans ce cas, et nécessitera davantage de paramètres. Pour cette raison, nous avons effectué nos choix d’hyperparamètres à $h = 1$ en 1D, et $h = 3$ environ en 2D : pour chaque valeur de la taille L du système, nous avons commencé par régler nos hyperparamètres pour obtenir une convergence suffisamment précise pour ce h dans la zone critique. Une fois le jeu d’hyperparamètres correctement réglé, nous avons lancé des optimisations pour chaque valeur de h souhaitée afin de tracer les courbes ci-après. Chaque couple de valeurs (h, L) correspond donc à un entraînement distinct d’un ansatz variationnel, contrairement à ce que nous présentons dans la partie avec désordre.

Il reste cependant une nuance à apporter : en 1D, augmenter α pour notre RBM augmente considérablement le temps de calcul nécessaire à l’entraînement de notre ansatz, car celui-ci a plus de paramètres. Pour ne pas rallonger inutilement les calculs, pour les plus grandes valeurs de L , nous prenons $\alpha = 5$ uniquement pour des valeurs de h proches de 1, et plus loin de $h = 1$ nous réduisons α .

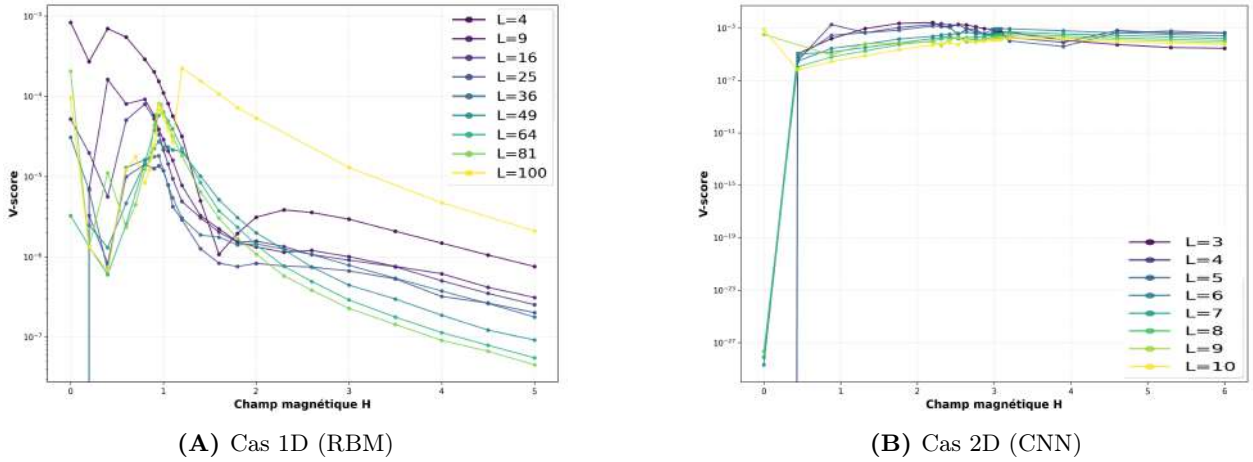


FIGURE 9 – Évolution du V -score (échelle logarithmique) en fonction de h pour différentes tailles de système. Toutes les simulations respectent le critère $V < 10^{-3}$.

L’observation comparée des Figures 9A et 9B appelle plusieurs commentaires :

- **Convergence validée** : pour chaque entraînement, le V -score reste inférieur à 10^{-3} , ce qui valide la convergence de nos algorithmes pour chaque point (h, L) que nous avons utilisé pour tracer la magnétisation au carré.
- **Pic au point critique** : On observe une augmentation du V -score autour des points critiques respectifs ($h = 1$ en 1D et vers $h = 3$ en 2D). C’est ce que nous avons souligné en 3.3.2 : près de la transition de phase, les corrélations deviennent complexes (longue portée) et sont plus difficiles à capturer pour le réseau de neurones.
- L’existence d’une solution analytique exacte dans le cas $h = 0$ entraîne les V -scores très faibles observés en ce point. Ce n’est pas le cas pour toutes les tailles, ce qui s’explique par les difficultés d’échantillonnage traitées en 4.3.

Un réseau entraîné ainsi est donc une approximation de la fonction d’onde $(\Psi_\theta(x_1, \dots, x_N))$ de l’état fondamental, et son appel sur un état de la base de Hilbert (x_1, \dots, x_N) renvoie la probabilité portée par cet état. La méthode détaillée en 2.2 permet alors d’évaluer les diverses observables pertinentes à notre étude.

3.4 Résultats physiques concernant la transition de phase sans désordre

3.4.1 Magnétisation et transition de phase en dimension 1

Une fois toutes ces optimisations réalisées selon le protocole exposé ci-dessus, nous disposons d’une approximation des états fondamentaux de notre modèle d’Ising 1D en champ transverse pour

différentes tailles L , et pour différentes valeurs de h . Nous étions alors en capacité de calculer la valeur moyenne de n'importe quelle observable à l'état fondamental, et en particulier du carré de la magnétisation selon l'axe z \hat{M}_z :

$$\hat{M}_z = \frac{1}{L} \sum_{i=1}^L \hat{S}_i^z$$

Nous avons ainsi obtenu la courbe présentée en Figure 10. Nous constatons, comme nous l'avions supposé, une transition de phase, la courbe étant d'autant plus raide que L est grand, entre un comportement ferromagnétique pour $h < h_c$, et un système paramagnétique pour $h > h_c$. Nous constatons également une convergence des courbes vers ce qui semble être une fonction limite, la transition de phase s'effectuant au paramètre critique $h_c \approx 1$.

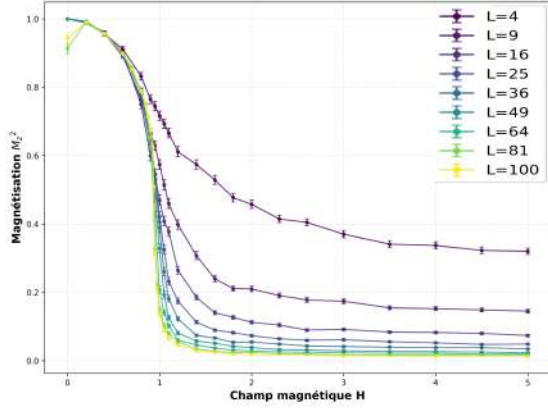


FIGURE 10 – Carré de la magnétisation $\langle \hat{M}_z^2 \rangle$ en fonction du champ magnétique transverse h pour différentes tailles de système L . On observe une transition de phase entre un régime ferromagnétique ($h < h_c$) et un régime paramagnétique ($h > h_c$).

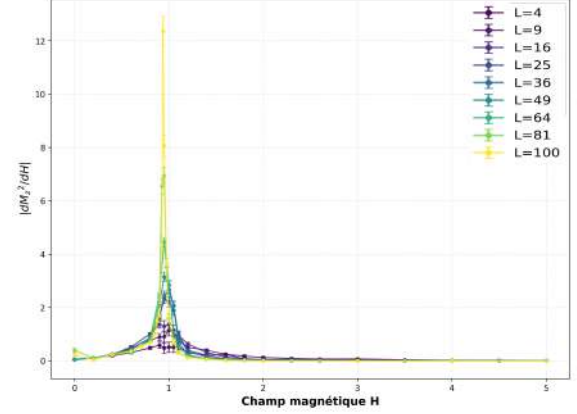


FIGURE 11 – Dérivée du carré de la magnétisation par rapport au champ magnétique $\frac{d\langle \hat{M}_z^2 \rangle}{dh}$ en fonction de h . La croissance du maximum avec L témoigne d'une transition de phase de plus en plus nette et visible.

3.4.2 Magnétisation et transition de phase en dimension 2

Nous avons tracé l'évolution du paramètre d'ordre, le carré de la magnétisation $\langle \hat{M}_z^2 \rangle$, en fonction du champ transverse h (Figure 12).

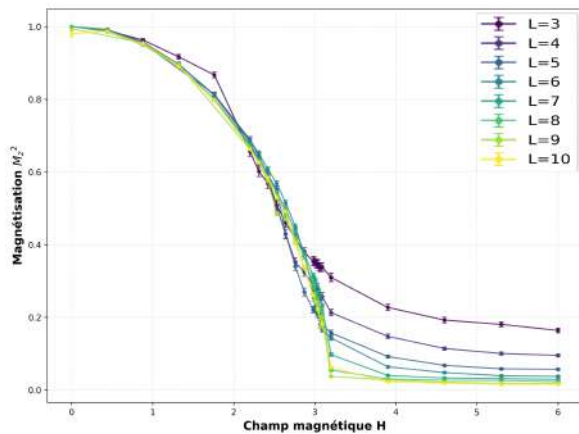


FIGURE 12 – Carré de la magnétisation $\langle \hat{M}_z^2 \rangle$ en fonction du champ magnétique transverse h pour différentes tailles de système L en 2D.

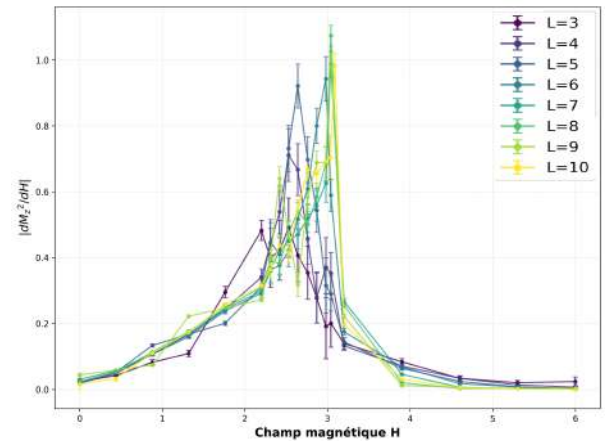


FIGURE 13 – Dérivée du carré de la magnétisation par rapport au champ magnétique $\frac{d\langle \hat{M}_z^2 \rangle}{dh}$ en 2D. Le pic de la dérivée indique la transition de phase (pseudo-point critique).

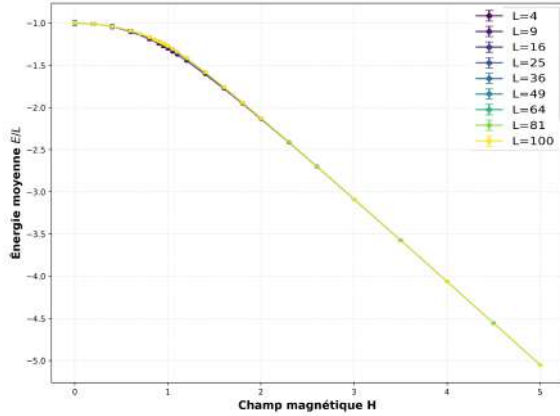
Nous observons que la courbe de la dérivée est bruitée à cause de la méthode de dérivation discrète utilisée.

Comme en 1D, on observe deux régimes asymptotiques distincts :

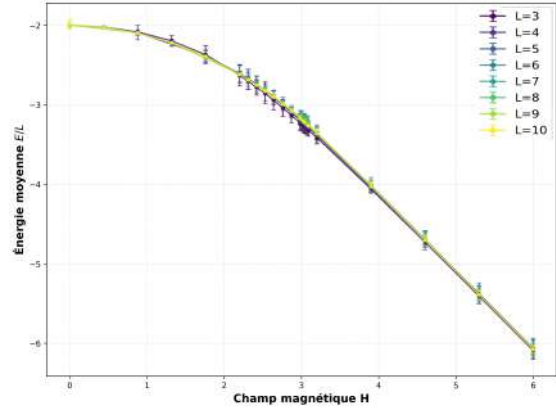
- Pour h faible, le système est dans une phase ferromagnétique ordonnée ($\langle \hat{M}_z^2 \rangle \approx 1$).
- Pour h grand, le champ transverse domine et impose une phase paramagnétique désordonnée ($\langle \hat{M}_z^2 \rangle \rightarrow 0$).

3.4.3 Énergie fondamentale : extensivité et linéarité

Nous nous sommes intéressés à l'énergie fondamentale du système par nombre de site, une grandeur intensive $\frac{E}{L}$. Les résultats pour le cas 1D (chaîne) et 2D (réseau carré) sont présentés côte à côte ci-dessous (Figure 14).



(A) Cas 1D : Énergie par nombre de site.



(B) Cas 2D : Énergie par nombre de site.

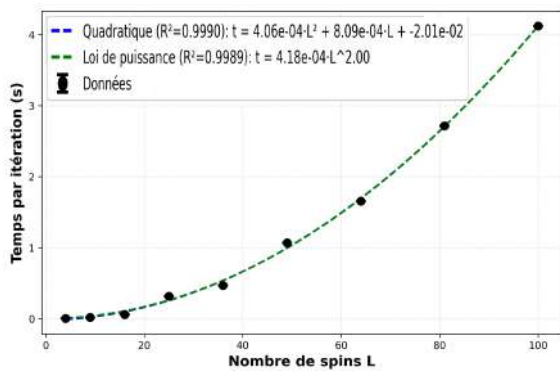
FIGURE 14 – Comparaison de l'énergie fondamentale par site E/L en fonction du champ magnétique h . L'extensivité est vérifiée dans les deux cas.

L'analyse de ces courbes met en évidence deux faits majeurs, valables en 1D comme en 2D :

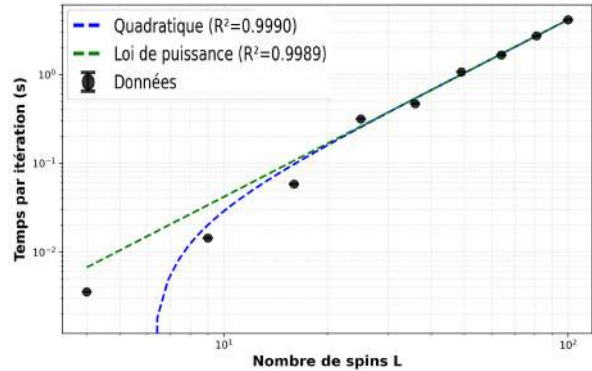
1. **Extensivité** : Toutes les courbes se superposent quelle que soit la taille L du système. Cela confirme que l'énergie fondamentale est bien proportionnelle au nombre de sites (ou linéaire en L).
2. **Linéarité en champ fort** : Pour h supérieur au champ critique ($h_c = 1$ en 1D, $h_c \approx 3$ en 2D), l'énergie par site devient une fonction linéaire de h . Cela correspond à la phase paramagnétique où les spins s'alignent avec le champ transverse, rendant l'énergie dominée par le terme $-h \sum \sigma^x$.

3.5 Coût computationnel

La réalisation de toutes les simulations permettant la construction de la Figure 10 a nécessité environ 48 heures de calcul cumulé sur GPU. Nous présentons ci-dessous l'évolution du temps de calcul par itération pour le point le plus coûteux ($h = 1$).



(A) échelle linéaire



(B) échelle log - log

FIGURE 15 – Temps de calcul par itération en fonction de la taille du système L , pour $h = 1$ (Cas 1D).

4 Méthodologie pour étudier le modèle d'Ising en champ transverse désordonné

Nous avons mis en évidence dans la partie précédente une transition de phase ferromagnétique-paramagnétique pour le modèle d'Ising en champ transverse sans désordre. Comprendre et réussir à quantifier les différents paramètres de cette transition de phase est intéressant puisque cela nous donne la valeur du champ transverse critique tel qu'en modifiant légèrement h autour de ce champ critique, les propriétés magnétiques du système sont complètement modifiées. L'état fondamental est tout d'abord celui d'un système régi par les interactions entre spins puis celui d'un système régi par le champ magnétique transverse extérieur.

Ensuite, nous avons voulu améliorer notre modélisation des matériaux magnétiques en prenant en compte la présence d'impuretés et de lacunes. **Ces impuretés et lacunes peuvent être modélisées par du désordre régnant dans le matériau.**

En effet, en pratique, à cause des impuretés du matériau, de l'agitation thermique et des lacunes, le champ magnétique ressenti par chaque site est différent. **Ainsi, le champ magnétique transverse n'est plus uniforme dans le matériau. Nous supposons cependant que la constante J d'interaction entre spins voisins n'est pas modifiée.**

Etudier l'influence du désordre sur la transition de phase ferromagnétique-paramagnétique nous permet de comprendre et de quantifier le comportement en fonction du champ transverse des matériaux utilisés en pratique dans les applications reposant sur les matériaux magnétiques. Nous souhaitons donc voir dans quelle mesure l'introduction de désordre va détériorer la transition de phase exhibée sans désordre.

4.1 Modélisation physique de la notion de désordre

Nous modélisons ces fluctuations locales de champ magnétique transverse par du désordre au sein du matériau. Pour chaque site du matériau, nous décidons de tirer sa valeur de champ magnétique transverse selon une loi normale d'espérance une certaine valeur h_0 (qui correspond au h dans l'ancien modèle et que nous ferons varier) et d'écart-type σ . L'hamiltonien étudié devient donc :

$$H = J \sum_{\langle i,j \rangle} \hat{S}_i^z \hat{S}_j^z + \sum_{i=1}^N h_i \hat{S}_i^x$$

où

$$(h_1, \dots, h_N)^T \sim \mathcal{N}((h_0, \dots, h_0)^T, \sigma^2 I_N)$$

avec I_N la matrice identité de taille $N \times N$. Comme précédemment, nous prenons la convention $J = -1$ dans la suite.

Nous avons choisi de représenter le désordre à l'aide d'une loi normale pour conserver l'idée que le désordre est une perturbation du matériau. Ce sont les lacunes et défauts du matériau qui font que chaque site ne "verra" pas le même champ transverse mais que en moyenne, les champs transverses "perçus" seront égaux à h_0 .

Ensuite, nous avons voulu utiliser une loi symétrique puisqu'il y a priori autant de chances que le champ perçu par le site en question soit plus faible ou plus élevé que le champ moyen. Enfin, Ce choix présente une motivation physique naturelle : lorsque les perturbations locales du champ transverse résultent de nombreuses contributions indépendantes (défauts de fabrication, impuretés diluées), le théorème central limite prédit précisément une distribution gaussienne. Ainsi, σ représente ainsi l'intensité du désordre régnant dans le matériau : plus il est grand, plus les valeurs de champ transverse h pourront être différentes en fonction des sites.

Ce modèle appartient à la classe des modèles d'Ising désordonnés en champ transverse, communément appelée Random Transverse Field Ising Model (RTFIM). Le RTFIM a été étudié depuis la fin des années 1990 par diverses méthodes numériques, en 1D par Fisher [8] de façon asymptotiquement exacte, et en 2D par plusieurs études QMC de référence [10, 20, 11]. Si les propriétés critiques du RTFIM 1D sont aujourd'hui considérées comme bien établies, le cas 2D reste moins consolidé numériquement et constitue le principal terrain d'étude actif. Toutes ces études utilisent des distributions uniformes (box) des champs transverses.

À notre connaissance, le choix d'une distribution gaussienne pour les champs trans-

verses dans le RTFIM 2D ferromagnétique n’a pas été exploré dans la littérature antérieure.

Dans la suite, nous appellerons h_0 l’espérance de la configuration de désordre dont il est question et h cette configuration de désordre.

4.2 Une nouvelle méthode computationnelle d’étude de la magnétisation du matériau

L’Hamiltonien que nous étudions en présence du désordre est :

$$H = J \sum_{\langle i,j \rangle} \hat{S}_i^z \hat{S}_j^z + \sum_{i=1}^N h_i \hat{S}_i^x$$

Ainsi, chaque valeur du vecteur de champs transverses correspond à un nouvel Hamiltonien. Notre modélisation du désordre implique que si nous avons appliqué la même méthode que dans la précédente partie, nous aurions été obligés d’entraîner le réseau pour chaque valeur différente du vecteur de champs transverse pour obtenir l’état fondamental de ce nouvel Hamiltonien.

Or, pour étudier correctement l’influence du désordre et garantir la reproductibilité de nos résultats, nous devons tirer beaucoup de configurations aléatoires du vecteur de champ transverses pour un seul h_0 . De plus, nous devons alors faire cela pour chaque valeur de h_0 que nous voulons avoir pour tracer la courbe de magnétisation.

En somme, **puisque notre puissance de calcul est limitée et qu’entraîner le modèle à chaque configuration de désordre est très coûteux comme exposé plus haut, nous devons utiliser de nouvelles méthodes numériques permettant d’approcher l’état fondamental de ce système pour plusieurs configurations de désordre à la fois.** Le réseau de neurones devra donc être beaucoup plus gros et avoir une architecture nouvelle.

En ce sens, nous avons utilisé les *Foundation neural-networks quantum states*, où états quantiques neuronaux de fondation. Il s’agit d’une approche extrêmement récente, présentée pour la première fois par Rende en 2025 [6], et appliquée à une version du modèle d’Heisenberg par Viteritti [7]. Ce sont les deux seuls articles sur le sujet à la date d’écriture de ce rapport. De plus, ils présentent l’idée théorique de la méthode et des exemples d’application, mais n’exposent pas certains détails techniques cruciaux pour sa mise en œuvre.

Notre approche s’inspire de ces travaux fondateurs mais s’en distingue sur plusieurs points. D’une part, nous appliquons les FNQS au RTFIM ferromagnétique avec désordre gaussien, une combinaison modèle/distribution non traitée dans la littérature existante. **D’autre part, notre paramétrisation du réseau permet de balayer continûment le champ moyen h_0 en plus des réalisations du désordre, ce qui autorise la traversée de la transition de phase par un unique réseau entraîné.** Enfin, nous mettons en évidence et exploitons une propriété de généralisation hors-distribution sur l’amplitude du désordre σ , qui rend computationnellement accessible l’étude systématique du comportement de la magnétisation en fonction des paramètres (h_0, σ) .

Ces spécificités nous ont conduits à développer un protocole expérimental dédié, que nous détaillons dans la suite de cette partie.

4.2.1 Le Modèle de Fondation

Notre problématique, sans considérer la notion de désordre dans un premier temps à : comment à partir de plusieurs h_0 couvrant un certain intervalle I nous pouvons obtenir, avec les mêmes paramètres du Transformer, les états fondamentaux de tous les Hamiltoniens tels que $h_0 \in I$? Pour notre modélisation de désordre, c’est le même questionnement puisque nous voulons qu’à partir de plusieurs configurations du désordre représentées par un vecteur h et distribuées autour de différents h_0 (formant un intervalle I), nous arrivions à donner l’état fondamental de n’importe quel configuration de désordre engendré par un h_0 appartenant à I . Ceci sans optimiser pour chaque h_0 voulu bien évidemment.

Nous utilisons pour cela le *Modèle de Fondation*.

$$\begin{aligned} \text{Réseau classique :} & \quad \Psi_{\theta}(\sigma) \quad \rightarrow \quad \text{un modèle par } h \\ \text{Fondational model :} & \quad \Psi_{\theta}(\sigma; h) \quad \rightarrow \quad \text{un seul modèle pour tout } h \in [h_{\min}, h_{\max}] \end{aligned}$$

Le champ magnétique transverse a donc changé de statut : ce n'est plus un hyperparamètre mais une entrée, un argument du réseau qui est maintenant fonction de h . Si nous avons inclus dans notre modélisation les fluctuations en J , J serait dans un modèle de fondation également devenu un paramètre.

En moyennant désormais sur la distribution de désordre de h , nous avons :

$$\langle O \rangle_h = \int p(h) dh \langle \psi_{\theta}(h) | \hat{O} | \psi_{\theta}(h) \rangle \approx \frac{1}{N_r} \sum_{h \in \{\text{ensemble d'entraînement}\}} \langle \psi_{\theta}(h) | \hat{O} | \psi_{\theta}(h) \rangle$$

avec $\psi_{\theta}(h)$ l'état fondamental avec un vecteur de champ transverse h , N_r le nombre de réalisations de h donc le cardinal de l'ensemble d'entraînement et $p(h)$ la distribution de probabilité des h (gaussienne ici).

La perte de notre réseau, qui est l'énergie de notre état, que nous essayons de minimiser est désormais égale à :

$$\mathcal{L}(\theta) = \mathbb{E}_{h \sim p(h)} [E_h(\theta)] = \int p(h) E_h(\theta) dh \approx \frac{1}{N_r} \sum_{h \in \{\text{ensemble d'entraînement}\}} \langle \psi_{\theta}(h) | \hat{H}(h) | \psi_{\theta}(h) \rangle$$

avec θ les paramètres du réseau.

Ainsi, nous nous retrouvons à vouloir trouver les paramètres qui minimisent la moyenne des hamiltoniens de l'ensemble d'entraînement et ayant donc les h dans l'ensemble d'entraînement comme configurations de désordre.

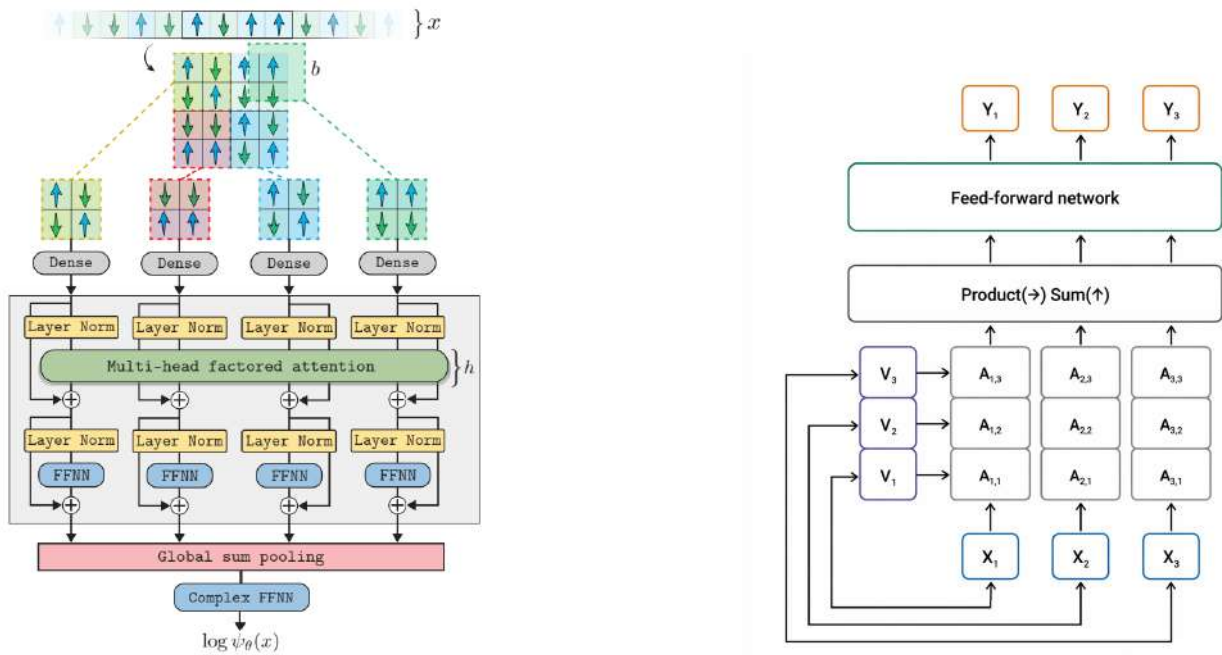
Hypothèse : Soit I l'intervalle des espérances h_0 des configurations de désordre utilisées pour l'entraînement. Avec cette approche, nous supposons donc que l'état fondamental d'une configuration de désordre donnée d'espérance $h_0 \in I$ n'est pas si loin des états fondamentaux trouvés pendant l'entraînement. Cela lui permet de généraliser son apprentissage à de nouvelles valeurs de désordre. Les modèles de fondation seront donc sûrement moins précis que d'optimiser le réseau pour chaque configuration de désordre h mais il nous permettra de ne pas être bloqué par la nécessité d'avoir une trop grande puissance de calcul pour réaliser des courbes de transition de phase avec désordre. En quelque sorte, les paramètres du réseau sont "partagées" pour des hamiltoniens proches.

4.2.2 Un nouvel ansatz : Le Transformer

Jusqu'à présent, lorsque nous étudions la transition de phase ferromagnétique-paramagnétique du système sans désordre, nous utilisons une machine de Boltzmann. Cependant, notre modèle doit ici être capable d'apprendre simultanément les configurations de l'état fondamental d'un grand nombre de configurations de champ transverse (avec ou sans désordre), et ceci malgré un entraînement sur un nombre restreint de valeurs. Il nous faut désormais utiliser un modèle bien plus puissant capable de comprendre efficacement la physique du problème et le lien entre les données pour pouvoir prédire correctement l'état fondamental pour une configuration de champs transverses non vue pendant l'entraînement. Nous avons décidé d'utiliser un *Transformer*.

La spécificité du Transformer, introduit en 2017 par des chercheurs de chez Google [12], par rapport aux autres architectures d'apprentissage réside dans le *mécanisme d'attention* modélisée dans notre architecture par la matrice A comme nous pouvons l'observer sur la figure 16 (B).

Avec les données d'entrée qui sont dans notre cas des petits paquets de spins, nous pouvons former des vecteurs abstraits de plus haute dimension avec des matrices "d'embedding" formées de paramètres optimisées pendant l'entraînement. Ensuite, à partir des vecteurs abstraits que nous notons X_1 , X_2 et X_3 , nous définissons une matrice de paramètres qui va être optimisée lors de l'entraînement : la matrice de valeurs V . Celle-ci apprend les valeurs qu'il faut accorder à chaque vecteur abstrait donc à chaque paquet de spin pour atteindre l'état fondamental.



(A) Structure globale d'un Transformer

(B) Mécanisme d'attention

FIGURE 16 – Architecture d'un Transformer (A) et sa particularité : le mécanisme d'attention (B)

D'autre part, comme nous l'avons introduit, nous définissons une seconde matrice de paramètres optimisés pendant l'entraînement : la matrice d'attention A . Avec les matrices V et A , nous pouvons définir les sorties du réseau. Chaque vecteur de sortie sera égal à

$$Y_i = \sum_j A_{i,j} V_j$$

En somme, la matrice d'attention nous dit (apprend) à quel point une entrée quelconque par exemple la 1 doit faire attention à l'entrée 2 et 3. Le vecteur valeur 1 est alors une combinaison linéaire des vecteurs valeurs 1, 2 et 3 pondérée par l'importance que doit accorder 1 à 2 et 3 respectivement.

L'architecture globale d'un Transformer correspond à la répétition de ce bloc d'attention couplée avec des processus de normalisation et des "Feed Forward" couches. Les couches de "Feed Forward" correspondent à des réseaux de neurones complètement connectés les uns avec les autres comme pour la machine de Boltzmann par exemple.

De manière générale, plus le réseau a de couches d'attention, plus il est capable de saisir des schémas et influences sophistiqués. Cependant, le nombre de paramètres croît alors aussi ce qui rend les calculs plus longs et compliqués. Finalement, à la fin du réseau, des couches de "Feed Forward" sont de nouveau utilisées pour obtenir $\log \psi(\sigma)$ correspondant à une configuration.

Plus physiquement, le Transformer représentait un ansatz idéal pour plusieurs raisons :

- **Par le mécanisme d'attention, notre ansatz n'encode pas seulement les corrélations locales comme dans notre précédente CNN (Convolutional Neural Network) mais est capable d'encoder pour chaque spin sa relation avec tous les autres, quelle que soit la distance.** Ceci est fondamental puisque proche de la transition, les comportements collectifs sont prépondérants et expliquent la présence ou non de transitions de phase.
- A travers la compréhension des liens entre les différents spins, le mécanisme d'attention est idéal pour la méthode de fondation puisqu'il permet au réseau d'apprendre une véritable représentation physique de l'Hamiltonien. Il ne se contente donc pas d'apprendre les états fondamentaux des vecteurs de champ transverses utilisés pendant l'entraînement mais il est capable de généraliser à des données non vues jusqu'à présents grâce à son apprentissage de la physique du problème.
- l'architecture de Transformer que nous utilisons est différente de celle utilisée en machine learning par les chercheurs de Google par exemple. En effet, leur architecture est invariante par

de permutations entre les entrées X_1 , X_2 et X_3 . Cependant, physiquement, nous ne désirons pas avoir cet effet puisque notre système n'est pas invariant par permutations mais par translations dans l'espace. **Nous construisons alors la matrice d'attention de sorte à garantir cette invariance de translation.**

4.2.3 Une première validation : retrouver la transition de phase sans désordre

Pour valider le principe des méthodes fondationnelles, nous avons d'abord voulu retrouver les courbes de la Figure 10 présentant la magnétisation au carré $\langle M_z^2 \rangle$ en fonction du champ transverse h_0 dans le cas sans désordre. Rappelons que chaque point de la Figure 10 a été obtenu en entraînant un réseau de neurones spécifique au couple (L, h_0) . Cette fois, nous avons entraîné un unique réseau de neurones pour chaque taille L , entraîné sur un jeu valeurs de h_0 correspondant aux points rouges sur la figure suivante, et ensuite capable d'approximer l'état fondamental du système pour de nouvelles valeurs de h_0 , représentées par les points bleus sur la figure ci-dessous.

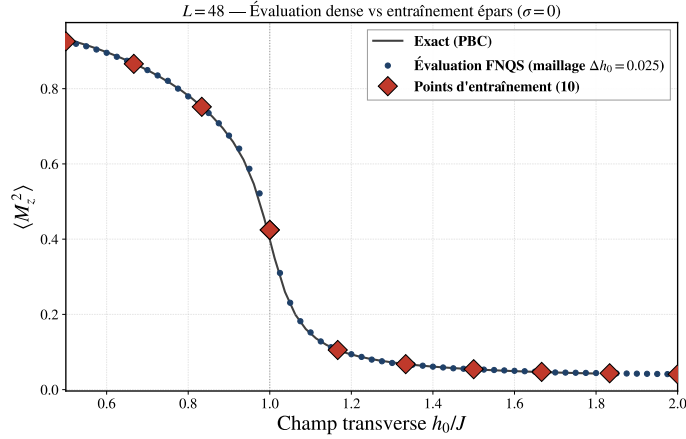


FIGURE 17

La puissance de calcul et la taille conséquente de l'architecture du Transformer utilisé pour produire cette courbe permettent d'être beaucoup plus précis que notre démarche produisant la Figure 10. Ainsi, nous constatons que les valeurs de $\langle M_z^2 \rangle$ obtenues grâce à notre modèle collent extrêmement bien à la courbe exacte.

4.2.4 Application du modèle de fondation à l'étude du désordre

Nous avons présenté ci-dessus le principe théorique de la méthode de fondation. Pour étudier l'état fondamental du modèle d'Ising en champ transverse désordonné, nous souhaitons l'utiliser pour entraîner des réseaux de neurones prenant comme paramètre d'entrée une configuration de champ transverse $(h_1, \dots, h_N)^T$, et renvoyant une approximation de l'état fondamental du système décrit par l'hamiltonien d'Ising dans cette configuration de champ transverse.

En pratique, nous avons composé le jeu de configurations d'entraînement de nos Transformers en choisissant une certaine grille de valeurs de champ moyen h_0 , et en tirant au hasard selon la loi $\mathcal{N}((h_0, \dots, h_0)^T, \sigma^2 I_N)$ un nombre N_c de configurations de désordre pour chaque valeur de h_0 . Très vite, nous avons rajouté la configuration homogène $(h_0, \dots, h_0)^T$ pour chaque valeur de h_0 , afin que notre réseau apprenne aussi l'état fondamental du système sans désordre.

Le choix de la grille de h_0 , de la valeur de l'intensité du désordre σ utilisée pour l'entraînement, et du nombre N_c de configurations par h_0 constituent des paramètres déterminants pour les performances de nos modèles, et dans la suite nous montrerons comment nous les avons faits varier. À titre indicatif, nous prenons la plupart du temps une dizaine de valeurs de h_0 , N_c entre 10 et 20, et σ de l'ordre de 0.1.

4.2.5 Vérifier la performance du réseau entraîné

Une fois un réseau entraîné, nous souhaitons être capable de vérifier ses performances. Pour ce faire, nous validons d'abord la convergence de l'entraînement, en traçant notre fonction de perte (la somme des énergies prédites par le réseau dans chaque configuration désordonnée $(h_1, \dots, h_N)^T$

d'entraînement) en fonction du nombre d'itérations de notre descente de gradient, et en vérifiant qu'elle était environ constante sur les dernières itérations. Sur l'exemple fourni par la Figure 18, on constate bien que notre fonction de perte (l'énergie cumulée) a convergé à la fin de l'entraînement, à faibles fluctuations d'échantillonnage près.

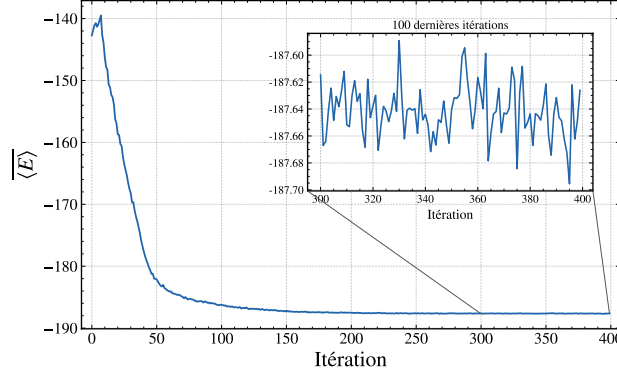
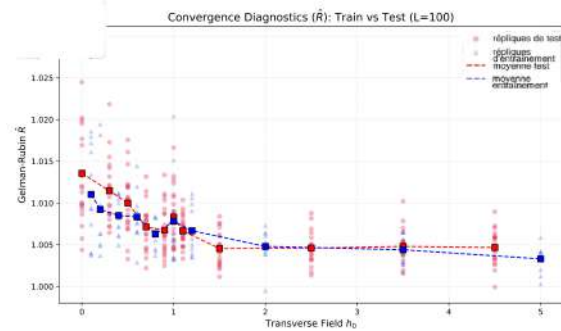


FIGURE 18 – Convergence de la fonction de perte (moyenne des énergies de toutes les configurations d'entraînement) lors de l'entraînement de notre modèle fondationel en 1D pour $L=100$

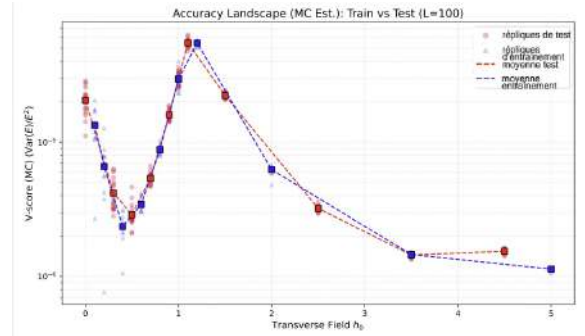
Ensuite, nous traçons des nuages de points représentant le V-score et le \hat{R} (que nous expliquons au paragraphe suivant) obtenus en évaluant notre modèle fondationel à la fois sur les configurations de champ transverse utilisées pour l'entraîner, et sur de nouvelles configurations, qu'il n'avait jamais vues à l'entraînement, pour vérifier ses capacités de généralisation.

Le \hat{R} , aussi connu sous le nom de la statistique de Gelman-Rubin, nous permet de vérifier que les chaînes de Markov utilisées pour évaluer le réseau sont bien convergentes. Le principe de cet indicateur est de comparer la variance intra-chaînes à celle inter-chaînes. En effet, si plusieurs chaînes ont convergé vers la même distribution cible, alors leur variance entre elles et leur variance interne devraient être presque égales. Heuristiquement, la convergence des chaînes est ainsi excellente entre des valeurs de \hat{R} entre 1.00 et 1.05 et elle n'est pas bonne à partir de 1.1.

Le V-score, comme vu précédemment, nous permet pour sa part de vérifier la convergence d'une approximation variationnelle d'un état fondamental.



(A) Nuage de points du \hat{R}



(B) Nuage de points du V-score

FIGURE 19 – Nuages de points du \hat{R} et du V-score pour évaluer les performances de notre modèle fondationel en 1D avec $L=100$ sur les configurations d'entraînement (points bleus), et sur de nouvelles configurations (points rouges).

Sur l'exemple de la Figure 19, on constate ainsi que le \hat{R} est inférieur à 1.03 partout, et donc que toutes les chaînes MC utilisées sont convergentes, donc fiables. De plus, le V-score est inférieur à $6 \times 10^{-5} \ll 10^{-3}$ partout, donc la convergence de la minimisation de l'énergie est très bonne, à la fois sur les configurations d'entraînement et de test.

On peut de plus constater deux phénomènes intéressants, liés à la physique du système. D'abord, le \hat{R} est d'autant moins bon que h_0 est petit. Ce phénomène est lié à la dégénérescence de l'état fondamental à $h_0 = 0$, et nous revenons dessus plus en détail dans la sous-partie abordant les problèmes rencontrés. Ensuite, on voit que le V-score se détériore autour de $h_0 = 1$, c'est-à-dire

autour de la transition de phase, car c'est là où la physique du système est la plus complexe, et donc la plus difficile à approcher.

4.2.6 Evaluer $\langle M_z^2 \rangle$ de façon numériquement efficace et rigoureuse : l'échantillonnage préférentiel (Importance Sampling)

A cette étape de la méthode nous disposons donc pour chaque taille d'un modèle fondationnel entraîné. Ce modèle peut alors être évalué en une valeur de paramètre quelconque et nous pouvons ainsi théoriquement accéder à la valeur d'une observable arbitraire dans l'état fondamental pour un paramétrage arbitraire de notre hamiltonien. Pour obtenir la valeur des observables, nous utilisons la méthode de Monte Carlo (détaillée dans la section 2.2). Cependant, cette méthode nécessite d'échantillonner l'amplitude de Born du réseau en le paramètre choisi. L'échantillonnage a un coût en $O(N_{\text{samples}}LC)$ où C est le coût d'une évaluation du réseau. Dans notre étude, nous cherchons à évaluer $\langle M_z^2 \rangle$ en un grand nombre de paramètres, en effet, pour être capable de dériver par rapport à h nous choisissons une grille fine, environ 100 points dans une fenêtre $h \in [0.3, 0.7]$. Nous allons pour chaque h , tirer 50 points pour chaque valeur du désordre σ (7 valeurs).

Cela représente donc des dizaines de milliers de choix de paramètres différents où évaluer $\langle M_z^2 \rangle$. C'est bien trop cher, ainsi, il est crucial d'utiliser la méthode d'*importance sampling*. Cette méthode permet d'échantillonner selon l'amplitude de Born du fondamental seulement pour certaines valeurs des paramètres, ici les configurations homogènes, et d'en déduire la valeur d'observable évaluée en d'autres jeux de paramètres suffisamment proches.

L'*importance sampling* repose sur l'identité suivante : pour toute distribution de probabilité q dont le support contient celui de la cible p ,

$$\langle \hat{O} \rangle = \sum_x p(x) O_{\text{loc}}(x) = \sum_x q(x) \frac{p(x)}{q(x)} O_{\text{loc}}(x) = \mathbb{E}_{x \sim q}[w(x) O_{\text{loc}}(x)], \quad w(x) := \frac{p(x)}{q(x)}. \quad (1)$$

Le changement de mesure est exact et les poids w absorbent l'écart entre q et p . En FNQS, p et q ne sont accessibles qu'à une constante de normalisation près (Z_p, Z_q inconnues), donc on passe aux poids non normalisés $\tilde{w}(x) := |\psi(x)/\phi(x)|^2$, et les constantes se simplifient dans un quotient :

$$\langle \hat{O} \rangle = \frac{\sum_x q(x) \tilde{w}(x) O_{\text{loc}}(x)}{\sum_x q(x) \tilde{w}(x)} = \frac{\mathbb{E}_q[\tilde{w} O_{\text{loc}}]}{\mathbb{E}_q[\tilde{w}]}, \quad (2)$$

$$\widehat{\langle \hat{O} \rangle}_{N_s}^{\text{SNIS}} = \frac{\sum_{s=1}^{N_s} \tilde{w}(x_s) O_{\text{loc}}(x_s)}{\sum_{s=1}^{N_s} \tilde{w}(x_s)}, \quad x_s \sim q. \quad (3)$$

C'est l'estimateur *self-normalized* (SNIS) : biaisé à N_s fini (rapport de deux moyennes empiriques corrélées, biais en $\mathcal{O}(1/N_s)$), mais consistant par la loi des grands nombres appliquée au numérateur et au dénominateur séparément. Sa variance asymptotique se contrôle par la taille d'échantillon effective (ESS) qui est N_{eff}/N_s où :

$$N_{\text{eff}} := \frac{(\sum_s \tilde{w}(x_s))^2}{\sum_s \tilde{w}(x_s)^2} \in [1, N_s], \quad (4)$$

qui estime le nombre d'échantillons i.i.d. tirés directement selon p qui fourniraient la même précision statistique que les N_s échantillons repondérés tirés selon q . Elle mesure ainsi la concentration des poids : $N_{\text{eff}} \rightarrow N_s$ quand $q \approx p$ (poids uniformes), et $N_{\text{eff}} \rightarrow 1$ quand un seul x_s domine la somme (dégénérescence des poids, explosion de la variance).

Dans notre étude, nous décidons de rejeter l'ensemble des points pour lesquels $ESS = N_{\text{eff}}/N_s \leq 20\%$. Cette borne de rejet, dépend du nombre d'échantillons N_s dont on dispose car le nombre d'échantillons effectifs est représenté par $N_{\text{eff}} = ESS \times N_s$.

En pratique, nous échantillonons uniquement les états en les configurations homogènes, i.e. en les vecteurs h homogènes) avec 32k échantillons. Ensuite, nous utilisons la méthode d'importance sampling pour en déduire les observables avec du désordre tiré autour de cette configuration homogène.

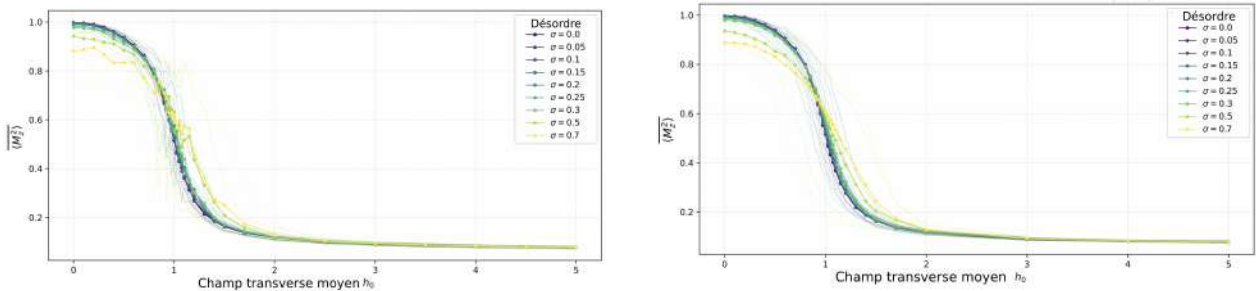
4.2.7 Choix pratiques pour l'évaluation de la magnétisation

L'introduction d'aléa dans notre hamiltonien génère naturellement du bruit sur les observables que nous étudions, car chaque configuration de désordre engendre des valeurs légèrement différentes. Pour dégager des comportements physiques globaux de ces données bruitées, nous avons donc dû trouver la bonne façon de moyennner nos grandeurs d'intérêt.

Pour tracer des courbes de magnétisation pour un système de L spins avec un désordre caractérisé par un écart type σ , notre première approche naïve fut de choisir un ensemble de valeurs du champ moyen h_0 , et pour chaque valeur de h_0 , tirer un nombre N_c de configurations du désordre, c'est-à-dire tirer N_c vecteurs aléatoires indépendants et identiquement distribués selon la loi gaussienne $\mathcal{N}(h_0 U_L, \sigma^2 I_L)$, où $U_L = (1, \dots, 1)^T \in \mathbb{R}^L$ est le vecteur ne contenant que des 1, et où I_L désigne la matrice identité de $\mathcal{M}_L(\mathbb{R})$. Ensuite, nous évaluons notre modèle fondationnel entraîné en taille L en chacune de ces configurations du désordre, puis nous effectuons une moyenne sur tous ces tirages du désordre.

Étant limités par le temps de calcul que nécessite l'échantillonnage selon la loi donnée par notre modèle fondationnel en chaque configuration du désordre, nous ne pouvions pas prendre N_c supérieur à 80 ou 100 maximum. Cette valeur est trop faible pour nous placer dans le domaine de validité de la Loi des grands nombres, donc les moyennes sur le désordre présentaient des décalages d'une valeur de h_0 à l'autre, ce qui donnait des courbes bruitées complètement inexploitables, d'autant plus que nous souhaitions ensuite les dériver.

Pour remédier à ce problème, nous avons modifié notre méthodologie pour choisir les configurations du désordre en lesquelles nous évaluons notre réseau : nous tirons désormais une seule fois un nombre N_c de vecteurs de perturbation, selon la loi $\mathcal{N}(0, \sigma^2 I_L)$, puis en chaque valeur de h_0 , nous ajoutons ces vecteurs de perturbation au vecteur homogène $(h_0, \dots, h_0)^T \in \mathbb{R}^L$. Autrement dit, au lieu de tirer de nouvelles configurations du désordre différentes en chaque h_0 , on tire une seule fois nos configurations du désordre, qu'on applique ensuite en toutes les valeurs de h_0 . Ceci peut introduire un léger biais sur la moyenne dans la mesure où N_c est fini, mais ce biais est identique en tout les h_0 , ce qui permet d'obtenir des courbes beaucoup plus propre, et ne modifie pas drastiquement la physique étudiée si on prend N_c assez grand.



(A) Courbes de magnétisation tracées avec des tirages du désordre différents en chaque valeur de h_0

(B) Courbes de magnétisation tracées avec les mêmes tirages du désordre en chaque valeur de h_0

FIGURE 20 – Comparaison de deux protocoles de choix des réalisations du désordre utilisées pour tracer les courbes de magnétisation. On constate que le second donne des courbes nettement plus lisses et exploitables.

Toujours dans le but d'obtenir des courbes aussi lisses que possible pour pouvoir les dériver numériquement sans trop de bruit, nous nous sommes aussi intéressés aux méthodes de lissage de courbe. Nous avons donc implémenté et appliqué à nos données les algorithmes de traitement du signal Gaussien et de Savitzky-Golay, pour les comparer à ce que donne la dérivée standard de nos données sans traitement. Cependant, ces méthodes donnaient chacune des résultats si différents que nous avons préféré abandonner cette piste, et continuer à étudier nos données pures plutôt qu'une version traitée, qui ne traduirait peut-être plus la même physique.

Enfin, nous avons voulu utiliser le maximum de la dérivée de la courbe de $\langle M_z^2 \rangle$ en fonction du champ moyen h_0 pour localiser un éventuel champ critique et caractériser une potentielle transition de phase. Une question s'est alors posée : devons-nous étudier le maximum de la dérivée de la

moyenne des $\langle M_z^2 \rangle$ sur le désordre, ou plutôt la moyenne sur le désordre des maxima des dérivées des courbes de $\langle M_z^2 \rangle$ en fonction de h_0 pour chaque configuration de désordre. Le maximum n'étant pas un opérateur linéaire, cela ne revient pas au même. En notant $\overline{\cdot}$ l'opération de moyenner sur les configurations du désordre, nous avons donc comparé

$$\max_{h_0} \frac{\partial}{\partial h_0} \overline{\langle M_z^2 \rangle} \quad \text{à} \quad \overline{\max_{h_0} \frac{\partial}{\partial h_0} \langle M_z^2 \rangle}$$

L'intérêt majeur de la seconde expression est que l'on peut facilement calculer une incertitude statistique associée à cette moyenne, ce qui donne des barres d'erreur sur les grandeurs calculées.

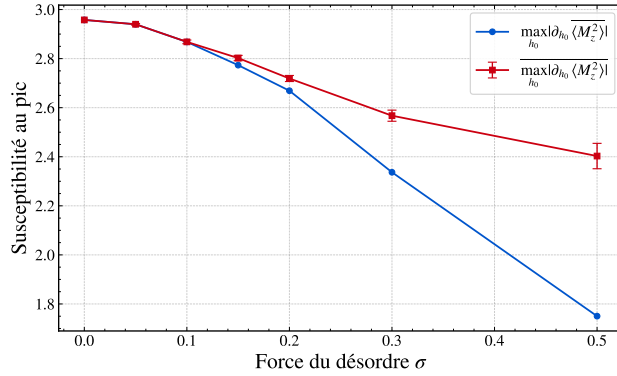


FIGURE 21 – Comparaison entre $\max_{h_0} \frac{\partial}{\partial h_0} \overline{\langle M_z^2 \rangle}$ et $\overline{\max_{h_0} \frac{\partial}{\partial h_0} \langle M_z^2 \rangle}$ sur une même courbe de magnétisation en 1D

On peut constater plusieurs choses sur cette figure. D'abord, la courbe bleue est en-dessous de la rouge, ce qui est attendu d'après l'inégalité de Jensen. Ensuite, on voit qu'à petit σ , les deux courbes se surperposent, ce qui s'explique par le fait suivant : lorsque le désordre est faible, les configurations du désordre sont toutes presque identiques, donc présentent un même maximum de susceptibilité magnétique en le même h_0 , donc la moyenne des maxima est environ égale au maximum de la moyenne. En revanche, à grand σ , le désordre est fort : chaque configuration a son propre h_0 critique, les pics sont dispersés, donc la moyenne lisse et abaisse le pic de la courbe moyennée. Enfin, on voit que l'erreur statistique sur $\overline{\max_{h_0} \frac{\partial}{\partial h_0} \langle M_z^2 \rangle}$ est nulle pour un désordre nul, et augmente avec σ , ce qui est là encore cohérent.

Finalement, la susceptibilité physiquement pertinente est la moyenne sur les réalisations du désordre du pic individuel de chaque configuration, $\max_{h_0} |\partial_{h_0} \langle M_z^2 \rangle|$, plutôt que le pic de la courbe moyennée. En effet, la courbe moyennée $\overline{\langle M_z^2 \rangle}(h_0)$ est un artefact statistique : aucun système réel ne correspond à la superposition de nombreuses réalisations du désordre, et son pic s'effondre à grand σ non pas parce que la transition s'affaiblit, mais parce que les différentes configurations transitent à des champs critiques $h_0^{(k)}$ dispersés. La grandeur rouge capture au contraire le maximum de susceptibilité typique d'un échantillon individuel, comme ce que l'on pourrait mesurer sur un matériau réel.

4.2.8 Sources d'incertitude et propagation

Avant de présenter nos résultats finaux, nous recensons les trois sources indépendantes d'incertitude affectant les estimations produites par la méthode exposée ci-dessus, en distinguant ce qui est quantifié dans nos barres d'erreur de ce qui ne l'est pas.

(a) Erreur variationnelle. Notre ansatz $|\psi_g\rangle$ diffère de l'état fondamental exact d'une distance variationnelle ε inconnue. Comme discuté en section 4.3.3, le V-score borne l'erreur sur l'énergie mais pas sur $\langle M_z^2 \rangle$. Cette incertitude est **systematique et non quantifiée** dans nos barres d'erreur.

(b) Erreur d'échantillonnage Monte Carlo. Pour chaque triplet $(h_0, \sigma, \text{réalisation du désordre})$, l'estimation de $\langle M_z^2 \rangle$ est une moyenne sur N_{samples} configurations tirées par MCMC. L'incertitude associée, quantifiée par l'écart-type empirique entre chaînes parallèles, évolue en $1/\sqrt{N_{\text{samples}}}$ et est contrôlée par le diagnostic \hat{R} , comme exposé en 4.3.

(c) Erreur statistique sur le désordre. La moyenne des observables sur $N_c = 80$ réalisations du désordre introduit une variance en $\sigma_{\text{désordre}}^2/N_c$, où $\sigma_{\text{désordre}}^2$ désigne la dispersion entre réalisations,

intrinsèque à l'existence de désordre. Cette contribution est calculée directement à partir de l'écart-type empirique sur les N_c tirages.

Les barres d'erreur reportées sur nos figures correspondent à la somme quadratique de (b) et (c). En pratique, l'erreur de désordre (c) domine de plusieurs ordres de grandeur l'erreur MC (b), car deux réalisations du désordre donnent des valeurs de $\langle M_z^2 \rangle$ très différentes tandis que deux runs MC sur une même réalisation donnent des résultats quasi identiques : c'est le comportement attendu et caractéristique d'un échantillonnage bien dimensionné. Les incertitudes sur (h_c, ν) extraites de l'effondrement de Binder reflètent donc essentiellement (c), sans capturer (a) ni les corrections de taille finie au-delà de l'ordre dominant.

4.3 Problématiques rencontrées lors de l'élaboration de cette méthode

4.3.1 Mise en place de techniques d'échantillonnage adaptées

En essayant d'entraîner puis de tracer les courbes correspondant à un entraînement avec désordre, nous obtenions des courbes avec des V-scores de 0.1 ou 0.2 pour des champs proches de 0.

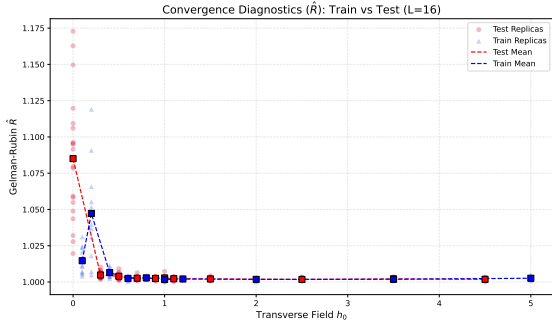
Face à ces courbes, une de nos hypothèses était la mauvaise convergence des chaînes de Monte Carlo, que ce soit pendant l'entraînement ou pendant l'évaluation des observables. En effet, si après l'entraînement, nous souhaitons évaluer une observable, nous ré-échantillons des chaînes de Markov convergeant vers la mesure cible pour évaluer l'observable en les échantillons et puis faire une moyenne des évaluations trouvées. **Ainsi, si la chaîne n'est pas arrivée à convergence ou si elle n'arrive pas à parcourir tout l'espace des configurations, l'évaluation des échantillons et donc l'évaluation de l'observable sur l'état cible sera biaisé, faussant la mesure.** Ce problème est omniprésent à faibles valeurs de champ transverse car l'état fondamental est dégénéré à cause de la symétrie par rapport à l'axe des z . L'état fondamental est donc une superposition des états tous les spins en haut et tous les spins en bas. Pour évaluer correctement cet état, il faut que l'échantillonneur parcourt tout l'espace des configurations. Or, passer d'une configuration tout haut à tout bas est très improbable dans l'algorithme de Monte Carlo, ce qui fait que l'échantillonneur est bloqué dans un sous-espaces de l'espace des configurations total.

Comme nous l'avons vu précédemment, un des outils utilisés pour évaluer la convergence des chaînes de Monte Carlo est le *R-hat*. Pour s'assurer de la bonne convergence de nos chaînes, celui-ci doit être entre 1.00 et 1.05.

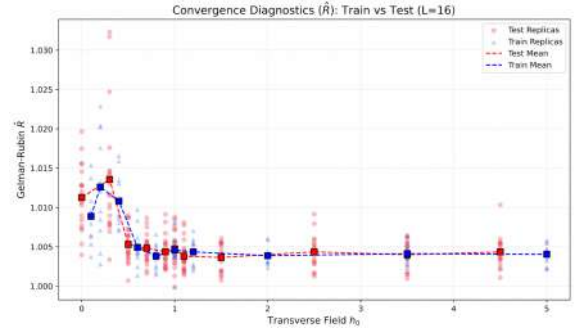
Afin de déterminer correctement le *R-hat* de nos chaînes, nous avons élaboré un protocole à deux étapes : nous entraînons notre modèle comme expliqué ci-dessus avec une liste de champ transverse h_0 . Ensuite, nous évaluons les *R-hat* pour les configurations de désordre générées autour de ces h_0 de train ainsi que pour des configurations générés par d'autres h_0 "test". Ceci nous permet de vérifier que la bonne convergence des chaînes n'est pas seulement pour des h_0 spécifiques utilisés pendant l'entraînement du modèle.

Dans un premier temps, nous obtenions des courbes comme celle à gauche. Puis, nous nous sommes rendus compte d'un problème avec notre traitement du désordre : pour des h_0 trop petits lorsque nous tirions des configurations avec un désordre trop important, certaines valeurs de champ transverse h pouvaient être négatives. Ces anomalies négatives faussaient nos évaluations d'observables ainsi que d'échantillonnage, l'état cible étant physiquement incorrect. Nous avons alors essayé de remplacer les valeurs négatives par des 0 mais cela créait une accumulation artificielle de 0. **Nous avons donc décidé de remplacer le tirage négatif par sa valeur absolue.**

Finalement, en constatant que le *R-hat* n'était toujours pas optimal pour $L = 16$, nous avons décidé de changer la règle d'échantillonnage. Avant nous utilisions une règle de Netket qui implémentait la règle de Metropolis en modifiant seulement un spin à chaque étape. Désormais, nous ajoutons une nouvelle étape qui nous permet avec une probabilité définie par l'utilisateur de retourner l'ensemble des spins d'un coup. **Cette nouvelle règle "La GlobalFlipRule" force ainsi les échantillons à visiter tout l'espace des configurations sans être bloqué dans un sous-espace comme ce qui se passait avant.** Enfin, elle converge toujours vers la distribution cible par symétrie de l'opération de retournement de tous les spins laissant la matrice d'acceptation inchangée.



(A) Courbe initiale

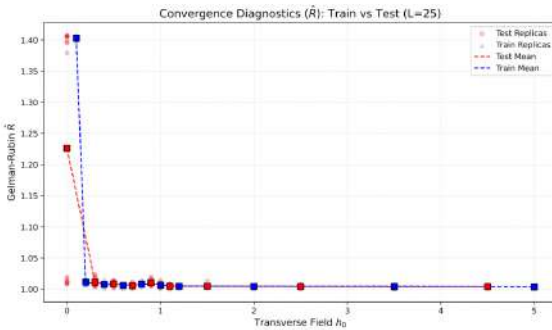


(B) En remplaçant par la valeur absolue du tirage et en utilisant la GlobalFlipRule.

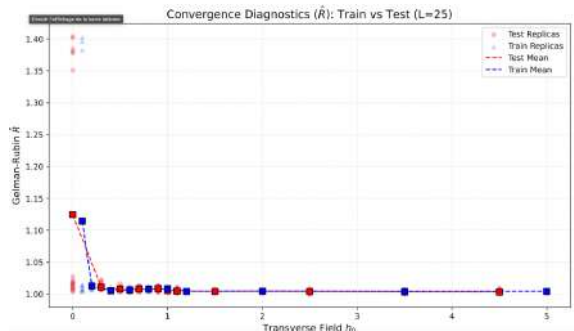
FIGURE 22 – Courbes représentant le R-hat en fonction des configurations de train et de test. Les carrés rouges et bleus sont les moyennes des configurations ayant le même h_0 moyen.

Malheureusement, en passant à grande taille, l'échantillonnage se re-détériore près de 0 malgré la GlobalFlipRule et la valeur absolue, comme que le montre la figure 23A. Ce n'est pas surprenant car à grandes tailles la probabilité de passer d'un état presque qu'avec des spins à un état qu'avec des spins bas est plus improbable qu'à $L = 16$.

Pour résoudre ce problème, nous avons trouvé un nouveau moyen de **forcer l'échantillonneur à visiter tout l'espace** : en lançant nos chaînes de Monte Carlo, nous modifions l'état de l'échantillonneur pour que la moitié des chaînes partent avec tous les spins vers la bas et la moitié des chaînes partent avec tous les spins vers le haut. En zéro, le problème souligné au-dessus est donc limité mais toujours présent (le R-hat étant supérieur à 1.10 pour des h_0 proches de 0).



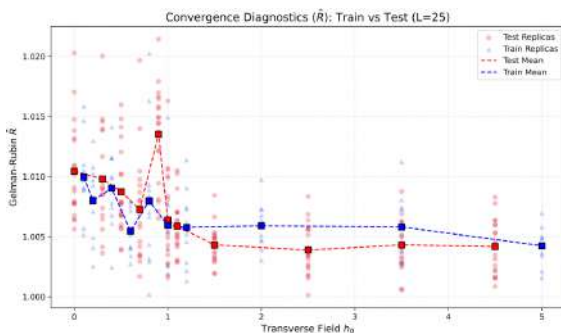
(A) Courbe initiale à $L=25$ avec GlobalFlipRule et la valeur absolue des configurations



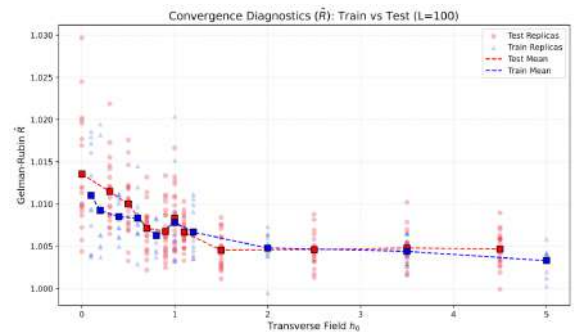
(B) En initialisant selon le protocole décrit ci-dessus.

FIGURE 23 – Courbes représentant le R-hat en fonction des configurations de train et de test.

Afin de résoudre ce problème d'échantillonnage en 0, nous essayons alors de jouer sur les hyperparamètres pertinent : la probabilité dans la GlobalFlipRule. Nous décidons de l'augmenter pour être sûr de visiter correctement l'espace des configurations. Nous obtenons alors des R-hat inférieurs à 1.015 pour tout h_0 ce qui est très bon. Ceci nous permet d'obtenir des V-score entre 10^{-4} et 10^{-5} .



(A) Courbe finale de R-hat avec toutes nos modifications pour $L=25$.



(B) Courbe finale de R-hat pour une taille de 100 spins.

FIGURE 24 – Courbes représentant le R-hat en fonction des configurations de train et de test.

4.3.2 Choix des hyperparamètres du Transformer

Il convient en premier lieu de détailler la manière dont est encodé un vecteur d'entrée, lequel, dans le cadre de la méthode FNQS, se compose d'une configuration de spins σ et d'un ensemble de paramètres h_i de l'hamiltonien. L'embedding de notre modèle encode nativement le fait que la configuration de spins et les paramètres sont des composantes de natures différentes : la matrice d'embedding, qui envoie chaque composante de l'entrée sur un vecteur de taille d_{model} , projette les configurations de spins sur les $d_{\text{model}}/2$ premières dimensions, et les paramètres de l'hamiltonien sur les $d_{\text{model}}/2$ suivantes. Ce choix se justifie empiriquement : l'étude reportée en figure 25 montre qu'un embedding séparé améliore significativement le v-score par rapport à un embedding partagé, à budget de paramètres comparable.

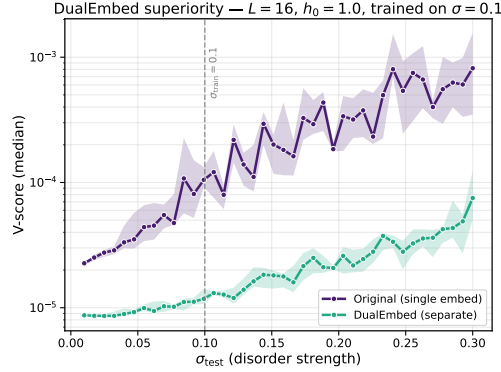


FIGURE 25 – Comparaison du V-score obtenu avec un embedding partagé (spins et paramètres projetés conjointement sur les d_{model} dimensions) et un embedding séparé (spins sur les $d_{\text{model}}/2$ premières dimensions, paramètres sur les suivantes), à nombre de paramètres entraînaibles équivalent. L'écart observé motive le choix de la séparation.

Regroupement en patches. Nous regroupons par ailleurs les spins en *patches* de taille b en 1D (resp. b^2 en 2D) lors de l'embedding, ce qui permet de réduire sensiblement le coût computationnel en raccourcissant la séquence d'entrée du Transformer. L'étude que nous avons menée sur ce paramètre (figure 26) indique que, dans les régimes de taille et de dimensionnalité que nous avons explorés, $b = 4$ réalise un bon compromis entre expressivité du modèle et coût de calcul ; c'est cette valeur que nous retenons pour l'architecture finale. Nous ne prétendons pas que ce choix soit universellement optimal : pour des systèmes sensiblement plus grands ou pour des hamiltoniens présentant des corrélations à très longue portée, une exploration spécifique du paramètre b resterait nécessaire.

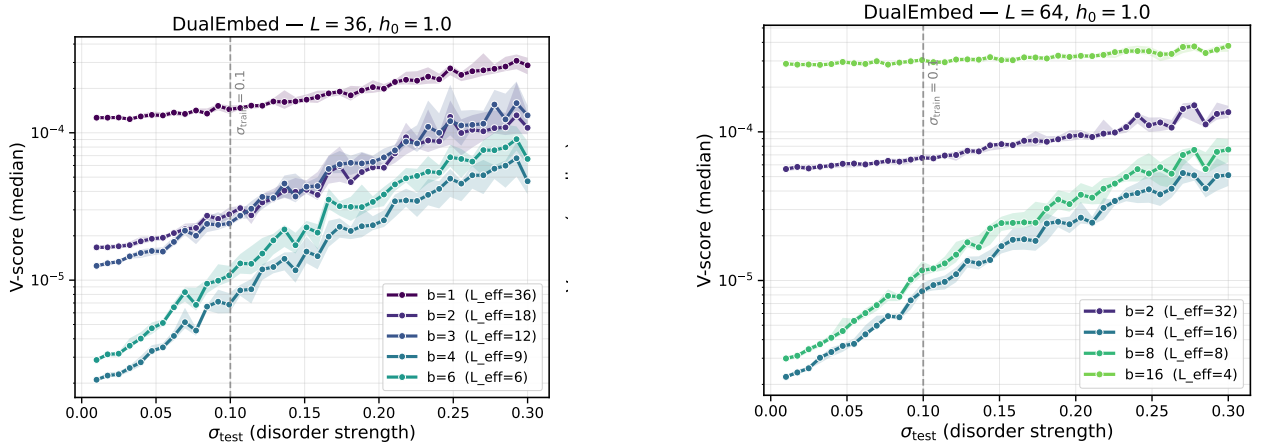


FIGURE 26 – Influence de la taille de patch b sur le v-score et le coût computationnel, à d_{model} et nombre de couches fixés. Le choix $b = 4$ réalise un compromis favorable entre qualité de l'approximation et temps de calcul dans les régimes étudiés. $n_{\text{samples}} = 1024$, $n_{\text{replicas}} = 120$

Dimension du modèle et nombre de têtes. Un autre paramètre fondamental est la dimension d_{model} de l'espace dans lequel est encodé le vecteur d'entrée. La littérature des modèles de fondation

utilise souvent $d_{model} = 72$ [7, 6], nous nous attendons à un gain sur le V-score en fonction de la taille d_{model} . Ce comportement est cohérent avec l'augmentation attendue de l'expressivité du modèle, mais suggère également que notre modèle reste sous-dimensionné relativement à la tâche : une exploration à plus grand d_{model} constituerait une suite naturelle de ce travail. Compte tenu de nos contraintes matérielles, nous avons retenu la valeur $d_{model} = 60$, maximale dans la limite de nos ressources. Cette valeur présente par ailleurs l'avantage d'être divisible par 2 (séparation spins / paramètres), par 6 (voir ci-dessous) et par 10, offrant ainsi une souplesse appréciable dans le choix des dimensions dérivées.

Le nombre de têtes d'attention est alors fixé heuristiquement par la relation

$$n_{heads} = \frac{d_{model}}{6}, \quad (5)$$

soit $n_{heads} = 10$ dans notre configuration.

FIGURE 27 – Évolution du v-score en fonction de la dimension d'embedding d_{model} , à taille de patch et nombre de couches fixés. La décroissance du v-score, sans saturation visible sur la plage accessible, motive le choix de $d_{model} = 60$, valeur maximale compatible avec nos contraintes matérielles.

Profondeur du réseau. Concernant le nombre de couches, nos contraintes de calcul nous ont imposé de le plafonner à $L = 4$; l'étude menée sur ce paramètre suggère qu'une profondeur plus importante aurait vraisemblablement conduit à de meilleures performances. Le goulot d'étranglement computationnel provient en effet du calcul et du stockage de la jacobienne intervenant dans l'optimisation par *stochastic reconfiguration* (cf. Section 2.4 "Descente de gradient naturel"), de taille $N_{samples} \times N_{params}$, où N_{params} désigne le nombre de paramètres entraînaibles du modèle fondationnel.

Avec les hyperparamètres retenus ci-dessus, on obtient $N_{params} \approx 2,0 \times 10^5$. Cet ordre de grandeur se situe dans la fourchette haute des tailles rencontrées dans la littérature NQS.

4.3.3 Difficultés de convergence aux grandes tailles

Comme expliqué précédemment, un des indicateurs sur lequel nous nous sommes appuyés pour évaluer la qualité de nos états variationnels FNQS est le V-score. Après entraînement du Transformer aux hyperparamètres décrits plus haut, les V-scores obtenus au voisinage de la transition étaient systématiquement inférieurs à 10^{-4} , ce qui suggérait une convergence satisfaisante en énergie. Nous avons alors procédé à l'évaluation de la magnétisation afin d'observer et de quantifier la transition ferromagnétique-paramagnétique. Une seconde étape de validation consistait à vérifier que le réseau FNQS, entraîné avec désordre, était capable de retrouver la physique du système sans désordre; en particulier, nous devons retrouver le *finite-size scaling* décrit à la section 4.2 sur les pics de la dérivée $\partial \langle M_z^2 \rangle / \partial h_0$.

Pour les petites tailles, nous observons clairement une croissance du pic de la dérivée conforme au scaling attendu; en revanche, à partir de $L = 48$ en 1D, la hauteur du pic stagnait (figure 28). Cette stagnation n'étant pas compatible avec ce que nous avons observé avec nos méthodes exposées dans la partie 3, nous avons mis en oeuvre plusieurs stratégies correctives, que l'on peut séparer en deux classes : celles agissant sur l'évaluation des observables, et celles agissant sur l'entraînement lui-même.

Tentatives sur l'évaluation. Notre premier réflexe a été d'augmenter la résolution de la grille en h_0 au moment de l'évaluation. Cela n'a pas amélioré le résultat : la dérivée numérique amplifie mécaniquement le bruit d'échantillonnage, et une grille plus fine ne fait que démultiplier ce bruit sur une quantité déjà bruitée. Nous avons ensuite augmenté le nombre d'échantillons utilisés pour l'estimation Monte Carlo de la magnétisation, sans succès notable aux grandes tailles. Nous avons enfin envisagé d'utiliser la susceptibilité de fidélité estimée via l'information quantique de Fisher (QFI) comme observable alternative, potentiellement plus sensible à la transition; cette piste n'a pas non plus permis de récupérer le scaling attendu.

La conclusion de cette série de tentatives est la suivante : un V-score de l'ordre de 10^{-4} garantit une très bonne convergence en énergie, mais ne constitue pas un certificat de qualité pour les autres observables. Cette dissymétrie est attendue théoriquement : pour

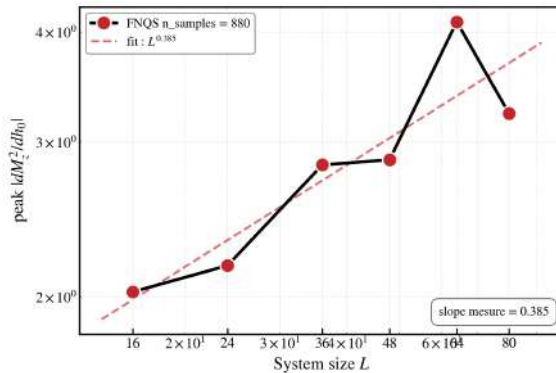


FIGURE 28 – Dérivée $\partial\langle m^2 \rangle / \partial h_0$ obtenue à partir du FNQS pour différentes tailles L en 1D, avec $N_{\text{samples}} = 880$ à l’entraînement. La hauteur du pic croît avec L pour les petites tailles mais stagne à partir de $L = 48$, en contradiction avec le *finite-size scaling* attendu.

un état variationnel à distance ϵ de l’état cible, l’erreur sur l’énergie se comporte en $\mathcal{O}(\epsilon^2)$ par le principe variationnel, tandis que l’erreur sur une observable générique reste en $\mathcal{O}(\epsilon)$ [21, 4]. Aucune amélioration du post-traitement ne pouvait donc compenser une fonction d’onde intrinsèquement insuffisamment convergée.

Tentatives sur l’entraînement. Nous avons alors cherché à modifier le protocole d’entraînement. Deux premières pistes ont été testées : resserrer la fenêtre des points d’entraînement en h_0 autour de la transition, puis densifier le nombre de points dans cette fenêtre. Aucune de ces deux modifications n’a significativement amélioré les résultats aux grandes tailles.

À ce stade, nous étions tentés de conclure que la stratégie d’entraînement avec désordre, non documentée dans la littérature au moment de ce travail et permettant pourtant d’étudier en un unique entraînement une large famille d’hamiltoniens désordonnés, était trop ambitieuse relativement à nos moyens. Un retour à la littérature FNQS, et en particulier à l’article [6], nous a cependant mis sur la piste du véritable problème.

Le paramètre critique est le nombre total d’échantillons N_{samples} utilisés lors de l’estimation du gradient de la fonction de perte à chaque étape d’optimisation. Notre valeur initiale, $N_{\text{samples}} = 1000$, se situait à un ordre de grandeur en dessous de la valeur recommandée (10^4). Or ce paramètre, bien que fondamental, est directement responsable du coût mémoire dominant : la jacobienne intervenant dans l’optimisation par *stochastic reconfiguration* est de taille $N_{\text{samples}} \times N_{\text{params}}$.

Nous avons dans un premier temps augmenté N_{samples} à 3000, valeur maximale compatible avec la mémoire des GPU disponibles à l’École (cf. section 4.3.4). Les courbes obtenues étaient nettement meilleures que les précédentes, mais conservaient un niveau de bruit trop important pour extraire un scaling comparable à celui fourni par d’autres méthodes numériques de référence. Pour atteindre la valeur recommandée de $N_{\text{samples}} = 10000$ sans sacrifier l’architecture du Transformer, nos ressources computationnelles devaient être significativement augmentées.

4.3.4 Ressources computationnelles

L’immense majorité des calculs de cette étude a été réalisée sur les GPU des salles informatiques de l’École polytechnique, équipées de NVIDIA RTX 4000 Ada Generation (20 Go de VRAM GDDR6). La procédure SLURM nous a permis de lancer un grand nombre de calculs en parallèle, ce qui a été déterminant pour l’évaluation rapide d’observables (80 GPU en parallèle pour certains calculs).

Cette infrastructure devient toutefois insuffisante dès lors que l’on souhaite porter N_{samples} à 10000 tout en conservant l’architecture du Transformer : la seule jacobienne, de taille $N_{\text{samples}} \times N_{\text{params}} \approx 10^4 \times 2,0 \times 10^5 = 2 \times 10^9$ coefficients en `float32`, représente à elle seule environ 8 Go, ce qui, combiné aux activations et aux états de l’optimiseur, excède largement les 20 Go de VRAM d’une RTX 4000. Le cluster de calcul du CPhT, qui aurait pu répondre à ce besoin, étant resté indisponible durant toute la durée du PSC, nous avons eu recours à une location ponctuelle de GPU cloud. L’entraînement final a ainsi été effectué sur 11 NVIDIA H200 (140 Go de VRAM chacune,

bande passante mémoire accrue) pendant environ 10 heures, la totalité de la capacité de calcul allouée étant exploitée.

Les résultats finaux présentés ci-après montrent que nous retrouvons le comportement physique attendu du système et valident la viabilité de l'entraînement avec désordre. La tendance observée entre $N_{\text{samples}} = 1000, 3000$ et 10000 , amélioration systématique du scaling aux grandes tailles, suggère qu'**une augmentation supplémentaire de ce paramètre, au prix d'une infrastructure encore plus conséquente, permettrait de réduire davantage le bruit résiduel et d'affiner l'estimation des exposants critiques.**

5 Résultats finaux concernant l'influence du désordre sur la transition de phase ferromagnétique-paramagnétique dans le modèle d'Ising en champ transverse

5.1 Mise en évidence des transitions de phase

Comme notre objectif est d'étudier l'influence du désordre sur la transition de phase, nous devons être capables d'établir l'existence d'une transition de phase plus rigoureusement que dans la partie sans désordre. Cependant, comme nous sommes en taille finie, toutes les observables resteront continues ce qui peut nous empêcher de bien déceler les transitions de phase. Nous devons donc utiliser d'autres méthodes.

Ces méthodes reposent sur la divergence de la longueur de corrélation, pour les transitions d'ordre 2, ce qui est notre cas, dans la limite thermodynamique :

$$\xi \sim |h - h_c|^{-\nu}$$

La première est le *Finite Size-Scaling* ou *mise à l'échelle pour des tailles finies* en français qui permet de faire une mise à l'échelle. Cette méthode, élaborée par Fisher et Barber [8] tire parti de la divergence de la longueur de corrélation pour obtenir une courbe vérifiée par toutes les tailles de système, d'où le nom de collapsing. Si toutes les courbes coïncident à des mises à l'échelle près, cela signifie que la transition est régie par la longueur de corrélation et son exposant critique, conformément à la théorie.

Pour cela, nous traçons le cumulants de Binder

$$U_L(h) = 1 - \frac{\langle m^4 \rangle_L}{3 \langle m^2 \rangle_L^2}, \quad (6)$$

en fonction de la variable réduite $(h - h_c) L^{1/\nu}$, pour différentes tailles L . Étant adimensionné, U_L ne nécessite aucun préfacteur de rescaling sur l'axe vertical : son ansatz de mise à l'échelle se réduit à

$$U_L(h) = \tilde{U}\left((h - h_c) L^{1/\nu}\right). \quad (7)$$

Les courbes $U_L(h)$ obtenues pour différentes tailles doivent ainsi se croiser en un point unique $h = h_c$, ce qui fournit une détermination indépendante du point critique, avant de s'effondrer sur une unique fonction d'échelle \tilde{U} une fois l'abscisse rescalée. Le fit de l'effondrement se fait selon les méthodes présentées dans [22, 23].*

Une seconde méthode qui est dérivée de la théorie du Finite Size Scaling (FSS) est le *Peak Scaling*. Cette méthode consiste à étudier pour chaque taille du système le maximum de la dérivée de $\langle M_z^2 \rangle$ par rapport au champ transverse. Ensuite, la théorie FSS prédit que la position du pic de la dérivée $h^*(L)$ converge vers h_c la champ critique selon la loi :

$$h^*(L) = h_c + A L^{-1/\nu}$$

De plus, la hauteur du pic peut également être exploitée pour vérifier l'existence de loi d'échelle :

$$\langle M_z^2 \rangle_{\text{max}}(L) \equiv \left. \frac{d}{dh} \langle M_z^2 \rangle(h, L) \right|_{h=h^*(L)} \sim L^{\rho/\nu}$$

où ρ est un exposant spécifique à l'observable $\langle M_z^2 \rangle$.

Ces deux techniques nous permettent donc de vérifier la présence de loi d'échelle, ce qui suggère la présence d'une transition de phase et le cas échéant permet d'obtenir le champ transverse critique.

Dans l'ensemble de la section, les courbes dans les teintes de rouge correspondent à la 1D et celles en teintes de bleu à la 2D.

5.2 Validation des entraînements du réseau

L'accès final aux GPU H200 et notre dernière architecture de Transformer nous a permis d'obtenir des foundation NQS suffisamment bien entraînés pour conduire une étude du modèle d'Ising 1D et 2D avec désordre.

Pour vérifier de la bonne convergence de nos états, nous utilisons la métrique V-score. Nous l'avons calculé en chaque point d'une grille bidimensionnelle (h_0, σ) . La Figure 29 confirme que le V-score reste largement inférieur à 10^{-3} dans la majorité de la zone étudiée et croît naturellement avec σ , et est le plus mauvais à h_c . Remarquons que nous avons une approximation correcte de l'énergie à grand σ alors même que l'entraînement n'a eu lieu qu'à une valeur fixée de $\sigma = 0.1$. **Le réseau semble donc capable de généraliser le comportement à plus grand désordre.** Cependant, si nous en avons eu les ressources computationnelles, une extension en σ de l'entraînement aurait pu représenter une piste d'amélioration.

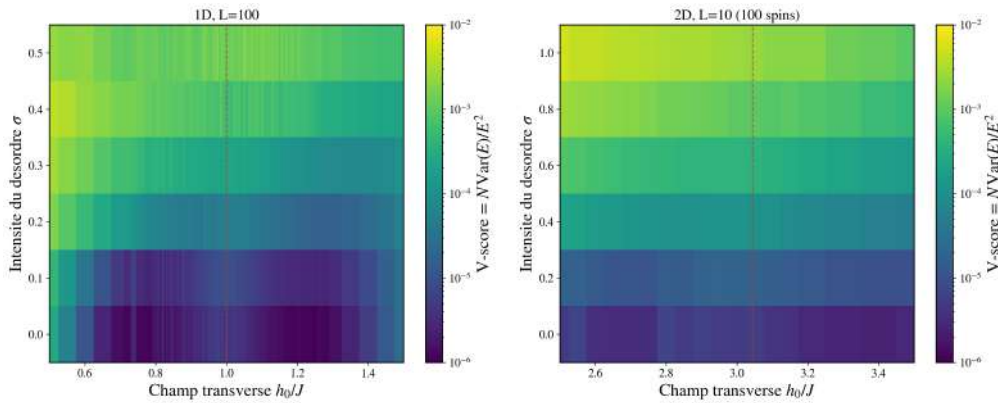
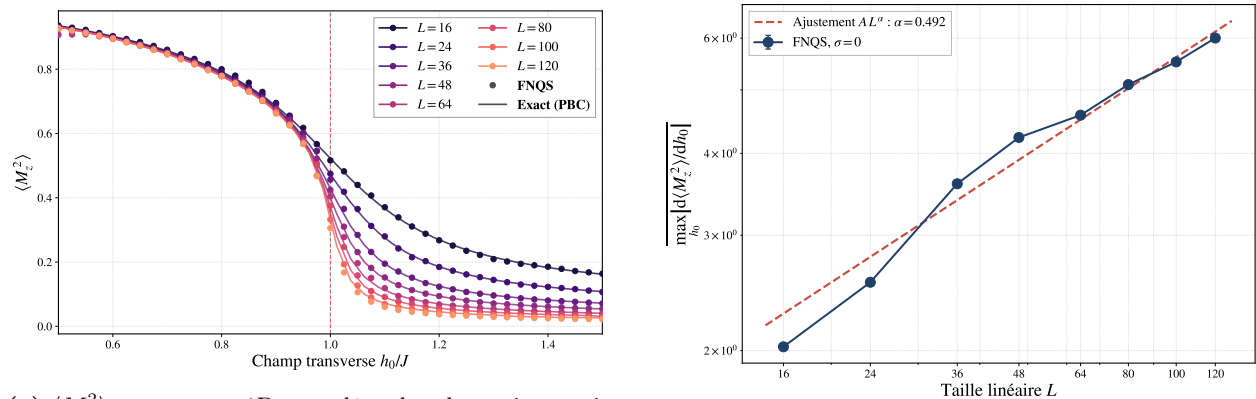


FIGURE 29 – V-score sur la grille (h_0, σ) . En chaque point de cette grille, l'évaluation du V-score moyen est fait sur du 10 tirages.

Une seconde validation de la très bonne qualité des états entraînés est le fait de retrouver le comportement de $\langle M_z^2 \rangle$ sans désordre en 1D. En 1D sans désordre, le comportement de $\langle \hat{M}_z^2 \rangle$ en fonction de h_0 est connu exactement via la transformation de Jordan-Wigner suivie d'une diagonalisation de Bogoliubov [24]. De plus, la Figure 30b.



(a) $\langle M_z^2 \rangle$ vs exact en 1D sans désordre; les traits continus correspondent aux valeurs exactes en chaque taille

(b) FSS en 1D sans désordre.

FIGURE 30 – Validation 1D sans désordre

En 2D, il n'existe pas de courbe exacte de la magnétisation, même sans désordre. Ainsi, le benchmark de nos états repose sur la valeur de l'exposant critique ν et du champ critique h_c . Nous utilisons la méthode d'effondrement du binder cumulant pour extraire ces deux valeurs. Nous observons un effondrement extrêmement net, les valeurs extraites concordent avec ce que l'on trouve

dans la littérature. Nous obtenons $\nu = 0,616 \pm 0.059$ et $h_c = 3,032 \pm 0.008$ où la littérature donne $h_c = 3.04438 \pm 0.00002$ et $\nu = 0.6301 \pm 0.0004$ [25]. De plus, le binder cumulant tracé en différentes tailles possède la propriété de se croiser au point critique de la transition de phase ferromagnétique-paramagnétique. On visualise parfaitement ce comportement sur la Figure 31a, en un point concordant avec la valeur théorique du champ critique h_c , ce qui confirme encore la validité de nos simulations.

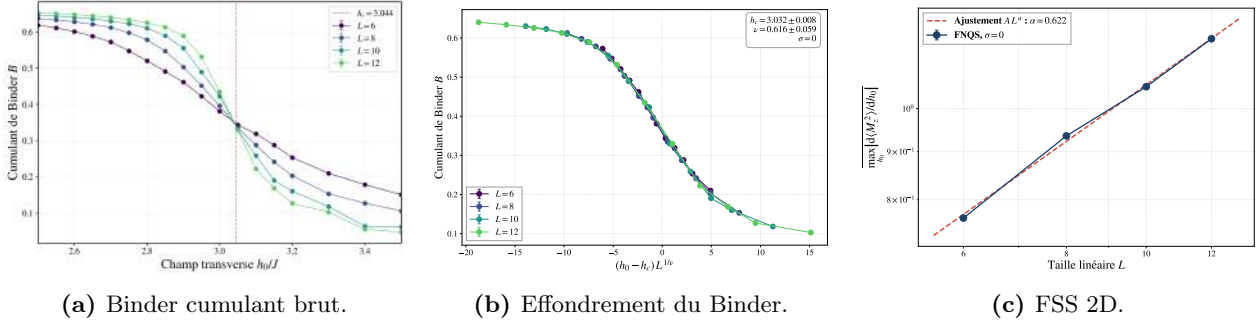


FIGURE 31 – Validation 2D sans désordre : effondrement du Binder cumulant

Nos états FNQS ayant une très bonne convergence en V-score avec et sans désordre et approximant de manière convaincante le résultat physique sans désordre, ils garantissent une étude sereine des résultats avec désordre.

5.3 Effet qualitatif du désordre sur le comportement du système

La Figure 32 permet de visualiser directement l'effet du désordre sur notre transition de phase à taille donnée : en 1D comme en 2D, la courbe de $\langle M_z^2 \rangle$ est de moins en moins raide quand σ augmente, ce qui montre que le désordre rend le passage du régime ferromagnétique au régime paramagnétique moins abrupte en fonction de h_0 .

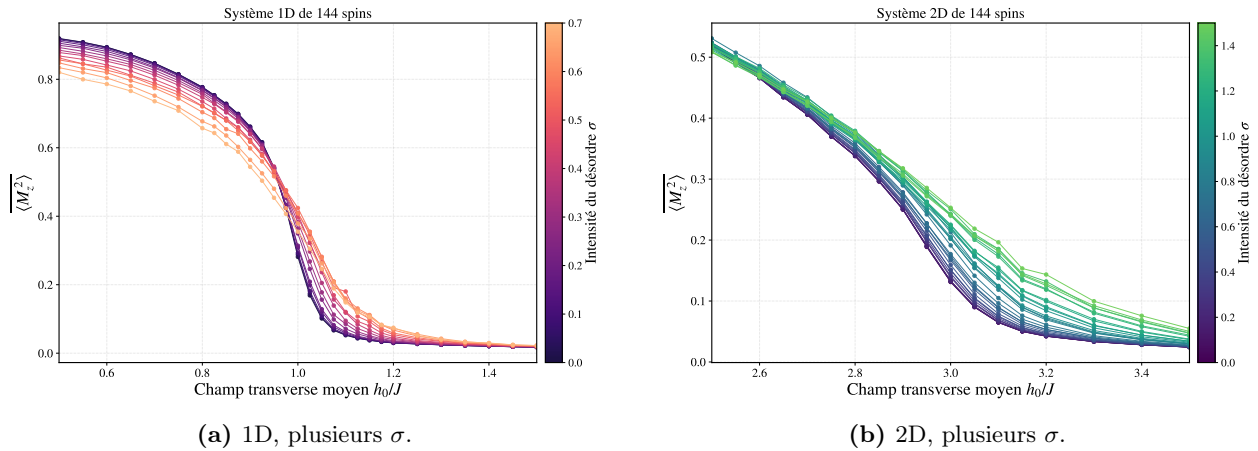
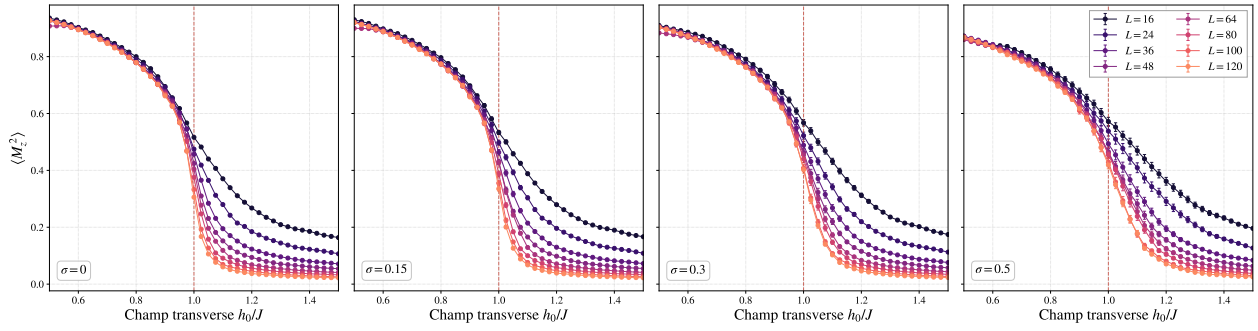
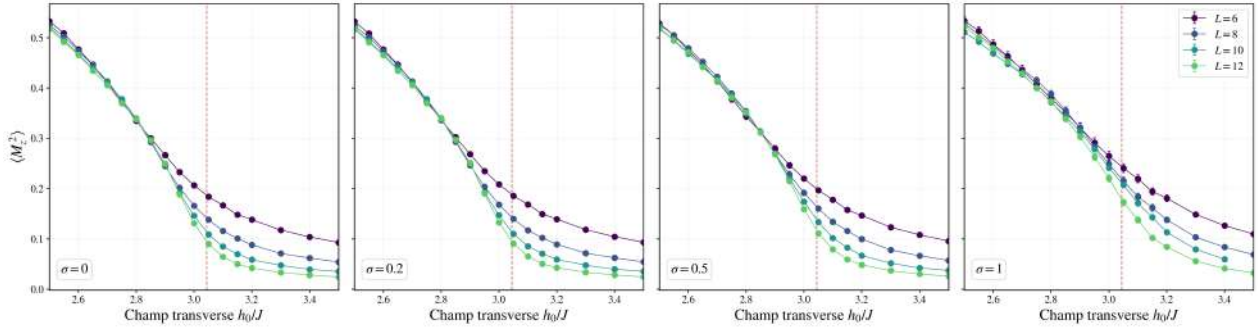


FIGURE 32 – $\langle M_z^2 \rangle$ (moyenné sur 80 réalisations du désordre) en fonction de h_0 à plusieurs intensités du désordre σ .

La question suivant naturellement est donc : existe-t-il encore une transition de phase à σ donné, même si à taille finie la courbe $\langle M_z^2 \rangle$ en fonction de h_0 est moins raide que pour celle à $\sigma = 0$? Les courbes de $\langle M_z^2 \rangle$ pour plusieurs L permettent visuellement et qualitativement de répondre à cette question. En effet, la pente doit s'accroître lorsque L augmente s'il y a une transition de phase.



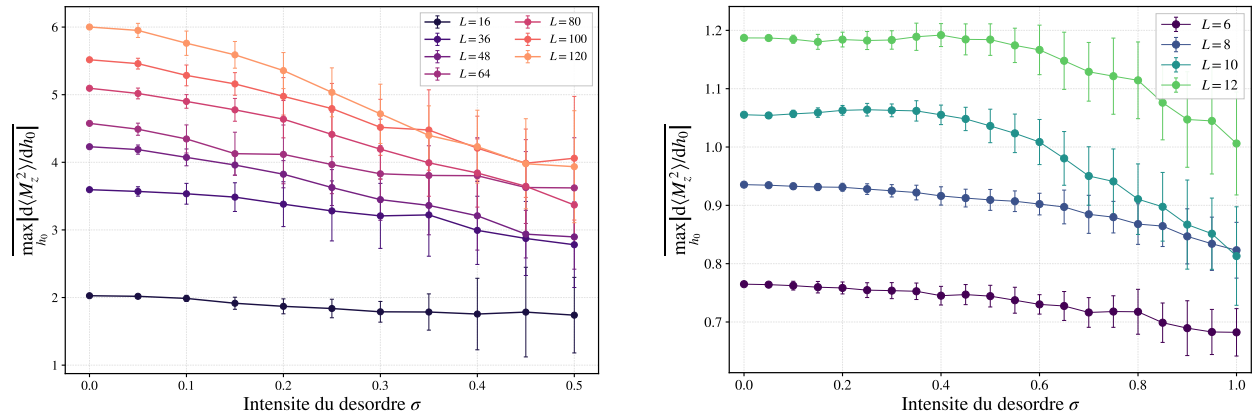
(a) 1D, plusieurs σ .



(b) 2D, plusieurs σ .

FIGURE 33 – $\overline{\langle M_z^2 \rangle}$ en fonction de h_0 pour plusieurs L , à différents désordres.

Nous observons sur la Figure 33 l'accentuation des pentes en fonction de L en 1D et en 2D, malgré l'ajout de désordre. La transition de phase semble donc subsister malgré le désordre. Cependant, l'accentuation semble s'adoucir au fur à mesure que σ augmente.



(a) 1D.

(b) 2D.

FIGURE 34 – Maximum de $|\partial_{h_0} \langle M_z^2 \rangle|$ en fonction de σ .

Le comportement qualitatif apparaît plus clairement sur la Figure 34, qui présente $\max_{h_0} |\partial_{h_0} \langle M_z^2 \rangle|$ en fonction de σ . On constate ainsi que plus le désordre est intense, plus la transition de phase s'adoucit, sa pente maximale étant plus faible. Elle continue néanmoins à croître avec L , ce qui suggère que la transition de phase persiste, même si elle change peut-être de nature : c'est ce que nous étudions plus en détails dans la partie suivante. L'étude précise de h_c et de l'exposant critique ν éclairera quantitativement ce point.

Notons par ailleurs que l'on voit une différence de comportement entre 1D et 2D aux grandes tailles : en 1D, même un très faible désordre commence à détériorer la raideur de la transition de phase, alors qu'en 2D, elle se maintient jusqu'à un désordre relativement important ($\sigma \approx 0,5$) avant de commencer à l'adoucir.

5.4 Effet quantitatif du désordre sur le comportement du système

Pour étayer l'existence d'une transition de phase (qui ne peut être complètement confirmée qu'à taille infinie), nous traçons l'effondrement optimal du cumulante de Binder à différentes valeurs de σ , comme présenté sur la Figure 35.

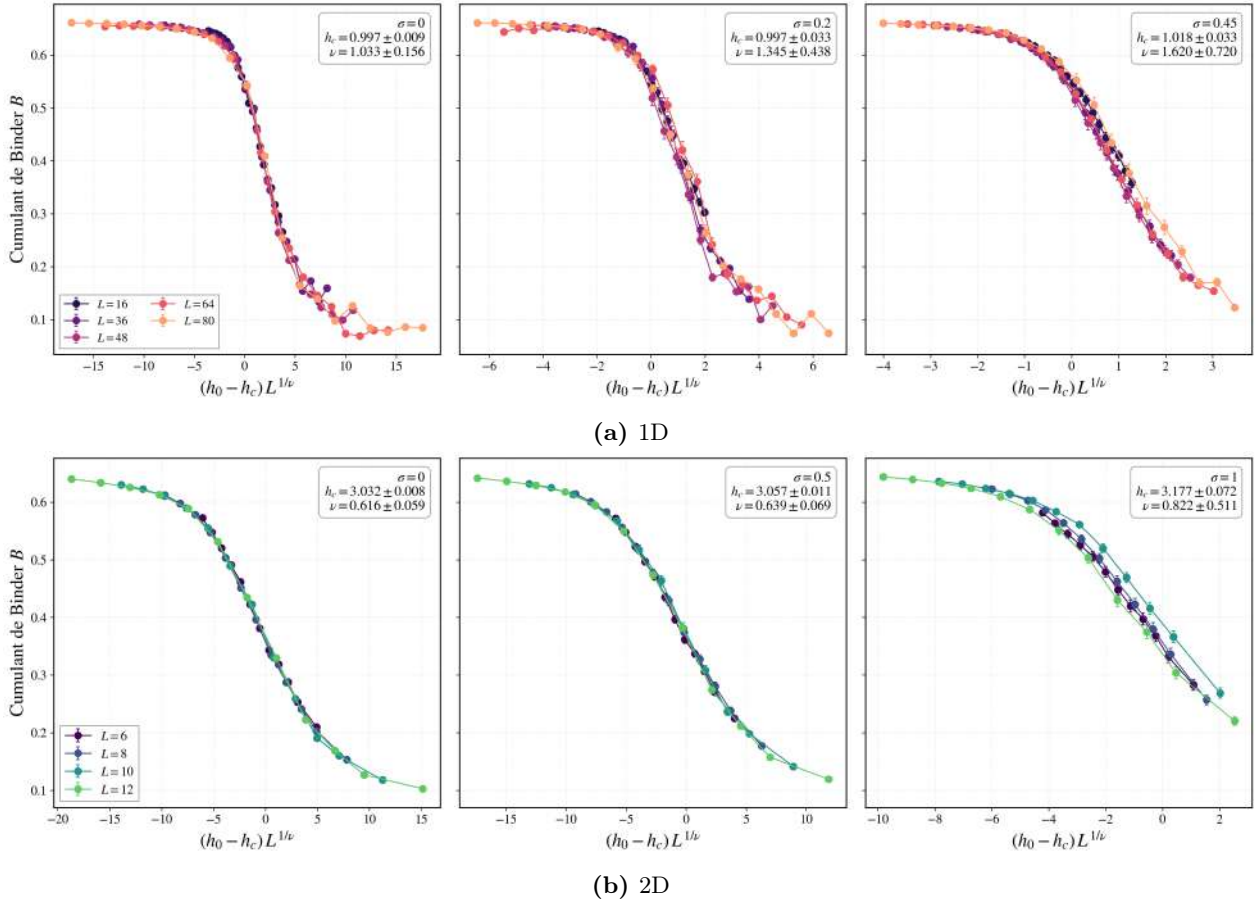


FIGURE 35 – Effondrement du cumulante de Binder pour différentes intensités du désordre σ , en 1D (a) et 2D (b). À gauche, $\sigma = 0$ (cas sans désordre) : l'effondrement est excellent et fournit une référence. Au centre et à droite, désordre croissant : l'effondrement se dégrade visiblement à grand σ .

La méthode est identique à celle employée sans désordre, à ceci près que nous moyennons les observables sur les réalisations du désordre avant de réaliser l'effondrement. La Figure 35 montre que l'effondrement reste excellent pour des valeurs de σ modérées, ce qui constitue un argument fort en faveur de la persistance de la transition de phase. À mesure que σ augmente, en revanche, l'effondrement se dégrade visiblement : nos réseaux atteignent alors la limite de leur portée de généralisation, et l'extraction de h_c et ν devient moins fiable, comme en témoignent les barres d'erreur croissantes sur la Figure 36, calculées par bootstrap sur les réalisations du désordre. L'exploitation systématique de cet effondrement à différentes valeurs de σ permet d'extraire les exposants critiques reportés sur la Figure 36.

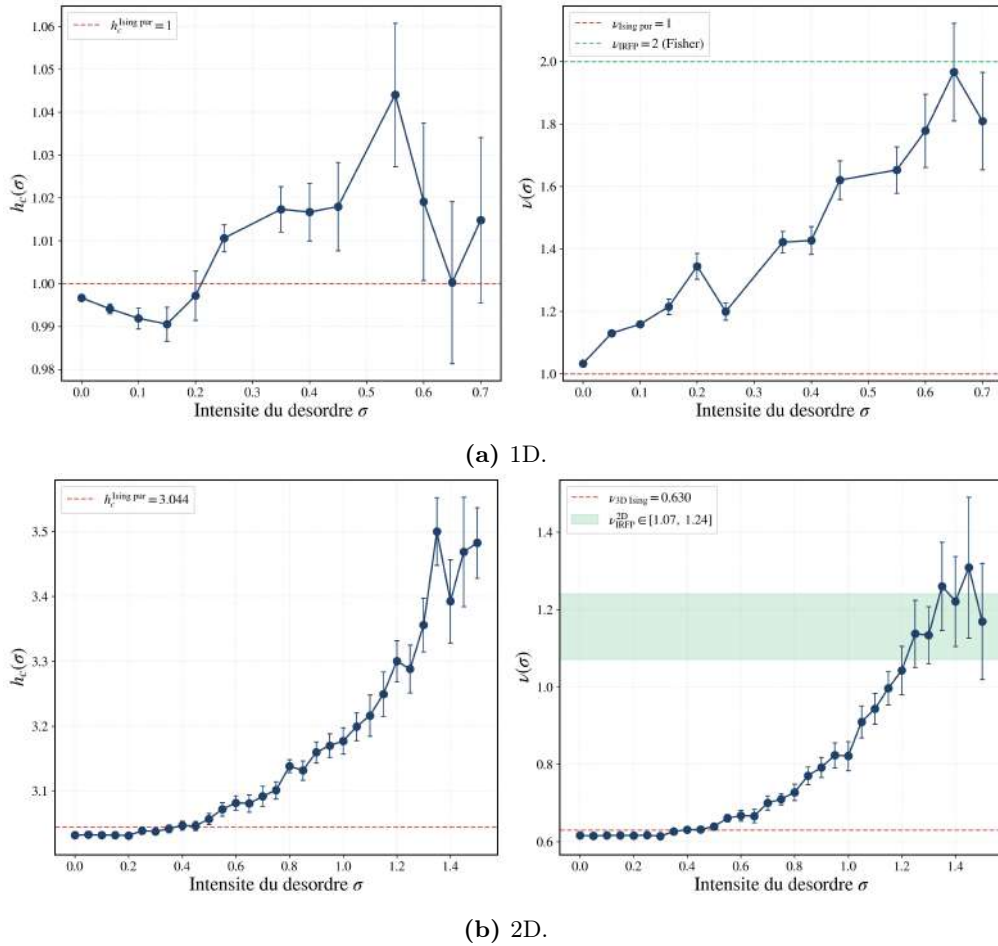


FIGURE 36 – Exposants critiques en fonction de σ .

Deux régimes se dessinent clairement en 2D : à faible sigma, la transition semble préservée de manière quantitative, ses exposants sont invariants et égaux à ceux du modèle Ising sans désordre tandis qu'à grand sigma, la transition voit ses exposants augmenter et ils semblent ensuite stagner.

Lorsqu'on introduit du désordre dans un système qui présente une transition de phase, deux scénarios sont possibles : soit la nature de la transition est préservée, soit le désordre la modifie qualitativement. Pour le TFIM, le critère de Harris [26] prédit que l'on est dans le second cas, en 1D comme en 2D.

Fisher [8] a montré en 1D que la transition est alors contrôlée par un nouveau régime universel, appelé *point fixe de désordre infini* (Infinite Randomness Fixed Point, IRFP), dans lequel le désordre domine la physique critique à toutes les échelles. Ce régime se caractérise par des valeurs précises des exposants ν , qui diffèrent de celles du modèle sans désordre et jouent le rôle de signature quantitative :

- en 1D : $\nu = 2$ [8] ;
- en 2D : $\nu \in [0,7; 1,24]$ [27, 9, 10].

Ainsi, l'observation de l'évolution des exposants pourrait correspondre à la signature de la transition du régime de modèle d'Ising sans désordre au régime de l'IRFP. En effet, ν semble se stabiliser dans les bornes connues dans la littérature [9, 10]. En 2D, la stagnation des exposants à grand désordre pourrait ainsi signifier l'entrée dans le régime de l'IRFP.

En 1D, dès $\sigma = 0.5$, l'exposant ν augmente. De plus, il atteint un plateau à $\nu = 2$ qui pourrait également être le signe d'une entrée dans le régime IRFP. [8].

Le flot (vitesse de convergence) plus lent en 2D qu'en 1D vers l'IRFP est un résultat connu de la littérature. [27, 9].

Cependant, nous devons mentionner que les résultats à grand σ sont à traiter avec précaution car les V-score se dégradent dans cette zone.

Enfin, la Figure 36 est l'occasion de faire une remarque sur nos travaux : on constate très nettement que la courbe en 1D est beaucoup moins propre qu'en 2D. Plus généralement, lors de la

résolution de nos problèmes en grande taille en 4.3.3, nous avons constaté que la dimension 1 y était beaucoup plus sujette que la 2D. Ainsi, à puissance de calcul donnée, nos méthodes semblent capables de donner des évaluations de $\langle M_z^2 \rangle$ précises à des tailles bien supérieures en 2D qu'en 1D, malgré des V-scores comparables. Nous n'avons pas trouvé d'explication convaincante à ce phénomène, qui mériterait d'être davantage investigué.

5.5 Prolongement potentiel de notre étude

Notre étude mériterait donc, avec une puissance de calcul accrue, d'être poussée davantage, cela permettrait notamment d'obtenir des résultats en plus grande taille et meilleurs à grand σ et éventuellement obtenir une estimation plus précise des exposants et de la transition vers l'IRFP.

5.5.1 Entraînement du réseau

Nous augmenterions l'architecture du réseau pour avoir un nombre de paramètres comparable ou supérieur au réseau proposé dans [7] (1 million de paramètres vs 200k dans notre réseau). Nous augmenterions éventuellement encore le nombre d'échantillons total à 16k [6]. Nous entraînerions aussi le réseau à plusieurs valeurs de sigma en même temps.

5.5.2 Evaluation du réseau

Pour l'évaluation du réseau, nous pourrions augmenter le nombre d'échantillons en les points où nous réalisons un échantillonnage MCMC. De plus nous pourrions faire plusieurs échantillonnages par h_0 à différents σ pour améliorer l'importance sampling. Enfin, lors de l'étude avec désordre il serait meilleur de moyenner sur davantage de réalisations du désordre pour limiter l'erreur statistique.

5.6 Potentiel d'exploitation et d'innovation

Notre étude a permis de montrer **l'efficacité des méthodes numériques utilisées couplant les réseaux de neurones et modèle de fondation**. Nous avons ainsi résolu des problèmes et développé des outils, des méthodes que d'autres chercheurs pourront réutiliser s'ils souhaitent explorer davantage cette question en prenant appui sur nos résultats. **La conception des notebooks (dont nous parlons dans la sous-partie 6.2) pour le Journal of Open Source répond également à cette logique de partage des connaissances que nous avons acquises pendant ce projet.** A la suite de cette étude, d'autres recherches pourraient être menées pour approfondir le sujet et arriver à quantifier plus précisément nos résultats. De même, d'autres modèles de matériaux magnétiques pourraient être étudiés avec ces méthodes comme le modèle d'Heisenberg par exemple.

Notre étude présente également plusieurs potentiels d'exploitation dans le monde hors académique.

Premièrement, **notre étude permet de modéliser le comportement de certains matériaux magnétiques réels**. En plus du CoNb_2O_6 que nous avons vu précédemment, nous pouvons également modéliser les aimants dilués $\text{LiHo}_x\text{Y}_{1-x}\text{F}_4$ comme le montre l'article de Coldea[28].

En outre, **l'étude de la transition de phase quantique du modèle d'Ising en champ transverse aléatoire avec un désordre distribué selon une loi normale présente un potentiel d'innovation significatif dans le domaine du calcul quantique** et, plus précisément, du recuit quantique. Ce dernier est une méthode d'optimisation qui exploite les fluctuations quantiques induites par le champ transverse pour explorer efficacement un paysage d'énergie complexe et atteindre son état fondamental, correspondant à la solution optimale d'un problème donné, avec des applications directes en logistique, finance, planification industrielle ou chimie computationnelle. Les processeurs de recuit quantique existants, tels que ceux développés par D-Wave, reposent sur un Hamiltonien dont la structure est formellement celle du modèle d'Ising, mais dont les paramètres couplages et champs transverses locaux sont inévitablement entachés d'imperfections de fabrication dont la distribution est typiquement gaussienne (d'écart-type σ). Notre étude a ouvert la voie à d'autres recherches permettant la caractérisation précise de la transition de phase en présence d'un tel désordre gaussien et qui pourraient essayer d'établir quantitativement jusqu'à quel rapport $\frac{\sigma}{\hbar}$ la transition de phase demeure suffisamment nette pour garantir l'efficacité du recuit, et fournir ainsi des critères rigoureux de tolérance de fabrication.

Finalement, comme nous l'avons vu lors du Hackathon de l'EPFL (dont nous parlons dans la partie 6), tout problème de combinatoire à valeurs entières peut se ramener à trouver l'état fondamental d'un Hamiltonien d'Ising en champ transverse. Ainsi, avec nos méthodes numériques qui permettent de s'affranchir des problèmes de stockage et de calcul que nous avons mentionnés dans la partie 1, **nous pouvons simuler et trouver les solutions de problèmes d'optimisation entières quelque soit le domaine d'application**. Notre étude illustre ainsi la puissance de ces méthodes numériques capables de fonctionner correctement même en présence de désordre.

6 Évènements et ressources produites en lien avec notre PSC

6.1 Pour reproduire nos résultats

Les états quantiques neuronaux de fondation que nous avons entraînés et qui ont produit les résultats présentés dans la cinquième partie sont disponibles ici : <https://github.com/MaxAngl/FoundationNeuralQuantumStatesDisorderedIsing/> Nous y retrouverons également l'échantillonnage des configurations homogènes permettant de reproduire l'évaluation d'observables par importance sampling.

6.2 Construction d'un outil pédagogique et de visualisation

Comprendre intuitivement l'optimisation d'un état quantique neuronal est difficile sans visualisation directe. Pour répondre à ce besoin, nous avons développé une interface web permettant de manipuler en temps réel les hyperparamètres du réseau et d'observer instantanément leur impact sur la dynamique des spins, l'énergie et la magnétisation. Cet outil, conçu pour être accessible à des physiciens non spécialistes de l'apprentissage machine, est disponible à l'adresse : <http://nnqs.anglade.org>

6.3 Participation à la publication de la bibliothèque *Netket Foundation* et de notebooks d'enseignement

Dans le cadre de notre étude, nous avons développé plusieurs heuristiques pour résoudre nos problèmes de simulation. Notre tuteur, jugeant ces approches utiles à tout étudiant souhaitant aborder la transition de phase ferromagnétique-paramagnétique avec désordre, nous a demandé de rédiger des notebooks pédagogiques retraçant les problématiques rencontrées ainsi que les solutions apportées.

Le premier notebook retrace notre méthode et l'explication de notre code utilisé pour étudier les matériaux magnétiques avec désordre tandis que le second notebook met en valeur des fonctions qui peuvent être utiles pour un projet de physique computationnelle mais qui ne sont pas encore implémentées dans la bibliothèque de Netket. Dans le second notebook, nous fournissons ainsi un code et une documentation pour l'Importance Sampling dont nous avons parlé dans la partie 4, pour s'assurer de la bonne convergence d'une chaîne de Monte Carlo (la thermalisation) et pour initialiser une chaîne de Monte Carlo sur un état quantique prédéfini.

Ces notebooks seront publiés en complément d'une autre étude dans le Journal of Open Source Software (JOSS).

6.4 Participation à un Hackathon d'informatique quantique à l'EPFL

Lors du week-end du 28 février, nous avons participé au Hackathon de l'EPFL, une compétition internationale et exigeante rassemblant des étudiants de master spécialisés en informatique quantique venus de toute l'Europe (Italie, Danemark, Turquie...). Nous étions la seule équipe de l'École polytechnique, malgré une découverte des bases de l'informatique quantique seulement quelques jours auparavant.

En 24 heures, nous avons travaillé sur un problème d'optimisation appliqué à l'assurance : allouer les ressources d'une compagnie traitant des milliers de dossiers, dont certains frauduleux, tout en respectant la contrainte de ressources d'inspection limitées. Nous avons proposé une modélisation originale en intégrant la notion de réseaux de fraude.

Nous avons reformulé le problème d'optimisation initial en Hamiltonien d'Ising via la méthode QUBO. Nous avons ensuite commencé la résolution grâce aux méthodes variationnelles et aux ansatz développés dans notre étude. C'est précisément cette approche qui nous a permis de nous démarquer : là où les autres équipes étaient limitées à des réseaux de 5 à 10 spins car se limitaient uniquement à des algorithmes quantiques, nos algorithmes hybrides quantique-classique nous ont permis de simuler des systèmes plus grands et de proposer une solution alternative et originale qui pourrait dès aujourd'hui passer à l'échelle. **Nous avons décroché la 2ème place de ce hackaton grâce à nos connaissances sur les NQS.**

7 Conclusion

En conclusion, **ce PSC a été extrêmement enrichissant** pour nous, et ce sur plusieurs plans.

D'abord, il nous a poussés à **acquérir énormément de nouvelles connaissances**, à la fois en Physique Quantique et Statistique, en Mathématiques, et en Informatique. Il nous a donné l'occasion de **prendre en main des outils à la pointe de la recherche actuelle dans le domaine des simulations de systèmes quantiques par des méthodes de réseaux de neurones quantiques.**

Ensuite, **il nous a beaucoup appris sur la méthodologie inhérente à la recherche scientifique**, de façon très concrète. Nous avons découvert des outils et acquis des automatismes qui nous seront très utiles dans la suite de notre parcours académique. Nous avons en outre été confrontés à de nombreux problèmes qui font partie du quotidien des chercheurs, et avons appris à chercher la bonne approche pour les résoudre. Ce projet nous a de plus ouvert de nombreuses portes : participation à la publication d'un article scientifique, podium au hackathon de l'EPFL, opportunités de stage.

Enfin, du point de vue scientifique, **nous avons produit une nouvelle étude du modèle d'Ising en champ transverse désordonné selon une loi gaussienne, en dimensions 1 et 2, dont nous n'avons pas trouvé d'équivalent dans la littérature.** Nous avons réussi, en mettant en œuvre des méthodes extrêmement pointues, à étudier de façon quantitative l'influence de perturbations sur la transition de phase ferromagnétique-paramagnétique dans ce modèle. Pour ce faire, **nous avons développé une méthode scientifique complète, qui applique des outils de Deep Learning à l'étude de systèmes quantiques désordonnés**, en tenant compte de leurs spécificités physiques. Cette méthode sera utile au-delà de ce PSC, puisqu'**elle sera conservée et utilisée par l'équipe du CPhT** au sein de laquelle nous avons réalisé notre projet.

Pour conclure, notre rapport ouvre sur **deux axes majeurs de poursuite de nos travaux.** D'abord, comme exposé tout au long du rapport, notre principal facteur limitant fut la puissance de calcul à laquelle nous avons accès. Ainsi, **utiliser des GPU plus puissants et en plus grand nombre** pourrait permettre d'augmenter la taille des Transformers employés ainsi que du nombre d'échantillons $N_{samples}$ utilisés à chaque étape de l'optimisation. Notre étude montre qu'on obtiendrait alors des résultats physiques encore plus précis, et ce pour des systèmes de spins encore plus grands ou à plus grand désordre, ce qui pourrait produire des résultats plus significatifs au sujet de régimes physiques mal connus, comme l'IRFP en 2D. Ensuite, la méthodologie élaborée tout au long de notre PSC n'est pas spécifique au modèle d'Ising en champ transverse désordonné. Nous avons choisi ce modèle comme premier sujet d'étude, ce qui nous a permis de développer et perfectionner de nombreux outils. À présent, il semble tout-à-fait possible de **les appliquer à d'autres modèles physiques désordonnés, plus élaborés que celui d'Ising**, ce qui pourrait fournir de nouveaux résultats physiques très intéressants. Le premier candidat naturel pour préciser la modélisation des propriétés magnétiques des matériaux pourrait être le modèle d'Heisenberg désordonné.

8 Remerciements

Nous exprimons notre profonde gratitude au Professeur Vicentini, notre tuteur de PSC, pour la qualité exceptionnelle de son enseignement et sa volonté constante de nous transmettre son savoir mais également le métier de chercheur. Sa pédagogie inspirante, sa patience et sa disponibilité sans faille ont transformé ce projet en une véritable leçon scientifique et humaine. Merci de nous avoir formés avec une si grande bienveillance.

Nous remercions également Antoine Misery, qui s'est également rendu disponible tous ces mois durant afin de nous former sur les outils et les méthodes du domaine, pour répondre à nos interrogations et pour partager des fous rires réguliers durant les réunions de cadrage.

Enfin, nous remercions le Binet Réseau qui nous a permis de configurer nos ordinateurs pour nous connecter aux postes des salles d'informatique de l'École et qui nous ont ainsi permis, dans les périodes d'abondance, de paralléliser nos calculs en monopolisant plus d'une centaine d'ordinateurs.

Nous remercions également cet élève héroïque et anonyme de l'EPFL, qui a appuyé durant près d'une heure sur le bouton pour allumer la lumière dans la pièce dans laquelle nous travaillions durant le hackathon, alors que la lumière s'éteignait automatiquement toutes les 5 min et qu'il nous a fallu ainsi réaliser plus de 200 allers retours pour s'éclairer et vaincre le sujet.

Nous souhaitons également adresser nos remerciements les plus sincères aux équipes pédagogiques et administratives de l'École polytechnique qui ont facilité nos démarches et ont assuré un suivi tout au long de notre PSC.

Références

- [1] W. L. McMillan. Ground state of liquid he^4 . *Physical Review*, 138(2A) :A442–A451, 1965.
- [2] Steven R. White. Density-matrix algorithms for quantum renormalization groups. *Physical Review B*, 48(14) :10345–10356, October 1993.
- [3] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac. Classical simulation of infinite-size quantum lattice systems in two spatial dimensions. *arXiv preprint*, 2007.
- [4] Federico Becca and Sandro Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 2017.
- [5] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325) :602–606, 2017.
- [6] Riccardo Rende, Luciano Loris Viteritti, Federico Becca, Antonello Scardicchio, Alessandro Laio, and Giuseppe Carleo. Foundation neural-networks quantum states as a unified ansatz for multiple hamiltonians. *Nature Communications*, 16(1) :7213, 2025.
- [7] Luciano Loris Viteritti, Riccardo Rende, Giacomo Bracci Testasecca, Jacopo Niedda, Roderich Moessner, Giuseppe Carleo, and Antonello Scardicchio. Quantum spin glass in the two-dimensional disordered heisenberg model via foundation neural-network quantum states. *arXiv preprint arXiv :2507.05073*, 2025.
- [8] Daniel S. Fisher. Critical behavior of random transverse-field Ising spin chains. *Physical Review B*, 51(10) :6411–6461, 1995.
- [9] István A. Kovács and Ferenc Iglói. Renormalization group study of the two-dimensional random transverse-field Ising model. *Physical Review B*, 82(5) :054437, 2010.
- [10] C. Pich, A. P. Young, H. Rieger, and N. Kawashima. Critical behavior and Griffiths-McCoy singularities in the two-dimensional random quantum Ising ferromagnet. *Physical Review Letters*, 81(26) :5916–5919, 1998.
- [11] Calvin Krämer, Jan Alexander Koziol, Anja Langheld, Max Hörmann, and Kai Phillip Schmidt. Quantum-critical properties of the one- and two-dimensional random transverse-field Ising model from large-scale quantum Monte Carlo simulations. *SciPost Physics*, 17 :061, 2024.
- [12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30 :5998–6008, 2017. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
- [13] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. Transformer variational wave functions for frustrated quantum spin systems. *Physical Review Letters*, 130(23) :236401, 2023.

- [14] Sandro Sorella. Green function Monte Carlo with stochastic reconfiguration. *Physical Review Letters*, 80(20) :4558–4561, 1998.
- [15] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4) :303–314, 1989.
- [16] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2) :251–257, 1991.
- [17] Alessandro Santini and Filippo Vicentini. Disordered ising models with neural quantum states. *arXiv preprint arXiv :2410.10720*, 2024.
- [18] Dian Wu, Riccardo Rossi, Filippo Vicentini, Nikita Astrakhantsev, Federico Becca, Xiaodong Cao, Juan Carrasquilla, Francesco Ferrari, Antoine Georges, Mohamed Hibat-Allah, Masatoshi Imada, Andreas M. Läuchli, Guglielmo Mazzola, Antonio Mezzacapo, Andrew Millis, Javier Robledo Moreno, Titus Neupert, Yusuke Nomura, Jannes Nys, Olivier Parcollet, Rico Pohle, Imelda Romero, Michael Schmid, J. Maxwell Silvester, Sandro Sorella, Luca F. Tocchio, Lei Wang, Steven R. White, Alexander Wietek, Qi Yang, Yiqi Yang, Shiwei Zhang, and Giuseppe Carleo. Variational benchmarks for quantum many-body problems. *Science*, 386(6719) :296–301, 2024.
- [19] R. Coldea, D. A. Tennant, E. M. Wheeler, E. Wawrzynska, D. Prabhakaran, M. Telling, K. Habicht, P. Smeibidl, and K. Kiefer. Quantum criticality in an Ising chain : Experimental evidence for emergent E_8 symmetry. *Science*, 327(5962) :177–180, 2010.
- [20] Jaechan Choi and Seung Ki Baek. Finite-size-scaling analysis for the two-dimensional random transverse-field Ising ferromagnet. *Physical Review B*, 108(14) :144204, 2023.
- [21] Roland Assaraf and Michel Caffarel. Zero-variance principle for Monte Carlo algorithms. *Physical Review Letters*, 83(23) :4682–4685, 1999.
- [22] O. Melchert. autoScale.py – a program for automatic finite-size scaling analyses : A user’s guide, 2009.
- [23] J. Houdayer and A. K. Hartmann. Low-temperature behavior of two-dimensional Gaussian Ising spin glasses. *Phys. Rev. B*, 70 :014418, 2004.
- [24] Pierre Pfeuty. The one-dimensional Ising model with a transverse field. *Annals of Physics*, 57 :79–90, 1970.
- [25] M. Hasenbusch. Finite size scaling study of lattice models in the three-dimensional Ising universality class. *Physical Review B*, 82(17) :174433, 2010.
- [26] A. B. Harris. Effect of random defects on the critical behaviour of Ising models. *Journal of Physics C : Solid State Physics*, 7(9) :1671–1692, 1974.
- [27] Olexei Motrunich, Siun-Chuon Mau, David A. Huse, and Daniel S. Fisher. Infinite-randomness quantum Ising critical fixed points. *Phys. Rev. B*, 61 :1160, 2000.
- [28] G. Aeppli D. Bitko, T. F. Rosenbaum. Quantum critical behavior for a model magnet. *Physical Review*, 1996.
- [29] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91 :045002, 2019.
- [30] C. Pich, A. P. Young, H. Rieger, and N. Kawashima. Critical behavior and Griffiths-McCoy singularities in the two-dimensional random quantum Ising ferromagnet. *Phys. Rev. Lett.*, 81 :5916, 1998.