



**Universidad**  
Zaragoza

# **Trabajo de fin de máster**

---

**Entrelazamiento en redes neuronales**

**Entanglement with neural networks**

**Departamento de Materia Condensada. Facultad de Ciencias**

06/2024

Alumno: Ramón Ferrández Gómez  
Tutor: David Zueco Láinez

# Índice

<b>1</b>	<b>Introducción</b>	<b>2</b>
<b>2</b>	<b>Modelos cuánticos de largo alcance</b>	<b>3</b>
2.1	Modelo de Ising cuántico antiferromagnético . . . . .	4
2.2	Entropía de entrelazamiento . . . . .	6
2.3	Transiciones de fase cuánticas . . . . .	8
<b>3</b>	<b>Redes neuronales</b>	<b>10</b>
3.1	Feed Forward networks . . . . .	10
3.2	Metropolis Montecarlo . . . . .	12
3.3	Descenso de gradiente con reconfiguración estocástica . . . . .	13
3.3.1	Backpropagation . . . . .	15
3.3.2	Triangular Schedule . . . . .	15
3.4	Transformer . . . . .	16
3.4.1	Atajos en la red . . . . .	18
3.5	V-score . . . . .	19
<b>4</b>	<b>Resultados</b>	<b>19</b>
4.1	N=20 . . . . .	19
4.2	N=50 . . . . .	23
<b>5</b>	<b>Conclusión</b>	<b>27</b>

# 1 Introducción

Los problemas de muchos cuerpos en mecánica cuántica son un problema abierto hoy en día cuya solución contribuiría enormemente a la ciencia. Uno de los problemas que nos encontramos considerando un sistema de  $N$  sitios es el crecimiento exponencial del espacio de Hilbert en el que vive una función de onda genérica del sistema, imposibilitando la solución del Hamiltoniano incluso numéricamente para sistemas de más de alrededor de 20 sitios. Por suerte, muchos estados cuánticos de interés en un sistema viven una porción reducida del espacio de Hilbert, lo que nos permite intentar encontrar soluciones sin necesidad de contemplar todo el espacio.

Recientemente, se han utilizado técnicas de *machine learning* para este cometido [1] [2]. Estas técnicas tienen principalmente dos funcionalidades, resolver problemas de clasificación y de optimización. Este segundo tipo es precisamente el que requerimos para la minimización de un Hamiltoniano dado, y así poder encontrar su estado fundamental por medio de un espacio de Hilbert reducido. Centrándonos concretamente en las redes neuronales, estas son capaces de almacenar la información de la función de onda en los parámetros que se definen, considerando al mismo tiempo correlaciones entre los componentes del sistema, y con la aparición de las arquitecturas *transformer* se ha aumentado aun más su capacidad resolutoria y de inclusión de estas correlaciones.

El avance y mejora de estas técnicas, nos incita a trabajar con sistemas con interacciones a largas distancias, es decir, debiendo tener en cuenta más componentes en la dinámica de cada uno al aumentar sus correlaciones entre ellos, lo que haremos complementando con trabajos previos del grupo de investigación del tutor del trabajo, y estudiando el efecto de la variación de estas distancias. Y, teniendo en cuenta que las mismas correlaciones mencionadas son las culpables de la aparición de transiciones de fase en los sistemas, esperamos poder encontrar diferentes evoluciones de las fases

Por tanto, el objetivo de este trabajo es estudiar las características de los modelos de largo alcance, para a partir de la elección de un modelo concreto buscar propiedades distintivas en sus correlaciones y en las transiciones de fase que pueda presentar. Entonces, a partir de los trabajos del grupo realizados sobre redes neuronales para problemas como este, ser capaz de comprenderla y optimizarla, comprobando si es capaz de resolver el problema elegido, y en cualquier caso examinar los resultados.

## 2 Modelos cuánticos de largo alcance

De forma general, los modelos cuánticos de largo alcance [3][4] en una red  $d$ -dimensional y de  $N$  sitios interactuantes se describen por

$$H_c = - \sum_{ij}^N J_{ij} \mathcal{C}_i \mathcal{C}_j \quad (1)$$

Siendo  $\mathcal{C}_i$  un operador hermítico local actuando sobre el sitio  $i$ , y las interacciones siguiendo una ley de potencias descrita por

$$J_{ij} = J \frac{\tilde{J}_{ij}}{\tilde{N}} \quad (2)$$

Definiendo  $J$  la intensidad de las interacciones de intercambio, que establece la escala de energía microscópica, y con

$$\tilde{J}_{ij} = \begin{cases} b & \text{si } i = j \\ |i - j|^{-\alpha} & \text{si no} \end{cases} \quad (3)$$

En la ecuación,  $b$  es un parámetro que desplaza el espectro de  $J_{ij}$  y  $\alpha$  establece el rango de interacciones. En este trabajo la red será unidimensional  $d = 1$ , y usaremos condiciones de contorno periódicas (PBC), haciendo el modelo traslacionalmente invariante, por tanto siempre se tomará la distancia de la anterior ecuación como la del camino más corto  $\min(|i - j|, N - |i - j|)$ .

Para los casos en que el rango de interacciones es menor que la dimensión del sistema  $\alpha < d$ , las interacciones decaen lentamente, y como los componentes del Hamiltoniano (1) contienen  $N^2$  términos la suma de estos crecerá más rápido que linealmente, lo que tendrá como consecuencia la pérdida de extensividad del sistema. La introducción de  $\tilde{N}$ , conocido como el factor de renormalización de Kac, recupera la extensividad, y con ello el comportamiento bien definido en el límite termodinámico, siendo este factor para  $\alpha < d$

$$\tilde{N} = \sum_j \tilde{J}_{ij} = b + \sum_{j=1}^N \frac{1}{j^\alpha} \quad (4)$$

A su vez, en este intervalo de  $\alpha < d$ , se rompe la aditividad, lo que significa que la energía de interacción entre los subsistemas o particiones que pudiéramos hacer del sistema no es despreciable en el límite termodinámico, es decir, el total no es igual a la suma de sus partes. Este intervalo descrito es a lo que se denomina régimen de muy largo alcance.

Para  $\alpha > d$ , el factor  $\tilde{N}$  alcanza un valor constante en el límite termodinámico, siendo simplemente un reescalado. En este régimen se pueden destacar otros subregímenes que manifiestan diferentes comportamientos, pues existe un  $\alpha_*$  tal que cuando  $\alpha < \alpha_*$  las interacciones de largo alcance siguen siendo relevantes, aunque se mantiene la aditividad, lo que se denomina régimen de largo alcance débil. En cambio, para  $\alpha > \alpha_*$  entramos en el régimen de corto alcance, y los exponentes críticos coinciden con los de primeros vecinos ( $\alpha \rightarrow \infty$ ).

Al mismo tiempo, existe otro valor  $\alpha_{MF}$  que define una segunda clasificación para los regímenes. En el caso en que  $\alpha < \alpha_{MF}$ , con  $d < \alpha_{MF} < \alpha_*$ , tenemos una región correspondiente a las soluciones predichas por las teorías de campo medio (MF). Los diferentes cambios en la descripción del modelo en la región de largo alcance débil provoca que la descripción de este régimen sea difícilmente abordable mediante métodos analíticos, y por tanto los métodos numéricos se hacen indispensables en su tratamiento.

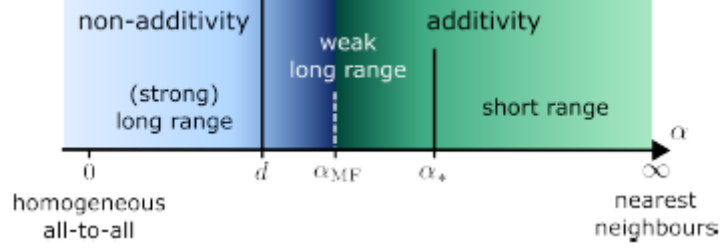


Figure 1: Regiones de comportamiento de modelos cuánticos de largo alcance. Imagen tomada de [5].

## 2.1 Modelo de Ising cuántico antiferromagnético

Para el estudio y distinción de los regímenes explicados, en este trabajo vamos a considerar el modelo de Ising cuántico como nuestro modelo de largo alcance a estudiar. Este modelo se formula con el siguiente Hamiltoniano

$$H = \sum_{ij}^N J_{ij} \sigma_i^z \sigma_j^z + h_x \sum_i \sigma_i^x + h_z \sum_i \sigma_i^z \quad (5)$$

Los grados de libertad cuánticos están representados por las matrices de Pauli  $\sigma^{x,z}$  de cada sitio que, en la base en la que  $\sigma^z$  es diagonal, son

$$\sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (6)$$

Los parámetros  $h_z$  y  $h_x$  representan la intensidad de los campos magnéticos longitudinal y transversal. En ausencia de estos campos, la intensidad de intercambio  $J$ , describiría una fase ordenada ferromagnética para  $J > 0$ , alineando los espines de la cadena de Ising, que por la presencia de simetría de paridad de esta interacción, este alineamiento se produciría equivalentemente en sentido *up* o *down* siendo indistinguible físicamente, es decir, si le das la vuelta al imán este no cambia (degeneración). De igual manera se describiría una fase ordenada antiferromagnética para  $J < 0$ , en la que los espines deben estar alineados en sentido opuesto a los que interacción, sin embargo, como en nuestro problema estamos tratando interacciones de largo alcance, no se va a poder satisfacer el ordenamiento antiferromagnético expuesto para todos los espines, lo que se conoce como frustración. A fin de evitar este fenómeno en nuestro estudio, y

tratar con una cadena antiferromagnética independientemente del rango de interacción, modificamos la matriz de interacción de la siguiente manera

$$J_{ij} = (-1)^{i+j} \frac{\tilde{J}_{ij}}{\tilde{N}} \quad (7)$$

Lo que orientaría los espines en sentido contrario para distancias impares y en el mismo sentido para pares, o lo que es lo mismo, podríamos tratar dos subredes ferromagnéticas de forma independiente, presentándose de nuevo la simetría y la indistinguibilidad descrita.

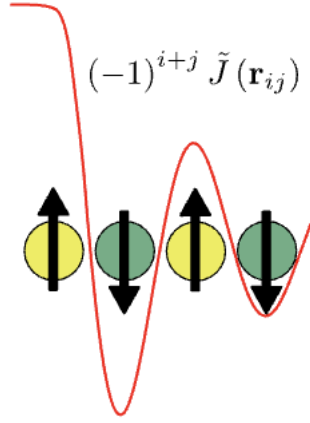


Figure 2: Ilustración del campo alternado en una cadena de espines. Imagen tomada de [6].

La introducción del campo longitudinal  $h_z$ , rompería la simetría presente en el caso ferromagnético, al decantarse por la orientación de espín paralela al campo. En cambio, el caso antiferromagnético va a presentar cierta resistencia al campo longitudinal hasta que se alcance un campo crítico, que producirá una transición de fase de primer orden para  $T = 0$  debido a que a esta temperatura la transición viene acompañada de un salto repentino del parámetro de orden, y una transición de fase de segundo orden para  $T > 0$  por las fluctuaciones térmicas que harán que el parámetro de orden cambie suavemente en el campo crítico [7].

El parámetro de orden de la interacción antiferromagnética descrita en (7) es la imanación alternada (*staggered*)

$$m_s = \frac{1}{N} \sum_{i=1}^N (-1)^i \langle \sigma_i^z \rangle \quad (8)$$

Y para otras descripciones, nos será de utilidad la imanación, que es el parámetro de orden ferromagnético

$$m = \frac{1}{N} \sum_{i=1}^N \langle \sigma_i^z \rangle \quad (9)$$

Por otro lado, como  $\sigma^x$  tiene elementos fuera de la diagonal, el campo transversal inducirá eventos túnel mecano-cuánticos invirtiendo la orientación del espín, lo que estudiaremos en más detalle en la siguiente subsección.

Este sistema posee solución analítica para el regime de largo alcance anteriormente descrito. Además, el sistema sufre una transición cuántica de segundo orden desde la fase ordenada a la fase paramagnética (desordenada) en un punto crítico  $J_c(\alpha)$ . Y, al conocer que  $\alpha_{MF} = 5/3$  y  $\alpha_* > 3$ , podemos estudiar numéricamente la calidad de las soluciones con nuestro ansatz variacional.

Las soluciones de este sistema las trataremos en la base computacional, que describe la orientación del espín en dirección  $z$ , siendo  $|0\rangle$  cuando el espín sea *down* y  $|1\rangle$  cuando sea *up*, esta base de dos niveles nos permite etiquetar a los estados con el número  $n$  en base decimal que forma la secuencia de unos y ceros en base binaria describiendo el estado, por ejemplo el estado  $|000\dots 0\rangle$  sería  $n = 0$ , y  $|111\dots 1\rangle$  sería  $n = 2^N - 1$ . Con esto, definimos un estado cuántico, autovalor del Hamiltoniano (5) como una superposición de estados  $|n\rangle$

$$|\psi\rangle = \sum_{n=0}^{2^N-1} \psi(n)|n\rangle \quad (10)$$

Con la función de onda:

$$\psi(n) = \exp(a + bi) \quad (11)$$

En nuestro caso, los números  $a, b \in \mathbb{R}$ , serán los resultados a encontrar a través de la red neuronal.

Para calcular el valor esperado del Hamiltoniano  $\langle H \rangle$ , es decir, la energía de nuestro sistema, necesitamos computar

$$\langle H \rangle = \langle \Psi | \hat{H} | \Psi \rangle = \frac{\sum_{n,m} \psi^*(m) \langle m | \hat{H} | n \rangle \psi(n)}{\sum_n |\psi(n)|^2} \quad (12)$$

Obtener la energía del estado fundamental (12) variacionalmente con nuestra red neuronal, nos permitirá optimizarla y llegar al estado cuántico correcto, pudiendo determinar algunas variables macroscópicas de importancia.

## 2.2 Entropía de entrelazamiento

El concepto de entrelazamiento nace de la necesidad de distinguir las correlaciones cuánticas que pueden surgir en los estados cuánticos de más de un cuerpo de las correlaciones clásicas [8]. Este concepto es de interés debido a que las transiciones de fase ocurren como consecuencia de las correlaciones intercorpúsculares [7]. El entrelazamiento se visualiza en un estado cuántico sabiendo que dos estados entrelazados no son separables, es decir, un estado separable no contiene entrelazamiento. Este último, en caso de que sea un estado puro de un estado bipartito se puede expresar como

$$|\psi\rangle_{AB} = |\omega\rangle_A \otimes |\gamma\rangle_B \quad (13)$$

O con matrices densidad

$$\rho_{AB} = \rho_A \otimes \rho_B \quad (14)$$

Donde  $\rho$  se define para un estado puro como el que nos ocupa:

$$\rho = |\psi\rangle\langle\psi| \quad (15)$$

Y las matrices de las particiones son

$$\rho_A = \text{Tr}_B(\rho_{AB}), \quad \rho_B = \text{Tr}_A(\rho_{AB}) \quad (16)$$

Donde  $\text{Tr}_{A/B}$  son las trazas solo de las particiones que indica.

En nuestro problema  $|\psi\rangle_{AB}$  sería el estado fundamental de la forma (10), y los particiones  $A$  y  $B$  serían distintos trozos de la cadena de Ising, no intersectantes y cuya unión recupera la cadena. Para ilustrar esto, poniendo como ejemplo  $|101010\rangle$ , podríamos tomar como partición  $A$   $|1\rangle, |10\rangle, |101\rangle\dots$  y  $B$  sería el restante de la cadena. Vemos que ese estado sí sería separable, pero no siempre será tan sencillo saberlo para un estado genérico (10). Y aquí es donde entra la entropía de von Neumann.

La entropía de von Neumann se define como

$$S(\rho) = -\text{Tr}(\rho \log_2(\rho)) \quad (17)$$

Para un estado puro, la matriz densidad solo va a tener un autovalor diferente de cero, lo que implica que  $S(\rho) = 0$ . Por tanto, para un estado bipartito, se puede comprobar

$$S(\rho_A) = S(\rho_B) \begin{cases} = 0 & \text{si } \rho_{AB} = \rho_A \otimes \rho_B \\ > 0 & \text{si no} \end{cases} \quad (18)$$

Y de la teoría de la información de Shannon, esto se interpreta como que en el caso de tener un estado puro, tenemos un completo conocimiento del estado que vamos a obtener al medir. No obstante, nuestro conocimiento de los estados (mixtos<sup>1</sup>) de las partes  $A$  y  $B$  es incompleto. Hay información en el estado total codificada en las correlaciones entre las partes del sistema, que no puede detectarse cuando miramos las partes por separado. Es decir, al medir una de las partes cuando tenemos el sistema completo, nos va a proporcionar información de la otra parte, aun no habiéndola medido y desconociendo su estado. El hecho de que la entropía nos va a decir cuánta información, por definición, hace que a esta entropía cuando se tratan sistemas bipartitos se le llame entropía de entrelazamiento. Los valores de la entropía por espín  $S/N$  van desde 0 a 1, siendo 0 estado separable, saber una parte no nos dice nada sobre la otra, y 1 máximamente entrelazado, saber una parte nos desvela la otra.

---

<sup>1</sup>Los estados mixtos son una mezcla estadística (clásica) de estados puros. Nuestro sistema viene descrito por un conjunto de estados puros  $\{|\psi_1\rangle, |\psi_2\rangle, \dots\}$ , con unas probabilidades asociadas  $\{p_1, p_2, \dots\}$ , estando la información del sistema incompleta.

## 2.3 Transiciones de fase cuánticas

Una transición de fase cuántica se produce por la presencia de un punto no analítico (un pico) en la evolución de la energía del estado fundamental. Este pico se puede dar principalmente mediante dos fenómenos, teniendo por un lado un cruce de niveles energéticos, haciéndose el estado excitado el nuevo estado fundamental, y por otro lado un evitamiento de un cruce de niveles fundamental-excitado, que es curvo para una cadena finita, pero se hace más picudo conforme nos acercamos al límite termodinámico. Estos sucesos describen una transición cuántica de primer y segundo orden respectivamente [9]. En nuestro Hamiltoniano (5) estos puntos no-analíticos vendrán determinados por los valores que toman los campos magnéticos.

Hemos descrito estos fenómenos con singularidades en el estado fundamental, lo que significa que estrictamente solo se van a dar a  $T = 0$ , y por ello la tarea de la teoría de transiciones de fase cuánticas es describir consecuencias de esta singularidad para  $T > 0$ .

Y es que en los modelos clásicos las transiciones de fase son provocadas únicamente por fluctuaciones térmicas, que desaparecen a  $T = 0$ . Sin embargo, los estados cuánticos también tienen fluctuaciones, debido al principio de incertidumbre, incluso para  $T = 0$ , lo que va a provocar las transiciones de fase en estos casos, y la ruptura espontánea de simetría de paridad previamente mencionada.

Cabe destacar, que las transiciones de fase normalmente vienen acompañadas por cambios cualitativos en la naturaleza de las correlaciones del estado fundamental, lo que nos lleva a pensar que esto deberá verse reflejado cuando calculemos la entropía de las fases.

En el TFG [6] se calculó semianalíticamente el diagrama de fases del modelo presentado en el régimen de largo alcance. Observándose las siguientes gráficas para las imanaciones

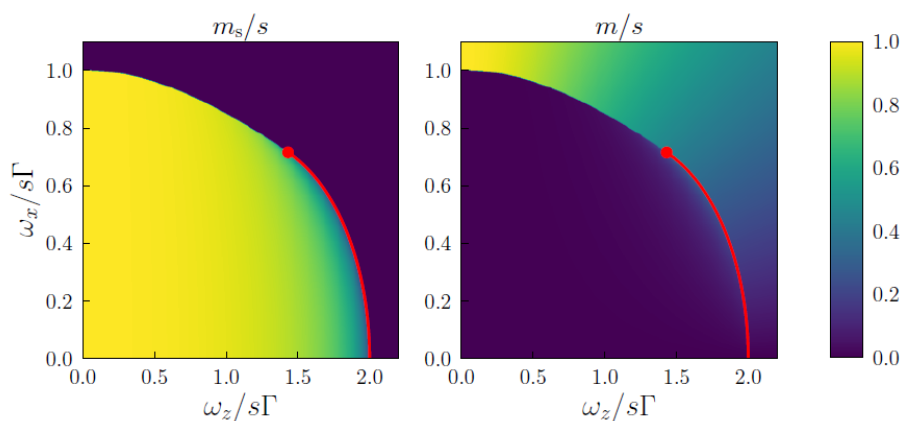


Figure 3: Diagramas de fases de la cadena de Ising antiferromagnética alternada para  $T = 0$  con interacciones de largo alcance. Los parámetros  $\omega$  simbolizan el campo magnético, estando intercambiados los ejes  $x, z$  en este trabajo,  $\Gamma$  siendo  $J$  y  $s = 2$  el espín. El primer diagrama muestra la imanación alternada y el segundo la imanación. Imagen tomada de [6].

Podemos visualizar en ambas figuras la presencia de una línea roja terminada en un punto rojo. Esto representa la transición de fase de segundo orden y el punto crítico calculadas por teoría de Landau. A su vez, vemos como los campos en los ejes están representados respecto al valor de  $J$ , y es que no va a importar el valor absoluto de estos sino el relativo a la intensidad de interacciones, lo que nos servirá para fijarlo en las simulaciones posteriores.

En la primera figura se muestra la imanación alternada, que será equivalente para todos los cuadrantes, la parte iluminada se corresponde con la fase antiferromagnética, que vemos que en las regiones más cercanas a la ausencia de campos se encuentra más ordenada, mientras que a campos longitudinales bajos a medida que se aumentaba el campo transversal se iban produciendo más inversiones de espín en la cadena hasta que se llegaba a la línea y se accedía a la zona oscura paramagnética. Esta progresión continua en el parámetro de orden hasta la línea de transición es característica para las transiciones de segundo orden. Esto muestra un buen acuerdo entre la teoría clásica de Landau y lo esperado con la matriz  $\sigma^x$ .

En la segunda imagen vemos la imanación y cómo, para campos transversales bajos, a medida que incrementamos el campo longitudinal llegamos a una discontinuidad de la imanación en la que toma un valor ordenado paramagnético, debiéndose a que se ha alcanzado un campo longitudinal tan intenso que vence a las interacciones antiferromagnéticas alineando los espines en el sentido del campo, viéndose un alineamiento de los espines en el mismo sentido, que se disipa poco a poco conforme introducimos campo transversal. Este salto que se observa en las dos imanaciones es característico de las transiciones de primer orden, pero no se calculó la línea con teoría de Landau por falta de convergencia analítica.

A su vez, en numerosas referencias se muestra un comportamiento similar para interacciones a primeros vecinos

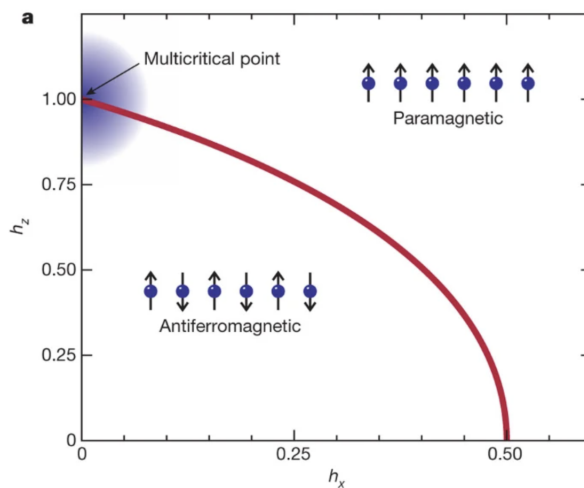


Figure 4: Digrama de fases de la cadena de Ising antiferromagnética a primeros vecinos para  $T = 0$  Imagen tomada de [10].

En esta imagen encontramos de nuevo la línea roja de segundo orden salvo cuando no hay campo transversal, donde se hace una transición clásica (Hamiltoniano diagonal) y de primer orden. Por la construcción de nuestro modelo, el comienzo y final de la línea es en  $(h_x = 2/3, h_z = 0)$  y  $(h_x = 0, h_z = 4/3)$ .

La evidente similitud entre las transiciones de fase para interacciones de largo y corto alcance, nos hace sospechar que para interacciones en el régimen de largo alcance débil, aumentando el alcance de la interacción (aumentando  $\alpha$ ) la región de primer orden se debe ir reduciendo desde la región mostrada en 3 hasta llegar al colpaso en un solo punto presenciado en el régimen de corto alcance en 4.

### 3 Redes neuronales

Para comprobar la evolución predicha de la transición de fase de primer orden y la forma que toman algunas magnitudes de importancia como la entropía de entrelazamiento o los parámetros de orden, buscaremos encontrar la solución del Hamiltoniano (5) para diferentes campos mediante simulaciones numéricas elaboradas usando redes neuronales y *transformer*. En la siguiente sección se muestra el diseño de la red neuronal construido mediante el uso de la librería "Netket" [11] en el lenguaje de programación *Python*.

Las redes neuronales por capas recibieron el nombre de "perceptrones" (*perceptrons*) cuando fueron estudiadas por primera vez por Rosenblatt. Consisten en un primer conjunto de terminales de entrada cuya función es proveer los patrones de entrada al resto de la red. Después encontramos una o más capas intermedias (la capa de entrada no se suele contar), desembocando en la capa de salida, que proporciona el resultado de la red. Cada capa está formada por un número a elegir de unidades que realizan las operaciones y ajustes que llevan a la entrada a dar la respuesta de salida correcta.

Las redes neuronales se pueden utilizar para modelar la energía de un sistema físico, ajustando los parámetros de la red para que la energía predicha sea lo menor posible con una configuración de red óptima. El tipo de aprendizaje de red que se usa para esta tarea es el aprendizaje no supervisado. Este consiste en aprender patrones o estructuras en los datos de entrada sin utilizar valores objetivo específicos, con el fin en este caso de minimizar la energía. Mediante este aprendizaje, la red desarrolla la habilidad de crear representaciones internas para codificar características de los valores de entrada [12].

Así, en esta sección nos vamos a dedicar a describir cómo creamos una red neuronal capaz de llevar a cabo esta tarea

#### 3.1 Feed Forward networks

En el caso de las *feed-forward networks* (FFN), la información fluye en una única dirección, de entrada a salida, lo que significa que no se establecen conexiones entre las unidades de una capa dirigidas hacia capas anteriores o entre las unidades de esa misma capa, ni tampoco se darán con capas más allá de la contigua [13]. A su vez, las capas estarán completamente conectadas, todas las unidades de una capa recibirán información de todas

las unidades de la anterior . Por esta razón, a las capas intermedias se les suele llamar capas ocultas, al no tener conexión directa con la entrada o la salida.

Las conexiones entre capas se representan por medio de matrices de conexión  $w_{ij}$ , que está formada por el vector de conexión de cada unidad. Estas matrices se encargan de conectar las unidades entre sí como hemos explicado antes, y se inicializan con números aleatorios menores que 1 en módulo, optimizándose en cada época (iteración) como veremos más adelante<sup>2</sup>. Así, cada unidad de una capa recibe la entrada procedente de la capa anterior en forma de vector de entrada, al cual se aplica el vector de conexión perteneciente a la unidad, o se le aplica la matriz si hablamos de forma global, y se hace pasar por una función de activación  $g(x)$  que determina la salida, pudiendo ser esta función lineal o no lineal. En cualquier caso, estas funciones de activación deben ser diferenciables, son a elección del usuario en la construcción de la red y puede ser diferentes para cada unidad, aunque lo más común es tomarlas todas iguales. Al mismo tiempo, aunque tanto las entradas como las salidas pueden ser booleanas o continuas, la naturaleza de las salidas queda determinada por la elección de la función de activación. Con esta descripción, la salida de cada capa es

$$O_i^{(k)} = g^{(k)}\left(\sum_j w_{ij}^{(k)} I_j^{(k)}\right) = g^{(k)}(v_i^{(k)}) \quad (19)$$

Donde los subíndices  $i, j$  representan cada unidad, de salida y de entrada respectivamente, y  $(k)$  el número de capa, con  $I^{(k+1)} = O^{(k)}$ . De manera que iterativamente tenemos para la capa  $k$ -ésima

$$O_i^{(k)} = g^{(k)}\left(\sum_j w_{ij}^{(k)} g^{(k-1)}\left(\sum_l w_{jl}^{(k-1)} I_l^{(k-1)}\right)\right) \quad (20)$$

El objetivo entonces, es crear una red neuronal que ajuste sus matrices de conexión de acuerdo a la minimización de la función coste, el Hamiltoniano en este caso, de manera que el patrón de entrada que introducimos en la red es una cadena binaria  $\xi_n$  de 1 y  $-1$  que describen el estado cuántico de espines  $n$  reemplazando los 1 y 0 de la cadena en base computacional, y esta en nuestro caso nos debe proveer como salida los números  $a, b$  de la función de onda (11). En nuestro caso, buscaremos estos dos números por medio de dos redes neuronales de una única salida, igualmente construidas pero con parámetros optimizados diferentes, usando cada una para la obtención de cada número. Siendo entonces la función de onda para una cadena:

$$\psi(\xi_n; w) = \exp(O_{\text{Re}}^{(-1)}(w_{\text{Re}}) + iO_{\text{Im}}^{(-1)}(w_{\text{Im}})) \quad (21)$$

Donde el exponente  $(-1)$  significa que es la última capa, y los términos "Re" e "Im" se refieren a la red para el número  $a$  y  $b$  respectivamente. De modo que cuando minimicemos el valor esperado (12), las funciones de las ecuaciones (11) y (21) sean iguales.

Para calcular el valor esperado del Hamiltoniano de manera exacta, necesitaríamos las funciones de onda de todos los estados del sistema, algo que debido al tamaño de este,

---

<sup>2</sup>El valor bias  $\theta_i$  está contenido en esta matriz, ya que podemos definir una entrada auxiliar  $I_0 = -1$ , quedando  $w_{i0}I_0 = -\theta_i$ .

es inviable de calcular. Es por ello que en la subsección siguiente describimos un método numérico que nos permite su obtención empleando un número reducido de los estados.

## 3.2 Metropolis Montecarlo

El algoritmo de Metropolis es un método Montecarlo basado en cadenas de Markov, que es un método de simulación de una distribución de probabilidad desconocida que produce una cadena de Markov ergódica cuya distribución estacionaria es la propia distribución desconocida [12]. Se puede demostrar que el siguiente algoritmo de Metropolis es una cadena de Markov ergódica, y además que una cadena de Markov va a alcanzar una distribución estacionaria tras sucesivos pasos estocásticos (transiciones determinadas por probabilidades). En este sentido, podemos visualizar el proceso que se procede a describir como si nos encontráramos en una solución fuera del equilibrio, e hiciéramos avanzar el tiempo hasta que el sistema alcanza el estado estacionario, lo cual, al ser ergódico, es equivalente al promedio de un conjunto de estados, que es lo que pretendemos calcular.

El algoritmo consiste en tomar  $N$  estados  $|n\rangle$ , aleatoriamente del conjunto  $2^L$  que consideramos representativos del total. Nuestro algoritmo Metropolis va a tener una regla de transición local, lo que significa que para cada estado escogido, se elige aleatoriamente un sitio  $i$  de la cadena del estado y se propone un cambio del valor de este sitio escogiendo de forma aleatoria entre sus posibles valores, en este caso 1 o  $-1$ , definiendo el estado  $|m\rangle$ . Teniendo en cuenta que las probabilidades del sistema de estar en el estado  $|n\rangle$  y  $|m\rangle$  son  $p_n = |\psi(n)|^2$  y  $p_m = |\psi(m)|^2$ , definiendo un número aleatorio un número aleatorio  $r \in [0, 1]$ , describimos los siguientes pasos del algoritmo

$$\frac{p_m}{p_n} \begin{cases} > 1 & |n\rangle \rightarrow |m\rangle \\ < 1 & \begin{cases} > r & |n\rangle \rightarrow |m\rangle \\ < r & |n\rangle \rightarrow |n\rangle \end{cases} \end{cases}$$

Resumidamente, establecemos que la probabilidad de transición  $|n\rangle \rightarrow |m\rangle$  es  $p_{n \rightarrow m} = \min(1, \frac{p_m}{p_n})$ . De esta manera, al poder transicionar de estado aun no siendo el de mayor probabilidad, podremos esquivar mínimos relativos del promedio, y junto a la suposición de que la mayoría de los  $2^N$  estados del sistema va contribuir despreciablemente al promedio al tener una probabilidad muy baja o nula<sup>3</sup>, el algoritmo convergerá de forma muy aproximada al promedio real del sistema sin necesidad de calcular todos los estados. Así que, realizando el proceso anterior un gran número de veces, utilizando los estados resultantes, y calculando un promedio temporal, obtenemos el valor esperado del Hamiltoniano. En nuestro caso, por defecto "Netket" barría cada cadena tantas veces como componentes tuviera ( $N$ ). Y en cada Montecarlo se usaron 1000 estados base.

---

<sup>3</sup>Como hemos explicado en la introducción, los autoestados suelen residir en una pequeña región del espacio de Hilbert.

### 3.3 Descenso de gradiente con reconfiguración estocástica

Una vez obtenido el valor esperado del Hamiltoniano, debemos ajustar los parámetros de la red para conducirlos a un valor esperado menor en la siguiente época.

Para ello utilizamos el método del descenso del gradiente. Este está basado en que la función multivariable que queremos minimizar  $\langle H \rangle(w)$ , suponiendo que esta función está definida y es diferenciable alrededor de los valores de los parámetros para los que se encuentra, decrece más rápido en sentido contrario al que marca el gradiente en ese punto.

$$w_{ij} \rightarrow w_{ij} - \eta \frac{\partial \langle H \rangle}{\partial w_{ij}} \quad (22)$$

Y la derivada parcial es, tomando (12)

$$\frac{\partial \langle H \rangle}{\partial w_{ij}^{(k)}} = 2 \left( \frac{\sum_{n,m} D_k^\dagger(m) \psi^*(m) \langle m | \hat{H} | n \rangle \psi(n)}{\sum_n |\psi(n)|^2} - \langle H \rangle \frac{\sum_m D_k^\dagger(m) |\psi(m)|^2}{\sum_n |\psi(n)|^2} \right) \quad (23)$$

Lo que es equivalente a

$$\frac{\partial \langle H \rangle}{\partial w_{ij}} = 2(\langle D_k^\dagger H \rangle - \langle H \rangle \langle D_k^\dagger \rangle) \quad (24)$$

Este promedio se calcula también por medio del método Montecarlo descrito en la sección anterior.

El problema de optimizar la red por medio de la ecuación (22) es que su convergencia puede ser muy lenta, y la presencia de distintos mínimos relativos en el espacio del Hamiltoniano puede evitar la convergencia al mínimo absoluto. Al mismo tiempo, la fuerte dependencia entre numerosos parámetros variacionales puede conllevar que la variación de algunos de ellos sea compensada casi por completo por la variación de otros parámetros, lo que también conduce a inestabilidades y lenta convergencia [14].

Para solventar el problema mencionado, recurrimos a método de reconfiguración estocástica (SR) (*Stochastic reconfiguration*). La idea es modificar los parámetros de la función test de forma que esta tienda al estado fundamental por medio de un camino dictado por la proyección  $\mathbb{I} - \tau H$ , con  $\tau > 0$  y pequeño, y  $\mathbb{I} - \tau H \geq 0$ . De esta forma, utilizamos más información de la función test que en el descenso de gradiente simple, lo que lleva a una optimización más rápida. Si descomponemos nuestra función test  $|\psi\rangle$  en una suma de su parte paralela al Hamiltoniano  $|\varphi\rangle$  y la parte ortogonal  $|\phi\rangle$  [15].

$$\mathbb{I} - \tau H |\psi\rangle = (\mathbb{I} - \tau H) |\varphi\rangle + |\phi\rangle \quad (25)$$

Nuestra intención es cambiar los parámetros  $w$  que describen  $|\psi\rangle$  a fin de hacer este estado lo más cercano posible a  $|\varphi\rangle$ . Como  $\tau$  es pequeño, tenemos que

$$|\varphi\rangle = |\psi + \delta\psi\rangle = |\psi\rangle + \tau \delta |\psi\rangle \quad (26)$$

Como buscamos la variación con respecto de los parámetros de red

$$\delta |\psi\rangle = \sum_k \delta w^{(k)} \sum_n \frac{\delta \psi(n)}{\delta w^{(k)}} |n\rangle = \sum_k \delta w^{(k)} \sum_n \frac{\partial \psi(n)}{\partial w^{(k)}} |n\rangle \quad (27)$$

Por la ecuación (21), nos queda

$$\delta|\psi\rangle = \sum_k \delta w^{(k)} \sum_n \frac{\partial \ln(\psi(n))}{\partial w^{(k)}} \psi(n)|n\rangle \quad (28)$$

A partir de aquí definimos el operador  $D_k = \frac{\partial}{\partial w^{(k)}}$ , y  $O_0 = \mathbb{I}$ . Y con esto la ecuación (25) queda:

$$\mathbb{I} - \tau H|\psi\rangle = \alpha_0|\psi\rangle + \tau \sum_{k=1} \alpha_k D_k|\psi\rangle + |\phi\rangle \quad (29)$$

Donde los coeficientes  $\alpha_k$  son las variaciones de los parámetros junto a los factores de normalización. Multiplicando ahora esta última ecuación por la izquierda por  $\langle\psi|$  y  $\langle\psi|D_k^\dagger$ , nos queda

$$\begin{aligned} 1 - \tau\langle H\rangle &= \alpha_0 + \tau \sum_k \alpha_k \langle D_k\rangle \\ \langle D_k^\dagger\rangle - \tau\langle D_k^\dagger H\rangle &= \alpha_0 \langle D_k^\dagger\rangle + \tau \sum_l \alpha_l \langle D_k^\dagger D_l\rangle \end{aligned} \quad (30)$$

Y podemos resolver para  $\alpha_0$  para llegar a

$$\sum_l S_{k,l} \alpha_l = -\frac{1}{2} \frac{\partial \langle H\rangle}{\partial w^{(k)}} \quad (31)$$

Que nos da la variación de los parámetros de la red en cada época a espensas de una constante, que podemos elegir como nuestro factor de aprendizaje  $\eta$ , y con

$$S_{kl} = \langle D_k^\dagger D_l\rangle - \langle D_k^\dagger\rangle \langle D_l\rangle \quad (32)$$

Así, nos queda

$$w^{(k)}(t+1) = w^{(k)}(t) - \eta S^{-1} \frac{\partial \langle H\rangle}{\partial w^{(k)}}$$

Sin embargo, la matriz  $S_{kl}$  se hace singular cuando  $\sigma = \frac{\lambda_N}{\lambda_1}$ , que es el cociente de su mayor y menor autovalor, se hace muy grande, produciendo inestabilidades numéricas [16]. Que  $\sigma$  sea grande se debe a la redundancia los parámetros de la red que no generan cambios en la función de onda dentro de una tolerancia dada  $\epsilon$ , que representa la distancia ente la función de onda normalizada exacta y la función aproximada para la que dejamos de iterar y optimizar la red. Por ello, se introduce un reescalado  $S \rightarrow S + \epsilon$  obteniendo una matriz bien condicionada con todos los autovalores mayores que  $\epsilon$ . Quedándonos así

$$w^{(k)}(t+1) = w^{(k)}(t) - \eta (S + \epsilon \mathbb{I})^{-1} \frac{\partial \langle H\rangle}{\partial w^{(k)}}(t) \quad (33)$$

Donde  $t$  simboliza la época de optimización. A continuación describiremos cómo aplicar esta actualización a cada capa de la red

### 3.3.1 Backpropagation

Una vez explicado el método de optimización, afrontamos el problema de cómo obtener la derivada de la función de onda respecto a los parámetros que aparece en (23). Para ello, recordando la ecuación (19), la derivada de una salida con respecto a las matrices de conexión de esa misma capa es

$$\frac{\partial O_l^{(k)}}{\partial w_{ij}^{(k)}} = \frac{\partial g^{(k)}(v_l^{(k)})}{\partial v_\alpha^{(k)}} \frac{\partial v_\alpha^{(k)}}{\partial w_{ij}^{(k)}} = \partial_i g^{(k)}(v_l^{(k)}) I_j^{(k)} \quad (34)$$

Donde  $\partial_i = \frac{\partial}{\partial v_i^{(k)}}$ . Ahora bien, teniendo en cuenta la forma de la función de onda (21), tenemos que

$$D_k \psi(n) = \psi(n) \frac{\partial \ln \psi(n)}{\partial w^{(k)}} = \psi(n) \left[ \frac{\partial (O_{\text{Re}}^{(-1)}(w_{\text{Re}}) + i O_{\text{Im}}^{(-1)}(w_{\text{Im}}))}{\partial w^{(k)}} \right] \quad (35)$$

Apreciando que cada vez que hacemos una optimización necesitamos el resultado de la otra red neuronal. Centrándonos en solamente una de las dos redes, llamando a la salida simplemente  $O^{(-1)}(w)$ . Vemos que la diferenciación será a la salida de la última capa con respecto a las conexiones de la capa  $k$ , y vamos a tener que recurrir a la regla de la cadena, yendo hacia atrás en la red (*backpropagation*). Usando que  $I^{(k+1)} = O^{(k)}$ , la derivada de una salidad con respecto a la salida anterior es

$$\frac{\partial O_l^{(k)}}{\partial O_p^{(k-1)}} = \frac{\partial O_l^{(k)}}{\partial I_p^{(k)}} = \frac{\partial g^{(k)}(v_l^{(k)})}{\partial v_\alpha^{(k)}} \frac{\partial v_\alpha^{(k)}}{\partial g^{(k-1)}(v_p^{(k-1)})} = \partial_\alpha g^{(k)}(v_l^{(k)}) w_{\alpha p}^{(k)} \quad (36)$$

Este resultado junto con el de la ecuación (34), nos permite encontrar la derivada de la salida de la red con a la matriz de conexión de cualquier capa de la red

$$\frac{\partial O^{(-1)}}{\partial w_{ij}^{(k)}} = \frac{\partial O_l^{(-1)}}{\partial O_p^{(-2)}} \frac{\partial O_p^{(-2)}}{\partial O_q^{(-3)}} \cdots \frac{\partial O_\gamma^{(k)}}{\partial w_{ij}^{(k)}} = \left[ \prod_{z>k} \partial g^{(z)} w^{(z)} \right] \partial g^{(k)} I^{(k)} \quad (37)$$

Cabe recordar, que el superíndice negativo explícito indica el número de capa contando desde el final, mientras que el positivo desde el principio. Así, vemos que para obtener esta derivada, vamos a necesitar las funciones y conexiones de todas las capas posteriores.

### 3.3.2 Triangular Schedule

El factor de aprendizaje  $\eta$  controla cuánto se ajustan las conexiones del modelo con respecto al error estimado cada vez que se actualizan. En un *triangular schedule*, Comenzar con una tasa de aprendizaje baja evita que el modelo haga ajustes demasiado bruscos al principio, lo que puede conducir a oscilaciones significativas y pérdida de convergencia, por ello, aumentar progresivamente la tasa de aprendizaje nos ahorra este problema, y permite atravesar rápidamente los puntos de silla o zonas de bajo gradiente que puedan decelerar el proceso de aprendizaje [17]. Tras un número de épocas, cuando ya suponemos que la optimización ha atravesado estas zonas problemáticas, el factor de aprendizaje

comienza a disminuir linealmente, permitiendo una exploración más controlada del espacio de parámetros y realizando ajustes más finos y precisos a medida que se acerca al mínimo, y previniendo la aparición de oscilaciones alrededor de este mínimo. Así, definiremos

$$\eta = \begin{cases} \eta_{\min} + \frac{t}{T}(\eta_{\max} - \eta_{\min}) & t \leq T \\ \eta_{\max} + \frac{t-T}{T}(\eta_{\min} - \eta_{\max}) & t \geq T \end{cases} \quad (38)$$

Por las mismas razones, hacemos lo mismo con la tolerancia (desplazamiento diagonal)  $\epsilon$ . Y es que en este caso, un valor bajo inicial nos asegura que no se produzcan ajustes demasiado bruscos, aumentar el valor va a asegurar que la matriz  $S$  esté bien condicionada, evitando problemas con la inversa de la matriz y mejorando la estabilidad numérica, y disminuyéndolo otra vez finalmente, conseguimos de nuevo que el modelo se adapte mejor a los datos al haber un menor impacto en la diagonal, ajustándose más al valor óptimo. Así que, de nuevo

$$\epsilon = \begin{cases} \epsilon_{\min} + \frac{t}{T}(\epsilon_{\max} - \epsilon_{\min}) & t \leq T \\ \epsilon_{\max} + \frac{t-T}{T}(\epsilon_{\min} - \epsilon_{\max}) & t \geq T \end{cases} \quad (39)$$

### 3.4 Transformer

El *transformer* es una arquitectura basada en auto-atención que consiste en bloques sucesivos, cada uno conteniendo auto-atención, capas *feedforward*, y otros componentes como capas de normalización, conexiones residuales, escalados de atención, y multi-head auto-atención [18].

La idea de utilizar el *transformer* nace de que aprender dependencias entre componentes de una cadena o secuencia que se encuentran a gran distancia relativa, es crucial para obtener buenos resultados en sus predicciones. La capacidad de una red para aprender estas dependencias depende de la longitud de los caminos que las señales de la red deben recorrer. Cuanto más cortos sean estos caminos, más fácil será para la red aprender las dependencias. Una capa de autoatención conecta todas las posiciones con un número constante de operaciones secuenciales, lo que significa que la cantidad de pasos necesarios no depende de la longitud de la secuencia, mientras que en otro tipo de redes depende directamente de la longitud de la secuencia [19].

El transformer comienza tomando la cadena de entrada  $\vec{\xi}$  y dividiéndola en particiones llamadas *tokens*  $\vec{\xi}_i$  con todas el mismo tamaño  $\rho$  para nuestro caso, en el espacio  $\mathbb{R}^\rho$ . A cada una de estas particiones, se le aplica un transformación lineal  $E$  llamada *embedding*, en el espacio  $\mathbb{R}^{d_e \times \rho}$ . La dimensión  $d_e$  debe establecerla el usuario según la calidad de los resultados de sus simulaciones. Por ello, a lo largo del cálculo que se desarrolla en la red, trabajaremos con la representación  $\vec{x}_i = E\vec{\xi}_i$ . En el ámbito de traducción de idiomas llama representación no-contextual, ya que es independiente del contexto de las palabras que le rodean, y de la misma manera nuestras particiones están incontextualizadas del

resto que les rodea, sin embargo, mediante transformaciones y computaciones posteriores, esta representación nos va a permitir encontrar clasificaciones y relaciones entre los componentes de la cadena que no hubiésemos podido en la dimensión de partida  $\rho$ , esto se consigue en el bloque de atención siguiente.

Para comenzar el bloque de atención, en primer lugar se normaliza cada representación  $\vec{x}_i$  separadamente mediante una operación LayerNorm[20]:

$$\text{LayerNorm}(\vec{x}_i) = \frac{\vec{x}_i - \text{Mean}(\vec{x})}{\text{Var}(\vec{x})} \quad (40)$$

Ahora describimos la autoatención de clave-consulta-valor (*key-query-value*). A fin de aumentar la capacidad de la autoatención, el bloque del *transformer* aplica un número  $h$  de sets de autoatenciones en paralelo, estos sets reciben el nombre de *heads*  $H^\mu$ , con  $\mu$  indicando el número de *head*. Para ello, debemos distribuir los componentes de las representaciones  $\vec{x}_j$  entre los *heads*, de forma que la dimensión de cada nueva partición es  $p = d_e/h^4$ . Normalmente se definen las matrices *key*  $K^\mu$ , *query*  $Q^\mu$  y *value*  $V^\mu$ , y utilizando las dos primeras se define  $a_{ij}$ , no obstante esta matriz se va a constreñir como explicaremos a continuación, y prescindiremos de la formulación con una función no lineal y  $Q$  y  $K$ . Por otro lado,  $V$  es una matriz  $p \times p$  que actúa sobre  $\vec{x}_j$ .

$$H_i^\mu = \sum_{j=1}^p a_{ij}^\mu V^\mu \vec{x}_j^\mu \quad (41)$$

La matriz  $a_{ij}$  es la matriz de pesos de atención, que representa cuánto de  $\vec{x}_j$  deber contribuir al *head* correspondiente, y cuya forma funcional puede ser elegida de acuerdo con el problema que buscamos resolver [21]. Esta matriz debe considerar la relevancia del ordenamiento de los componentes de nuestra cadena de entrada, y para lograr esto hay dos vías, implementar la dependencia del ordenamiento en  $\vec{x}$  o hacerlo en los pesos. Para incluir esto, como el problema trata con cadenas de espín con condiciones de contorno periódicas, forzamos la simetría traslacional fijando  $a_{ij}$  a ser una matriz circulante. Y por tanto tiene la siguiente forma

$$a_{ij} = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_1 \\ a_1 & a_0 & a_1 & \cdots & a_2 \\ a_2 & a_1 & a_0 & \cdots & a_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1 & \cdots & a_2 & a_1 & a_0 \end{pmatrix} \quad (42)$$

Cuyos vectores propios siempre que tenga la misma dimensión la matriz van a ser los mismos, y sus valores solo depende de la distancia a la que se encuentren sus elementos  $a_{ij} = a_{|i-j|}$ .

---

<sup>4</sup>El hecho de haber reducido las dimensiones, hace que el coste computacional vaya a ser similar al de una única *head* con la dimensión completa.

Ahora, con el resultado de cada *head*  $\mu$ , lo concatenamos obteniendo de nuevo unas representaciones  $\vec{y}_i$  de dimensión  $d_e$ .

$$\vec{y}_i = \text{Concat}(H_i^1, \dots, H_i^h) \quad (43)$$

Aunque hay cruce de características al hacer actuar la matriz de autoatención sobre las particiones, esto tiene una capacidad limitada y por tanto recurrimos a una nueva fase que complementa este cruce. Los vectores  $\vec{y}_i$  se hacen pasar por un perceptrón multicapa (MLP), siguiendo los esquemas descritos en las primeras secciones. De forma que la multicapa está compuesta por tres capas que siguen la siguiente secuencia: Dense, que es una capa lineal, LayerNorm, y capa no lineal Swish. Esta última es una capa como las descritas, con función de activación  $g(x)$ :

$$\text{swish}(x, \beta = 1) = \frac{x}{1 + e^{-x}} \quad (44)$$

Ya que parece tener un mejor rendimiento que las convencionales sigmoides o ReLU [22]. De forma que las capas serían

$$O_i = \text{swish}(\text{LayerNorm}(w_{ij}I_j)) \quad (45)$$

El siguiente paso es realizar el *pooling*  $\Sigma$ , el cual consiste en obtener un único vector  $\vec{x}$ , realizando la media de los vectores  $\vec{x}_i$ , y posteriormente le aplicamos la función  $\log(\cosh(\vec{x}))$ , que por alguna razón parecía mejorar los resultados.

Una vez concluido el *pooling*, pasamos nuestro resultado por una MLP de post-procesamiento como (45) con un número de capas a elegir y en forma de embudo, reduciéndose la dimensión del vector resultante en cada capa.

Finalmente, una última capa lineal actúa sobre nuestro vector resultante, proveyendo un valor único, nuestro resultado.

### 3.4.1 Atajos en la red

A medida que en una red neuronal se incluyen más capas llega un momento que su precisión tiende a saturarse, no mejora por más capas que añadamos.[23].

Una forma de solucionar esto es introducir una función residual, como la identidad, de forma que cuando introducimos el atajo en la capa  $k$  con salida  $O^{(k)}$ , hasta la capa  $k + l$  con salida  $O^{(k+l)}$ , a esta última salida habrá que sumarle la del inicio del atajo, resultando:

$$O^{(k+l)} \rightarrow O^{(k+l)} + O^{(k)} \quad (46)$$

Y vemos que si la salida original de  $O^{(k+l)}$  fuera nula, seguiríamos conservando la de  $K$ . Este tipo de atajos son convenientes no introducen más parámetros ni complejidad computacional, y se incluyen cada cierto número de capas, concretamente en nuestra red después del *embedding* hasta el MLP, antes del MLP hasta pasado el MLP, y previo al post-procesamiento hasta finalizar este.

### 3.5 V-score

Una vez construida la red neuronal, necesitamos una métrica que nos permita determinar cuándo hemos alcanzado el mínimo de la energía. En este trabajo se usó la V-score [24], definida como

$$V_{\text{score}} = \frac{\text{Var}(\langle H \rangle)}{\langle H \rangle^2} \quad (47)$$

La aptitud de esta métrica se debe a que en optimizaciones variacionales, la varianza de la energía escala linealmente con la discrepancia entre la energía calculada y la real, lo que significa que cuanto menor sea el V-score, más cerca estaremos del estado fundamental, siendo exacto cuando esta métrica es cero. Por ello, de entre todos los estados calculados durante la simulación, nos quedaremos con el del mínimo V-score.

## 4 Resultados

Con la red neuronal presentada, resolvimos el sistema para cadenas de  $N = 20$  y  $N = 50$  espines, para  $\alpha$ 's en todas los regímenes mencionados. La razón de esta selección es que para  $N = 20$  es posible la resolución por diagonalización exacta. Esto viene implementado en la librería "Netket", lo que nos permitió comparar en nuestras soluciones la fidelidad del resultado. Calculándose la fidelidad como:

$$F = |\langle \psi | \psi(w) \rangle| \quad (48)$$

Que es una comparativa de cuánto se parecen el autovector real y el calculado con los parámetros de red  $w$ . Sin embargo, para  $N = 50$  esto ya no es posible, la cantidad de estados posibles hace inviable la diagonalización. De esta manera, podremos comparar los resultados con solución previa y sin ella.

A continuación se muestran los resultados obtenidos.

### 4.1 N=20

Para la longitud de cadena de 20 espines, en primer lugar, contemplamos la fidelidad obtenida. Podemos distinguir cómo la fidelidad es completa para gran parte de los puntos de la gráfica, concentrándose la mayoría de puntos discrepantes en la región antiferromagnética, concretamente donde el campo transversal es nulo, y por ende el Hamiltoniano es diagonal, y es que el modelo empleado es demasiado complejo para la sencillez del problema en esta región, ya que en esta zona el estado fundamental estará formado por uno o un número pequeño de estados base, y la red no siempre será capaz de determinarlo mediante las simulaciones Montecarlo. Al mismo tiempo, se aprecia que los errores y las zonas donde se acumulan son mayores cuanto mayor es la distancia de interacción entre espines. Estas zonas deficientes se arrastrarán en las representaciones del resto de magnitudes de relevancia, como vemos seguidamente.

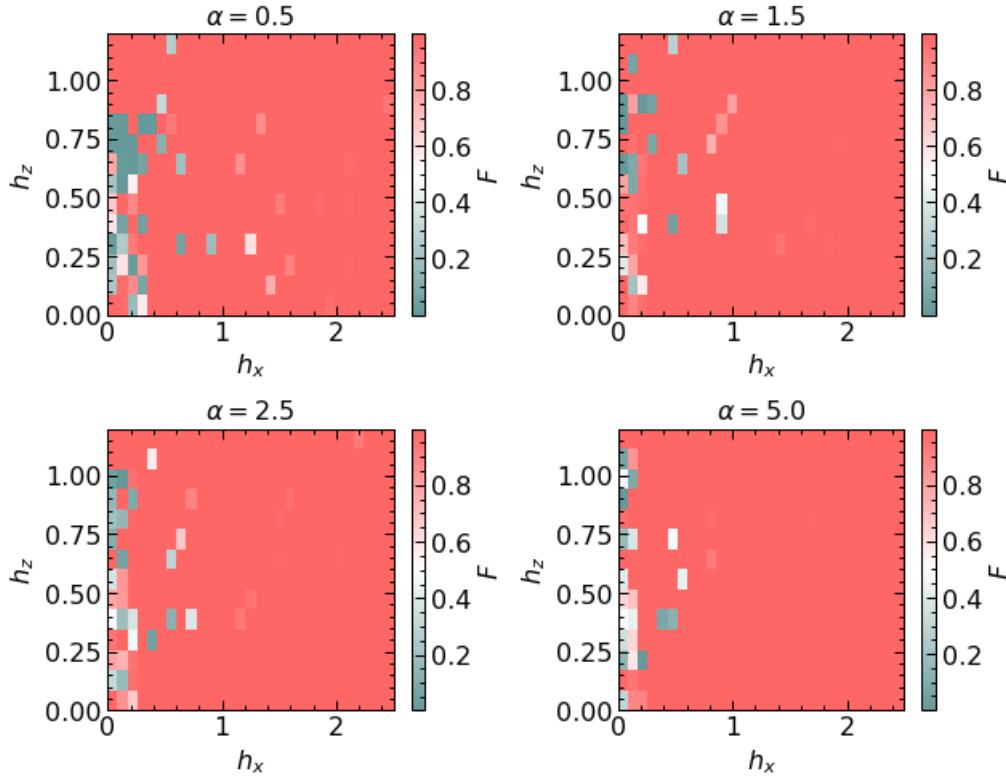


Figure 5: Representación de la fidelidad entre el estado obtenido por diagonalización y el obtenido por la red para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 20$ .

En las gráficas de la imanación alternada se muestran valores en la fase antiferromagnética que no son nulos y tampoco máximos ( $m_s = 1$ ), además encontramos valores tanto negativos como positivos. Esto se debe a que en esta fase, el estado fundamental se encuentra degenerado entre los estados simétricos por paridad a falta de un campo más intenso que lo decante por uno de los dos estados, como ya se que comentó en las primeras secciones. Esta degeneración, es una superposición de ambos estados no equilibrada, es decir, no tienen los mismos coeficientes  $\psi$ , siendo unas veces superior uno y otras otro, por ello la presencia de ambos signos en la imanación. Esto se soluciona representando las fluctuaciones  $\langle m_s^2 \rangle$  en lugar de la imanación, como veremos en el caso  $N = 50$ .

En cualquier caso, en estas gráficas se puede comprobar el comportamiento que sospechábamos en la sección 2.3. Para  $\alpha = 0.5$ , estamos en el régimen de largo alcance, y ya se aprecia una forma similar a la de la imagen 3, para  $\alpha = 5$  es la región de corto alcance, y la forma de nuevo es como buscábamos. Entre estas dos gráficas, la forma se va amoldando para llegar a la otra, estirándose conforme las interacciones se acortan como nos diría la intuición, las correlaciones deben ser mayores para mayores alcances, y va a ser más difícil romper el ordenamiento mediante un campo. A la misma vez, se observa una difuminación cuando aumentamos el campo transversal a campo longitudinal bajo, como debe ocurrir en la transición de segundo orden, y un corte en el caso de campo transversal bajo y longitudinal alto. Todas estas características se verán con mayor claridad en el caso  $N = 50$ , al disponer de un mayor número de puntos.

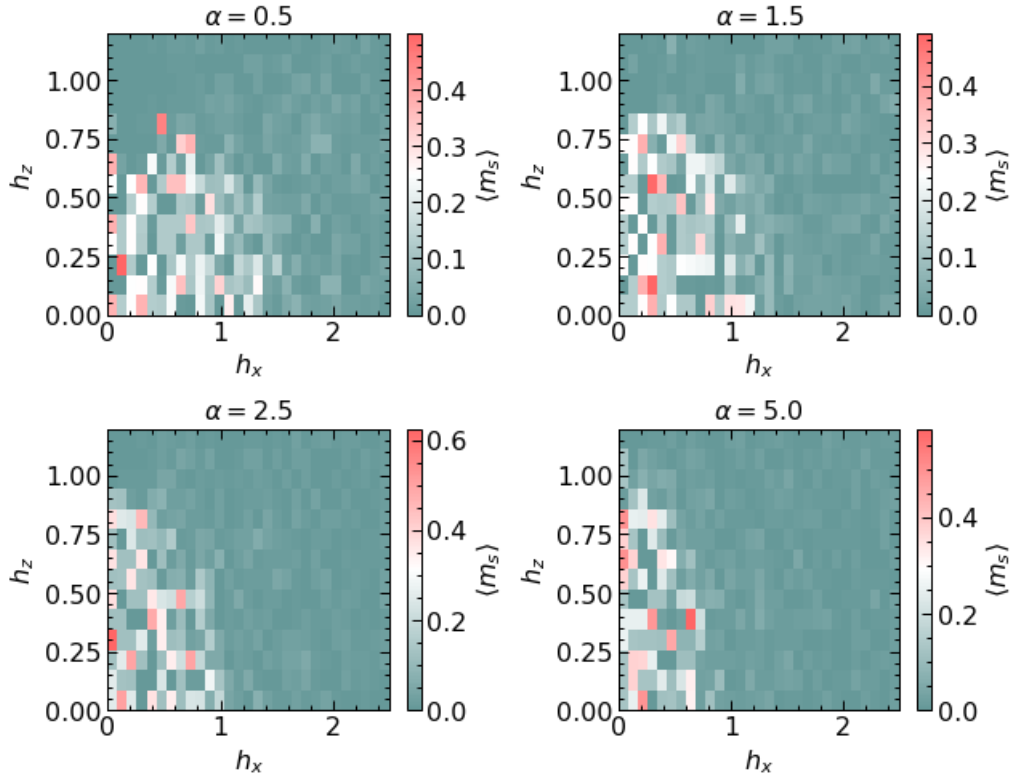


Figure 6: Representación de la imanación alternada  $m_s$  para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 20$ .

Con la imanación no encontramos el problema de la degeneración anterior, ya que aun teniendo los dos estados degenerados, ambos son antiferromagnéticos, y la suma de sus espines va seguir siendo cero, y por tanto siempre va salir cero. En la zona donde predomina el campo longitudinal, se va a producir alineamiento de los espines, y el resto de las zonas se va a comportar también a como anticipamos en la sección 2.3.

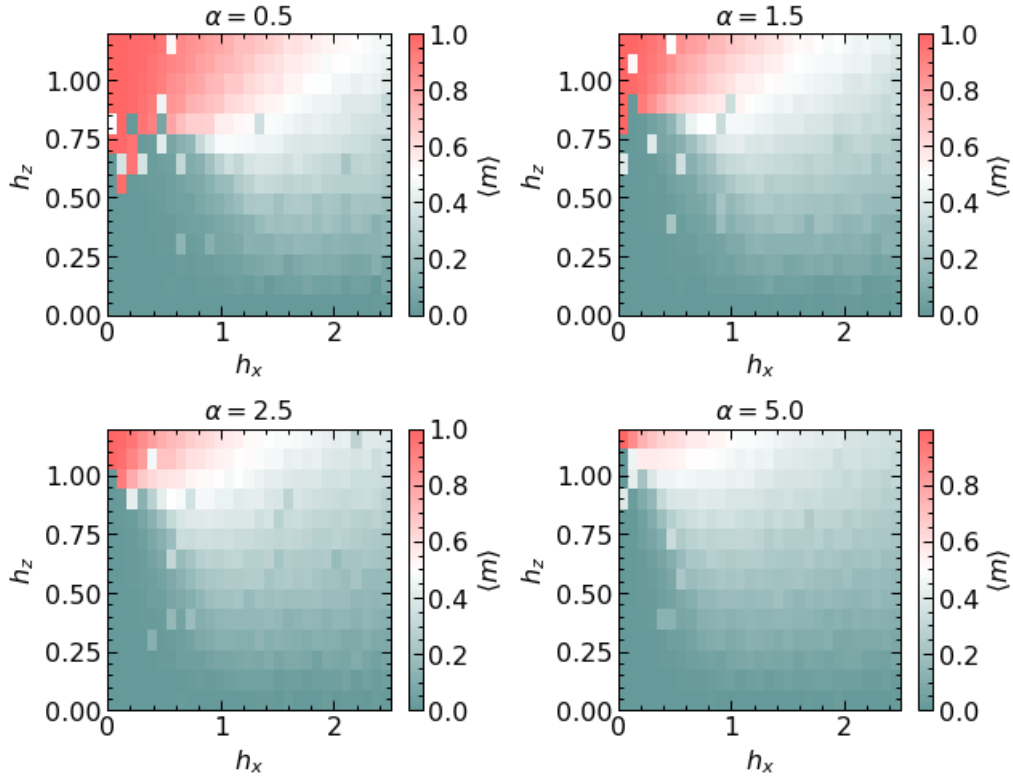


Figure 7: Representación de la imanación  $m$  para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 20$ .

Por último, visualizamos la entropía, que se calculó para ambos tamaños tomando las particiones como la mitad de la cadena, y donde cabe aclarar que los puntos más iluminados se corresponden con bajas fidelidades. Como ya sospechábamos, vemos un comportamiento similar al de las dos figuras anteriores, distinguiéndose mejor la transición de segundo orden en este caso. Con la entropía de entrelazamiento, se muestra la presencia de este en las fases antiferromagnéticas, y como era de esperar una mayor región cuanto mayor alcance tienen las interacciones. Por otro lado vemos que en la zona paramagnética la entropía es siempre cero, ya que en la zona donde hay imanación, no vamos a obtener información de una partición porque ya sabemos que la otra también se va encontrar con los espines alineados, resultando en entropía nula. Para la parte sin ningún tipo de imanación, o de imanación baja, los espines van a estar completamente desordenados y aunque conozcamos una parte de la cadena vamos a seguir sin saber cómo se encuentra la otra.

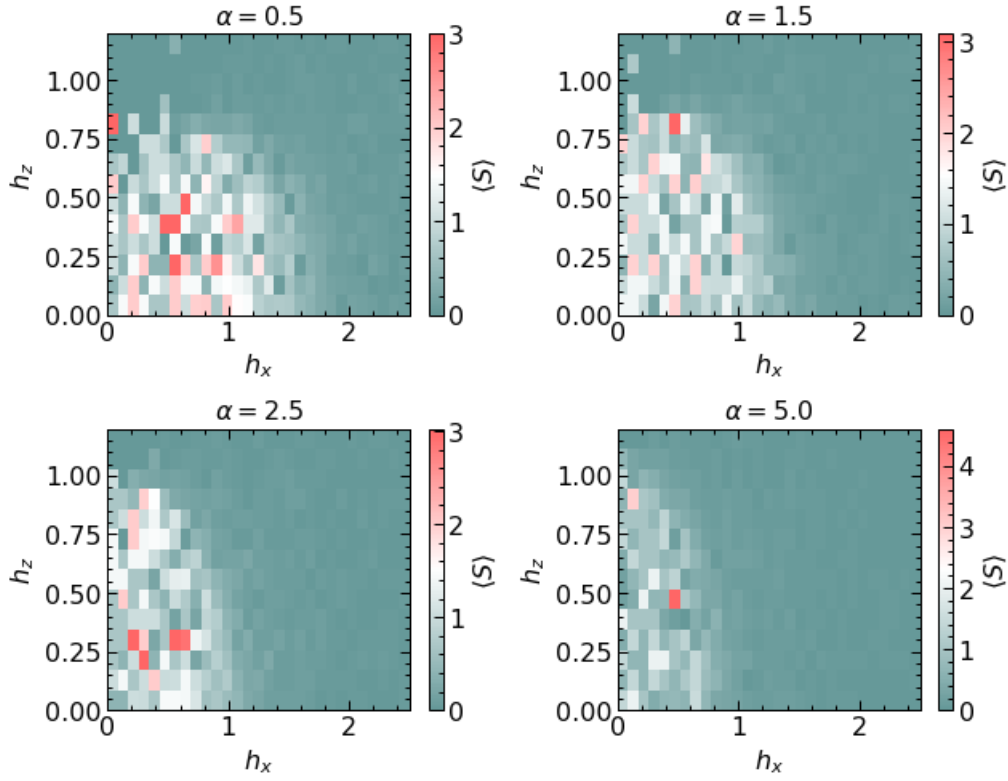


Figure 8: Representación de la entropía de entrelazamiento  $S$  para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 20$ .

## 4.2 $N=50$

Tras realizar de nuevo los cálculos para un tamaño de cadena mayor, se observaron unas gráficas similares, con una mayor claridad debido al incremento de puntos utilizados.

En la siguiente gráfica se observa que efectivamente las fluctuaciones nos ahorran los problemas mencionados y proporcionan una mejor descripción cualitativa de las fases. Apreciándose ahora completamente la evolución de la transición de fase de primer orden y del posicionamiento del punto crítico.

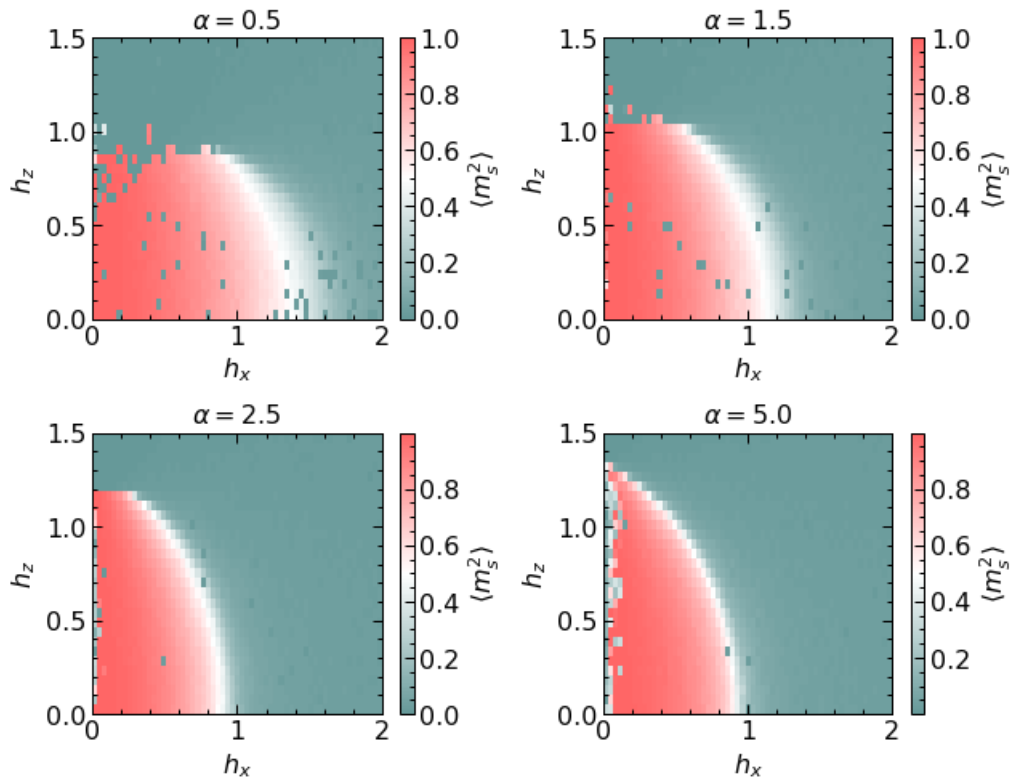


Figure 9: Representación de las fluctuaciones  $\langle m_s^2 \rangle$  para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 50$

Ahora, observando la imanación, vemos que posee la misma forma que la gráfica anterior, con una gráfica prácticamente idéntica a la gráfica derecha de 3.

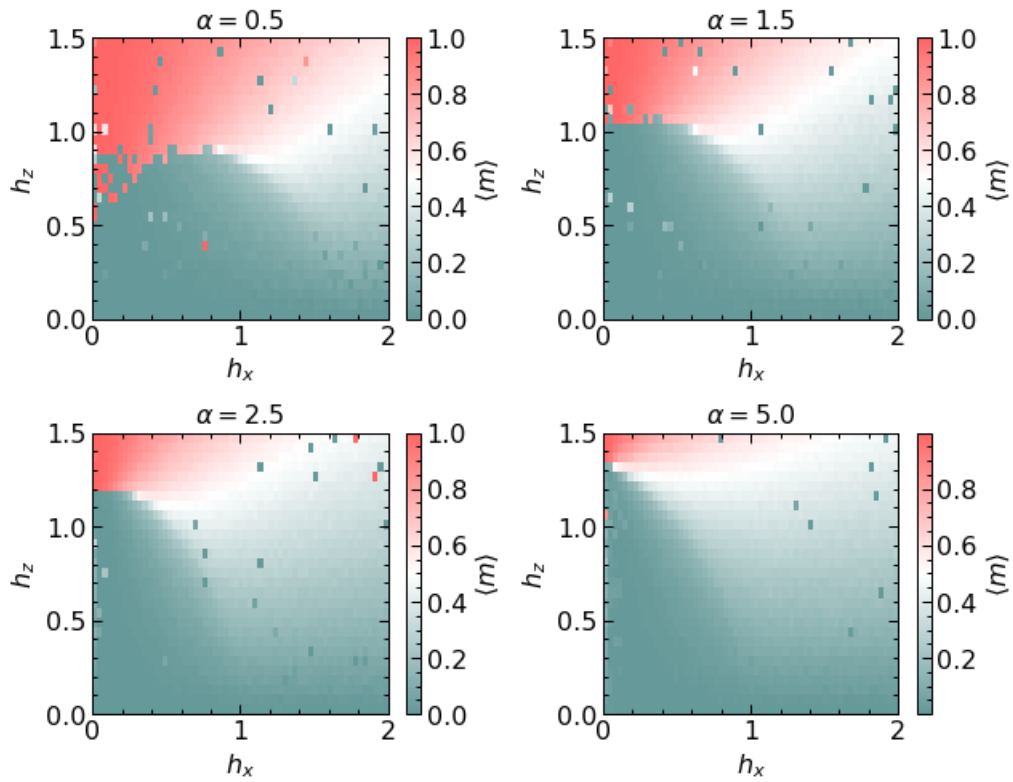


Figure 10: Representación de la imanación  $m$  para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 50$

En cuanto a la entropía, al tener ahora más puntos, se puede discernir que en la transición de fase de segundo orden la entropía decrece de forma continua hacia la fase desordenada, como debe ocurrir en una transición de fase de segundo orden. Mientras que se observa una discontinuidad en la transición de primer orden de la figura de la izquierda, como también debe ocurrir en las transiciones de primer orden.

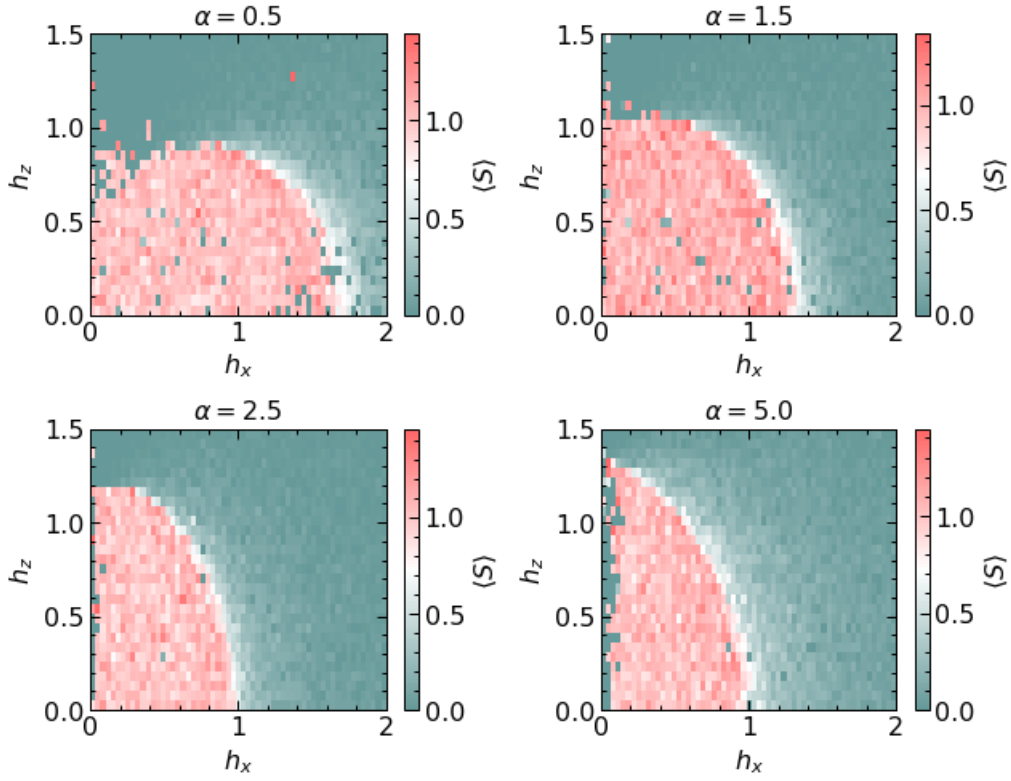


Figure 11: Representación de la entropía de entrelazamiento  $S$  para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 50$

Ahora también incluimos el V-score. De estas gráficas inferimos que la zona donde más complicado resulta encontrar los estados, aparte de para campo transversal cero, es la frontera de la transición de segundo orden, pudiéndose comprobar en la gráfica anterior, que el entrelazamiento no es la causa de este suceso. Por otro lado, en la frontera de primer orden no ocurría esto.

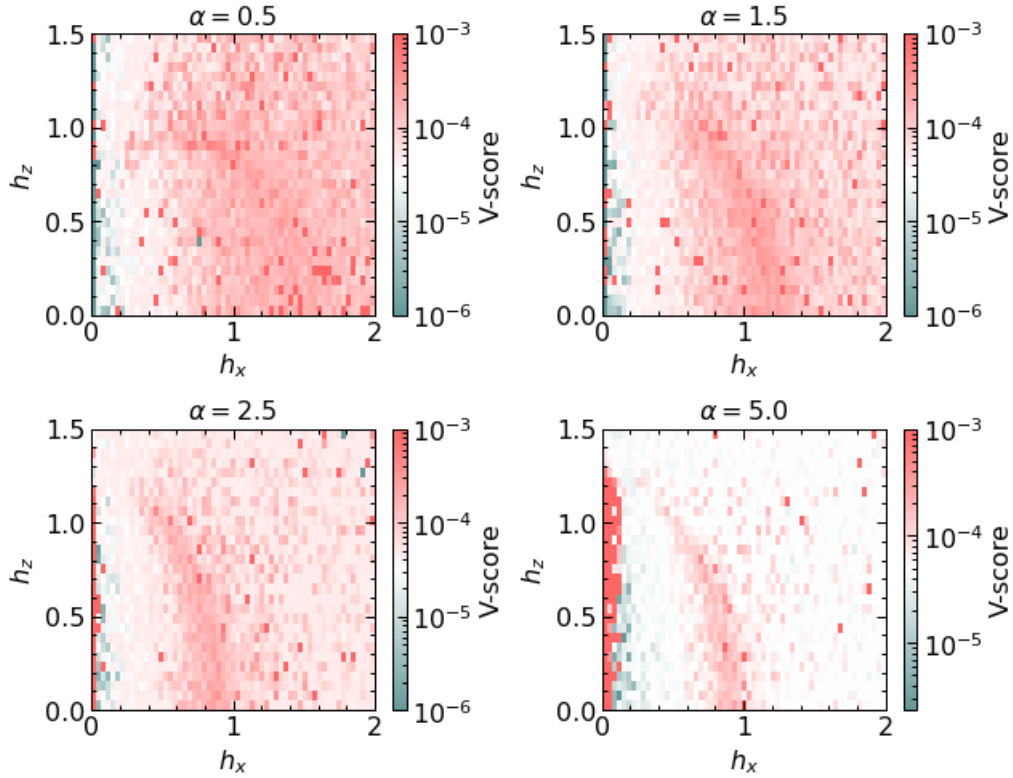


Figure 12: Representación del V-score para cuatro  $\alpha$ 's en las mismas regiones de campos aplicados con  $N = 50$

## 5 Conclusión

En este trabajo hemos estudiado los modelos cuánticos de largo alcance, donde hemos diferenciado los diferentes regímenes de comportamiento según el valor que presentara el alcance de las interacciones, e indicando algunas propiedades como la aditividad o la extensividad, solucionando el problema con esta última, y comentando algunos resultados. Posteriormente definimos el caso concreto de largo alcance del modelo de Ising cuántico antiferromagnético, siendo el caso que íbamos a resolver, donde se enfatizaba principalmente la presencia del campo transversal que es lo que hace al modelo cuántico, y presentando en qué condiciones ocurre una transición de fase cuántica, y mostrando los regímenes extremos del modelo de Ising resueltos que fueron los que nos sirvieron de base para estudiar el régimen intermedio y comprobar nuestros resultados. A su vez establecíamos la imanación, la imanación alternada y la entropía como las magnitudes principales a medir en las simulaciones.

Posteriormente se presentaba las redes neuronales y su utilidad en la resolución de problemas variacionales de este tipo, ya que al ser capaz de codificar los valores que se introducen, era posible encontrar representaciones de soluciones con un tamaño de espacio menor al de Hilbert, entonces se realizó un repaso de sus características más esenciales y mostrando la construcción de un *transformer* y sus distintas componentes, donde aprendimos que debido a que la cantidad de pasos para la conexión de una capa

con las demás no dependía de la longitud de la secuencia, lo que implicaba que la red captaría mejor las interacciones de largo alcance, siendo ideal para el problema a resolver.

Por último, se analizaron los resultados, encajando estos con las predicciones realizadas para la aparición de una transición de fase de segundo orden para campos longitudinales bajos y transversales altos, y una de primer orden para transversales bajos y longitudinales bajos, observando que la evolución de estas transiciones para distintas distancias de interacción donde, aunque algunos puntos en la zona clásica no eran satisfactorios, globalmente se mostraba a través de los parámetros de relevancia el comportamiento de las fases y sus transiciones, haciendo presente la importancia de las correlaciones con la entropía de entrelazamiento, y encajando finalmente con los resultados de trabajos anteriores.

Desde mi punto de vista, el hecho más importante a destacar del trabajo es haber conseguido encontrar la solución de un sistema con dimensión  $2^{50}$ , mediante el uso de tan solo 1000 estados base en el método Montecarlo, pudiendo aun así obtener los resultados correctos en unos tiempos satisfactorios.

## References

- [1] Hiroki Saito. “Solving the Bose–Hubbard Model with Machine Learning”. In: *Journal of the Physical Society of Japan* 86.9 (2017), p. 093001. DOI: 10.7566/JPSJ.86.093001.
- [2] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. “Quantum Entanglement in Neural Network States”. In: *Physical Review X* 7.2 (May 2017). ISSN: 2160-3308. DOI: 10.1103/physrevx.7.021021. URL: <http://dx.doi.org/10.1103/PhysRevX.7.021021>.
- [3] Nicolò Defenu et al. “Long-range interacting quantum systems”. In: *Rev. Mod. Phys.* 95 (3 Aug. 2023), p. 035002. DOI: 10.1103/RevModPhys.95.035002. URL: <https://link.aps.org/doi/10.1103/RevModPhys.95.035002>.
- [4] Juan Román-Roche, Víctor Herráiz-López, and David Zueco. “Exact solution for quantum strong long-range models via a generalized Hubbard-Stratonovich transformation”. In: *Phys. Rev. B* 108 (16 Oct. 2023), p. 165130. DOI: 10.1103/PhysRevB.108.165130. URL: <https://link.aps.org/doi/10.1103/PhysRevB.108.165130>.
- [5] Sebastián Roca-Jerat et al. “Transformer Wave Function for Quantum Long Range models”. En redacción.
- [6] Víctor Herráiz López. *Modelos cuánticos de muy largo alcance: solución general y su aplicación a cadenas de Ising*. TFG Universidad de Zaragoza. 2023.
- [7] Dimo I. Uzunov. *Theory of Critical Phenomena. Mean field, Fluctuations and Renormalization*. 2nd ed. World Scientific, 2010.
- [8] Giuliano Benenti et al. *Principles of Quantum Computation and Information*. WORLD SCIENTIFIC, 2019. DOI: 10.1142/10909.
- [9] Subir Sachdev. *Quantum Phase Transitions*. 2nd ed. Cambridge University Press, 2011.
- [10] Jonathan Simon et al. “Quantum simulation of antiferromagnetic spin chains in an optical lattice”. In: *Nature* 472.7343 (Apr. 2011), pp. 307–312. ISSN: 1476-4687. DOI: 10.1038/nature09994. URL: <http://dx.doi.org/10.1038/nature09994>.
- [11] The Netket authors. *NetKet: The Machine Learning Toolbox for Quantum Physics*. Último acceso 11/06/2024. 2019-2021. URL: <https://netket.readthedocs.io/en/latest/>.
- [12] Simon Haykin. *Neural Networks and Learning Machines*. 3rd ed. Pearson, 2008.
- [13] John Hertz, Anders Krogh, and Richard Palmer. *Introduction to the Theory of Neural Computation*. Taylor & Francis Group, 2018.

- [14] Sandro Sorella, Michele Casula, and Dario Rocca. “Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods”. In: *The Journal of Chemical Physics* 127.1 (July 2007). ISSN: 1089-7690. DOI: 10.1063/1.2746035. URL: <http://dx.doi.org/10.1063/1.2746035>.
- [15] Chae-Yeun Park and Michael J. Kastoryano. “Geometry of learning neural quantum states”. In: *Physical Review Research* 2.2 (May 2020). ISSN: 2643-1564. DOI: 10.1103/physrevresearch.2.023232. URL: <http://dx.doi.org/10.1103/PhysRevResearch.2.023232>.
- [16] Michele Casula, Claudio Attaccalite, and Sandro Sorella. “Correlated geminal wave function for molecules: An efficient resonating valence bond approach”. In: *The Journal of Chemical Physics* 121.15 (Oct. 2004), pp. 7110–7126. ISSN: 1089-7690. DOI: 10.1063/1.1794632. URL: <http://dx.doi.org/10.1063/1.1794632>.
- [17] Leslie N. Smith. *Cyclical Learning Rates for Training Neural Networks*. 2017. arXiv: 1506.01186 [cs.CV].
- [18] John Hewitt. *Note 10: Self-Attention and Transformers*. 2023.
- [19] Ashish Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL].
- [20] Richard E. Turner. *An Introduction to Transformers*. 2024. arXiv: 2304.10557 [cs.LG].
- [21] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. “Transformer Variational Wave Functions for Frustrated Quantum Spin Systems”. In: *Physical Review Letters* 130.23 (June 2023). ISSN: 1079-7114. DOI: 10.1103/physrevlett.130.236401. URL: <http://dx.doi.org/10.1103/PhysRevLett.130.236401>.
- [22] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. *Searching for Activation Functions*. 2017. arXiv: 1710.05941 [cs.NE].
- [23] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV].
- [24] Dian Wu et al. *Variational Benchmarks for Quantum Many-Body Problems*. 2023. arXiv: 2302.04919 [quant-ph].
- [25] Matthew B. Hastings et al. “Measuring Renyi Entanglement Entropy in Quantum MonteCarlo Simulations”. In: *Physical Review Letters* 104.15 (Apr. 2010). ISSN: 1079-7114. DOI: 10.1103/physrevlett.104.157201. URL: <http://dx.doi.org/10.1103/PhysRevLett.104.157201>.