

TRABAJO DE FIN DE GRADO

INTELIGENCIA ARTIFICIAL APLICADA A LA MECÁNICA CUÁNTICA

Curso 2023-2024

DEPARTAMENTO DE FÍSICA DE LA MATERIA CONDENSADA

Autor: Alberto Albericio Chueca

Tutor: Luis Martín Moreno

Cotutor: Eduardo Sánchez Burillo



Índice

1. Introducción	2
2. Marco teórico	4
2.1. El modelo de Hubbard bosónico	4
2.2. Introducción a las redes neuronales	5
2.2.1. Neuronas	5
2.2.2. Arquitectura de la red	6
2.2.3. Proceso de entrenamiento. Función de coste y optimizador.	7
2.2.4. Sobreentrenamiento	8
2.2.5. Definición de hiperparámetros	9
3. Estado Fundamental	9
3.1. Reducción del número de parámetros	10
3.2. Estudio de la fidelidad según el tamaño del conjunto de entrenamiento	11
3.3. Predicción de la función de onda para un número de partículas diferente	12
3.4. Predicción de los estados fundamentales en función del parámetro de Hubbard	13
4. Estado excitado	15
4.1. Simetría de la función de onda	15
4.2. Red para la predicción del primer excitado	17
4.3. Predicciones para diferente número de bosones	18
5. Transfer-Learning	19
5.1. Objetivos y procedimiento	19
5.2. Resultados del método. Relación con la ratio de proporción (PR)	20
5.3. Transición de fase cuántica	22
6. Conclusiones	24
Referencias	25
7. Anexo	26
7.1. Condiciones de contorno	26
7.2. Funciones de activación	27
7.3. <i>Back Propagation</i>	29
7.4. El modelo de Anderson	29

1. Introducción

Una de las descripciones más exitosas de los electrones en los sólidos es la conocida como “teoría de bandas”. Esta se basa en la reducción de las interacciones de muchos cuerpos a una descripción efectiva de un solo cuerpo. Sin embargo, existen diversas situaciones de interés físico en las que la teoría de bandas falla. El ejemplo más destacado son los aislantes de Mott, los cuales con un número impar de electrones de valencia por celda elemental presentan un comportamiento aislante, en contraposición a la predicción de la teoría de bandas. Para este tipo de situaciones se usan modelos más sofisticados, como el modelo de Hubbard, el cual es capaz de captar comportamientos que la teoría de bandas no puede describir [1].

Tal y como se introduce en el libro de texto de Ashcroft y Mermin [2], el modelo de Hubbard es un “modelo muy simplificado” para electrones que interactúan fuertemente en un sólido. Es un tipo de modelo mínimo que tiene en cuenta el movimiento cuántico de los electrones y la interacción repulsiva no lineal entre ellos. Hay pocas dudas de que el modelo es demasiado simple para describir fielmente los sólidos reales. Sin embargo, es uno de los modelos más importantes en la física teórica [3].

A pesar de ser un modelo sencillo, el espacio de Hilbert asociado a la función de onda de los estados del hamiltoniano adopta una dimensionalidad muy alta al aumentar el número de partículas o el tamaño de la red que simula nuestro sólido. Con ello, los costos computacionales de los cálculos cuánticos precisos aumentan exponencialmente con el tamaño del sistema, alcanzando un límite de rendimiento con solo decenas de átomos, convirtiendo la simulación de sistemas grandes en algo imposible [4]. Debido a este factor limitante, se nos requiere la idea de aplicar estrategias computacionales que nos permitan reducir complejidad al problema. En este momento es cuando surge la idea de la implementación de la inteligencia artificial para poder ayudarnos. Nuestro objetivo es modesto, ya que utilizaremos casos que son tratables por diagonalización exacta, pero no por ello menos útil. Se pretende reflejar cómo la inteligencia artificial nos permite encapsular la función de onda en un objeto de menos parámetros, permitiéndonos así ahorrar información y en muchos casos, también nos permite ahorrar tiempo de computación.

Recientemente hemos contemplado cómo esta rama de la computación ha entrado en auge, hasta el punto de que se ha convertido en algo habitual en nuestra sociedad. Resulta ciertamente útil para la resolución de problemas en los que los límites de nuestra comprensión comienzan a fallar. Esto se da sobre todo en problemas numéricos de alta complejidad como por ejemplo, análisis de datos, reconocimiento de patrones o simulaciones numéricas, como es nuestro caso, superando con creces a los algoritmos tradicionales.

De este modo, puesto que nuestro principal problema para la simulación del modelo de Hubbard reside en sistemas grandes, la finalidad principal de este trabajo es tratar de estudiar la implementación de redes neuronales para la predicción de estados del modelo a través del entrenamiento de un sistema más asequible computacionalmente.

Esto se realizará tanto para el estado fundamental del sistema como para el primer excitado, entrenando una red para un cierto número de bosones con la que trataremos de averiguar si es capaz de predecir estados con un número diferente.

También estudiaremos características físicas del modelo, en concreto, el cambio de la función de onda del estado fundamental según el parámetro de interacción Hubbard. Análogamente trataremos de implementar redes neuronales para descubrir si es capaz de reconocer el proceso de cambio y realizar predicciones en función de este.

Para finalizar, aprovechando que tenemos un conocimiento previo, esto es, un modelo entrenado para el estado fundamental, trataremos de realizar el método conocido como *transfer-learning*, o conocimiento por transferencia, para utilizarlo en favor del estudio del estado excitado. El objetivo de esto es tratar de determinar si de algún modo el conocimiento de la estructura del estado fundamental ayuda al aprendizaje del estado excitado, haciendo este más sencillo, es decir, utilizando menos parámetros entrenables.

Como es sabido, en los modelos cuánticos como el de Hubbard, el estado excitado normalmente está relacionado con el fundamental a través de pequeñas modificaciones locales. Sin embargo, en ocasiones, al variar un parámetro del sistema, un estado altamente excitado puede convertirse en fundamental. Cerca de esta transición se encuentran casi degenerados, y el fundamental se convierte en el primer excitado, perdiendo la relación cercana anterior entre fundamental y excitado. Esto es lo que conocemos como una transición de fase cuántica. De este modo, puesto que la correlación entre fundamental y excitado se rompe en las transiciones de fase, nuestro objetivo será tratar de utilizar el *transfer-learning* como una herramienta para detectarlas, ya que el aprendizaje del excitado debería de ser más complicado y ofrecer peores resultados en estos casos.

Cabe destacar que todo este proceso requiere de un estudio minucioso de los hiperparámetros que definen nuestra red neuronal, así como de un conocimiento de los procesos que esta lleva a cabo internamente durante su aprendizaje, por lo que a lo largo de todo el trabajo se irán indicando las configuraciones aplicadas, así como el procedimiento llevado a cabo para la obtención de los valores concretos que nos permiten una optimización de los resultados.

En resumen, ya hemos visto como la expansión de la inteligencia artificial ha ido implementándose en todos los ámbitos de trabajo y de la vida en general. La física, en concreto, la física cuántica, por ende, no iba a ser una excepción. Con este trabajo, esperamos que quede constancia de las capacidades que nos ofrecen las redes neuronales de ayudarnos en el estudio de sistemas complejos como lo son los sistemas cuánticos, proporcionándonos herramientas para la reducción de la dimensionalidad del problema objetivo a través de la compresión de información, así como otros muchos usos potenciales para la detección de características físicas que aún nos quedan por descubrir.

2. Marco teórico

2.1. El modelo de Hubbard bosónico

Como ya hemos mencionado anteriormente, el modelo de Hubbard nos permite el estudio de la física de sólidos para propiedades que la teoría de bandas colapsa como, por ejemplo, las transiciones de Mott. Esta es más fácil de estudiar en sistemas bosónicos, por lo que el modelo que se va a emplear es el conocido como modelo de Bose-Hubbard.

Consideraremos nuestro sólido como una red fija donde pueden situarse las partículas y cada una de ellas tiene la posibilidad de saltar a sus sitios próximos, lo cual se ilustra esquemáticamente en la figura 1.

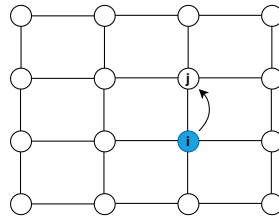


Figura 1: Salto de partícula de posición 'i' a posición 'j' en la red con el operador $c_j^\dagger c_i$.

De este modo, al ser las partículas bosónicas, estas pueden ocupar el mismo estado cuántico de modo que podrán encontrarse juntas en un mismo espacio de la red. Además, nuestro trabajo se centra en el estudio unidimensional del modelo, por lo que definiremos la base del espacio de Hilbert de nuestro sistema como:

$$\mathcal{H} = \left\{ |n_1, n_2, \dots, n_M\rangle \mid \sum_i n_i = N \right\}, \quad (1)$$

donde N es el número total de bosones del sistema el cual debe mantenerse constante, M es el número de sitios de la red y n_i representa el número de bosones que se encuentran en el sitio 'i' de la red.

Una vez definida la base, lo próximo es obtener la expresión del Hamiltoniano que, en el caso del modelo de Bose-Hubbard, viene dado por la expresión [5]:

$$H = -t \sum_{\langle i,j \rangle} (c_i^\dagger c_j + c_j^\dagger c_i) + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1) + \sum_i E_i \hat{n}_i, \quad (2)$$

donde $\hat{n}_i = c_i^\dagger c_i$ es el operador número, E_i es la energía de la partícula en el sitio 'i', $-t$ es el término de "hopping" que representa la probabilidad de salto entre lugares vecinos de la red y U es el parámetro que modula la interacción de Hubbard entre los bosones que se encuentran en un mismo sitio, representando una interacción efectiva repulsiva si $U > 0$ y atractiva si $U < 0$.

A su vez, c_i^\dagger y c_j son los operadores de creación y destrucción de bosones en los sitios i, j de la red respectivamente, que actúan sobre los estados de la base de la siguiente manera:

$$\begin{aligned}
 c_i |n_1, n_2, \dots, n_i, \dots, n_M\rangle &= \sqrt{n_i} |n_1, n_2, \dots, n_i - 1, \dots, n_M\rangle \\
 c_i^\dagger |n_1, n_2, \dots, n_i, \dots, n_M\rangle &= \sqrt{n_i + 1} |n_1, n_2, \dots, n_i + 1, \dots, n_M\rangle.
 \end{aligned}
 \tag{3}$$

Además, es importante destacar que a lo largo de todo el trabajo se trata a toda la red con condiciones periódicas de contorno (se ilustra en la sección 1 del anexo), así como los valores de los parámetros que se van a utilizar son $t = 1$, $\{E_i\} = 1 \forall i$ y $U = 2$ salvo en los apartados en los que se realice un estudio en función de este último.

2.2. Introducción a las redes neuronales

2.2.1. Neuronas

Antes de adentrarnos en la dinámica de una red neuronal, tenemos que definir cómo es cada uno de los elementos que componen esta: las neuronas. Cabe destacar que las neuronas pueden definirse con multitud de formas. Sin embargo, nos centraremos en las más elementales [6]:

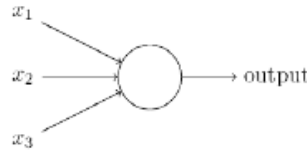


Figura 2: Esquema de una neurona.

- Perceptrones:

Definiremos un perceptrón como un elemento con múltiples inputs ($\{x_i\}$) a los cuales se les asocia una cantidad, un “peso” ($\{w_i\}$), y un output binario que su valor dependerá de la activación de los inputs si cumplen la condición de activación:

$$output = \begin{cases} 0 & \text{si } \sum_j x_j w_j \leq threshold \\ 1 & \text{si } \sum_j x_j w_j > threshold \end{cases}
 \tag{4}$$

siendo “*threshold*” un valor asociado a cada neurona que podríamos interpretar como un “voltaje de activación”.

- Neuronas sigmoides:

A diferencia de los perceptrones, las neuronas sigmoides pueden devolver en el output un valor real entre 0 y 1, en vez de un valor binario. La arquitectura es la misma que los perceptrones salvo que la función que devuelve a la salida es la siguiente:

$$output = \sigma(z) = \frac{1}{1+e^{-z}} \quad ; \quad z = \sum_j x_j w_j + b \quad ; \quad b := \text{“Threshold”}.
 \tag{5}$$

La ventaja sobre los perceptrones es que $\sigma(z)$ es derivable, lo que permite el aprendizaje por “descenso de gradiente” (sección 2.2.3) y *back-propagation* (anexo, sección 3). A su vez, también ofrecen una interpretación probabilística en problemas de clasificación binaria y nos permiten abordar problemas de regresión siempre y cuando los datos estén normalizados entre 0 y 1, como es nuestro caso con la función de onda del estado fundamental de nuestro sistema cuántico. De este modo, definiremos la función no lineal que se aplica a la salida de la neurona como “función de activación”. Existe una gran variedad de funciones de activación posibles para aplicar. Todas las que se utilizan a lo largo del trabajo quedan recogidas y explicadas en la sección 2 del anexo.

2.2.2. Arquitectura de la red

Una vez definidos los elementos básicos de la red, las neuronas, tenemos que realizar la estructura de conexiones entre ellas para definir nuestra red. A continuación se explica la arquitectura de red que se va a utilizar en la mayoría de las secciones.

En primer lugar, definimos una capa de entrada con un número de neuronas equivalente al tamaño de nuestra muestra de datos. Como nuestras entradas serán los estados de la base del espacio de Hilbert dados en (1), nuestra capa de entrada tendrá 'M' neuronas, correspondiéndose con cada uno de los sitios de nuestra red y como input el número de bosones correspondiente en cada sitio.

A continuación, definiremos capas ocultas con un número variable de neuronas (n_1, n_2, \dots) que formarán una red completamente conexa, es decir, cada neurona está unida a todas las neuronas de la capa precedente. Cada capa de neuronas tendrá una función de activación determinada, las cuales se especificarán en cada problema concreto.

Para finalizar, tendremos en la capa de salida una única neurona cuyo valor de salida se corresponderá con el del coeficiente de la función de onda asociado al estado de la base dado en la capa de entrada. En la figura 3 se muestra una imagen de la arquitectura de red.

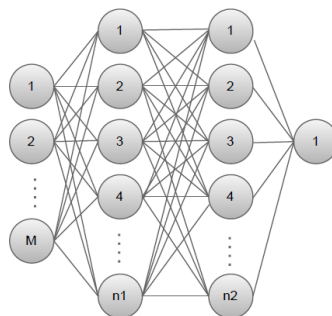


Figura 3: Esquema de la red completamente conexa.

2.2.3. Proceso de entrenamiento. Función de coste y optimizador.

Una vez definida la estructura de la red, lo siguiente es conocer la dinámica que sigue el proceso de aprendizaje de esta. El objetivo es entrenar a la red con datos ($\{x_i\}$) de resultados ya conocidos ($\{y_i\}$) de modo que los pesos y los voltajes de activación de las neuronas se ajusten de la mejor manera posible para reproducir los resultados. Para ser capaces de cuantificar cómo de bien se está ajustando la red para predecir los resultados dados, es típicamente usado en problemas de regresión la función de coste:

$$C = \frac{1}{2n} \sum_{i=1}^n (y_i - a_i(x_i))^2, \quad (6)$$

siendo $a_i(x_i)$ la predicción de la red para el dato x_i y n el número total de entradas.

De este modo, el objetivo del entrenamiento es variar los pesos y los sesgos de modo que se minimice esta función de coste. Para optimizar este proceso, se utiliza el método de “descenso por gradiente”, el cual se explica a continuación.

Puesto que la función de coste depende de todos los pesos y voltajes de las neuronas de la red, es una función de muchas variables y resolver analíticamente el mínimo de la función es directamente imposible. Por este motivo es que se crea el método de descenso por gradiente. Por ello, partimos gracias al cálculo diferencial que el cambio en la función de coste viene dado por:

$$\Delta C \approx \sum_i \frac{\partial C}{\partial v_i} \Delta v_i. \quad (7)$$

Definiendo los vectores:

$$\nabla C = \left(\frac{\partial C}{\partial v_1}, \dots, \frac{\partial C}{\partial v_n} \right)^T ; \quad \Delta v = (\Delta v_1, \dots, \Delta v_n)^T, \quad (8)$$

la ecuación (7) puede reescribirse como:

$$\Delta C \approx \nabla C \cdot \Delta v \quad (9)$$

Lo interesante de esto, es que nos permite escoger los cambios en los parámetros ($\{v_i\}$) de modo que el cambio en la función de coste siempre sea negativo (dirigiéndonos hacia el mínimo). En concreto, tomaremos $\Delta v = -\eta \nabla C$, de modo que $\Delta C \approx -\eta |\nabla C|^2 < 0$. Lo cual se traduce en que las variables (los pesos y voltajes en la red) cambiarán según la expresión:

$$v \rightarrow v' = v - \eta \nabla C, \quad (10)$$

siendo η el parámetro conocido como “tasa de aprendizaje”, el cual es necesario ajustar para que la red aprenda de la mejor manera posible.

Puesto que el cálculo del gradiente de la función coste conlleva el cómputo de multitud de derivadas, es un proceso muy costoso. Sin embargo, gracias a la técnica de *back propagation*, los cálculos se agilizan considerablemente. Esta técnica consiste en la obtención de las derivadas de la función coste respecto a los parámetros de la red desde la capa de salida hacia la capa de entrada, calculando las derivadas de capas anteriores de manera recursiva. Esta técnica se explica de manera detallada en la sección 3 del anexo.

Puesto que las dimensiones que puede alcanzar el espacio de Hilbert son muy altas, el tamaño de nuestras muestras es demasiado grande como para utilizar el método anterior de descenso por gradiente. En su defecto, para estos casos con un número de datos tan elevado, se utiliza el conocido como descenso de gradiente estocástico. La única diferencia con el anterior es que elegiremos de manera aleatoria un conjunto de 'm' datos de entrenamiento X_1, X_2, \dots, X_m al que llamaremos *mini-batch*, mediante el cual estimaremos el gradiente de la función coste a través del valor medio de los conjuntos:

$$\nabla C \approx \frac{1}{m} \sum_{j=1}^m \nabla C_{X_j}. \quad (11)$$

Una vez realizado todo este proceso, diremos que se ha completado una “época”. Después de completarla, se comienza de nuevo y así sucesivamente hasta finalizar el entrenamiento.

2.2.4. Sobreentrenamiento

El primer paso a dar antes de realizar predicciones con nuestra red neuronal es determinar los parámetros que conducen a un mejor funcionamiento de esta. Para ello, tenemos varias herramientas para analizar. Una de ellas es la monitorización de la función de coste a lo largo del proceso de entrenamiento de la red, ya que es una buena herramienta para determinar el *overfitting* o sobreentrenamiento.

Este hecho se da cuando la red se ajusta demasiado a los datos de entrenamiento y no es capaz de generalizar correctamente a nuevos datos. Es decir, los modelos sobreajustados tienden a memorizar todos los datos, incluido el ruido inevitable en el conjunto de entrenamiento, en lugar de aprender la disciplina oculta detrás de estos [7]. En el caso de las redes neuronales, el sobreajuste se aprecia principalmente cuando los costes en el conjunto de validación aumentan en función de la época mientras que los costes en el conjunto de entrenamiento continúan disminuyendo. En la figura 4 se muestra este efecto.

Para tratar de reducir el sobreentrenamiento se pueden aplicar varias medidas, entre ellas se destacan el uso de regularizadores (introducen un término adicional en la función de coste que penaliza ciertas configuraciones de pesos), la reducción del número de parámetros del sistema (lo cual será, nuestra principal medida) o la utilización de *early-stopping* [7], un *callback* que detiene el proceso de entrenamiento cuando los costes en el conjunto de validación comienzan a aumentar, lo cual también será implementado en nuestra red.

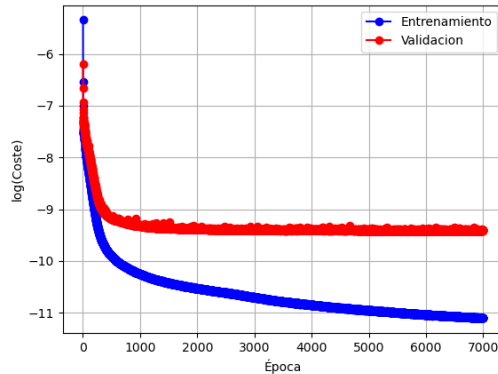


Figura 4: Observación del sobreentrenamiento con los logaritmos de las funciones de coste en el conjunto de entrenamiento y en el de validación. Destacar que el logaritmo utilizado para la representación es el neperiano, lo cual se mantiene a lo largo de todo el trabajo.

2.2.5. Definición de hiperparámetros

Para ser capaces de determinar los valores de los hiperparámetros de la red, vamos a basarnos en una variable que definimos como la fidelidad de la predicción[8]:

$$f = \frac{|\langle \psi | \psi_{exact} \rangle|^2}{|\langle \psi | \psi \rangle|^2}, \quad (12)$$

siendo ψ_{exact} la función de onda obtenida con la diagonalización del Hamiltoniano y ψ la función de onda construida por la red neuronal.

Con esto, tomaremos en un comienzo el 80 % de los datos como conjunto de entrenamiento y el 20 % restante como conjunto de validación. Con estos datos, vamos a observar la variación de la fidelidad de las predicciones según el número de neuronas de las capas intermedias y la tasa de aprendizaje, de tal modo que escogeremos los hiperparámetros de manera que se maximice la fidelidad de las predicciones.

3. Estado Fundamental

Una vez introducidos en la arquitectura y los procesos que lleva a cabo la red neuronal para ajustarse a los datos, comenzamos estudiando el estado fundamental del modelo de Hubbard unidimensional para un caso de 7 bosones y 7 huecos de red. Se han considerado estos valores ya que la dimensión del espacio de Hilbert es lo suficientemente alta como para aportarnos un número considerable de datos de entrenamiento y validación ($dim(\mathcal{H}) = 1716$) y que no sea demasiado costoso computacionalmente, ya que requiere la diagonalización de una matriz de dimensiones $dim(\mathcal{H}) \times dim(\mathcal{H})$.

3.1. Reducción del número de parámetros

Dentro del objetivo de elección de hiperparámetros de manera global, es crucial tratar de reducir el número de neuronas de nuestra red lo máximo posible para reducir así el número de parámetros y comprimir la información. Para ello, realizamos un estudio de la fidelidad de las predicciones en función del número de neuronas en las dos capas. La tasa de aprendizaje se ha fijado en $\eta = 0,55$ que es el valor con el que se obtuvieron mejores resultados, así como el 80% de datos de entrenamiento y el tamaño del *mini-batch* es 10. Sin embargo, para tratar de optimizar el aprendizaje de la red se ha implementado el *early-stopping*. De este modo, podemos entrenar la red con un número de épocas indefinidamente hasta que deje de entrenar correctamente y evitar el *overfitting*. Con estas nuevas características, el estudio de la fidelidad en función del número de neuronas en ambas capas ocultas se ve reflejado en la figura 5.

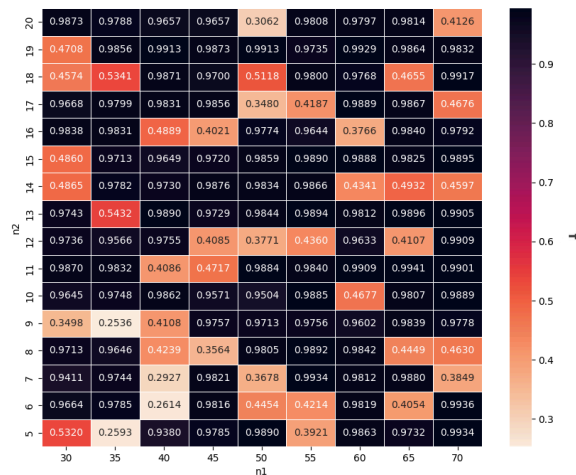


Figura 5: Mapa de calor de la fidelidad en función del número de neuronas en ambas capas.

En esta ocasión, el caso máximo lo alcanzamos con los parámetros:

$$f = 0,9941 \quad n_1 = 65 \quad n_2 = 11. \quad (13)$$

La función de coste asociada a este aprendizaje se recoge en la figura 6.

Por tanto, en esta sección hemos conseguido una fiabilidad alta reduciendo el número de parámetros a 1258, es decir, un 27.7% menos que la dimensión del espacio de Hilbert.

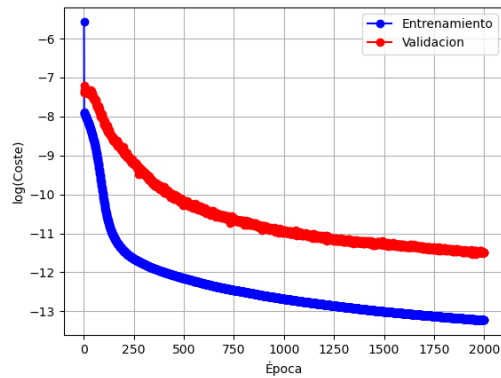


Figura 6: Función de coste de la red 7-65(ReLU)-11(ReLU)-1(sigmoide).

3.2. Estudio de la fidelidad según el tamaño del conjunto de entrenamiento

Puesto que podemos entrenar la red con el número de datos que queramos nos aborda la siguiente cuestión: ¿Cuál es el número mínimo de datos que necesitaríamos para considerar que nuestra red está aprendiendo correctamente? Es por este motivo que es interesante recoger la fidelidad de las predicciones según el tamaño de nuestro conjunto de entrenamiento para observar cómo es la tendencia.

Tal y como puede apreciarse en la figura 7, los resultados no varían prácticamente a partir del 80% del tamaño del conjunto total, por lo que consideramos que hemos hecho una buena elección escogiendo este tamaño en los apartados anteriores.

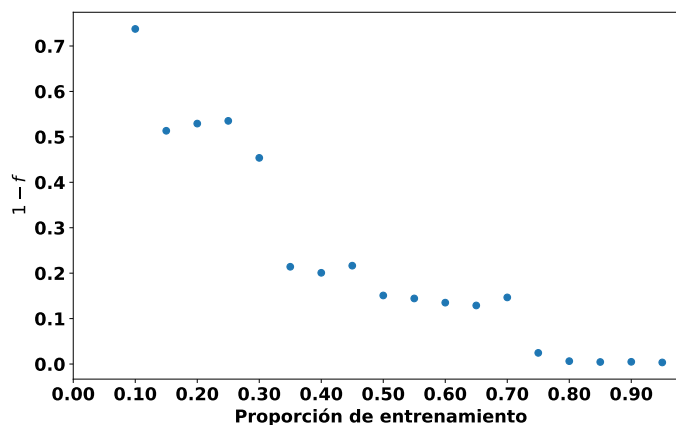


Figura 7: Fidelidad de las predicciones según el tamaño del conjunto de entrenamiento.

3.3. Predicción de la función de onda para un número de partículas diferente

Una vez tenemos nuestra red entrenada para el caso $N = 7$, nos planteamos la cuestión de si esta es capaz de predecir los estados fundamentales para un número diferente de bosones. Para ello, guardamos la configuración de pesos de esta y tomamos como input los estados de la base con un valor general de N , y mandamos al modelo predecir los coeficientes de la función de onda asociados a cada uno de ellos. De este modo, calculamos la fidelidad tomando las predicciones y las funciones de onda obtenidas directamente con el hamiltoniano. Las fidelidades de las predicciones se muestran en la figura 8.

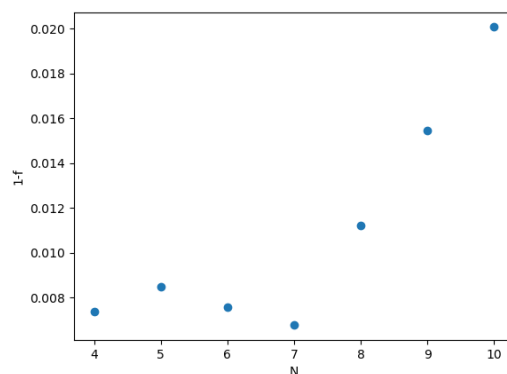


Figura 8: Fidelidad de las predicciones de la función de onda para un número de partículas diferente utilizando la red de la sección 3.1.

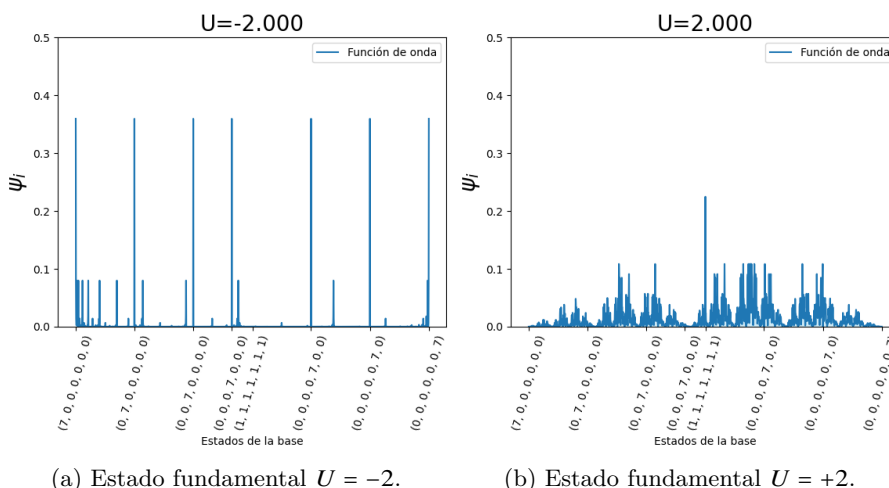
Tal y como puede observarse, la red es capaz de predecir los estados fundamentales con una fiabilidad asombrosa, llegando a ser del 98 % en el peor de los casos con $N = 10$. Por tanto, es de destacar que la red neuronal entrenada con $N = 7$ ha sido capaz de extraer las correlaciones necesarias para reproducir la función de onda del estado fundamental para valores del número de bosones más altos. Además, es apreciable como las predicciones son peores para casos de $N > 7$, al contrario para las de $N < 7$. Esto es causado ya que la dimensión del espacio de Hilbert aumenta rápidamente con el número de partículas ($\dim(\mathcal{H}) = \binom{N+M-1}{N}$), por lo que la complejidad del estado fundamental es mucho mayor que en los casos de menor dimensión y, por ende, la red realiza peores predicciones para estos casos.

En conclusión, podemos afirmar que el entrenamiento de una red con un valor de N bajo nos permite reproducir fielmente los estados fundamentales del modelo de Hubbard de dimensiones del espacio de Hilbert mucho mayores, siendo esta una muy buena herramienta a tener en cuenta para reducir complejidad computacional en simulaciones del modelo.

3.4. Predicción de los estados fundamentales en función del parámetro de Hubbard

A continuación, nos planteamos la cuestión de cómo cambia la función de onda del estado fundamental en función del parámetro de interacción Hubbard. Como es sabido, valores del parámetro de Hubbard negativos se corresponden con un efecto de interacción atractiva, por lo que las amplitudes de los estados en los que todos los bosones se encuentran en el mismo sitio deberían ser más altas que el resto. Por el contrario, ante valores positivos, la interacción repulsiva tiende a que la mayor amplitud se encuentre en el estado donde los bosones se encuentran cada uno en un sitio diferente.

Para ilustrar este efecto, se muestran dos imágenes de las funciones de onda para nuestro valor inicial de $U = 2$ y otra con $U = -2$, mostrando las etiquetas de los estados de todos los bosones juntos y todos separados en la figura 9.



(a) Estado fundamental $U = -2$.

(b) Estado fundamental $U = +2$.

Figura 9: Variación de las amplitudes en la función de onda según el parámetro de Hubbard

Una vez mostrado este efecto, nos aborda la cuestión de si una red es capaz de aprender la representación del estado fundamental a partir del parámetro de Hubbard. Para ello, obtenemos todos los estados fundamentales de la transición desde $U = -2$ hasta $U = +2$ en saltos de 0.025. A continuación, vamos a definir una red la cual la vamos a entrenar aleatoriamente con el 80 % de los datos obtenidos en este rango, tomando como capa de entrada una única neurona, la cual recogerá el parámetro de Hubbard, y una capa de salida con un número de neuronas igual a la dimensión del espacio de Hilbert que recogerá cada uno de los coeficientes de la función de onda.

Tras una inspección de los hiperparámetros que optimizan las predicciones, los mejores resultados se han obtenido para una red con tres capas ocultas de la forma: 1(input)-20(tanh)-80(relu)-10(relu)-dim(H)(sigmoide), con un tamaño del mini-batch de 10 y un valor de la tasa de aprendizaje de $\eta = 0,2$.

La función de coste asociada al entrenamiento se ilustra en la figura 10.

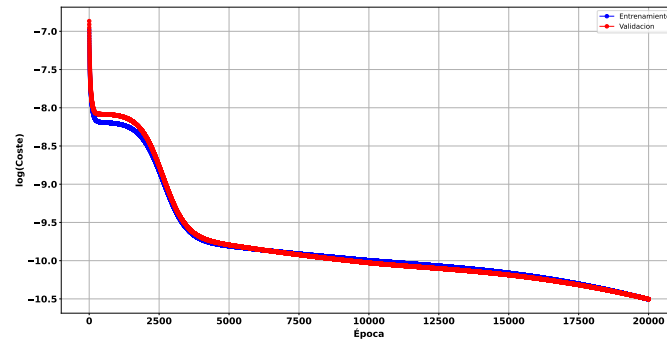


Figura 10: Función de coste asociada al entrenamiento para las predicciones del estado fundamental en función del parámetro de Hubbard. El número máximo de épocas se fijó en 20.000. A pesar de que puedan parecer muchas, se observa que el proceso de aprendizaje podría haber continuado. Tal vez la observación sugiera que hemos sido conservadores con el parámetro de aprendizaje, pero estudiando otros valores se llegó a la conclusión de que este era el valor más óptimo que nos producía mejores resultados en un tiempo de computación razonable.

Con todo ello, se recogen en la figura 11 las fidelidades de las predicciones escogidas aleatoriamente en el rango (el 20 % que no se han usado para entrenamiento). Tal y como puede apreciarse en ellas, las fidelidades son muy altas, la menor de ellas es cercana al 90 %, y se da para un valor de $U = -0,725$. Podemos observar además cierta simetría en las predicciones, ya que parece presenciarse un pico donde estas son peores centrado en el $U \approx -0,5$. Esto es debido a que sobre estos valores es donde se produce la transición de la forma de la función de onda mostrada en la figura 9. Este hecho se justifica ya que hemos tomado el parámetro de hopping como $t = 1$, por lo que este favorece el salto de bosones en la red, que se ve compensado con los valores negativos del parámetro de Hubbard, que producen una tendencia atractiva haciendo que el centro se vea desplazado un poco hacia estos. La compensación entre los dos parámetros hace que el estado fundamental sea impredecible, ya que sería equivalente a que los bosones tuvieran completa libertad para saltar aleatoriamente a sus sitios vecinos, haciendo todos los estados igualmente probables.

Por otro lado, también observamos una pérdida de fidelidad para valores extremos de U , tanto positivos como negativos. Esto es causado ya que al ser el parámetro de Hubbard tan alto en comparación con el parámetro de hopping, la mayoría de las componentes de la función de onda son nulas, ya que para el caso de U muy negativo los estados en los que todos los bosones están juntos conllevan la mayor parte de la amplitud de la función de onda, de igual modo que el estado con todos los bosones en sitios diferentes para valores de U altamente positivos. De esta manera, la mayor parte del conjunto de entrenamiento son valores nulos, por lo que se reduce altamente el número de datos útiles para que la red pueda captar la forma de la función de onda. Este hecho se explica más a detalle en la sección 5.2.

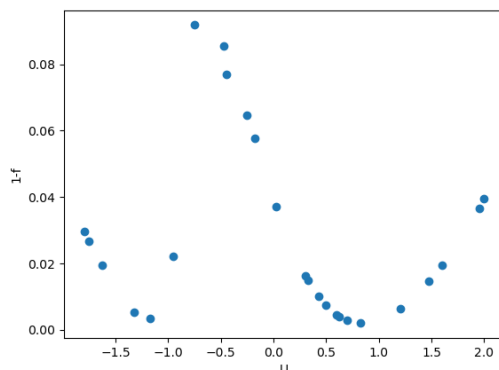


Figura 11: Fidelidad de las predicciones en función del parámetro de Hubbard.

De todos modos, podemos concluir que una red neuronal ha sido capaz de aprender la transición de la forma de la función de onda del estado fundamental, pudiéndonos aportar así información sobre la estructura de esta para valores de U desconocidos, lo cual es un resultado de gran relevancia porque nos permitiría reducir un gran esfuerzo computacional, ya que para obtener cada estado fundamental para cada valor del parámetro de Hubbard es necesario diagonalizar la matriz del hamiltoniano.

4. Estado excitado

En esta nueva sección, se propone el estudio del primer excitado del sistema. Sin embargo, se nos presenta un problema y es que este está degenerado. Como bien sabemos, una combinación lineal de estados con la misma energía es solución, pero como no conocemos la subrutina del proceso de diagonalización, no podemos saber si mezcla los dos estados excitados de manera aleatoria para diferentes valores de U , ya que entonces la red neuronal sería incapaz de aprender esta aleatoriedad. Con ello, para poder escoger un método de estudio de este, nos basaremos en la simetría de la función de onda para encontrar una combinación lineal simétrica de los dos.

4.1. Simetría de la función de onda

Como es sabido, la función de onda correspondiente a un estado del modelo de Hubbard debe ser simétrica en torno a traslaciones de la posición de los bosones en la red. Esto es debido a las condiciones periódicas de contorno, ya que los estados en los que los bosones han sido desplazados el mismo número de sitios serían equivalentes. Por ejemplo, en un caso más sencillo de 3 bosones en 3 sitios, los siguientes estados serían equivalentes:

$$|2, 1, 0\rangle \equiv |0, 2, 1\rangle \equiv |1, 0, 2\rangle \rightarrow |\psi_{|2,1,0}\rangle|^2 = |\psi_{|0,2,1}\rangle|^2 = |\psi_{|1,0,2}\rangle|^2. \quad (14)$$

De este modo, las amplitudes de la función de onda asociadas a estos estados deberían ser iguales, ya que el sistema debe tener la misma probabilidad de encontrarse en todos ellos, al ser indistintos.

De igual manera, puesto que las energías en cada sitio son iguales ($\{E_i\} = 1\forall i$), el sistema no tiene una dirección privilegiada, por lo que también serían equivalentes los estados simétricos. Volviendo al ejemplo anterior, esto sería:

$$|2, 1, 0\rangle \equiv |0, 1, 2\rangle \quad |0, 2, 1\rangle \equiv |1, 2, 0\rangle \quad |1, 0, 2\rangle \equiv |2, 0, 1\rangle. \quad (15)$$

Mencionado esto, puesto que necesitamos escoger un criterio de selección que nos permita elegir un estado excitado a partir de los dos para continuar con su estudio, vamos a considerar la combinación lineal simétrica de los estados respecto a la posición central del sistema.

Para ilustrar esta característica, se muestra un gráfico en la figura 12 donde se ordenan los coeficientes del estado fundamental de modo que los estados que son simétricos respecto al centro estén dispuestos a una distancia igual al eje de simetría. De este modo puede apreciarse como el estado fundamental cumple perfectamente la condición de simetría (figura 12).

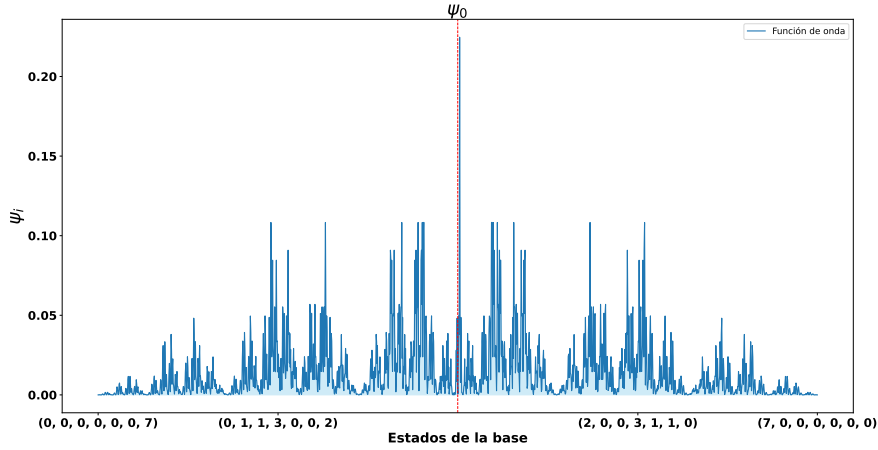


Figura 12: Simetría del estado fundamental.

Por otro lado, esto no se cumple para los estados excitados por separado al estar degenerados (figura 13), por lo que, según nuestro criterio de selección, buscamos encontrar la combinación lineal de estos que cumpla la condición de simetría y trabajar con este.

Mediante un algoritmo de ajuste obtenemos que los coeficientes que nos dan este resultado son:

$$\psi_{exc} = \frac{a}{\sqrt{a^2+b^2}}\psi_{exc}^1 + \frac{b}{\sqrt{a^2+b^2}}\psi_{exc}^2 \quad a \approx 0,3292 \quad b \approx 0,6707. \quad (16)$$

De este modo, obtenemos una combinación lineal simétrica de los estados excitados, lo cual queda reflejado en la figura 14.

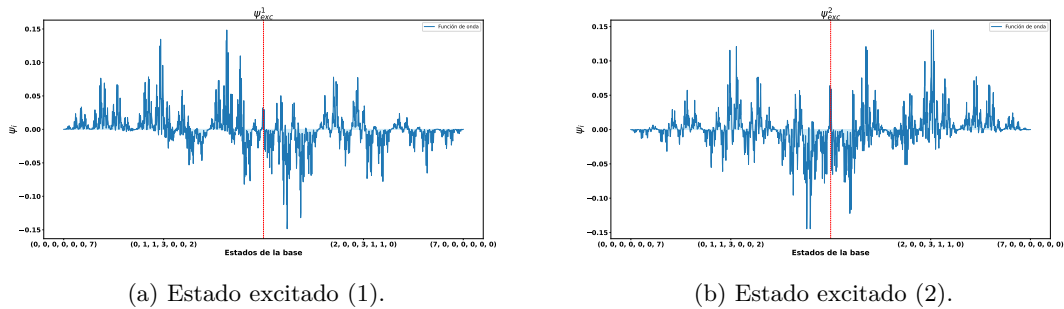


Figura 13: Simetría de los estados excitados

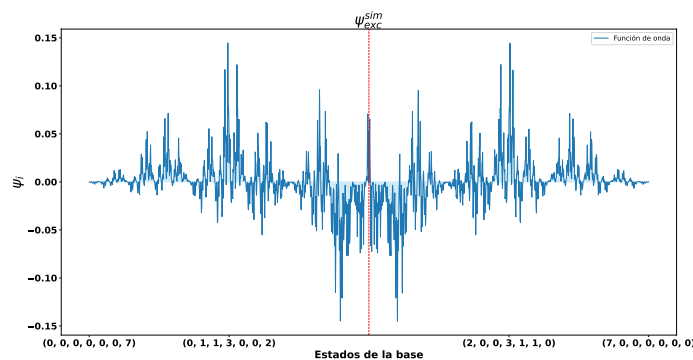


Figura 14: Simetría del estado excitado combinado.

Cabe destacar que se ha comprobado manualmente que los dos estados excitados cumplen la condición (14), así como también la combinación lineal simétrica tomada en (16) y su respectiva combinación antisimétrica asociada.

Al contrario que el estado fundamental, el estado excitado presenta coeficientes tanto positivos como negativos, por lo que la estructura es mucho más complicada. Además, no nos servirá la función de activación sigmoide a la salida como en el fundamental. Teniendo en cuenta que ahora los valores se dan en el rango $[-1, 1]$, la opción que vamos a utilizar como función de activación en la salida es una tangente hiperbólica.

4.2. Red para la predicción del primer excitado

La cuestión que pretendemos abordar a continuación es si es posible que una red neuronal aprenda a predecir el primer excitado del modelo de Hubbard. Para ello, seguimos el mismo procedimiento que hicimos con el estado fundamental, por lo que nos vamos a limitar a mostrar los mejores resultados obtenidos y no todo el proceso de optimización de los hiperparámetros.

De igual modo que antes, tomamos el 80% de los datos como entrenamiento y el 20% restante como validación, el tamaño del mini-batch sigue siendo 10, la tasa de aprendizaje la tomamos en 0.08 y el número máximo de épocas en 10000.

Respecto a la estructura de las capas ocultas, la configuración más óptima se ha encontrado con 3 capas. La primera de ellas con función de activación “*ReLU*”, la segunda con “*tanh*”, y la tercera con “*softsign*”. La fidelidad máxima obtenida fue $f = 0,9874$, con una estructura de red: 7(input)-75(relu)-15(tanh)-5(softsign)-1(tanh). Sin embargo, una configuración 7(input)-55(relu)-10(tanh)-3(softsign)-1(tanh) nos permitía obtener una fidelidad similar ($f \approx 0,9832$) y reducir el número de parámetros entrenables hasta los 1038, por lo cual, podemos concluir que una red neuronal es capaz de aprender el primer estado excitado del modelo de Hubbard con un número de parámetros entrenables relativamente bajo. La función de coste asociada a este proceso se muestra en la figura 15.

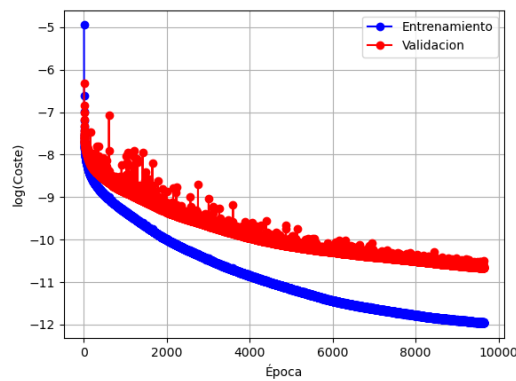


Figura 15: Función de coste de la red entrenada para el estado excitado

4.3. Predicciones para diferente número de bosones

De igual modo que hicimos con el estado fundamental, nuestra finalidad es observar si una red es capaz de extrapolar los datos utilizados durante su entrenamiento para predecir estados con un número de partículas diferente. Los resultados obtenidos en este caso se recogen en la figura 16.

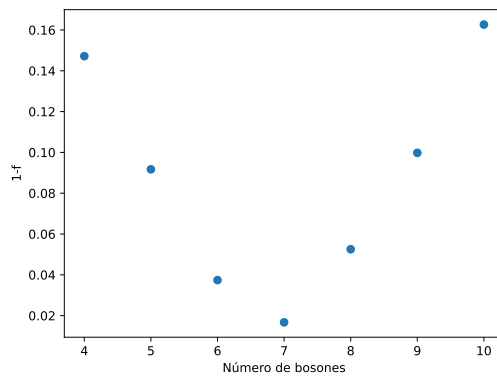


Figura 16: Predicción de la función de onda del estado excitado para diferente número de bosones con la red entrenada para $N = 7$.

Como puede observarse, las predicciones son ciertamente buenas para valores de N más próximos, tanto para menor como para mayor, siendo la fidelidad decreciente cuanto más se aleja el número de bosones respecto al valor con el que se ha entrenado la red ($N = 7$). Si comparamos estos resultados respecto a los obtenidos para las predicciones del estado fundamental, estos son peores, lo cual es comprensible dado que la función de onda del estado excitado es mucho más compleja que la del fundamental. Sin embargo, puesto que la fidelidad de las predicciones es cercana al 90% en los valores de N con distancia 2 respecto al centro, podemos considerar que nuestra red es capaz de predecir el estado excitado fielmente aunque en un rango más limitado. De todos modos, esto es un buen resultado ya que nos ahorra muchos procesos de trabajo, ya que para obtener el estado excitado, a parte de diagonalizar la matriz del hamiltoniano, requiere también encontrar la combinación lineal simétrica de los estados degenerados.

5. Transfer-Learning

5.1. Objetivos y procedimiento

Los algoritmos comunes de aprendizaje automático tradicionalmente abordan tareas aisladas. El aprendizaje por transferencia (*transfer-learning*) intenta cambiar esto desarrollando métodos para transferir el conocimiento aprendido en una o más tareas fuente y usarlo para mejorar el aprendizaje en una tarea objetivo relacionada [9].

En nuestro caso, consiste en la introducción de capas ya entrenadas en un modelo para tratar de conseguir extrapolar resultados nuevos entrenando un número de parámetros menor. De este modo, si la introducción de la primera capa entrenada para nuestro estado fundamental seguida de un par de capas más pequeñas es capaz de predecir la configuración de la función de onda del estado excitado con un número relativamente pequeño de parámetros, podríamos considerar el hecho como que el estado fundamental contiene información que ayuda a la red a aprender este nuevo estado.

Para ello, el procedimiento comienza con el entrenamiento de una red neuronal para el estado fundamental dado un valor del parámetro de Hubbard. La configuración utilizada se muestra esquemáticamente en la figura 17.

Una vez tengamos la red entrenada para nuestro estado fundamental, aplicaremos el *transfer-learning*. Para ello, tomaremos la primera capa de esta de modo que no se entrenará, es decir, su configuración de pesos se congela. A continuación, anexionaremos un par de pequeñas capas ocultas las cuales sí serán entrenables, dando como resultado una red con un número de parámetros entrenables mucho menor pero que aprovecha la configuración del fundamental, figura 18.

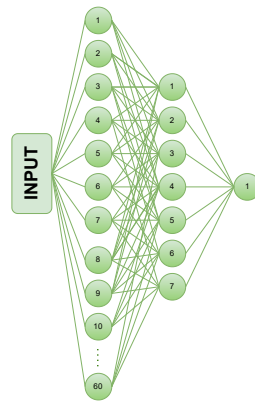


Figura 17: Imagen esquemática de la red entrenada para el estado fundamental dado un valor de U . La arquitectura consta de dos capas ocultas 60(ReLU)-7(ReLU). El tamaño del mini-batch es 10, la tasa de aprendizaje es de 0.1 y el valor máximo de épocas 1500. Esta configuración resulta en un número de parámetros entrenables de 915.

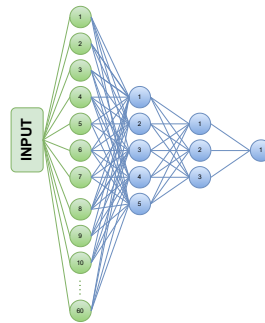


Figura 18: Imagen esquemática de la red de *transfer-learning*. Los parámetros entrenados para el estado fundamental los cuales no se van a modificar se representan en verde. Por el contrario, los parámetros entrenables para el estado excitado se representan en azul. Estructura de capas ocultas de la red: 60(no entrenable)-5(tanh)-3(ReLU). Tamaño del mini-batch de 10, el valor de la tasa de aprendizaje es de 0.1 y el número máximo de épocas 5000. Esta configuración da como resultado un número de parámetros entrenables de 327.

5.2. Resultados del método. Relación con la ratio de proporción (PR)

Teniendo todo lo anterior en mente, obtendremos la fidelidad del estado fundamental con la primera red y la del excitado con la segunda aplicando el transfer-learning. A modo de ilustrar la competencia de este método, vamos a obtener también las fidelidades del excitado sin aplicarlo, utilizando la red de la sección 4.2. Sin embargo, cuando hacemos esto nos damos cuenta de que las fidelidades siguen una tendencia conforme al parámetro Hubbard, siendo estas menores para valores muy negativos y aumentando rápidamente después. Además de que sus valores son bajos, también se puede apreciar mucho ruido en esta zona. Esto es debido a que para valores muy negativos del parámetro Hubbard la mayoría de las componentes de la función de onda son nulas, reduciendo el número de datos útiles del conjunto de entrenamiento y empeorando por consecuencia los resultados.

A modo de justificar este hecho, se ha relacionado estos resultados con el parámetro *participation ratio* (PR) [10], que da cuenta del número de estados de la base del espacio de Hilbert que tienen peso en la función de onda. Este parámetro se obtiene del modelo de Anderson, el cual se explica detalladamente en la sección 4 del anexo, y se define del modo siguiente:

$$PR = \frac{\sum_i |\psi_i|^4}{\sum_i |\psi_i|^2}. \quad (17)$$

Con esto, si la mayoría de los coeficientes de la función de onda son nulos, el inverso de la ratio de participación será más bajo. Por el contrario, si los coeficientes son equiprobables, el inverso alcanzará un pico. En consecuencia, si mostramos las fidelidades junto a este parámetro, deberíamos apreciar como estas siguen la misma tendencia. Este efecto se muestra en la figura 19.

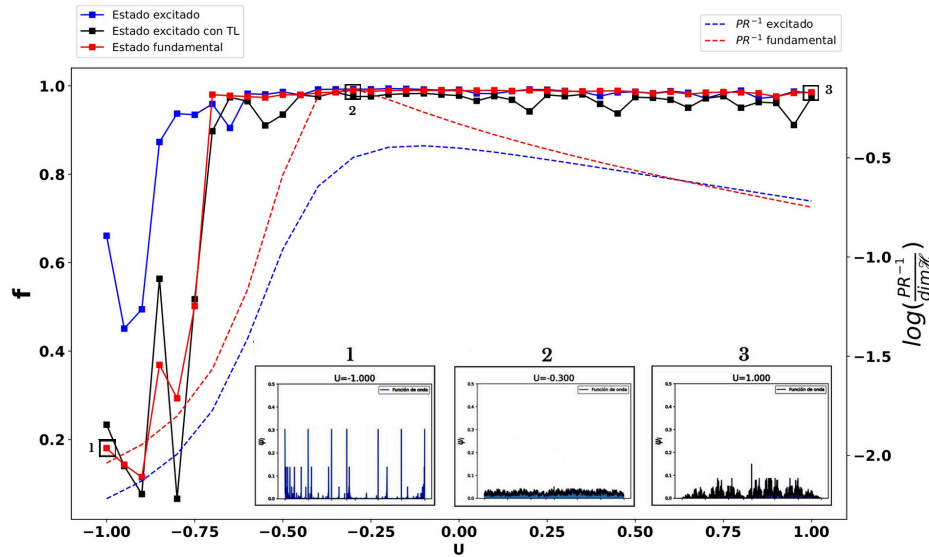


Figura 19: Ilustración de las fidelidades de los estados fundamental y excitado (con *transfer-learning*) junto a la tendencia de la ratio de participación. El parámetro PR ha requerido de un re-escalado el cual se muestra en el eje derecho, tomando el logaritmo del cociente entre el inverso de este y la dimensión del espacio de Hilbert. También se incluyen imágenes de la función de onda durante la transición para ilustrar el efecto.

Con la observación de esta gráfica podemos destacar múltiples cosas. La primera de ellas, es que se muestra evidencia del funcionamiento del método de *transfer-learning* ya que nos ha permitido, en la gran mayoría de los casos, obtener una fidelidad del estado similar utilizando menos parámetros, concretamente, un 64.26% menos. Además, se han monitorizado los tiempos de computación utilizando el *transfer-learning* y sin él para valores de $U = \{-1, 0, 1\}$ y en todos los casos ha resultado en una disminución, siendo del 23%, 65% y 85% respectivamente.

Por otra parte, también nos permite afirmar que la tendencia en las fidelidades es causada por la ratio de participación de la función de onda en su respectivo valor del parámetro de Hubbard, así como este también explicaría la zona de mayor ruido en $U < -0,75$, ya que la fidelidad dependería fuertemente de la aleatoriedad del algoritmo a la hora de escoger el conjunto de entrenamiento.

Finalmente, podemos apreciar la aparición de ciertas caídas de fidelidad con el *transfer-learning* en puntos como el $U = -0,55$, $U = 0,20$ y en adelante. Esto nos ha sugerido que quizás pueda ser un indicativo de algún proceso físico, en vez de simplemente ruido. El proceso físico en el que hemos pensado es que puede estar dándose una transición de fase cuántica, de modo que el transfer-learning pueda servir como indicador de este hecho.

5.3. Transición de fase cuántica

Puesto que, por norma general, en modelos de interacción entre partículas como el de Hubbard, el primer excitado presenta una configuración similar al fundamental pero con un cambio mínimo, podemos tratar el *transfer-learning* ya no solo como un procedimiento para aumentar la eficiencia del entrenamiento de la red sino como un método para poder determinar cómo de diferente es el estado excitado del fundamental, llevándonos a la idea de que podría resultar útil para la detección de transiciones de fase cuánticas, donde la correlación entre el excitado y el fundamental podría romperse.

A modo de averiguar si las caídas en la fidelidad realmente indican algo significativo sobre la configuración de los estados o son simplemente ruido del proceso de aprendizaje, hemos considerado primeramente la opción de la toma de estadística sobre estos.

Para ello, se han tomado cinco medidas de los puntos críticos y se han sustituido los puntos por su valor medio, incluyendo además su desviación. Los resultados obtenidos se ilustran en la figura 20.

Como puede observarse, al representar la media todos siguen manteniendo la forma, eso sí, con un error muy alto, lo cual sugiere que únicamente es ruido del aprendizaje y no representan ninguna característica propia del sistema.

A pesar de ello, si nos fijamos en la caída de $U = 0,45$, vemos como el tercer punto tiene un error muy bajo, en contraposición con el próximo. A la hora de plantearnos el porqué, nos caben dos posibilidades. La primera de ellas es que, efectivamente, algo está pasando en el punto con el error amplio que interfiere en el aprendizaje que en el anterior no lo hace. La segunda, y la más probable, teniendo en cuenta el resto de caídas, es que simplemente sea una coincidencia estadística.

Para salir de dudas, hemos considerado la representación de los seis primeros autovalores del sistema en función del parámetro de interacción Hubbard. De este modo, si hubiera una transición de fase que afecte al primer excitado, la curva de este debería cruzarse con otra o, al menos, producirse un cambio consistente en la tendencia de este. La representación se muestra en la figura 21.

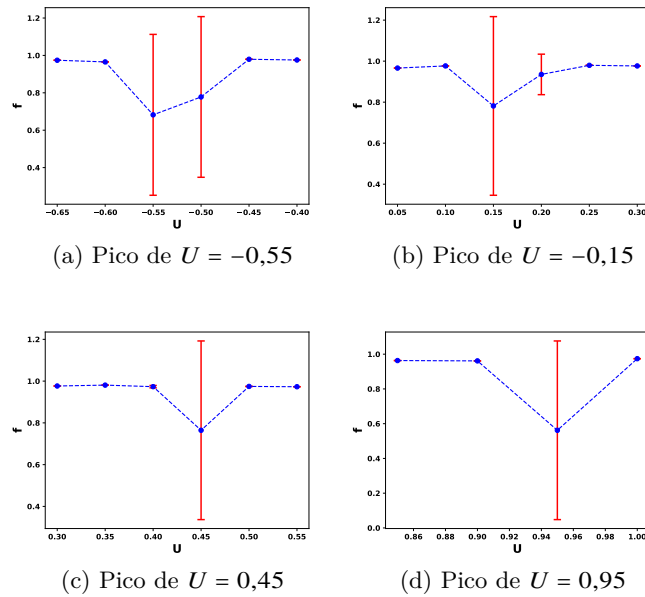


Figura 20: Análisis estadístico de los picos. Los puntos que no presentan barra de error se han incluido únicamente para una mejor visualización, no se ha tomado estadística sobre ellos.

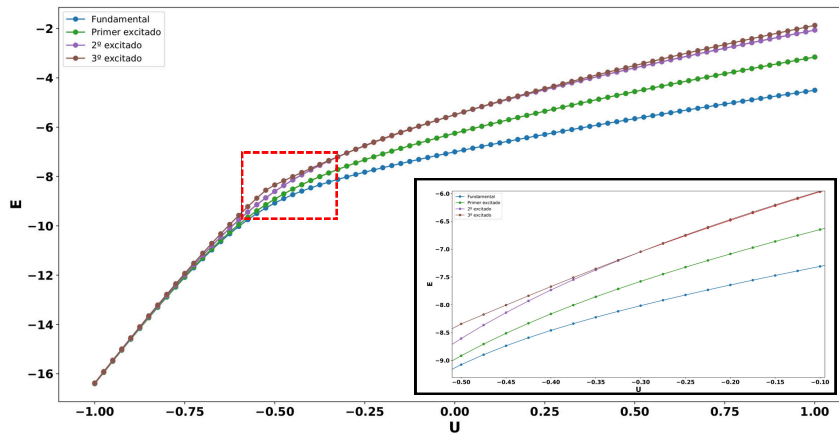


Figura 21: Representación de los 6 primeros autovalores del sistema en función del parámetro de Hubbard. Se incluye también un zoom a la transición entre el segundo y tercer excitado.

Podemos observar que no existe transición de fase del primer excitado del sistema. Sí que puede apreciarse una entre el segundo y el tercero, pero en nuestro caso no nos afecta para nada. Por lo tanto, esto confirma nuestras sospechas anteriores de que las caídas de fidelidad no eran más que ruido. De todos modos, a pesar de que nuestra hipótesis no ha resultado, plantearnos esta cuestión nos ha derivado hacia una comprobación de qué transiciones se dan en este rango de parámetros, dejando la puerta abierta para el estudio de si esta técnica sería útil en otros casos donde se den.

6. Conclusiones

La realización de este trabajo ha resultado lucrativa en muchos aspectos.

Primeramente, nos ha ofrecido una visión introductoria pero suficiente para comprender el funcionamiento de las redes neuronales, tanto la arquitectura como los mecanismos y procesos que estas llevan a cabo para realizar el aprendizaje.

En segundo lugar, nos ha acercado un poco al modelo cuántico de Hubbard que, como ya se ha mencionado en la introducción, si bien es un modelo bastante sencillo para entender el efecto de interacciones en sólidos, su estudio es una pieza clave en física teórica.

Sobre la intersección de los dos anteriores es donde se ha formado el núcleo de nuestro trabajo, dando como resultado una fusión beneficiosa entre ambos de acuerdo a los resultados mostrados en cada sección.

En la primera de ellas hemos comprobado que una red neuronal es capaz de aprender la función de onda del estado fundamental y realizar predicciones de esta para otro número diferente de partículas, así como también es capaz de reconocer una transición en función del parámetro de Hubbard.

En la segunda, hemos hecho lo propio pero con el primer excitado, solventando el problema que este presentaba al estar degenerado, escogiendo un criterio de selección para sus dos componentes.

Finalmente, en la última sección hemos implantado el método de *transfer-learning* con nuestros datos, el cual ha resultado en una optimización del proceso de aprendizaje. Además, hemos conseguido relacionar la fidelidad de nuestras predicciones con el parámetro PR, el cual proviene de un modelo físico (el modelo de Anderson, sección 4 del anexo) pero le hemos encontrado una utilidad dentro del *machine-learning*. El desarrollo del *transfer-learning* durante esta sección del trabajo nos ha llevado al estudio de la implementación de este en transiciones de fase cuánticas del modelo de Hubbard, lo cual, a pesar de no haber tenido éxito en este aspecto, consideramos que ha sido un buen acercamiento y planteárnoslo ha resultado en una mejor comprensión de estas, así como nos ha dejado la puerta abierta para el uso de esta implementación como objeto de estudio.

En resumen, haciendo un poco de retrospectiva a lo largo de todas las secciones, en todas ellas se ha conseguido el propósito principal con el que se inició este trabajo, que sería la compresión de información, es decir, mostrar como las redes neuronales nos iban a permitir reproducir con menos parámetros los estados del hamiltoniano de Hubbard, reduciendo así enormemente el esfuerzo computacional necesario para el tratamiento de sistemas grandes, como es nuestro caso. Sin embargo, a pesar de que el principal objetivo era este, a lo largo de su desarrollo se ha ido requiriendo un estudio más físico del modelo, especialmente a partir de la degeneración del nivel excitado, lo cual no ha apartado nuestra mirada directamente de las redes neuronales, sino que ha conseguido que prestemos una visión más conjunta del problema, resultando en un equilibrio entre física y *machine-learning* que consideramos perfecto.

Referencias

- [1] Fabian HL Essler, Holger Frahm, Frank Göhmann, Andreas Klümper, and Vladimir E Korepin. *The one-dimensional Hubbard model*. Cambridge University Press, 2005.
- [2] Neil W. Ashcroft and N. David Mermin. *Festkörperphysik*. Oldenbourg Wissenschaftsverlag, 2001.
- [3] Hal Tasaki. The hubbard model-an introduction and selected rigorous results. *Journal of Physics: Condensed Matter*, 10(20):4353, 1998.
- [4] Tetiana Zubatiuk and Olexandr Isayev. Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. *Accounts of Chemical Research*, 54(7):1575–1585, 2021.
- [5] N Elstner and H Monien. Dynamics and thermodynamics of the bose-hubbard model. *Physical Review B*, 59(19):12184, 1999.
- [6] Michael A Nielsen. *Neural networks and deep learning*, volume 25. Determination press San Francisco, CA, USA, 2015.
- [7] Xue Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019.
- [8] Hiroki Saito and Masaya Kato. Machine learning technique to find quantum many-body ground states of bosons on a lattice. *Journal of the Physical Society of Japan*, 87(1):014001, 2018.
- [9] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010.
- [10] Cécile Monthus and Thomas Garel. Anderson localization of phonons in dimension $d= 1, 2, 3$: Finite-size properties of the inverse participation ratios of eigenstates. *Physical Review B*, 81(22):224208, 2010.

7. Anexo

7.1. Condiciones de contorno

En el caso de una red fija, existen dos maneras diferentes de modular las interacciones entre elementos de esta:

- Condiciones de contorno no periódicas:

En este caso consideramos la red como “finita” de modo que los elementos que se encuentren a los bordes de esta solo pueden interactuar con sus vecinos interiores. A continuación se muestra una imagen que representa este tipo de interacción:



Figura 22: Red unidimensional con condiciones de contorno no periódicas

- Condiciones de contorno periódicas:

En este otro caso la red no tiene “límites” definidos de modo que todos los elementos de la red interactúan con el mismo número de vecinos. En el caso de una red unidimensional sería equivalente a decir que el último nodo está conectado con el primero, tal y como se muestra en la siguiente imagen:

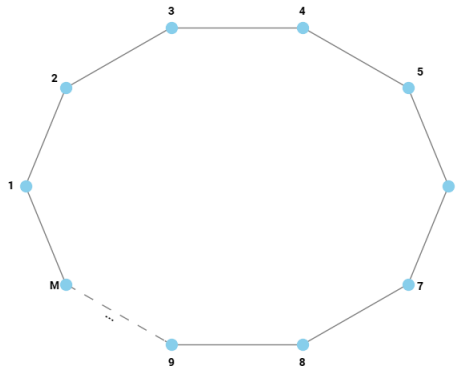


Figura 23: Red unidimensional con condiciones periódicas de contorno

7.2. Funciones de activación

En esta sección se recogen las funciones de activación utilizadas a lo largo del trabajo.

- Activación sigmoide:

La función de activación sigmoide se define como:

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (18)$$

dando como salida un intervalo continuo en el rango $[0, 1]$ (figura 24).

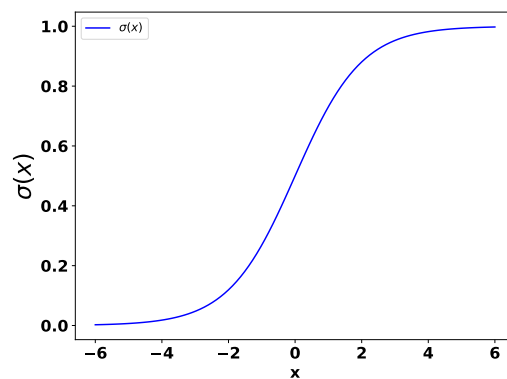


Figura 24: Activación sigmoide

- “ReLU”:

La función de activación “ReLU” se define como:

$$ReLU(x) = \max(0, x), \quad (19)$$

dando como resultado un rango continuo entre $[0, +\infty)$ (figura 25).

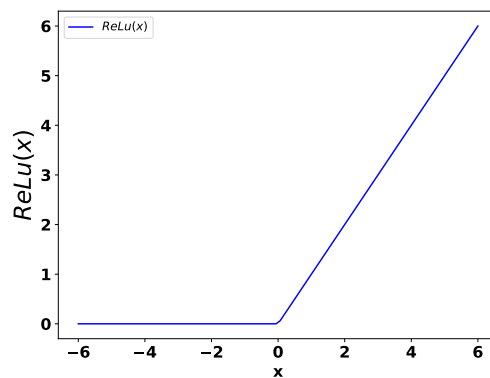


Figura 25: Función ReLU(x)

- Tangente hiperbólica:

La función de activación tangente hiperbólica sigue la expresión:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (20)$$

dando como salida un valor continuo en el rango $[-1, 1]$ (figura 26).

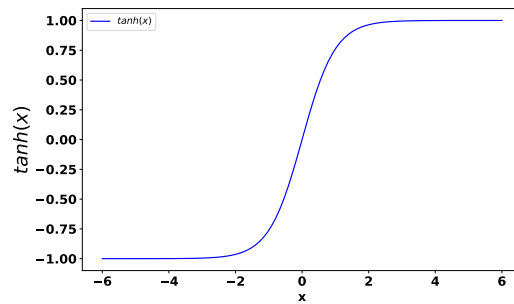


Figura 26: Función tangente hiperbólica

- *Softsign*:

La función de activación *softsign* sigue la expresión:

$$\text{Softsign}(x) = \frac{x}{1 + |x|}, \quad (21)$$

dando como salida un valor continuo en el rango $[-1, 1]$ (figura 27).

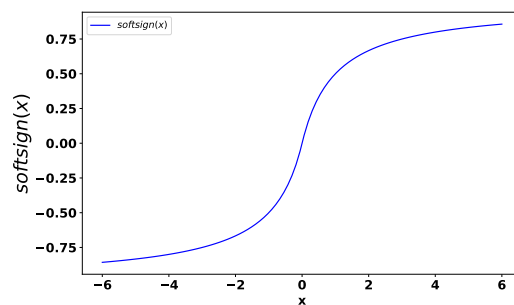


Figura 27: Función “Softsign”

7.3. Back Propagation

Como se ha mencionado anteriormente, la técnica de *Back Propagation* consiste en el cómputo de derivadas de la función coste respecto a los parámetros de la red de manera recursiva, desde las capas de salida hacia las de entrada. Para ilustrar esta técnica, tomamos un esquema de red como el de la figura 28.

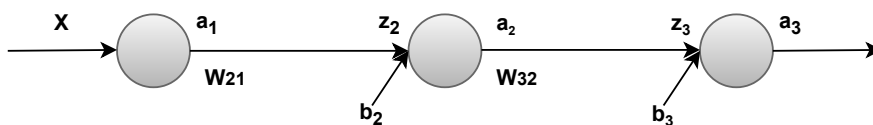


Figura 28: Imagen esquemática de una red neuronal lineal. 'X' se corresponde con el valor de input, los $\{a_i\}$ se corresponden con la activación de la neurona i -ésima, los $\{W_{i,j}\}$ el peso correspondiente a la unión entre la neurona j -ésima y la i -ésima, $\{b_i\}$ se corresponden con los *bias*, y los $\{z_i\}$ se definen como: $z_i = W_{i,i-1}a_{i-1} + b_i$.

Teniendo en cuenta el esquema de la figura anterior, podremos escribir las ecuaciones de activación de las neuronas:

$$\begin{aligned}
 z_2 &= W_{21}a_1 + b_2 \\
 a_2 &= \sigma(z_2) \\
 z_3 &= W_{32}a_2 + b_3 . \\
 a_3 &= \sigma(z_3) \\
 c &= C(a_3)
 \end{aligned}
 \tag{22}$$

De modo que si queremos hallar las derivadas de la función coste respecto a las variables W_{21} y b_2 :

$$\begin{aligned}
 \frac{\partial C}{\partial W_{21}} &= \frac{\partial C}{\partial z_2} \frac{\partial z_2}{\partial W_{21}} = a_1 \frac{\partial C}{\partial z_2} \\
 \frac{\partial C}{\partial b_2} &= \frac{\partial C}{\partial z_2} \frac{\partial z_2}{\partial b_2} = \frac{\partial C}{\partial z_2}
 \end{aligned}
 \tag{23}$$

podemos comenzar partiendo de las derivadas respecto a los parámetros de más a la derecha y las vamos usando para calcular las siguientes.

7.4. El modelo de Anderson

El modelo de Anderson es un modelo cuántico usualmente utilizado para describir sistemas de fermiones pesados. Este hamiltoniano produce el efecto conocido como “localización de Anderson”, también conocido como localización fuerte, que consiste en la ausencia de difusión de ondas en un medio desordenado, de modo que un electrón podría localizarse en el interior de un semiconductor si el grado de aleatoriedad de las impurezas o defectos es suficientemente grande. De este modo, este modelo puede utilizarse en la descripción del movimiento electrónico en semiconductores dopados.

En comparación con el modelo de Hubbard en una dimensión, el hamiltoniano sería equivalente a este pero con un único fermión, y donde los niveles de energía $\{E_i\}$ son aleatorios. A modo de conseguir simular este “medio aleatorio”, por donde el electrón debe moverse:

$$H_{Anderson} = -t \sum_{\langle i,j \rangle} (c_i^\dagger c_j + c_j^\dagger c_i) + \sum_i E_i \hat{n}_i \quad (U = 0), \quad (24)$$

y el espacio de Hilbert asociado:

$$\mathcal{H}_{Anderson} = \left\{ |n_1, n_2, \dots, n_M\rangle \mid n_i = 0, 1 \wedge \sum_i n_i = 1 \right\}, \quad (25)$$

que en este caso adoptará una dimensión del tamaño de la red (M).

Para sistemas donde el fermión se mueve de manera aleatoria, podemos considerar que todos los estados son igualmente probables, de modo que los coeficientes de la función de onda seguirían la relación:

$$\psi_i \sim \frac{1}{\sqrt{M}}, \quad (26)$$

con lo que la ratio de participación quedaría de la forma siguiente:

$$PR = \frac{\sum_i |\psi_i|^4}{\sum_i |\psi_i|^2} \sim \frac{M}{M^2} = \frac{1}{M} \rightarrow PR^{-1} \sim M, \quad (27)$$

por lo que el inverso del parámetro tendría una dependencia lineal con el número de sitios de la red. Sin embargo, si el fermión está fuertemente localizado, podríamos considerar que la amplitud de la función de onda está concentrada únicamente en un sitio, por ejemplo, el primero:

$$\psi_i = \delta_{i,0}, \quad (28)$$

de modo que en esta ocasión la ratio de participación quedaría:

$$PR = \frac{\sum_i |\psi_i|^4}{\sum_i |\psi_i|^2} = \frac{\sum_i |\delta_{i,0}|^4}{\sum_i |\delta_{i,0}|^2} = 1. \quad (29)$$

Teniendo en cuenta esto, la ratio de participación nos ofrece una medida de la localización de un sistema cuántico, de modo que cuanto mayor sea el inverso de este, nos indica que la amplitud de la función de onda está repartida de manera más equitativa entre todos los estados de la base del espacio de Hilbert.