

# Sintonización de parámetros en una cadena de puntos cuánticos mediante aprendizaje automático

por

**Emanuel Vazquez**

Presentado ante la Facultad de Matemática, Astronomía, Física y Computación  
como parte de los requerimientos para la obtención del grado de

**Licenciado en Física**

de la

Universidad Nacional de Córdoba



Septiembre 2023

©FAMAF – UNC 2023

**Directores:** Dr. Omar Osenda, Dr. Hernán L. Calvo



Sintonización de parámetros en una cadena de puntos cuánticos mediante aprendizaje automático, por Emanuel Vazquez, se distribuye bajo una Licencia Creative Commons Attribution-NonCommercial-ShareAlike 4.0 Internacional.



# Resumen

Debido al avance de la computación cuántica, la cual se construye en base al Qubit, una superposición de dos estados (1's y 0's), los puntos cuánticos ganaron terreno como una manera de construir estos sistemas físicos.

Más concretamente, un punto cuántico es un sistema estructurado artificialmente que se puede llenar con electrones (o huecos) y además puede acoplarse con barreras de túnel a reservorios macroscópicos, con los que se puede intercambiar electrones.

Si bien la física que rige estos mecanismos de transporte de electrones es sencilla para los casos más simples, su complejidad aumenta de manera considerable para sistemas donde el número de puntos cuánticos se incrementa. Esto se debe principalmente a los muchos parámetros de los que depende el sistema.

Esto llevó a que en los últimos años se explore el sintonizado de estos sistemas utilizando aprendizaje automático. Si bien en la mayoría de estos esfuerzos se centraron en utilizar los algoritmos para clasificar los diferentes estados del sistema, en este trabajo se busca hacer una regresión y así lograr la predicción de valores para los parámetros.

En este trabajo se busca sintonizar el peso de los voltajes en los niveles de energía utilizando redes neuronales convolucionales. Para esto se creó una base de datos simulando diferentes condiciones la cual se utilizó en el entrenamiento del algoritmo. La misma se construyó generando mapas de corriente (vector de entrada) utilizando diferentes pesos en los voltajes (vector de salida). Para lograr la simulación de corrientes se utilizó la teoría de transporte en tiempo real, reformulada en término de los super operadores de Liouville en la cual se evalúa una expansión perturbativa en el acople entre el sistema local y los reservorios anexados. Si bien a lo largo del trabajo se trabaja con un punto cuántico doble en serie, queda el camino desarrollado para plantear cualquier tipo de sistema que se desee resolver.

Finalmente se exploró el comportamiento de los resultados agregando ruido a las corrientes.

**Palabras Clave:** Puntos Cuánticos, Teoría de transporte, Super operadores de Liouville, Aprendizaje Automático, Parametrización



# Abstract

Due to the advance of quantum computing, which is built based on the Qubit, a superposition of two states (1's and 0's), quantum dots gained ground as a way to build these physical systems.

More specifically, a quantum dot is an artificially structured system that can be filled with electrons (or holes) and can also be coupled with tunnel barriers to macroscopic reservoirs, with which electrons can be exchanged.

Although the physics that governs these electron transport mechanisms is simple for the simplest cases, its complexity increases considerably for systems where the number of quantum dots increases. This is mainly due to the many parameters that the system depends on.

This has led to the tuning of these systems using machine learning being explored in recent years. Although most of these efforts focused on using algorithms to classify the different states of the system, this work seeks to do a regression and thus achieve the prediction of values for the parameters.

This work seeks to tune the weight of the voltages in the energy levels using convolutional neural networks. For this, a database was created simulating different conditions which was used in training the algorithm. It was built by generating current maps (input vector) using different weights on the voltages (output vector). To achieve the simulation of currents, real-time transport theory was used, reformulated in terms of the Liouville super operators in which a perturbative expansion is evaluated in the coupling between the local system and the attached reservoirs. Although throughout the work we work with a double quantum dot in series, the path remains developed to propose any type of system that we wish to solve.

Finally, the behavior of the results was explored by adding noise to the currents.

**Keywords:** Quantum Dots, Transport Theory, Liouville Super Operators, Machine Learning, Parametrization



*Para Arnaldo, mi abuelo, ojalá que desde algún lado puedas leer este trabajo.  
Te quiero y te extraño todos los días.*

# Índice general

<b>1. Introducción</b>	<b>6</b>
<b>2. Sistemas de puntos cuánticos</b>	<b>8</b>
2.1. Puntos cuánticos	8
2.2. Bloqueo de Coulomb	9
2.2.1. Diamantes de Coulomb	11
2.3. Bloqueo de Pauli	12
2.4. Puntos cuánticos dobles	12
<b>3. Teoría de transporte</b>	<b>14</b>
3.1. Introducción	14
3.2. Modelo Hamiltoniano	14
3.3. Ecuación maestra generalizada	15
3.3.1. Evolución del operador densidad reducido	16
3.4. Punto cuántico doble	18
3.5. Cálculo del kernel	20
3.6. Ocupaciones y corriente de carga en un DQD	21
<b>4. Algoritmos de aprendizaje automático</b>	<b>25</b>
4.1. Redes Neuronales	25
4.2. Redes Neuronales Convolucionales (RNC)	30
<b>5. Análisis del modelo de puntos cuánticos utilizando aprendizaje automático</b>	<b>32</b>
5.1. Modificación del modelo	32
5.2. Base de datos	33
5.2.1. Algoritmo para la creación del dataset	34
5.2.2. Ejecución del algoritmo	34
5.3. Entrenamiento del modelo	35
5.3.1. Primer modelo (CNN1)	36
5.3.2. Segundo modelo (CNN2)	37
5.4. Ruido en las mediciones	39
5.4.1. Generación de ruido	39
5.4.2. Importación del ruido	39
<b>6. Conclusiones</b>	<b>41</b>
<b>A. Desacople entre ocupaciones y coherencias en el régimen de acoplamiento fuerte</b>	<b>43</b>
<b>Bibliografía</b>	<b>44</b>

# Capítulo 1

## Introducción

En la actualidad, la computación cuántica ha adquirido una gran relevancia, al punto que muchos esfuerzos por parte de la comunidad están dirigidos a la construcción del primer ordenador basado en este tipo de principios. La unidad de información en este tipo de dispositivos son los qubits, que constituyen una versión cuántica de los ya conocidos bits clásicos conformados por 0's y 1's. Por lo tanto, parte de esa atención se redirige hacia el desarrollo de sistemas físicos apropiados donde se pueda crear y controlar un sistema de qubits. Los puntos cuánticos (QDs), que son estructuras cero dimensionales de gran confinamiento electrónico, han ganado popularidad como candidatas a bloques de construcción para dispositivos cuánticos del estado sólido. Si bien la física que rige estos mecanismos de transporte de electrones es sencilla para los casos más simples, su complejidad aumenta de manera considerable para sistemas donde el número de puntos cuánticos se incrementa. Esto se debe principalmente a los muchos parámetros de los que depende el sistema, como los voltajes de compuerta individuales, los cuales deben ser dispuestos en forma cuidadosa para lograr confinar un número determinado de electrones.

Más concretamente, un punto cuántico es un sistema estructurado artificialmente que se puede llenar con electrones (o huecos) y además puede acoplarse con barreras túnel a reservorios macroscópicos, con los que se puede intercambiar electrones. Debido a las dimensiones y disposiciones del mismo y si se le suma a ellos la electrostática básica de los gases de electrones de dimensionalidad reducida (2DEG), es posible explicar los fenómenos básicos que tienen lugar en este tipo de estructuras. El funcionamiento de los QDs se basa en aplicar diferencias de potencial entre los reservorios (o contactos) que les rodean además de controlar los voltajes de compuerta internos que permiten modificar sus niveles de energía. Utilizando conceptos tales como el bloqueo de Pauli y el bloqueo de Coulomb, se puede explicar la aparición de corriente carga para ciertos valores en los parámetros del sistema. Todo esto, junto a la extensión de un sistema de punto cuántico doble y la descripción de fenómenos característicos de estos sistemas, tales como los diamantes de Coulomb, se desarrollan en el capítulo 2. Este capítulo está fuertemente basado en el libro "Semiconductor Nanostructures: Quantum States and Electronic Transport", de Thomas Ihn [1].

Si bien estos conceptos permiten describir los sistemas mencionados, por sí solos no son suficientes a la hora de trabajar con estructuras más generales. En la actualidad existe un gran número de formalismos que permiten abordar este tipo de problemas, tales como las aproximaciones de Landauer-Büttiker [2, 3] para el caso de transporte electrónico en el régimen balístico, o bien los cálculos de corriente a partir de funciones de Green de no-equilibrio, derivados por Meir, Wingreen y Jauho [4, 5]. El marco teórico que aquí empleamos, permite dar en forma relativamente sencilla una ecuación maestra para la resolución de estructuras complejas basadas en QDs dentro del régimen de bloqueo de Coulomb.

En una segunda parte del trabajo, más precisamente en el capítulo 3, se describe el marco teórico de la teoría de transporte en tiempo real, reformulada en términos de superoperadores de Liouville para evaluar una expansión perturbativa en el acople entre el sistema local (conformado por uno o más puntos cuánticos) y los reservorios. Esto permite llegar a un sistema efectivo de ecuaciones para las ocupaciones del sistema, cuyos coeficientes quedan expresados post-desarrollo. Estos coeficientes se definen en base a una sumatoria en donde contribuye cada valor de expectación de transporte entre los estados de muchos cuerpos propios del sistema. Otra de las relaciones obtenidas permite el cálculo de corriente de partículas, para el cual se necesitan las ocupaciones de manera explícita. El desarrollo teórico que permite llegar a las expresiones descritas en el capítulo 3 se basa principalmente en la tesis doctoral de M. Leijnse [6] donde, si el lector lo desea, se pueden

encontrar más detalles asociados a su derivación.

Este conjunto de ecuaciones queda plasmado de manera tal que invita a la resolución computacional de los mismos. En la sección 3.4 se utiliza un sistema conformado por dos puntos cuánticos para ilustrar el camino a seguir a la hora de plantear estos sistemas utilizando la teoría planteada en el capítulo anterior y luego resolviéndolos de manera numérica. Esto permitió comprender el funcionamiento de estos sistemas, en donde se pudo replicar situaciones conocidas tales como los diamantes de Coulomb y los diagramas de estabilidad de carga presentados en el capítulo 2.

En los últimos años, algoritmos de aprendizaje [machine learning (ML)] y, específicamente, redes neuronales convolucionales (CNNs), han surgido como un buen camino para la clasificación automática de imágenes. Trabajos en los cuales se usan estos algoritmos para clasificar configuraciones de carga en sistemas de puntos cuánticos han tenido su desarrollo en diferentes publicaciones [7, 8]. Sin embargo, estos algoritmos también pueden utilizarse para aproximar valores de parámetros a través de imágenes que bien podrían obtenerse experimentalmente. Aquí es donde entra el objetivo de este trabajo, el cual se basa en brindar un camino de resolución de sistemas basados en puntos cuánticos tomando como herramienta principal algoritmos de aprendizaje automático.

En el capítulo 4 se desarrolla la teoría matemática que rigen las redes neuronales y los algoritmos con los cuales se procede al entrenamiento de las mismas. Esta parte está basada fuertemente en las clases del curso de Redes Neuronales que se dicta como especialidad en la carrera de Licenciatura en Física. En caso de querer profundizar los conocimientos teóricos de esta sección, se pueden consultar los libros [9–11]. En este capítulo también se desarrolla la extensión a las redes neuronales convolucionales ya que la estructura geométrica de estas permiten mantener la relación espacial entre los datos de entrada, característica de interés al utilizar imágenes o series temporales en estos sistemas.

En vistas a posibles escenarios experimentales propios de estos sistemas, este trabajo busca aplicar estas técnicas de aprendizaje para el sintonizado de parámetros de una cadena de puntos cuánticos. En particular, se utiliza el modelo desarrollado en la sección 3.4 de punto cuántico doble utilizando los pesos de los voltajes de compuerta sobre los niveles de energía del sistema como parámetros a sintonizar. El primer paso en este camino fue crear una base de datos en donde el vector de entrada consiste en un mapa de corriente y el vector de salida es la matriz de correlación  $D$ . Con esta base de datos se entrenan ciertos modelos de CNNs utilizando diferentes arquitecturas. En un último paso, se agrega ruido en la generación de corrientes con diferentes pesos y se estudia la influencia de los mismos en los entrenamientos de la red neuronal.

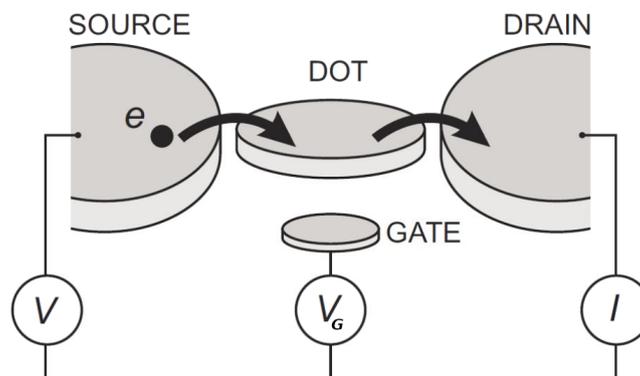
## Capítulo 2

# Sistemas de puntos cuánticos

En este capítulo describiremos los sistemas físicos que serán objeto de estudio a lo largo de este trabajo. En particular, nos enfocaremos en la fenomenología de los puntos cuánticos dentro del régimen de transporte conocido como “bloqueo de Coulomb”.

### 2.1. Puntos cuánticos

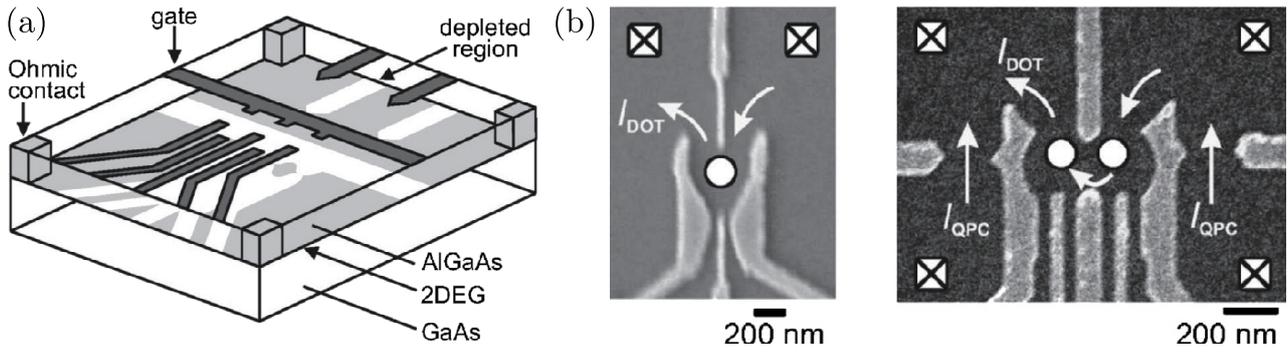
Un punto cuántico (QD) es un sistema estructurado artificialmente que se puede llenar con electrones (o huecos). El término “punto” hace referencia a que el confinamiento en estos sistemas es tan fuerte que, en esencia, se comporta como un objeto cero-dimensional, de modo que los portadores de carga solo pueden entrar o salir de dicha región del espacio, sin poder moverse libremente en su interior. Los sistemas de puntos cuánticos pueden acoplarse mediante barreras túnel a los reservorios, con quienes se intercambian electrones.



**Figura 2.1:** Representación esquemática lateral de un punto cuántico. En la figura se pueden observar los reservorios (source y drain), el punto cuántico(dot) y la compuerta (gate). Figura tomada de Ref. [12].

Como se muestra en la figura 2.1, al aplicar un voltaje “bias” entre los reservorios, es posible medir las propiedades de transporte del sistema examinando, por ejemplo, la corriente de carga que se establece entre ambos contactos. El QD también está acoplado a un electrodo de compuerta, que se utiliza para sintonizar el potencial electrostático del sistema con respecto a los reservorios.

Los puntos cuánticos más comunes para estudios experimentales se construyen a partir de heteroestructuras de GaAs y AlGaAs cultivados por disposición de haces moleculares. Dopando la capa de AlGaAs con Si se pueden introducir electrones libres, formando así un gas de electrones bidimensional debajo de la superficie. Con el voltaje de compuerta ( $V_g$ ) se puede sintonizar la densidad de electrones entre los reservorios.



**Figura 2.2:** Dispositivo de punto cuántico definido por electrodos de superficie metálica. a) Vista esquemática. (b) Micrografías electrónicas de barrido de un dispositivo de un solo punto (izquierda) de pocos electrones y un dispositivo de doble punto (derecha). Figura tomada de Ref. [12].

Las propiedades electrónicas de los puntos cuánticos están dominadas por dos efectos. En primer lugar, la repulsión de Coulomb entre los electrones, que conduce a un gasto de energía adicional cada vez que se desee agregar un electrón al QD. Debido a este hecho, el tunelamiento de electrones hacia o desde los contactos suele suprimirse a bajas temperaturas. Este fenómeno se denomina bloqueo de Coulomb y nos centraremos en él más adelante. El segundo efecto es que el confinamiento en las tres direcciones conduce a efectos cuánticos que influyen en la dinámica de los electrones, dando como resultado un espectro de energía discreto.

El modelo de interacción constante es comúnmente utilizado para describir las propiedades electrostáticas de los puntos cuánticos. Se basa en dos suposiciones: Primero, las interacciones entre los electrones del QD y los del entorno están parametrizadas por una capacitancia constante, que es la suma de las capacitancias entre el punto, la fuente, el drenaje y la compuerta, es decir,  $C = C_S + C_D + C_G$ . La segunda suposición es que el espectro de niveles de energía de una sola partícula es independiente de estas interacciones y, por lo tanto, del número de electrones. Bajo estas suposiciones, la energía total  $U(N)$  de un QD con  $N$  electrones en el estado fundamental, con voltajes  $V_S$ ,  $V_D$  y  $V_G$  aplicados a la fuente, drenaje y compuerta, respectivamente, viene dada por la ecuación:

$$U(N) = \frac{(-|e|(N - N_0) + C_S V_S + C_D V_D + C_G V_G)^2}{2C} + \sum_{n=1}^N E_n(B), \quad (2.1)$$

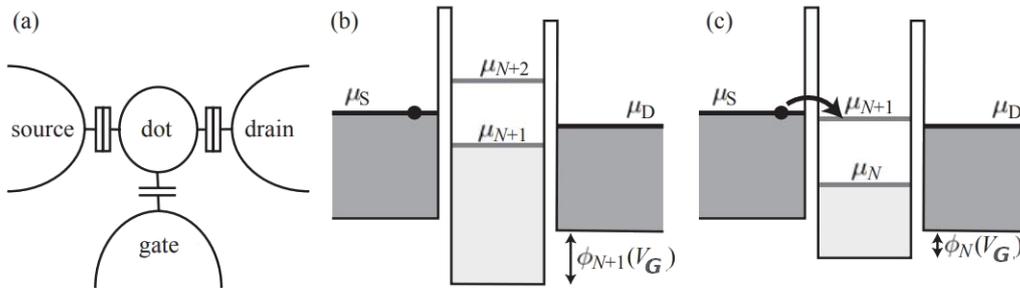
donde  $N_0|e|$  es la carga en el QD que compensa la carga de fondo positiva que se origina a partir de los donantes en la heteroestructura y  $B$  es el campo magnético aplicado. Los términos  $C_S V_S$ ,  $C_D V_D$  y  $C_G V_G$  se pueden modificar en forma continua y representan una carga inducida efectiva que cambia el potencial electrostático del QD. El último término de la ecuación es una suma de los niveles de energía de las partículas individuales ocupadas  $E_n(B)$  que dependen de las características del potencial de confinamiento. Entonces, el potencial electroquímico  $\mu(N)$  del punto cuántico se define como:

$$\mu_N = U(N) - U(N - 1) = \left(N - N_0 - \frac{1}{2}\right) E_C - \frac{E_C}{|e|} (C_S V_S + C_D V_D + C_G V_G) + E_N \quad (2.2)$$

Entonces, se puede ver como el potencial necesario para agregar el  $N$ -electrón depende de la capacitancia y voltajes, tanto de los reservorios como de la compuerta. Además el coste se vuelve lineal con  $N$  como se puede observar en el primer término. Por otro lado,  $E_N$  contiene las contribuciones de las interacciones electrón-electrón en el punto cuántico, además de la energía de confinamiento.

## 2.2. Bloqueo de Coulomb

El principal efecto que encontramos durante el estudio de los puntos cuánticos es el bloqueo de Coulomb. A nivel cualitativo, el efecto de bloqueo de Coulomb se puede describir de una manera muy intuitiva y general.



**Figura 2.3:** (a) Representación esquemática de un punto cuántico acoplado (vía tunneling) a los reservorios y capacitivamente a un voltaje de compuerta. (b) Estructura de niveles de energía del sistema en el bloqueo de Coulomb. (c) Posición de los niveles de energía que permite que la corriente fluya entre la fuente y el drenaje si se aplica un voltaje bias muy pequeño. Figura tomada de Ref. [1].

Las figuras 2.3(b) y (c) muestran la situación energética en los tres subsistemas: fuente, drenaje y QD. A baja temperatura, los niveles electrónicos en el contacto de la fuente (drenaje) se llenan desde la parte inferior de la banda de conducción hasta el potencial electroquímico  $\mu_S$  ( $\mu_D$ ). En el punto cuántico también podemos definir un potencial electroquímico. Este describe la energía necesaria para añadir un electrón al QD, dado que éste se encuentra tanto inicialmente como después de la adición en su estado fundamental. Por ejemplo, si consideramos inicialmente un punto cuántico con  $N - 1$  electrones, definimos el potencial electroquímico para agregar el  $N$ -ésimo electrón como:

$$\mu_N(V_G) = E_N^{(0)}(V_G) - E_{N-1}^{(0)}(V_G), \quad (2.3)$$

en donde  $E_N^{(0)}$  es la energía del estado fundamental del sistema para  $N$  electrones. El voltaje de compuerta  $V_G$  nos permite cambiar los niveles  $\mu_N(V_G)$  en energía. Encontramos para la dependencia de voltaje del potencial electroquímico en el punto cuántico la relación lineal:

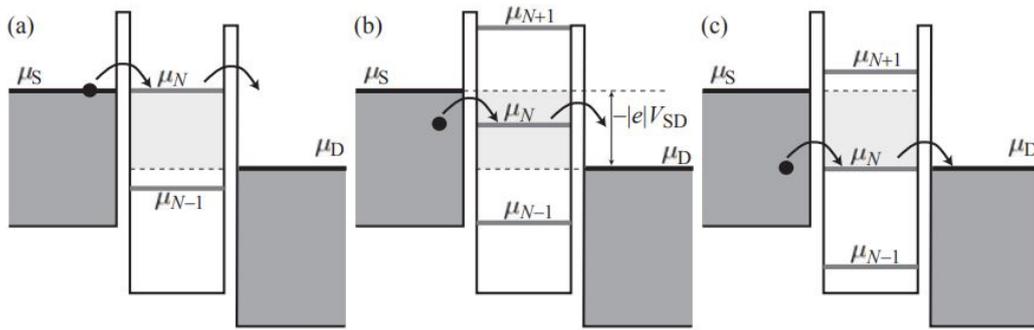
$$\mu_N(V_G) = \mu_N(V_G^{(0)}) - |e|\alpha_G \Delta V_G, \quad (2.4)$$

la cual es independiente del número de electrones  $N$ . La constante  $\alpha_G$  se llama “palanca” de la compuerta, y determina la respuesta del potencial  $\mu_N$  frente a cambios en  $V_G$ .

Controlando el voltaje de compuerta podemos sintonizar los potenciales electroquímicos del punto cuántico en la posición que se muestra en el lado derecho de la imagen de arriba (figura 2.3), donde tenemos  $\mu_S \simeq \mu_{N+1}(V_G) \simeq \mu_D$ . En este caso, la ganancia de energía  $\mu_S$  al quitar un electrón del contacto de la fuente es exactamente igual a  $\mu_{N+1}(V_G)$ , la energía requerida para agregar un electrón al punto. Una vez que el electrón está en el QD, la ganancia de energía  $\mu_{N+1}(V_G)$  para removerlo nuevamente es exactamente igual a la energía  $\mu_D$  requerida para agregarlo al contacto de drenaje. Por lo tanto, es posible obtener transporte elástico de electrones a través del punto cuántico y la medición muestra un pico de corriente en los respectivos voltajes de compuerta (pico de conductancia). Sin embargo, los electrones solo pueden tunear uno tras otro a través del QD, porque la diferencia de energía  $E_{N+2}^{(0)}(V_G) - E_N^{(0)}(V_G)$  para agregar dos electrones al mismo tiempo es significativamente mayor que la energía de un solo electrón. Por lo tanto, hablamos de tunelamiento secuencial de un solo electrón. La situación del bloqueo actual se muestra en la figura 2.3(b), así como en las figuras 2.4(a),(b) y (c) para los niveles  $\mu_{N+1}$  y  $\mu_{N-1}$ . A este voltaje de compuerta, el punto se llena con  $N + 1$  electrones. Para colocar el electrón  $N + 2$  se requiere más energía que la ganancia de energía al quitar un electrón del contacto de la fuente, es decir,  $\mu_{N+2}(V_G) > \mu_S, \mu_D$ . Por lo tanto, el flujo de corriente está bloqueado y hablamos de bloqueo de Coulomb. La separación  $\Delta V_G$  de resonancias de conductancia vecinas se encuentra dada como:

$$\Delta V_G = \frac{\mu_{N+1}(V_G^{(0)}) - \mu_N(V_G^{(0)})}{|e|\alpha_G}. \quad (2.5)$$

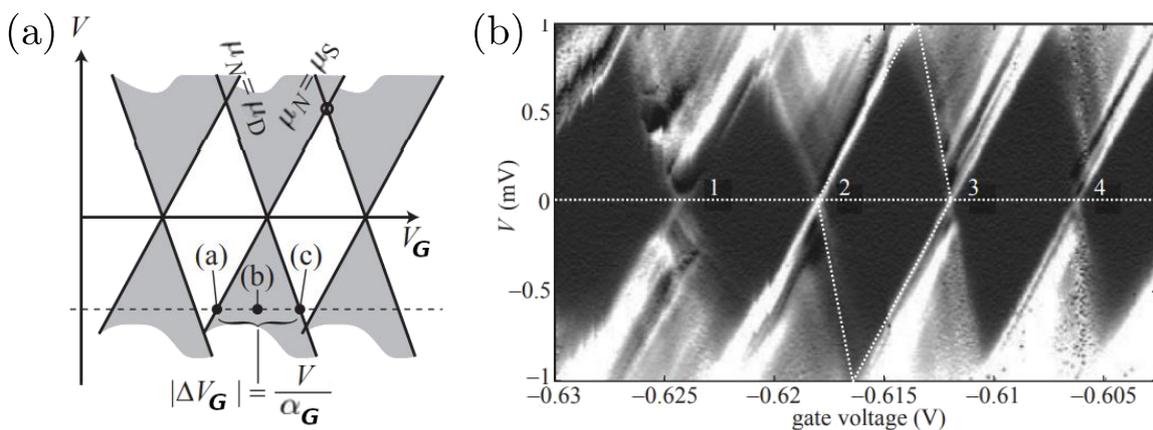
Para voltajes bias pequeños esta diferencia muestra, típicamente, la región de voltaje de compuerta donde el transporte se encuentra bloqueado. A su vez, el costo energético para agregar un electrón adicional al punto cuántico puede relacionarse, en muchos casos, con la interacción Coulombiana entre los portadores de carga dentro del sistema.



**Figura 2.4:** Representación esquemática de un sistema de puntos cuánticos con un voltaje de polarización fijo para varios voltajes de compuerta. La región de energía en gris claro representa la llamada ventana de polarización. Las flechas indican la transferencia de electrones. (a) Inicio de corriente para  $\mu_S = \mu_N(V_G)$ . (b) Situación con  $\mu_S > \mu_N(V_G) > \mu_D$  (región de flujo de corriente). (c) Inicio de corriente en  $\mu_D = \mu_N(V_G)$ . Figura tomada de Ref. [1].

### 2.2.1. Diamantes de Coulomb

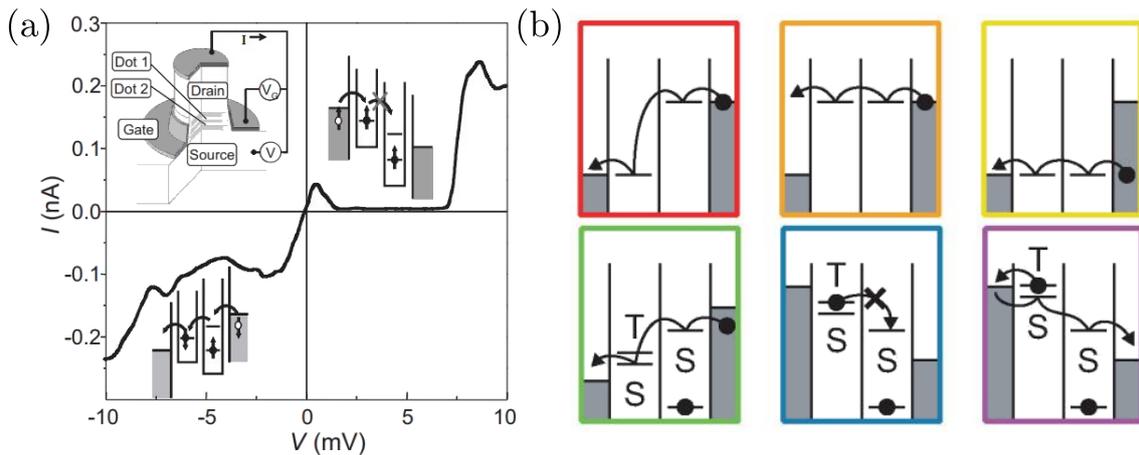
La medida de los diamantes de bloqueo de Coulomb, como se muestra en la figura 2.5, se puede entender de inmediato sobre la base de la terminología anterior y el hallazgo empírico expresado por la ecuación (2.4). Una medición de diamante corresponde a realizar muchos barridos de voltaje de compuerta para varios valores de voltaje bias. La aplicación de un  $V$  finito a un QD corresponde a la apertura de la llamada ventana de bias, que denota la región energética entre  $\mu_S$  y  $\mu_D$ , como se muestra en la región gris claro en la figura 2.4. El ancho de esta ventana de energía viene dado por  $\mu_S - \mu_D = -|e|V$ . Para  $V$  finito, una corriente puede fluir como resultado de la transferencia de electrones entre la fuente y el drenaje (ver las flechas en la figura 2.4), siempre que  $\mu_S \geq \mu_N(V_G) \geq \mu_D$ . Al aumentar el voltaje de compuerta, la corriente aumentará de cero a un valor finito cuando  $\mu_S = \mu_N(V_G^0)$  [figura 2.4(a)]. Entonces, la corriente permanecerá hasta que  $\mu_D = \mu_N(V_G^{(0)} + \Delta V_G) = \mu_N(V_G^{(0)}) - |e|\alpha_G \Delta V_G$ , donde vuelve a caer a cero [ver figura 2.4(c)]. El ancho en el voltaje de compuerta de la situación actual se obtiene de la diferencia de las dos ecuaciones anteriores para ser  $|\Delta V_G| = V/\alpha_G$ . En otras palabras, hay un rango finito de voltaje de compuerta que permite que fluya una corriente, cuyo tamaño aumenta proporcionalmente al voltaje de polarización. Esta ecuación define regiones triangulares que emergen de las resonancias en la conductancia que parten desde  $V = 0$  y se extienden hasta un valor finito, como se muestra en la figura 2.5. La constante de proporcionalidad es  $\alpha_G^{-1}$ , el cual se puede leer directamente a partir de las medidas del diamante de bloqueo de Coulomb.



**Figura 2.5:** (a) Diamantes de bloqueo de Coulomb esquemáticos. La corriente puede fluir en las regiones triangulares de color gris claro. Las líneas negras indican la alineación entre uno de los potenciales electroquímicos de los reservorios ( $\mu_S$  ó  $\mu_D$ ) con el potencial electroquímico  $\mu_N$  en el QD. Los puntos etiquetados (a)-(c) se refieren a las situaciones representadas esquemáticamente en la figura 2.4. El punto dentro de un círculo marca la situación en la que  $\mu_{N+1}(V) = \mu_S$  y  $\mu_N(V_G) = \mu_D$ . (b) Diamantes de bloqueo de Coulomb medidos en un punto cuántico. Figura tomada de Ref. [1].

### 2.3. Bloqueo de Pauli

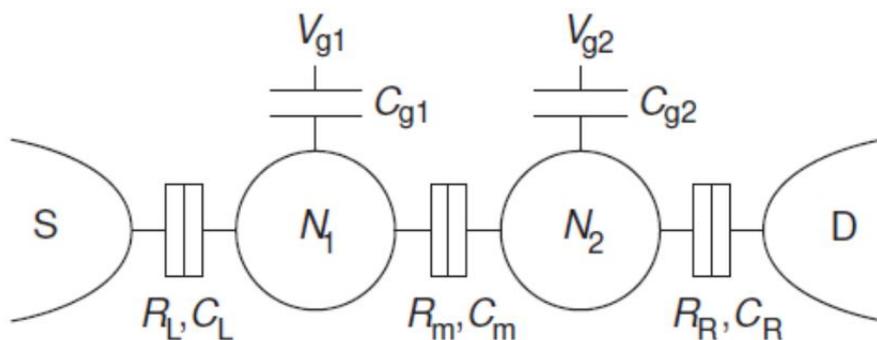
El bloqueo de Pauli es un efecto cuántico que encontramos en los sistemas de puntos cuánticos y es el resultado del principio de exclusión de Pauli. Para el caso de un punto cuántico doble con voltaje bias negativo, los electrones se transfieren a través del dispositivo en la secuencia  $(0, 1) \rightarrow (0, 2) \rightarrow (1, 1) \rightarrow (0, 1)$ . En este ciclo, el QD de la derecha siempre contiene un solo electrón. Supongamos que este electrón tiene espín positivo. Luego, en la transición  $(0, 1) \rightarrow (0, 2)$  el QD de la derecha solo puede aceptar un electrón de espín negativo de los contactos debido a exclusión de Pauli, y entre ambos se forma un estado singlete  $S(0, 2)$ . De manera similar, solo se puede agregar un electrón con espín positivo si el primer electrón tiene espín negativo. Desde  $S(0, 2)$ , un electrón puede tunear hacia el QD izquierdo y finalmente hacia el conductor izquierdo. Por el contrario, cuando el voltaje bias es positivo, el transporte de carga procede en la secuencia  $(0, 1) \rightarrow (1, 1) \rightarrow (0, 2) \rightarrow (0, 1)$  y el QD de la izquierda se puede llenar del mar de Fermi con ya sea un electrón con espín positivo o negativo, independientemente del espín del electrón en el QD de la derecha. Si los dos electrones forman un estado singlete  $S(1, 1)$ , el electrón en el QD de la izquierda puede transferirse al QD de la derecha formando  $S(0, 2)$ . Sin embargo, si los electrones forman uno de los estados de triplete  $T(1, 1)$ , el electrón en el punto de la izquierda no podrá pasar al punto de la derecha porque  $T(0, 2)$  tiene una energía demasiado alta. El sistema permanecerá entonces bloqueado en un estado de carga  $(1, 1)$  hasta que el espín del electrón se relaje. Dado que el tiempo de relajación  $T_1$  puede aproximarse a los milisegundos, la corriente en esta dirección es insignificante y se dice que el punto está en bloqueo de espín. Debido a que es el principio de exclusión de Pauli el que prohíbe que los electrones hagan una transición de un estado  $T(1, 1)$  a  $S(0, 2)$ , este bloqueo también se conoce como bloqueo de Pauli. Este efecto conduce a una asimetría de la corriente en función del voltaje bias, por lo que puede observarse fácilmente en un experimento.



**Figura 2.6:** Corriente  $I$  medida como función del voltaje bias  $V$  en un sistema de punto cuántico doble. La corriente distinta de cero se mide en todo el rango de voltaje negativo. Para bias positiva, sin embargo, la corriente se bloquea en el rango  $2 < V < 7$  mV. A voltajes superiores a los 7 mV, el estado triplete  $(0, 2)$  se vuelve accesible y se levanta el bloqueo de Pauli. (b) Diferentes situaciones de transporte y efecto de bloqueo de Pauli. Figuras tomadas de Refs. [13, 14].

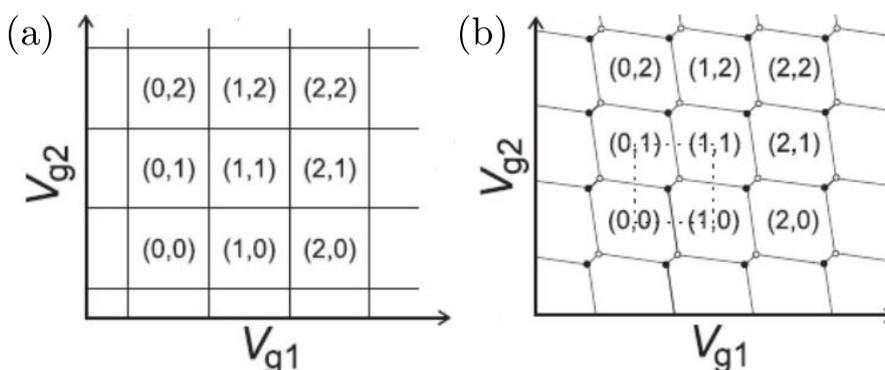
### 2.4. Puntos cuánticos dobles

Ahora describiremos un sistema donde existen dos puntos cuánticos (punto 1 y punto 2), cuyos potenciales electroquímicos son controlados independientemente por los voltajes de compuerta  $V_{G,1}$  y  $V_{G,2}$ , respectivamente. Este tipo de arreglo, denominado punto cuántico doble en serie, es lo que tomaremos como ejemplo de estudio a lo largo de este trabajo.



**Figura 2.7:** Representación esquemática del circuito formado por dos puntos cuánticos acoplados en serie. Los puntos cuánticos se representan por regiones circulares, y se acoplan de forma capacitiva a las compuertas  $V_{g,1}$  y  $V_{g,2}$ . Figura tomada de Ref. [15].

Si los puntos están completamente desacoplados ( $C_m \rightarrow 0$ ), su diagrama de estabilidad,  $(N_1, N_2)$  en función de  $V_{G,1}$  y  $V_{G,2}$ , donde  $N_1$  y  $N_2$  son el número de electrones en el punto 1 y 2, respectivamente, se muestra en la figura 2.8(a). Las líneas indican los valores de los voltajes de compuerta en los que cambia el número de electrones en el estado fundamental. Vemos que estas líneas son exactamente horizontales y verticales en el diagrama de la izquierda, ya que el potencial electroquímico en cada punto es independiente de la carga en el otro QD y cada voltaje de compuerta solo afecta a uno de los puntos.



**Figura 2.8:** Diagrama de estabilidad de carga  $(N_1, N_2)$  en función de los potenciales de compuerta para un sistema de dos puntos cuánticos, con (a) sistema desacoplado y (b) sistema acoplado. Figura tomada de Ref. [15].

Si ahora los QDs están acoplados capacitivamente, la adición de un electrón en uno de ellos afecta la energía electrostática del otro. Además, el voltaje de compuerta de un QD también tiene un acoplamiento capacitivo directo con el otro, efecto que vamos a desarrollar mas adelante en esta tesis. El diagrama de estabilidad de carga resultante se muestra en la figura 2.8(b). Vemos que cada punto de cruce en la figura se divide en los llamados “puntos triples”. El tipo de estructura que se forma entre las regiones de carga puede verse como una red hexagonal o panal. El hecho es que en un punto triple, se degeneran energéticamente tres estados de carga diferentes y la distancia entre los puntos triples está determinada por la capacitancia entre los dos puntos cuánticos.

## Capítulo 3

# Teoría de transporte

*El objetivo de este capítulo es presentar una teoría sistemática para describir el transporte electrónico fuera del equilibrio a través de puntos cuánticos que interactúan fuertemente, y que se encuentran débilmente acoplados a varios reservorios macroscópicos.*

### 3.1. Introducción

En la actualidad existe un amplio abanico de mecanismos que permiten abordar este problema como la aproximación de Landauer-Büttiker [2, 3] o la implementación de métodos computacionales basados, por ejemplo, en la Teoría del Funcional de la Densidad (DFT) combinada con funciones de Green de no-equilibrio. En particular, los métodos de ecuaciones maestras generalizadas (GME) continúan desempeñando un papel importante, y por buenas razones. Estos enfoques se basan en derivar una ecuación para la matriz de densidad reducida en estado estacionario del sistema conformado por los puntos cuánticos. La ecuación trata la condición de no equilibrio impuesta por el voltaje bias de manera exacta y se deriva en base a los estados propios de muchos cuerpos del punto cuántico aislado, haciéndolo exacto también en todas las interacciones locales, mientras que el acoplamiento túnel a los reservorios será tratado en forma perturbativa.

La forma general de GME se conoce desde hace mucho tiempo y se puede encontrar utilizando muchos enfoques diferentes pero equivalentes. Si el acople a los contactos es mucho más pequeño que la temperatura, es suficiente una teoría de perturbaciones a orden más bajo en el acople, que se reduce a ecuaciones de velocidad clásicas [16, 17] para las ocupaciones (elementos diagonales de la matriz de densidad) si el sistema está libre de degeneraciones. Si hay degeneraciones entre estados no asociados con las reglas de selección, se deben tener en cuenta las coherencias (elementos fuera de la diagonal de la matriz densidad), véase, por ejemplo, Refs. [18, 19]. Si el acople se vuelve comparable o mayor que la temperatura, se debe ir más allá del orden principal. Estos dos últimos casos, si bien podrían ser abordados dentro de las GMEs, van más allá del enfoque de este trabajo.

Aquí usaremos la teoría de transporte en tiempo real [20, 21], reformulada en términos de superoperadores de Liouville [22], para evaluar una expansión perturbativa que incluye las contribuciones de orden principal en el acoplamiento con los reservorios. Obtenemos expresiones explícitas para los kernels de transporte que ingresan en las ecuaciones maestras, válidas para una gran clase de sistemas de puntos cuánticos. Por conveniencia, usamos unidades naturales, es decir, establecemos  $e = \hbar = k = 1$  de aquí en adelante y en el resto de este trabajo, a menos que se indique lo contrario.

### 3.2. Modelo Hamiltoniano

Consideramos un sistema de puntos cuánticos, conectado a una serie de reservorios macroscópicos etiquetados por  $r$ . Dentro de la aproximación, se asume que los electrones en los reservorios no interactúan, pero las interacciones en el sistema local son tomadas de forma exacta. La principal limitación del método es el número de grados de libertad representados en el hamiltoniano de QDs, que debe ser diagonalizado exactamente para obtener los autoestados y autoenergías de muchos cuerpos. A su vez, este hamiltoniano incluye los efectos de los voltajes aplicados, así como la renormalización electrostática de los niveles de energía y la

energía de carga (no relacionada con el tunelamiento a los contactos). El sistema entero está descrito por  $\hat{H}_{\text{total}} = \hat{H} + \hat{H}_{\text{res}} + \hat{H}_{\text{tun}}$ , donde  $\hat{H}_{\text{res}} = \sum_r \hat{H}_r$ .

$$\hat{H} = \sum_a E_a |a\rangle\langle a|, \quad (3.1)$$

$$\hat{H}_r = \sum_{\sigma} \int d\omega \omega \hat{c}_{r\sigma-\omega} \hat{c}_{r\sigma+\omega}, \quad (3.2)$$

$$\hat{H}_{\text{tun}} = \sum_{r\sigma N} \sum_{\eta=\pm 1} \sum_{a,a'} \eta \int d\omega T_{r\sigma\eta}^{aa'} |a\rangle\langle a'| \hat{c}_{r\sigma\eta\omega}. \quad (3.3)$$

En el hamiltoniano  $\hat{H}$  del punto cuántico,  $|a\rangle$  denota un autoestado general del sistema de muchos cuerpos con energía  $E_a$ , la cual ya incluye el efecto de los voltajes aplicados. Se asume que se puede clasificar estos estados con el número de electrones de exceso  $N$  en el QD. Se denota por  $N_a$  el número de electrones en el estado  $a$ , donde se asume que  $\hat{N} |a\rangle = N_a |a\rangle$ , con  $\hat{N}$  el operador número de partículas en el QD. Esto implica que el número de partículas en el QD está bien definido.

Los reservorios están descritos por los hamiltonianos  $\hat{H}_r$ , e involucran la integral en  $\omega$  a lo largo de sus bandas de energía. Los operadores de campo  $\hat{c}_{r\sigma\eta\omega}$ , con  $\eta = \pm 1$  el índice partícula-hueco, se relacionan con los operadores de creación y aniquilación usuales según:

$$\hat{c}_{r\sigma+\omega} = \sum_k \frac{1}{\sqrt{\rho_{r\sigma}}} \delta(\epsilon_{r\sigma k} - \omega) \hat{c}_{r\sigma k}, \quad \hat{c}_{r\sigma-\omega} = \sum_k \frac{1}{\sqrt{\rho_{r\sigma}}} \delta(\epsilon_{r\sigma k} - \omega) \hat{c}_{r\sigma k}^{\dagger}, \quad (3.4)$$

donde  $\hat{c}_{r\sigma k}^{\dagger}$  ( $\hat{c}_{r\sigma k}$ ) crea (aniquila) un electrón en el reservorio  $r$  con proyección de espín  $\sigma$  y estado  $k$ . En estas ecuaciones,  $\rho_{r\sigma}$  representa la densidad de estados en el reservorio  $r$ .

El hamiltoniano  $\hat{H}_{\text{tun}}$  describe el tuneo desde o hacia el sistema de puntos cuánticos, involucrando un cambio en el estado de  $a'$  hacia  $a$  y sobre la carga del QD de  $N_{a'}$  a  $N_a = N_{a'} \pm 1$ . Los elementos de matriz relevantes están dados por las superposiciones entre los elementos de la matriz de tunneling de una partícula  $t_{\ell r\sigma}$  y las amplitudes de transición entre estados de muchos cuerpos del QD, es decir:

$$T_{r\sigma+}^{aa'} = \sqrt{\rho_{r\sigma}} \sum_{\ell} t_{\ell r\sigma} \langle a | \hat{d}_{\ell\sigma}^{\dagger} | a' \rangle, \quad (3.5)$$

$$T_{r\sigma-}^{aa'} = \sqrt{\rho_{r\sigma}} \sum_{\ell} t_{\ell r\sigma}^* \langle a | \hat{d}_{\ell\sigma} | a' \rangle = (T_{r\sigma+}^{a'a})^*. \quad (3.6)$$

Aquí  $\ell$  etiqueta una base de una única partícula para el QD, y el operador correspondiente  $\hat{d}_{\ell\sigma}^{\dagger}$  ( $\hat{d}_{\ell\sigma}$ ) crea (aniquila) en el sistema de puntos cuánticos un electrón con proyección de espín  $\sigma$  en el estado  $\ell$ . Notar que la densidad de estados se incorpora en los elementos de tunneling  $T_{r\sigma\eta}^{aa'}$ . Las densidades espectrales:

$$(\Gamma_{r\sigma})_{a'b'}^{ab} = 2\pi T_{r\sigma(N_a - N_{a'})}^{aa'} T_{r\sigma(N_{a'} - N_a)}^{b'b}, \quad (3.7)$$

representan así el conjunto de escalas de energía relevantes para el tunelamiento. Tanto  $\sqrt{\rho_{r\sigma}}$  como  $t_{\ell r\sigma}$  se asumen que son independientes de la energía. Como se puede ver en [6], los operadores de campo de los reservorios y del punto cuántico conmutan, lo cual simplifica el cálculo de signos cuando estos aparecen en las ecuaciones.

### 3.3. Ecuación maestra generalizada

Un sistema microscópico de puntos cuánticos acoplado a reservorios macroscópicos se describe completamente por su operador densidad reducido  $\hat{\rho}(t)$ , obtenido al promediar el operador de densidad total  $\hat{\rho}(t) = \text{tr}_{\text{res}}[\hat{\rho}(t)]$ , donde  $\text{tr}_{\text{res}}[\bullet]$  denota la traza sobre los grados de libertad del reservorio. El operador densidad reducido evoluciona en el tiempo de acuerdo con una ecuación maestra generalizada (GME) (o ecuación cinética). La presencia de efectos de no equilibrio (transporte no lineal) y fuertes interacciones locales (Coulombiana, vibración electrónica, etc.) hace que el cálculo de las tasas de transporte que ocurren en esta ecuación sea una

tarea engorrosa. Nuestro objetivo en este trabajo es derivar expresiones explícitas para las tasas de transporte a orden más bajo en tunneling en términos de los parámetros  $E_a$  y  $T_{r\sigma\eta}^{aa'}$  de los hamiltonianos dados en las ecuaciones (3.1) y (3.3), junto con las propiedades estadísticas de los electrodos, dadas por la temperatura  $T$  y los potenciales electroquímicos  $\mu_r$ .

### 3.3.1. Evolución del operador densidad reducido

La teoría de transporte en tiempo real, desarrollada en [20, 21, 23] y posteriormente ampliada por varios grupos [24–26], proporciona reglas sencillas para el cálculo de las tasas de transporte usando una representación esquemática en el contorno de Keldysh, evitando cualquier problema de regularización espuria. Esta técnica se ha simplificado aún más mediante el uso de superoperadores especiales de Liouville y los diagramas correspondientes en el contexto de un enfoque de grupo de renormalización fuera del equilibrio [22, 27, 28]. El punto de partida es la evolución temporal del operador densidad del sistema total, QD más reservorios, que en la representación de Schrödinger viene dada por:

$$\hat{\rho}(t) = e^{-i\check{L}_{\text{total}}t} \hat{\rho}(0). \quad (3.8)$$

Aquí, los superoperadores de Liouville en  $\check{L}_{\text{total}} = \check{L} + \check{L}_R + \check{L}_T$  actúan sobre un operador arbitrario  $\hat{A}$  formando un conmutador con el hamiltoniano de la manera  $\check{L}\hat{A} = [\hat{H}, \hat{A}]$ . Se asume que el sistema está desacoplado al tiempo inicial, el cual se supone de manera arbitraria como  $t = 0$ , tal que el operador densidad se factoriza como  $\hat{\rho}(0) = \hat{p}(0) \otimes \hat{\rho}_{\text{res}}$ , en donde  $\hat{\rho}_{\text{res}} = \bigotimes_r \hat{\rho}_r$  y  $\hat{\rho}_r$  describe el reservorio  $r$ . Al tiempo inicial, se asume que cada reservorio se encuentra en un equilibrio interno independiente y es descrito por un ensamble gran canónico. Cuando el tunelamiento comienza, el voltaje bias aplicado hace que los potenciales químicos de los diferentes reservorios difieran entre sí. Esto conlleva a una condición de borde no homogénea sobre el operador densidad del QD, lo que lo saca del estado de equilibrio. Si ahora tomamos la transformada de Laplace de la ecuación (3.8) y tomamos la traza sobre los reservorios, se pueden derivar una expresión para la ecuación maestra generalizada según [6]:

$$0 = (-i\check{L} + \check{W})\hat{p}, \quad (3.9)$$

donde  $\check{W}$  es un superoperador de energía propia (o kernel de evolución) y  $\hat{p}$  es el operador densidad reducido para tiempos estacionarios ( $t \rightarrow \infty$ ). Para evaluar  $\check{W}$ , se utiliza la expansión:

$$\check{W}(z) = \sum_{k=1}^{\infty} \check{W}^{(2k)}(z), \quad (3.10)$$

la cual es útil ya que nos permite elegir el orden a resolver de la ecuación. Claramente, a medida que se aumenta el orden del kernel, aumenta la complejidad de evaluación. Sin embargo, como mencionamos anteriormente, solo nos quedaremos con el orden más bajo en la expansión. En general, los elementos de matriz de un superoperador  $\check{S}$  que actúa sobre un operador  $\hat{A}$  vienen dados por:

$$(\check{S}\hat{A})_{ab} = \langle a | \check{S}\hat{A} | b \rangle = \sum_{a',b'} \langle a | (\check{S} | a' \rangle \langle a' | \hat{A} | b' \rangle \langle b' |) | b \rangle = \sum_{a',b'} S_{a,b}^{a',b'} A_{a',b'}. \quad (3.11)$$

Se pueden encontrar más detalles sobre los superoperadores en [29, 30]. En particular, en la base de los autoestados de muchos cuerpos del sistema de QDs, el Liouvilliano del punto cuántico toma la forma:

$$L_{ab}^{a'b'} = \langle a | ([\hat{H}, |a' \rangle \langle b'|]) | b \rangle = (E_a - E_b) \delta_{a,a'} \delta_{b,b'}. \quad (3.12)$$

Notar, por ejemplo, que los elementos diagonales de este superoperador, dados por  $a = b$  y  $a' = b'$ , resultan nulos. El objetivo principal es calcular el valor de expectación de la corriente que fluye desde los reservorios hacia los puntos cuánticos. Como se puede ver en [6], este valor de expectación puede ser obtenido utilizando la expansión del kernel de evolución  $\check{W}$  y el operador densidad reducido. En el caso estacionario:  $I_r = \text{tr}[\check{W}_{I_r} \hat{p}]$ , donde el kernel de corriente contiene el subconjunto de procesos de tunelamiento descritos por  $W$  que contribuyen a la corriente de estado estacionario a través del reservorio  $r$ . Entonces, ya podemos escribir de forma

generalizada la ecuación maestra:

$$0 = \sum_{a'b'} [-iL_{ab}^{a'b'} + W_{ab}^{a'b'}] p_{a'b'}, \quad (3.13)$$

$$1 = \sum_a p_{aa}, \quad (3.14)$$

$$I_r = \sum_a \sum_{a'b'} (\tilde{W}_{I_r})_{aa}^{a'b'} p_{a'b'}. \quad (3.15)$$

Aquí hemos tomado el orden más bajo de los kernels en el acople túnel. Esto corresponde a tomar  $k = 1$  en la ecuación (3.10), tal que los kernels contienen dos veces al superoperador  $\check{L}_{\text{tun}}$ , dando cuenta la aniquilación de un electrón en el reservorio  $r$  y la creación en el QD, o bien el proceso inverso. Notar que los órdenes impares en la expansión se anulan al trazar sobre los reservorios, ya que el hamiltoniano de tunneling es lineal en los operadores de campo de reservorios. Las ecuaciones (3.13-3.15) formulan de forma compacta el problema de transporte.

Como anticipamos anteriormente, en este trabajo los efectos de las coherencias sobre las ocupaciones pueden ser despreciados en una primer aproximación, ya que vamos a considerar un régimen donde los estados de muchos cuerpos no están degenerados (más allá de la degeneración trivial de espín, ver Apéndice A). En este sentido, resulta conveniente describir las ocupaciones mediante el vector  $\mathbf{p}$ , el cual colecta los elementos diagonales de  $\hat{p}$ . De esta forma, obtenemos la siguiente ecuación maestra:

$$\mathbf{W}\mathbf{p} = \mathbf{0}, \quad (3.16)$$

junto con la condición de normalización  $\mathbf{e}^\top \mathbf{p} = 1$ , con  $\mathbf{e}^\top = (1, 1, \dots)$  el operador traza en esta nueva notación. El sistema de ecuaciones representado en la ecuación (3.16) no es invertible, ya que  $\mathbf{p}$  puede ser visto como un autovector de  $\mathbf{W}$  cuyo autovalor es nulo. Sin embargo, se puede evitar este problema si incorporamos la condición de normalización dentro de (3.16), tal que podemos reescribir:

$$\tilde{\mathbf{W}}\mathbf{p} = \mathbf{v} \implies \mathbf{p} = \tilde{\mathbf{W}}^{-1}\mathbf{v}, \quad (3.17)$$

donde definimos un pseudo-kernel invertible  $\tilde{\mathbf{W}}$  a partir de  $\tilde{W}_{ab} = W_{ab} - W_{aa}$  y  $v_a = -W_{aa}$ . Resolviendo entonces esta última ecuación, encontramos las ocupaciones estacionarias en el QD, y junto con ellas podemos calcular otros observables como la corriente de carga:

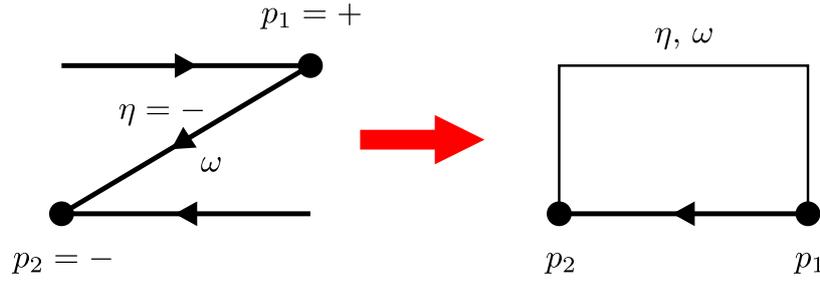
$$I_r = \mathbf{e}^\top \mathbf{W}_{I_r} \mathbf{p}, \quad (3.18)$$

donde el kernel de corriente  $\mathbf{W}_{I_r}$  puede calcularse a partir del kernel de evolución  $\mathbf{W}_r$  asociado al reservorio  $r$  únicamente según la siguiente expresión

$$[\mathbf{W}_{I_r}]_{ab} = (N_a - N_b)[\mathbf{W}_r]_{ab}, \quad \text{con} \quad \sum_r \mathbf{W}_r = \mathbf{W}. \quad (3.19)$$

De este modo, el análisis de las ocupaciones en el sistema de puntos cuánticos y las corrientes que lo atraviesan requiere del cálculo de los elementos de matriz del kernel  $\mathbf{W}$  que ahora vamos a detallar.

Considerando que la evolución temporal del operador densidad involucra una evolución hacia adelante y otra evolución hacia atrás en el tiempo, en los diagramas de doble contorno de Keldysh que dan lugar a los elementos del kernel  $\check{W}$  aparecen estados que tienen lugar en cada una de estas evoluciones. Para dar cuenta de esta estructura, introducimos una notación para los estados de acuerdo al contorno de Keldysh  $p = \pm 1$  al que estos pertenecen, i.e.,  $a_{i,p}$ , y su diferencia de energía  $E_{a_i} \equiv E_{a_{i,+}} - E_{a_{i,-}}$ . Como ilustración, en la figura 3.1 mostramos un ejemplo de un diagrama asociado al kernel de evolución, tanto en la representación de doble contorno de Keldysh (izquierda) como en la representación de superoperadores (derecha).



**Figura 3.1:** Ejemplo de diagrama asociado al kernel de evolución  $\check{W}$  para el orden más bajo en el acople túnel. Izquierda: diagrama de doble contorno de Keldysh, los índices  $p_1$  y  $p_2$  indican en qué contorno de Keldysh se produce el evento de tunneling. Derecha: diagrama de superoperadores asociado.

Haciendo los respectivos cálculos en la expansión a primer orden, la expresión más general para los elementos del kernel de evolución son [6]:

$$(W)_{a_{2,+}, a_{2,-}}^{a_{0,+}, a_{0,-}} = i \sum_{p_2, p_1} \sum_{r \eta \sigma} \sum_{a_{1\pm}} p_2 p_1 T_{r\sigma}^{a_2, p_2, a_1, p_2} T_{r\sigma}^{a_1, p_1, a_0, p_1} \delta_{a_2, \bar{p}_2, a_1, \bar{p}_2} \delta_{a_1, \bar{p}_1, a_0, \bar{p}_1} F_{r, \eta}^{p_1}(E_{a_1}). \quad (3.20)$$

En esta ecuación,  $p_1$  y  $p_2$  corresponden a los índices de contorno, y utilizamos las abreviaciones  $\bar{p}_i = -p_i$  y  $\bar{\eta} = -\eta$ . El signo general  $p_2 p_1$  surge de varias contribuciones asociadas a las reglas diagramáticas para este elemento,  $\eta$  toma los valores  $\pm 1$  y para nuestro caso tomamos  $r = \{L, R\}$ , que da cuenta de un reservorio a izquierda y otro a derecha. La función  $F_{r, \eta}^{p_1}(E_{a_1})$  proviene de la integral en energía del elemento irreducible bajo la aproximación de banda ancha para la densidad de estados en los reservorios, y tiene por resultado [31]:

$$F_{r, \eta}^{p_1}(E_{a_1}) = p_1 \phi \left( \frac{E_{a_1} - \eta \mu_r}{k_B T} \right) + i \pi f \left( p_1 \frac{E_{a_1} - \eta \mu_r}{k_B T} \right), \quad (3.21)$$

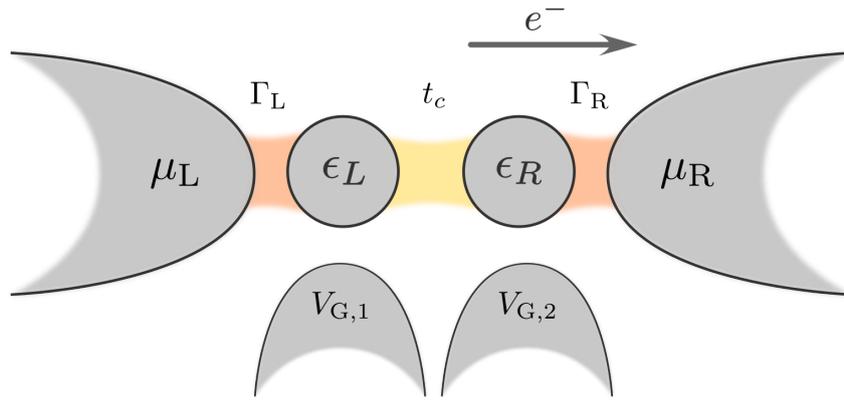
donde  $f$  es la función de distribución de Fermi-Dirac, definida como  $f(x) = [1 + \exp(x)]^{-1}$ , mientras que

$$\phi(x) = -\text{Re} \psi \left( \frac{1}{2} + i \frac{x}{2\pi} \right) + \ln \frac{D}{2\pi k_B T}, \quad (3.22)$$

con  $\psi$  la función digamma, y  $D$  el ancho de banda asociado a los reservorios. Si bien esta última expresión puede parecer un tanto complicada, veremos más adelante que para elementos diagonales del kernel, i.e.  $a_{0,+} = a_{0,-}$  y  $a_{2,+} = a_{2,-}$ , los términos asociados a la función  $\phi$  terminan cancelándose entre sí, de modo que en la ecuación (3.20) solo sobrevive la función de ocupación  $f$ .

### 3.4. Punto cuántico doble

A lo largo de esta tesis trabajaremos con un punto cuántico doble acoplado en serie (DQD), lo que corresponde a una estructura de dos puntos cuánticos en fila, donde el QD-L se acopla al reservorio de la izquierda, mientras que el QD-R está acoplado al reservorio de la derecha, ver figura 3.2.



**Figura 3.2:** Esquema de un punto cuántico doble en serie y acoplado a los reservorios L y R. Los círculos denotan la posición de cada punto cuántico, donde sus energías de sitio  $\epsilon_i$  pueden modificarse a través de los voltajes de compuerta  $V_{G,i}$ . Los factores  $\Gamma_L$  y  $\Gamma_R$  determinan las tasas de tunneling desde y hacia cada reservorio. Los puntos cuánticos se acoplan entre sí mediante el término  $t_c$ .

Entre ambos hay, además, un acople  $t_c$ , de modo que los electrones pueden saltar de un QD al otro sin cambiar el número total de partículas dentro del sistema. Para simplificar el análisis, se restringe el DQD a ocupación simple en cada QD a partir de la suposición de que la energía de interacción entre electrones en un mismo QD está por fuera de las escalas utilizadas. A su vez, vamos a trabajar en el régimen de acople fuerte donde  $t_c \gg \Gamma$  el cual, como se discute en el Apéndice A, nos permite desacoplar las coherencias de las ocupaciones del sistema. El hamiltoniano del DQD se puede expresar entonces de la siguiente manera:

$$\hat{H} = \sum_{\ell=\{L,R\}} \epsilon_\ell \hat{d}_{\ell\sigma}^\dagger \hat{d}_{\ell\sigma} + U \hat{n}_L \hat{n}_R + \frac{t_c}{2} \sum_{\sigma} (\hat{d}_{L\sigma}^\dagger \hat{d}_{R\sigma} + \text{h.c.}), \quad (3.23)$$

donde  $\epsilon_\ell$  es la energía de sitio asociada al punto cuántico  $\ell$ ,  $\hat{d}_{\ell\sigma}^\dagger$  ( $\hat{d}_{\ell\sigma}$ ) es el operador de creación (aniquilación) de un electrón en el QD  $\ell$  con proyección espín  $\sigma$  y  $\hat{n}_\ell = \sum_{\sigma} \hat{d}_{\ell\sigma}^\dagger \hat{d}_{\ell\sigma}$  es el operador número para el QD  $\ell$ .  $U$  es la energía de interacción de los electrones entre los QDs y  $t_c$  es la amplitud de salto entre ellos.

Definimos los estados del sistema de acuerdo a la base anterior para los operadores fermiónicos, de modo que se tiene  $|0\rangle$  para el estado vacío,  $|\ell, \sigma\rangle = \hat{d}_{\ell\sigma}^\dagger |0\rangle$  para los estados de ocupación simple, y  $|\sigma, \sigma'\rangle = \hat{d}_{R\sigma'}^\dagger \hat{d}_{L\sigma}^\dagger |0\rangle$  para los estados de doble ocupación. De esta manera se puede dar la forma matricial de este hamiltoniano en la base “local”. A su vez, por conservación de la carga en el DQD, se puede dividir a esta matriz en bloques dados por un mismo número de carga, es decir:  $N = \{0, 1, 2\}$ . Para el caso  $N = 0$ , se puede ver a través del hamiltoniano (3.23) que  $\hat{H} |0\rangle = 0$ , y por lo tanto este bloque es una matriz  $1 \times 1$  igual a cero, es decir  $\mathbf{H}_{N=0} = 0$ . Para el bloque  $N = 1$ , podemos aprovechar que el hamiltoniano conserva la proyección de espín, por lo que en la base  $\{|L, \sigma\rangle, |R, \sigma\rangle\}$  asociada a cada proyección se tiene el siguiente bloque de matriz:

$$\mathbf{H}_{N=1,\sigma} = \begin{pmatrix} \epsilon_L & t_c/2 \\ t_c/2 & \epsilon_R \end{pmatrix}, \quad (3.24)$$

mientras que para el bloque  $N = 2$  tenemos una matriz diagonal de la forma  $\mathbf{H}_{N=2} = (\epsilon_L + \epsilon_R + U) \mathbf{I}_4$ , donde  $\mathbf{I}_4$  es la matriz identidad de dimensión 4. Si volvemos a la ecuación (3.1), notamos que el sistema debe estar diagonalizado para la ejecución del procedimiento propuesto, por lo cual se deben calcular los autoestados para el bloque de carga  $N = 1$ . Como este es independiente de la proyección de espín, solo debemos diagonalizar la matriz en la ecuación (3.24), donde obtenemos los siguientes autovalores:

$$\epsilon_{b,a} = E \mp \frac{1}{2} \sqrt{\epsilon^2 + t_c^2}, \quad (3.25)$$

con  $E = (\epsilon_L + \epsilon_R)/2$  y  $\epsilon = \epsilon_L - \epsilon_R$ . Es decir, para el bloque de carga  $N = 1$ , el sistema tiene dos estados posibles por cada proyección de espín: “enlazante”, con energía  $\lambda_b$  y “anti-enlazante”, con energía  $\lambda_a$ . Estos

estados pueden escribirse a partir de los estados anteriores según [32]:

$$|b/a, \sigma\rangle = \frac{1}{\sqrt{2}} \left[ \sqrt{1 \mp \frac{\epsilon}{\sqrt{\epsilon^2 + t_c^2}}} |L, \sigma\rangle \pm \sqrt{1 \pm \frac{\epsilon}{\sqrt{\epsilon^2 + t_c^2}}} |R, \sigma\rangle \right]. \quad (3.26)$$

Con esta nueva base para los estados de ocupación simple, y notando que para los bloques correspondientes a  $N = 0$  y  $N = 2$  no se necesitan cambios de base, estamos en condiciones de evaluar los elementos de matriz del kernel a partir de la ecuación (3.20). En particular, la ecuación anterior puede resultar útil a la hora de evaluar los elementos de tunneling dados en las ecuaciones (3.5) y (3.6), ya que podemos definir los operadores de creación (y aniquilación) con respecto a esta nueva base, es decir  $d_{b\sigma}^\dagger$  y  $d_{a\sigma}^\dagger$ .

### 3.5. Cálculo del kernel

En las siguientes líneas derivamos los elementos de matriz del kernel para el DQD discutido anteriormente. Para ello utilizaremos propiedades generales como propias al modelo en cuestión. En primer lugar, tomamos las sumas sobre los índices de contorno  $p_1$  y  $p_2$ . Además, como solo vamos a considerar los elementos diagonales de la matriz densidad, nos interesará calcular los elementos del kernel que cumplan  $a_{0,-} = a_{0,+} = a_0$  y  $a_{2,-} = a_{2,+} = a_2$ , lo que nos permite omitir el rótulo de los mismos con  $p_1$  y  $p_2$ . Tenemos entonces los siguientes términos:

$$\begin{aligned} (W)_{a_2, a_2}^{a_0, a_0} = & + i \sum_{r\eta\sigma} \sum_{a_1} |T_{r\sigma\eta}^{a_1, a_0}|^2 [F_{r\eta}^+(E_{a_1, a_0}) + F_{r\bar{\eta}}^-(E_{a_2, a_1})] \delta_{a_2, a_0} \\ & - i \sum_{r\eta\sigma} [T_{r\sigma\eta}^{a_2, a_0}]^2 [F_{r\eta}^+(E_{a_2, a_0}) + F_{r\bar{\eta}}^-(E_{a_0, a_2})], \end{aligned} \quad (3.27)$$

donde usamos la abreviación  $E_{a,b} = E_a - E_b$  y podemos utilizar la siguiente propiedad de las funciones auxiliares  $F_{r\eta}^p(E_{a,b})$ :

$$F_{r\bar{\eta}}^p(E_{a,b}) = -[F_{r\eta}^p(E_{b,a})]^* \quad (3.28)$$

donde usamos que  $\phi(x)$  es una función par. Con esta propiedad, tenemos entonces que la ecuación (3.27) se reduce a:

$$(W)_{a_2, a_2}^{a_0, a_0} = \sum_{r\eta\sigma} \sum_{a_1} 2\pi |T_{r\sigma\eta}^{a_1, a_0}|^2 f\left(\frac{E_{a_1, a_0} - \eta\mu_r}{k_B T}\right) (\delta_{a_1, a_2} - \delta_{a_2, a_0}), \quad (3.29)$$

donde asumimos que los elementos de matriz de tunneling son reales, es decir, que el sistema a considerar respeta simetría de reversión temporal.

En base a esta expresión, podemos ahora calcular, por ejemplo, el elemento diagonal asociado al estado vacío  $|0\rangle$ . En este caso, se tiene:

$$(W)_{0,0}^{0,0} = - \sum_{r\eta\sigma} \sum_{a_1} 2\pi |T_{r\sigma\eta}^{a_1, 0}|^2 f\left(\frac{E_{a_1, 0} - \eta\mu_r}{k_B T}\right), \quad (3.30)$$

donde, claramente, solo sobrevivirán términos con  $\eta = +$ , ya que partiendo del estado vacío no es posible eliminar electrones en el QD. Entonces, para la suma en  $a_1$ , solo entrarán estados de una partícula, donde podemos tomar  $a_1 = \{\ell, \gamma\}$ , con  $\ell = \{b, a\}$  y  $\gamma = \pm 1$  la proyección de espín. De este modo, el término anterior se reduce a:

$$(W)_{0,0}^{0,0} = - \sum_{r,\ell} 2\Gamma_{r,\ell} f\left(\frac{\epsilon_\ell - \mu_r}{k_B T}\right), \quad (3.31)$$

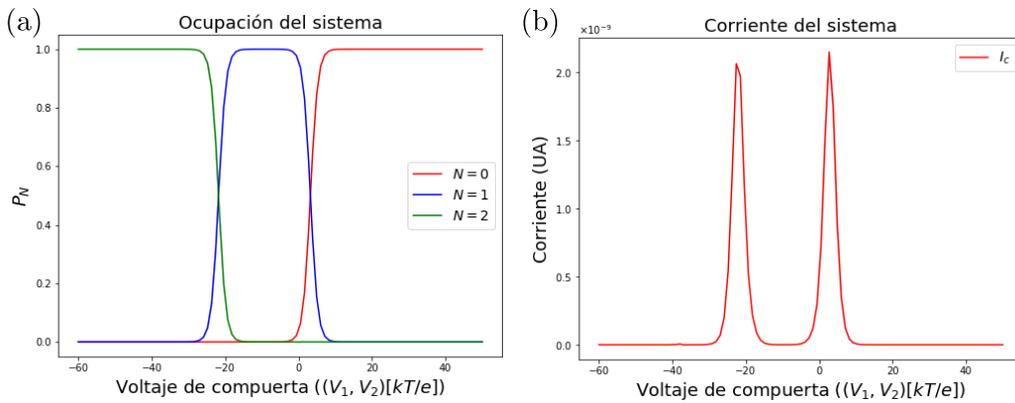
donde el factor 2 corresponde a la degeneración en espín, ya que asumimos que las tasas de tunneling no dependen de esta cantidad.

### 3.6. Ocupaciones y corriente de carga en un DQD

En esta sección ilustramos los resultados obtenidos en las secciones precedentes, mediante el cálculo de los estados de carga y corrientes en un punto cuántico doble acoplado en serie. Utilizando la forma del kernel de evolución dada en la ecuación (3.29), se plantea el sistema de ecuaciones para el modelo (3.23) y se los resuelve en forma numérica. Para todas las situaciones, se utiliza el valor  $k_B T$  como referencia de la escala de energía, mientras que los siguientes parámetros permanecen fijos:  $t_{r,\ell} = 0.1\sqrt{k_B T/8\pi}$ ,  $U = 20k_B T$  y  $t_c = 5k_B T$ . En un primera instancia, se establece una diferencia de potencial entre los contactos dada por  $\mu_L = -\mu_R = 1 \times 10^{-3}k_B T/2$ , y se barren los voltajes de compuerta en ambos puntos cuánticos mediante un mismo valor. Estos voltajes de compuerta modifican las energías de sitio asociadas a los puntos cuánticos según:

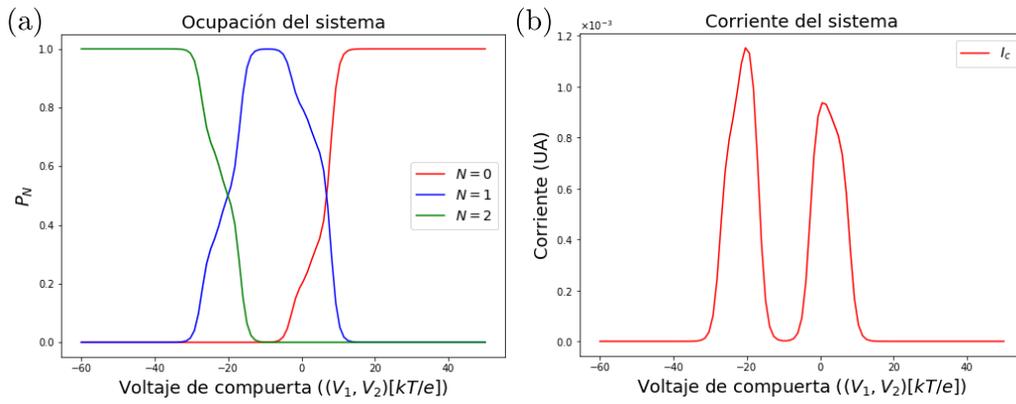
$$\epsilon_L = eV_{G,1}, \quad \epsilon_R = eV_{G,2}, \quad (3.32)$$

y, por simplicidad, consideramos el caso simétrico donde  $V_{G,1} = V_{G,2} = V_G$ . De acuerdo a la ecuación (3.25), esto implica que las energías asociadas a los estados enlazante y anti-enlazante son, respectivamente,  $\epsilon_{b/a} = E \mp t_c$ , con  $E = eV_G$ .



**Figura 3.3:** (a) Cálculo de la probabilidad de ocupación para cada bloque de carga. (b) Corriente de carga, en unidades arbitrarias, para  $\mu_L = -\mu_R = 1 \times 10^{-3}k_B T/2$ .

En la figura 3.3(a) se muestran las probabilidades de ocupación de acuerdo al bloque de carga. Es decir, las sumas  $\sum_{\alpha} p_{\alpha}$ , donde  $N_{\alpha} = \{0, 1, 2\}$ . Puede verse en la figura que si  $V_G$  toma valores muy grandes, entonces el DQD se encuentra vacío, por lo que  $p_0$  se aproxima a 1. En este caso, la diferencia de energía asociada a la transición  $0 \rightarrow 1$  es demasiado grande comparada con los potenciales de los reservorios, de forma tal que los electrones en los contactos no pueden ingresar al DQD. Si ahora disminuimos  $V_G$ , en un punto el DQD se comienza a poblar con un electrón, debido a que la diferencia de energía en la transición  $0 \rightarrow 1$  ya es accesible a los potenciales de los contactos. En este caso, podemos notar que la región de cambio de ocupación tiene lugar alrededor del valor de  $t_c$ , ya que para el nivel enlazante se tiene  $\Delta\epsilon_{0,b} = eV_G - t_c$ . Dado que la condición de resonancia se da para  $\mu_L \simeq \mu_R \simeq \Delta\epsilon_{0,b}$ , y como los potenciales de los reservorios son aproximadamente cero, se tiene entonces que el cambio ocurre para  $eV_G \simeq t_c$ . Si seguimos disminuyendo el voltaje de compuerta, nos encontramos con otro cambio en la ocupación, donde ahora el DQD se encuentra doblemente ocupado. Esto nos dice que la diferencia de energía dada por la transición  $1 \rightarrow 2$ , que resulta  $\Delta\epsilon_{a,2} = eV_G + U - t_c$  es ahora accesible por los reservorios, de modo que un segundo electrón puede ingresar al sistema. Respecto a la figura 3.3(b), notamos que la corriente electrónica tiene lugar cada vez que se produce un cambio en la ocupación del DQD, de modo que para este caso se observan dos picos, asociados a las transiciones  $0 \rightarrow 1$  y  $1 \rightarrow 2$ , respectivamente. Es de esperar que la distancia entre los picos tenga que ver con la energía de interacción  $U$ , ya que si tomamos la diferencia de energía entre ambos casos se tiene  $\Delta\epsilon_{a,2} - \Delta\epsilon_{0,b} = U$ . Fuera de estos picos, el transporte claramente se encuentra bloqueado.

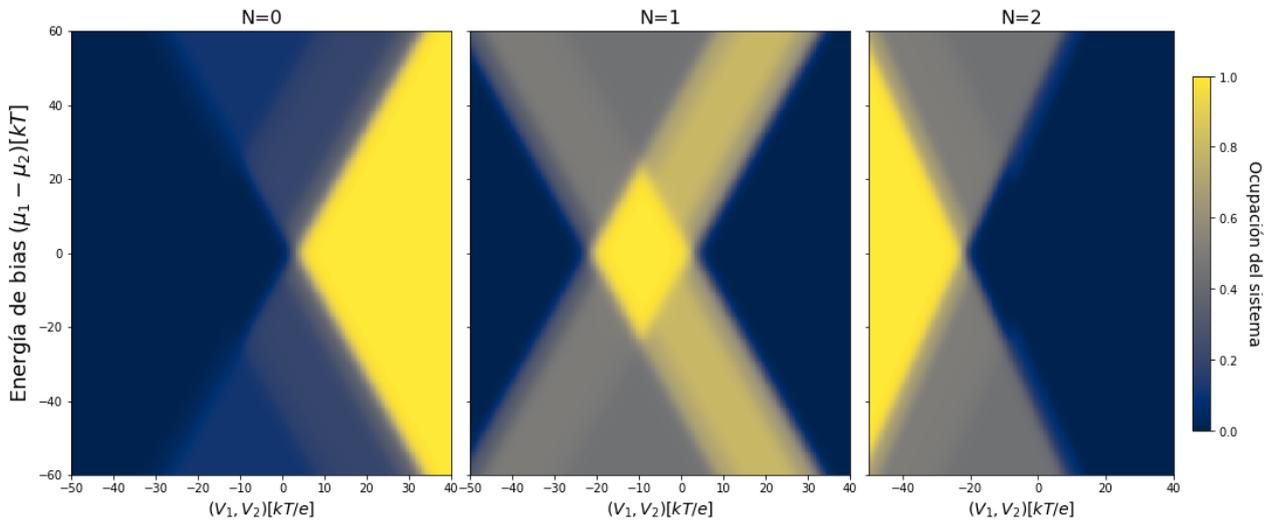


**Figura 3.4:** (a) Probabilidad de ocupación para cada bloque de carga. (b) Corriente de carga, en unidades arbitrarias, para  $\mu_L = -\mu_R = 10k_B T/2$ .

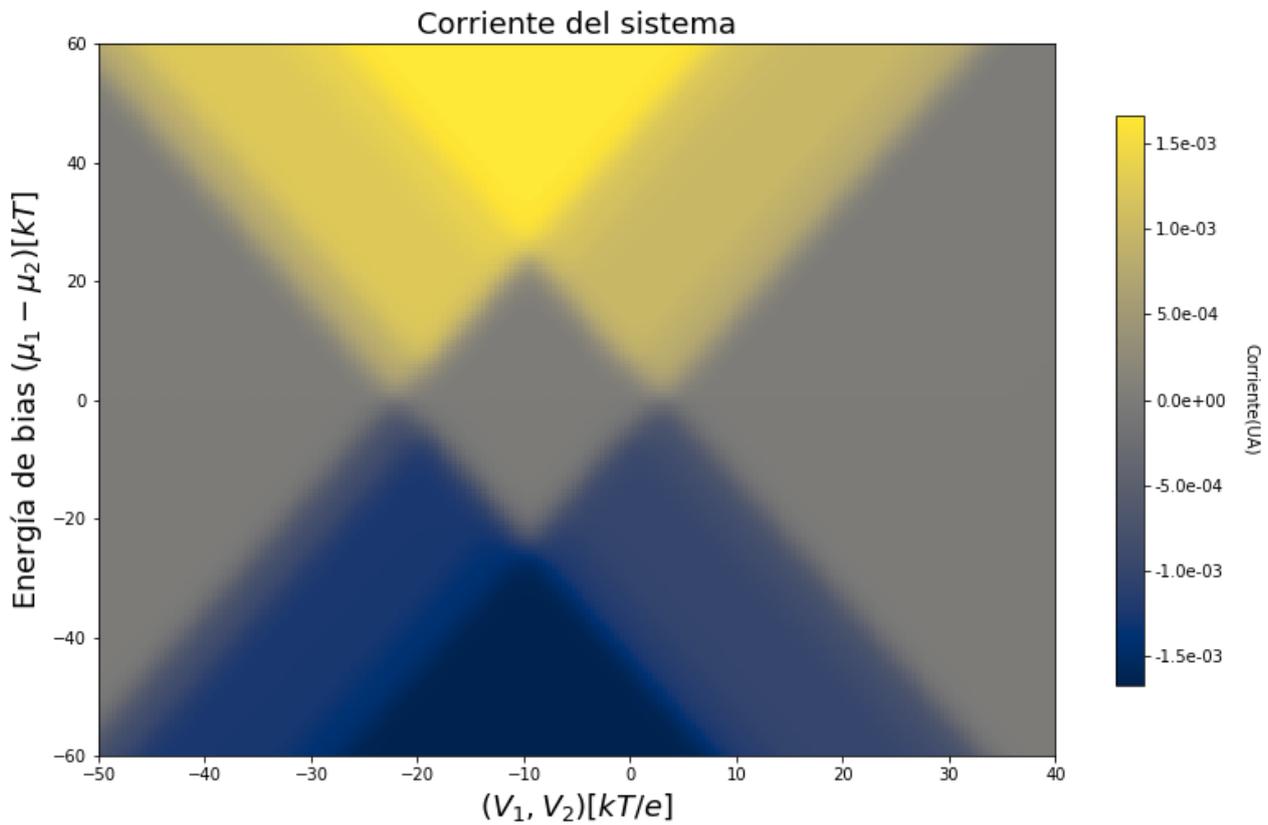
Si aumentamos el voltaje entre los reservorios, como en la figura 3.4 donde se toma  $\mu_L = -\mu_R = 10k_B T/2$ , se puede ver que las ocupaciones y en la corriente se dan de manera más suave, debido a que la ventana de bias es mayor, y con ello el rango de energías en donde dan las transiciones entre estados con diferente número de carga. Esto también provoca que los picos de corriente se ensanchen. En particular, se puede relacionar el ancho de estos picos con el valor utilizado para  $V$  si se vuelve a la figura 2.5, donde se puede ver que para cierto valor de bias, el ancho de los picos viene dado por  $|V_G| = \frac{\mu_L - \mu_R}{e\alpha_G}$ , donde aquí tomamos  $\alpha_G = 1$ . En definitiva, para estos valores de bias, las regiones de bloqueo se van reduciendo gradualmente. Otra característica que aparece en este régimen de parámetros donde la ventana de bias es comparable o mayor a  $\Gamma$  es la asimetría en los picos de corriente. Esto tiene origen en el hecho de que la cantidad de estados asociada a cada transición de carga no es la misma: Si pensamos en la transición  $0 \rightarrow 1$ , entonces para el bloque  $N = 0$  tenemos un solo estado ( $|0\rangle$ ), y para el bloque  $N = 1$  se tienen 2 estados dados por el nivel enlazante (debido a degeneración de espín). Por otro lado, en la transición  $1 \rightarrow 2$  se tiene que el bloque  $N = 1$  posee 2 estados dados por el nivel anti-enlazante, mientras que el bloque  $N = 2$  posee degeneración 4, también debido a espín. A su vez, si el voltaje de bias es grande, dentro de cada pico de corriente pueden haber variaciones, debido a que estados de la misma cantidad de carga se vuelven accesibles en la transición.

El variar un único parámetro puede brindar un conocimiento limitado sobre el modelo físico a investigar. Barrer en forma simultánea dos variables nos permite acceder a mayor información acerca de las características del sistema. En particular, aquí nos centraremos en mapas de voltaje, donde se calculan las ocupaciones y corrientes a través del DQD a medida que se varía (1) un mismo potencial de compuerta  $V_G$  para los dos QDs, junto con el voltaje bias, o bien (2) dejando fijo el bias, pero moviendo en forma independiente ambos potenciales de compuerta.

El escenario (1) permite tener un problema equivalente a un único punto cuántico con 2 orbitales, por lo que al variar el voltaje de bias, se pueden reproducir los diamantes de Coulomb como se observó en la sección 2.2.1. De esta forma, resolvemos el sistema de manera numérica, calculando las probabilidades de ocupación y la corriente a lo largo del mismo. La figura 3.5(a) muestra este caso, donde puede verse claramente la estructura de diamantes y las regiones de estabilidad para cada bloque de carga. La forma simétrica de estos diamantes tiene que ver con la elección particular  $\alpha_G = 1$ . A su vez, podemos notar pequeñas variaciones en las ocupaciones, que corresponden a cambios en la cantidad de estados que cuya energía de adición cae por debajo de uno de los potenciales de los reservorios.



**Figura 3.5:** Ocupaciones en el DQD en función del voltaje bias  $V$  y voltaje de compuerta  $V_G$ .

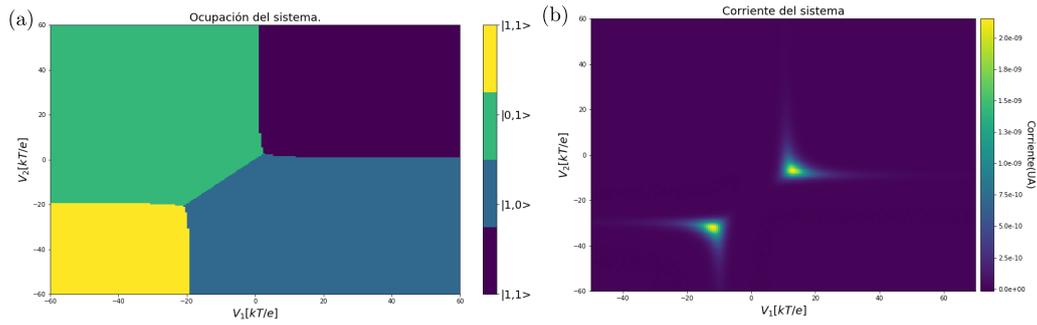


**Figura 3.6:** Corriente de carga a través del sistema.

En cuanto a la corriente, que puede verse en la figura 3.5(b), notamos una forma antisimétrica respecto al cambio de signo en el bias, que está relacionada en forma directa con la simetría de inversión del sistema. En nuestro caso dicha simetría se preserva, ya que los elementos de matriz de tunneling son independientes de los reservorios, es decir,  $t_{r,\ell} = t_\ell$ .

Como dijimos, también se puede considerar el escenario (2) donde se fija  $V$  y se varían en forma independiente los voltajes de compuerta  $V_{G,1}$  y  $V_{G,2}$ . Se resuelve este sistema utilizando el valor  $V = 1 \times 10^{-3} k_B T / e$ . Puede resultar interesante en este caso referir las ocupaciones a la base local, dada por los estados asociados a cada puntos cuántico, en vez de los estados enlazante/anti-enlazante del sistema completo. De esta manera podemos recuperar el diagrama de estabilidad planteado para un sistema acoplado como se mostró en la figura

2.8(b).



**Figura 3.7:** (a) Diagrama de ocupación en la base local para diferentes voltajes de compuerta  $V_{G,1}$  y  $V_{G,2}$ . (b) Corriente a través del DQD.

Como se puede ver en la figura 2.2.1(b), el diagrama de estabilidad de carga adquiere la forma romboidal típica en donde cada región corresponde a un bloque de carga. Se puede observar la carga de cada QD de manera individual, que asociamos a los índices  $(N_L, N_R)$ . Para el caso de la corriente de carga, que es independiente de la base elegida, la figura 3.7(b) muestra cómo esta toma valores finitos en las regiones donde cambia la ocupación del sistema, en concordancia con la situación de punto triple que se muestra en la figura 2.8(b). En este sentido, se puede ver que esto ocurre cuando cambia el número *total* de carga. En cambio, si los voltajes cambian de tal manera que se desocupa un punto cuántico pero se ocupa el otro (límite entre  $(0, 1)$  y  $(1, 0)$ ), manteniendo  $N = N_L + N_R$  constante, entonces dicha corriente no se observa.

Con estos ejemplos finalizamos la introducción acerca de las propiedades de los sistemas conformados por dos puntos cuánticos. Volveremos a este sistema en el capítulo 5, donde tomaremos una dependencia arbitraria en las energías de cada punto cuántico con respecto a los voltajes  $V_{G,1}$  y  $V_{G,2}$ .

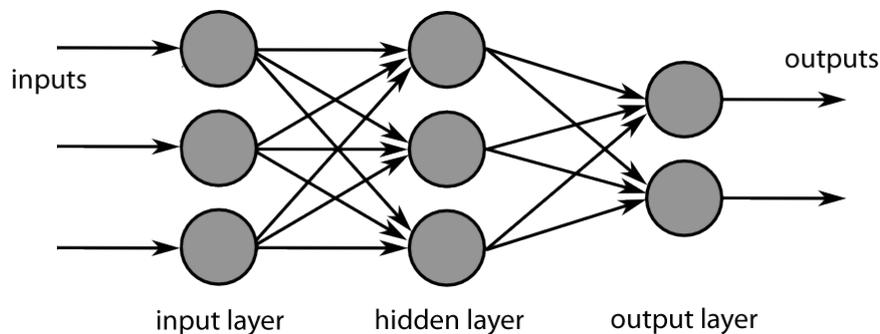
## Capítulo 4

# Algoritmos de aprendizaje automático

*El aprendizaje automático es una técnica de inteligencia artificial por medio de la cual podemos estimar el valor de una función arbitraria sin conocer su expresión funcional, a partir de un conjunto de valores conocidos de la función llamado conjunto de entrenamiento.*

### 4.1. Redes Neuronales

En una red neuronal *feed-forward* las neuronas se organizan en capas sucesivas. Como se puede ver en la figura 4.1 la información ingresa en la llamada capa de entrada, que no es una capa de neuronas. Desde la capa de entrada la información viaja a la primera capa de neuronas. Una vez procesada la información por las neuronas de la primera capa, estas la envían a las neuronas de la segunda capa, hasta llegar a la última capa o capa de neuronas de salida. No puede haber intercambio de información dentro de una capa, ni hacia atrás ni saltar capas hacia adelante. Esta arquitectura tan ordenada facilita la elaboración y la programación de los algoritmos.



**Figura 4.1:** Representación esquemática de una red neuronal.

Podemos descomponer la red feed-forward en tantos perceptrones simples como neuronas en la capa de salida tengamos, como se muestra en la figura 4.2. Para esto llamamos  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  al vector de entrada que pueden tomar valores discretos o continuos acotados o no. El rango de valor que puede tomar la salida  $y$  depende exclusivamente de la función de activación  $g(x)$ . Si esta es discreta, la red aprenderá a clasificar; y si es continua, aprenderá a hacer una regresión. Luego,  $w_i$  representa el peso sináptico.

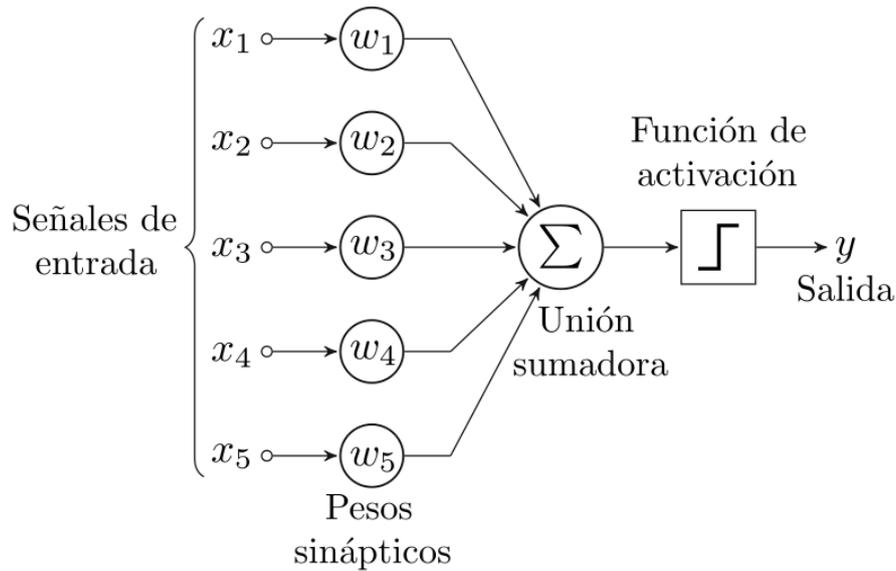


Figura 4.2: Funcionamiento de un perceptrón simple.

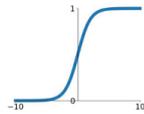
La función activación puede tomar varias formas, las más usuales se presentan en la figura 4.3. Esta función está regulada por un umbral de activación  $\mu$ , es decir, toma la forma:

$$y = g(\mathbf{w} \cdot \mathbf{x} - \mu) \tag{4.1}$$

### Activation Functions

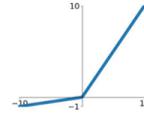
**Sigmoid**

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



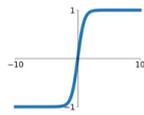
**Leaky ReLU**

$$\max(0.1x, x)$$



**tanh**

$$\tanh(x)$$

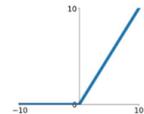


**Maxout**

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

**ReLU**

$$\max(0, x)$$



**ELU**

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

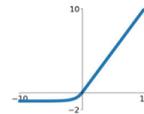


Figura 4.3: Funciones de activación utilizada en los perceptrones.

Supongamos que deseamos modelar cierto proceso que asigna a vectores en  $\mathbb{R}^n$  un único valor  $y$ . No tenemos ninguna información sobre la expresión matemática de dicha asignación, pero podemos imaginarla como una función real que toma una  $N$ -upla de números reales y devuelve un número real. Llamaremos  $f$  a esa relación. Entonces:

$$f(\mathbf{x}) = y \quad f : \mathbb{R}^n \mapsto \mathbb{R}. \tag{4.2}$$

Supongamos que la única información que disponemos es el valor de asignación que esta función asigna a cierto conjunto de valores de entrada. Supongamos que son  $p$  los datos de entrada cuya solución se conoce y los enumeramos con la letra griega  $\alpha$ .

$$\mathbf{x}^\alpha = (x_1^\alpha, x_2^\alpha, \dots, x_n^\alpha) \mapsto y^\alpha. \tag{4.3}$$

Entonces  $\{\mathbf{x}^\alpha, y^\alpha\}$ , con  $\alpha = \{1, \dots, p\}$ , se denomina el conjunto de entrenamiento. Nuestro desafío es encontrar un conjunto posible de pesos sinápticos  $w_i$  y un valor de umbral  $\mu_i$  que sean capaces de hacer que nuestra red asigne correctamente a las entradas del conjunto de entrenamiento las respectivas salidas.

Si el conjunto de entrenamiento es lo suficientemente grande y variado, esta red, que aprendió el conjunto de entrenamiento, será capaz de asignar valores aproximados a casos que no corresponde al conjunto de entrenamiento, fenómeno denominado generalización.

Esto es similar al problema de ajustar los parámetros de una función modelo mediante cuadrados mínimos. Observemos que podemos pensar a los pesos sinápticos, que pueden tomar cualquier valor real, como un vector en  $\mathbb{R}^n$ . Es decir, ambos vectores,  $x$  y  $w$ , viven en el mismo espacio. Si tenemos  $N$  entradas o *features* y  $M$  neuronas de salida, podemos aprender una función arbitraria  $f : \mathbb{R}^N \mapsto \mathbb{R}^M$ .

Como en la mayoría de los casos no se puede aprender perfectamente el conjunto de entrenamiento, se fija una tolerancia que limita el error que comete la red cuando asigna valores sobre el conjunto de entrenamiento, por ello se deja de buscar mejores valores para las pesos sinápticos cuando el error de entrenamiento sea menor a la tolerancia fijada. Además, si buscamos perfeccionar el modelo para el conjunto de entrenamiento provocaremos un sobre ajuste, lo que hará que el modelo funcione impreciso para el conjunto de testeo.

### El descenso por el gradiente

Para la mayoría de casos en los cuales se desea entrenar una red neuronal utilizando aprendizaje supervisado se utiliza el método del descenso por el gradiente en diferentes versiones adaptadas y sofisticadas. Se define una función error, ó *loss*, inspirada en el problema de regresión por cuadrados mínimos:

$$E(w) = -\frac{1}{2} \sum_{\mu=1}^P \sum_{i=1}^M (y_i^\mu - O_i^\mu)^2, \quad (4.4)$$

en donde se suma sobre los elementos del conjunto de entrenamiento y sobre todas las neuronas de la capa de salida con  $y_i^\mu$  los valores de entrenamiento y  $O_i^\mu$  los valores entregados por la red. La idea es comenzar con un vector de entrada aleatorio sobre el conjunto de entrenamiento y presentarle a la red secuencialmente ejemplos del conjunto de entrenamiento. Ante cada presentación calculamos el error o función loss. Entonces la función loss depende de  $M \times N$  coeficientes para un problema dado, por lo que el desafío consiste en encontrar estos  $M \times N$  coeficientes con los cuales el perceptrón lineal resuelve bien todos los ejemplos.

El método del descenso por el gradiente consiste en calcular la dirección de mayor crecimiento para la función loss en función de los coeficientes  $\omega_i$ , para esto debemos calcular las derivadas parciales, utilizando la expresión 4.1 pueden escribirse como:

$$\frac{\partial E}{\partial \omega_{ik}} = - \sum_{\mu=1}^P (y_i^\mu - O_i^\mu) g'(h_i^\mu) x_k^\mu \quad (4.5)$$

con  $h_i^\mu = \sum_{k=1}^N \omega_{ik} x_k^\mu$ . De esta forma, se calcula los nuevos coeficientes  $\omega_{ik}$  utilizando un hiperparámetro llamado "learning rate"

$$\omega_{ik}^{nuevo} = \omega_{ik}^{anterior} - \eta \frac{\partial E}{\partial \omega_{ik}}. \quad (4.6)$$

### Formalizando el problema de aprendizaje automático.

Todo problema de aprendizaje automático supervisado comienza con los mismos ingredientes, sin importar el abordaje.

- El primer ingrediente es el conjunto de datos (dataset)

$$\mathbb{D} = (\vec{x}^\mu, \vec{y}^\mu) \quad \mu = 1, 2, \dots, p \quad (4.7)$$

- El segundo ingrediente es un modelo que explica la distribución del conjunto de datos  $f(\vec{x}, \omega)$ , donde  $\omega$  representa el conjunto de parámetros  $\omega = \{\omega_1, \omega_2, \dots, \omega_q\}$ .

$$f(\vec{x}, \omega) = \vec{O} \quad f : \vec{x} \longrightarrow \vec{O} \quad (4.8)$$

- El tercer ingrediente es la función costo, error o loss:

$$C(\vec{x}, f(\vec{x}; \omega)) \quad (4.9)$$

que nos permite evaluar cuán bien funciona el modelo  $f$  para explicar los datos  $\vec{x}$

En el aprendizaje supervisado, cualquiera sea el método que usemos, dividiremos siempre al conjunto de datos disponibles  $X$  en dos subconjuntos:

$\mathbb{D}_{train}$  : Conjunto de entrenamiento

$\mathbb{D}_{test}$  : Conjunto de testeo

Estos conjuntos son mutuamente excluyentes y la división debe realizarse con mucho cuidado y aleatoriamente. Una regla general es dedicar el 90% de elementos al conjunto de entrenamiento y un 10% para el conjunto de testeo. Es importante aclarar que el conjunto de acoplamientos que obtendremos minimiza localmente la función loss sobre el conjunto de entrenamiento. Aquel modelo que alcance al final el menor valor para la función costo sobre el conjunto de entrenamiento será el que mejor aprendió. El modelo que alcance el menor valor para la función costo sobre el conjunto de testeo será el modelo que mejor funcione, o sea, el que mejor generalice resolviendo bien aquellos casos sobre los cuales no fue entrenado.

**Optimización del descenso por el gradiente.** En esta sección se presentan técnicas y algoritmos que optimizan los resultados del descenso por el gradiente.

- **Minibatch:** Esta técnica consiste en dividir el conjunto de entrenamiento en mini grupos por lo que en cada instancia de entrenamiento se hace una reasignación aleatoria de todos los ejemplos de manera que  $Qm = p$  (se tienen  $p$  elementos del conjunto de entrenamiento divididos en  $m$  grupos de tamaño  $Q$ ). Con esta técnica se pretende introducir aleatoriedad en el método del descenso por el gradiente al sortear época a época los elementos del conjunto de entrenamiento. En general se encuentra que la cantidad de minibatches debe ser lo suficientemente alta para lograr introducir la aleatoriedad deseada, sin embargo si la cantidad de particiones es demasiado alta también lo será el costo computacional pues el descenso por el gradiente se realizará muchas veces. Cuando los batch son muy grandes (muchos elementos de entrenamiento) la función costo tiende a converger hacia mínimos más agudos (pendientes altas). Por su parte, dividir los ejemplos en conjuntos de elementos más pequeños para analizar produce una convergencia del sistema hacia mínimos más planos (pendientes menores). Esto quiere decir que los batch más grandes tienden hacia mínimos de los cuales es difícil despegarse mientras que los batch más pequeños, al ser atraídos hacia mínimos con formas más planas tienen la capacidad de escapar de las regiones de mínimos locales.
- **Dropout:** Para reducir el *overfitting* y permitir que una red neuronal artificial responda correctamente cuando es testada, se aplica el *dropout*. Este algoritmo implica suprimir un cierto número de neuronas de manera aleatoria por lo que cuando se calcula la salida no se considera toda la información. Desde este punto de vista, cada neurona presente en las capas ocultas son omitidas al azar con una probabilidad  $p$  tal que estas neuronas no dependen del resto. Con esto se pretende evitar el establecimiento de adaptaciones complejas donde cada característica a detectar es útil y posible sólo en el contexto planteado por todas las demás. Es decir, cada neurona aprende a detectar una característica que permite producir la respuesta correcta en una multiplicidad de contextos internos posibles y variados en los cuales opera la red. En sentido gráfico, el método del descenso por el gradiente no es capaz de salir de mínimos locales y requiere tiempos muy largos para salir de un punto de inflexión, tiempo que crece exponencialmente a medida que aumenta el número de componentes en la cual es plano. Este problema se mitiga con la aplicación del *dropout*, al suprimir una de las dimensiones donde podría estar ubicado un mínimo local. Es posible encontrar un paralelismo entre este mecanismo de la red artificial (dropout) con el funcionamiento del cerebro. Las neuronas cerebrales requieren de un tiempo para recuperar la energía luego de la emisión de señales a neuronas vecinas de manera que, por un tiempo, se podría decir que permanecen apagadas.

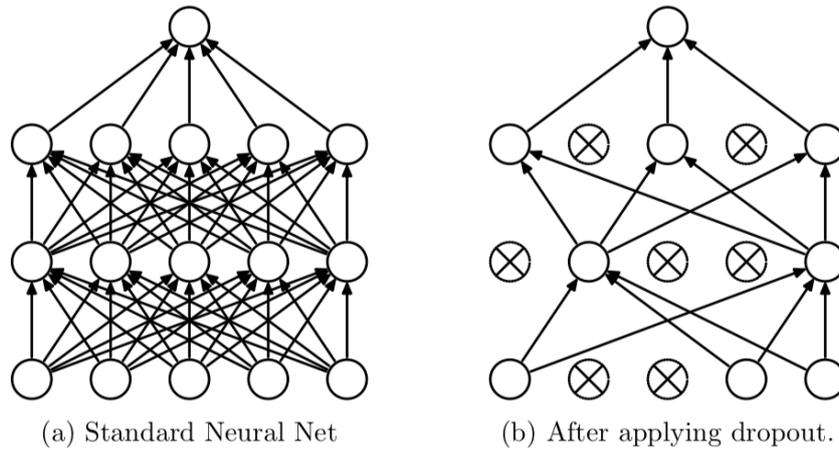


Figura 4.4: Esquemización del dropout en una red neuronal

- Entropía cruzada como función pérdida:** La función error se ha definido como la suma de los cuadrados de la diferencia entre los valores deseados y la salida obtenida por la red. Por la naturaleza de esta función y dado que el método del descenso por el gradiente se definió proporcional a  $\frac{\partial E}{\partial \omega}$ , cuanto más próximos estemos al mínimo donde sea que  $E=C=0$  menor será la velocidad (pendiente) con la que nos acerquemos. Si esta función costo fuera reemplazada por una función logarítmica, el descenso por el gradiente sería cada vez mayor. Esto se suele usar como una métrica para saber cuán cerca estamos de la solución correcta. En especial, para los casos en los que se necesita clasificar un conjunto se utiliza la entropía cruzada como función perdida.

La entropía cruzada (o *cross-Entropy*) mide la entropía relativa entre dos distribuciones de probabilidad diferentes,  $P(x)$  y  $Q(x)$  definidas sobre el mismo conjunto de eventos  $x$ .

$$H(P, Q) = - \sum_x P(x) \log_2(Q(x)) \tag{4.10}$$

Veamos ahora como usar la entropía cruzada en inteligencia artificial a través de un ejemplo de clasificación.

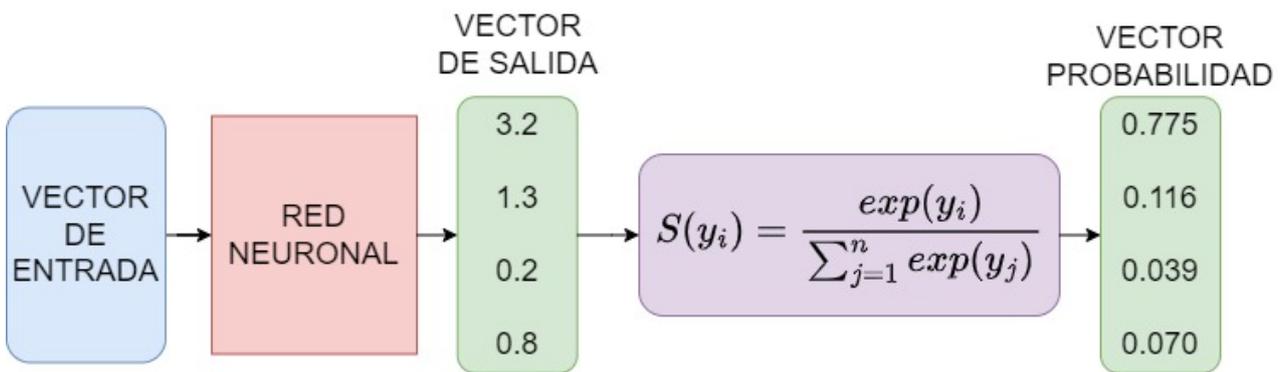


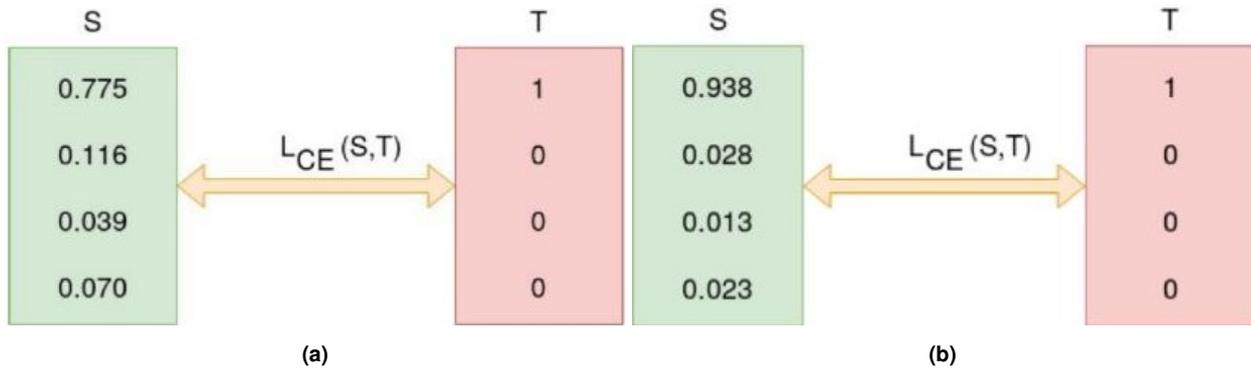
Figura 4.5: Entropía cruzada como función pérdida en una red neuronal.

En la entrada tenemos imágenes de 4 tipos de animales (perros, gatos, caballos y chitas). Una red neuronal feed-forward calcula valores para las 4 neuronas de salida (*logit* =  $\frac{\lg(x)}{\lg(1-x)}$ ), y una función simple llamada *softmax* opera sobre las cuatro variables de salida de forma tal que:

$$0 \leq S(y_i) \leq 1$$

$$\sum_i S(y_i) = 1$$

Estos valores no son estrictamente probabilidades, sino estimadores. Como el experto que etiquetó el conjunto de entrenamiento sabe que es un perro, entonces conocemos la verdadera distribución (T) del experimento. Supongamos que le mostramos la misma foto pero varias épocas después y obtenemos:

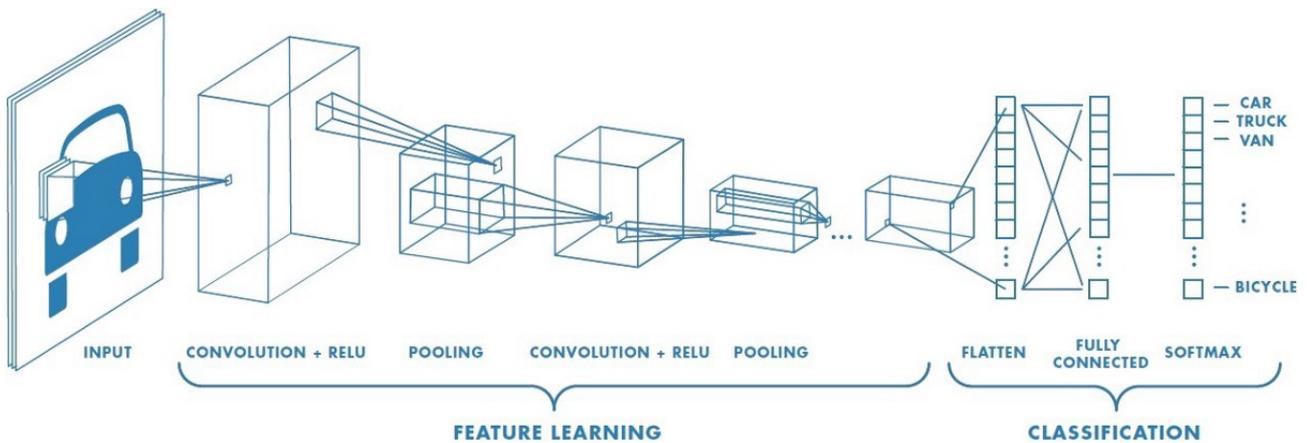


**Figura 4.6:** Vectores de salida para diferentes épocas. a) Distribución de probabilidades para la primera época. b) Distribución de probabilidades para una época mas avanzada.

Entonces para el primer caso tenemos  $L_{CE} = -\log(0.775) = 0.3677$  y para el segundo caso  $L_{CE} = -\log(0.938) = 0.095$ , es decir, a medida que la distribución de probabilidad se aproxima a la correcta, el valor de la entropía cruzada disminuye.

## 4.2. Redes Neuronales Convolucionales (RNC)

Las RNC, a diferencia de las redes feed-forward multi-capas, están diseñadas para conservar la relación espacial de los datos. Esto es muy importante en el caso de procesamiento de imágenes y de series temporales, donde la proximidad en los datos de entrada es un dato de vital importancia.



**Figura 4.7:** Esquemización de una red neuronal convolucional.

Las RNC están organizadas para recibir de entrada objetos tridimensionales, caracterizadas por el ancho  $W$ , el alto  $H$  y la profundidad  $D$ . La necesidad de tres dimensiones surge por ejemplo de la necesidad de considerar diferentes canales en las imágenes digitales, como por ejemplo la que usa en la convención RGB (red, green, blue). En nuestro caso las imágenes tendrán profundidad  $D=1$  que representará la corriente medida.

La arquitectura de las RNC se caracteriza por tener Capas Ocultas Convolucionales, que difieren de las capas densas (propias de las redes feed-forward), en el hecho de que puede detectar cierto patrón en una región local de dimensión  $d$  (la dimensión de la entrada). El nombre de estas capas guarda relación con la operación de convolución matemática que denotamos, como siempre se hace, con el símbolo  $*$ .

**Definición: Convolución discreta:** Sean  $f$  y  $g$  dos funciones enteras, continuas por tramos en  $\mathbb{R}$ . Definimos la **convolución** unidimensional de  $f$  y  $g$  como:

$$(f * g)(m) = (f(n) * g(n))(m) = \sum_n f(n)g(m - n) \tag{4.11}$$

Las capas convolucionales de las redes convolucionales aplican la llamada correlación cruzada a un tensor de entrada en común que una RNC tenga muchas capas convolucionales.

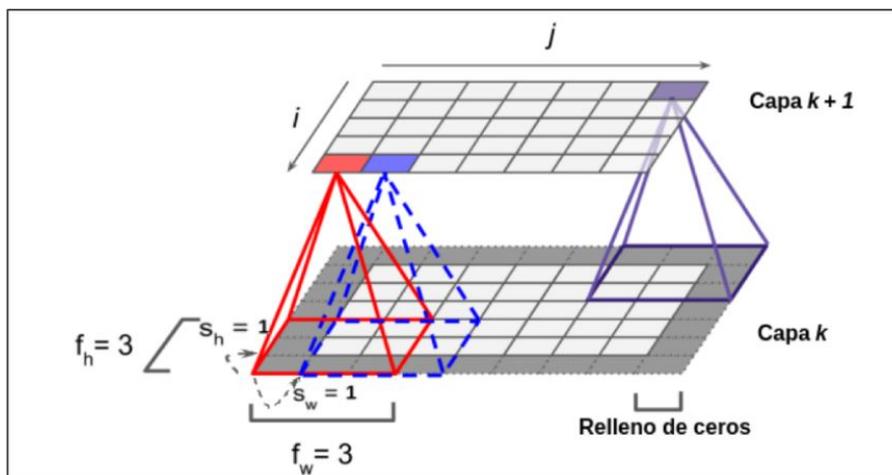
**Definición de filtro de convolución:** Un filtro de convolución bidimensional es un tensor de tamaño  $f_W \times f_H \times D$ , donde  $f_W$  y  $f_H$  son el ancho y el alto del núcleo (kernel) y  $D$  es la profundidad, que en general coincidirá a lo largo de la RNC con la profundidad  $D$  de la entrada.

Las entradas de los filtros son los pesos de las conexiones que varían entre cortes frontales. Como las imágenes son representadas por tensores de orden 3 y cada corte frontal es de tamaño  $W \times H$ , donde cada entrada es el valor del pixel correspondiente, se aplica la correlación cruzada discreta en dos dimensiones entre los pixeles seleccionados de la imagen y los pesos de un filtro. Esto es, se superponen el  $l$ -ésimo corte frontal del filtro sobre el  $l$ -ésimo corte frontal de la entrada de capa y se suman las multiplicaciones lugar a lugar:

$$(f * g)(m_1, m_2) = \sum_{n_1} \sum_{n_2} f(n_1, n_2)g(m_1 + n_1, m_2 + n_2) \tag{4.12}$$

Supongamos que tenemos una imagen cuadrada de ancho  $W$ . Si aplicamos un filtro bidimensional de  $f_W$  (o sea,  $f_H = f_W$  y  $D = 1$ ), entonces perderemos  $f_W - 1$  filas y  $f_W - 1$  columnas, a las cuales no le aplicamos el filtro.

Las neuronas de la primera capa convolucional no están conectadas a cada uno de los pixeles de la imagen de entrada, sino solo a los pixeles de sus campos receptivos. Durante la propagación hacia adelante, se realiza la correlación cruzada de cada filtro a través del ancho y alto del volumen de entrada y se produce un mapa de características bidimensionales con los resultados de esta operación. No hay un único filtro, sino muchos, y entonces la salida de la red convolucional apila muchas salidas.



**Figura 4.8:** Funcionamiento del kernel en una red neuronal convolucional.

Al apilar estos mapas a lo largo de la profundidad, se obtiene el volumen de salida. Todas las neuronas dentro de un mapa de características (*feature map*) comparten los mismos pesos sinápticos (parámetros), lo que reduce significativamente el número de parámetros del modelo, pero además las neuronas de los diferentes mapas de características usan distintos parámetros. El tamaño del volumen de salida de cada capa convolucional depende del tamaño del núcleo, la profundidad  $D$ , el paso o stride ( $S_W$  y  $S_H$ ) y el relleno de ceros o zero padding ( $P$ ).

## Capítulo 5

# Análisis del modelo de puntos cuánticos utilizando aprendizaje automático

Si bien cada sistema de puntos cuánticos/reservorios está relacionado con un modelo teórico, muchas veces se desea conocer los parámetros reales que lo determinan una vez construido. Entre los parámetros tenemos voltajes de compuerta, la temperatura, el voltaje de bias, energía de interacción, etc. A medida que el sistema contiene más elementos, más compleja se vuelve la tarea de calcular estos valores.

He aquí el núcleo de este trabajo, en donde la idea es utilizar la teoría desarrollada en el Capítulo 3 para simular sistemas. Variando los parámetros del sistema se crea una base de datos, la cual permite entrenar un modelo de aprendizaje automático, con el cual se podrán inferir el valor de los parámetros para un sistema real del cual el modelo teórico es una aproximación. Dando como entrada el mapa de corriente producido por un sistema real la red neuronal devuelve los parámetros del mismo.

Es claro que en caso de contar con una base de datos conformada por mapas de corriente experimental se puede utilizar un modelo entrenado para identificar regiones de interés y sintonizar los valores de los parámetros de control para que el sistema experimental funcione en el régimen deseado.

Se trabaja con un modelo de punto cuántico doble en donde cada dato de la base consiste en un mapa de corriente variando los voltajes de compuerta ( $V_1, V_2$ ).

### 5.1. Modificación del modelo

El Hamiltoniano para un punto cuántico doble en serie (Ec. 3.23) contiene la energía de cada punto,  $e_i = V_i e$ , en donde  $e_i$  es la energía del punto,  $V_i$  es el voltaje de compuerta y  $e$  la carga del electrón. Sin embargo, debido a las dimensiones de estos sistemas ( $\sim 1\mu m$ ), es lógico pensar dos cosas:

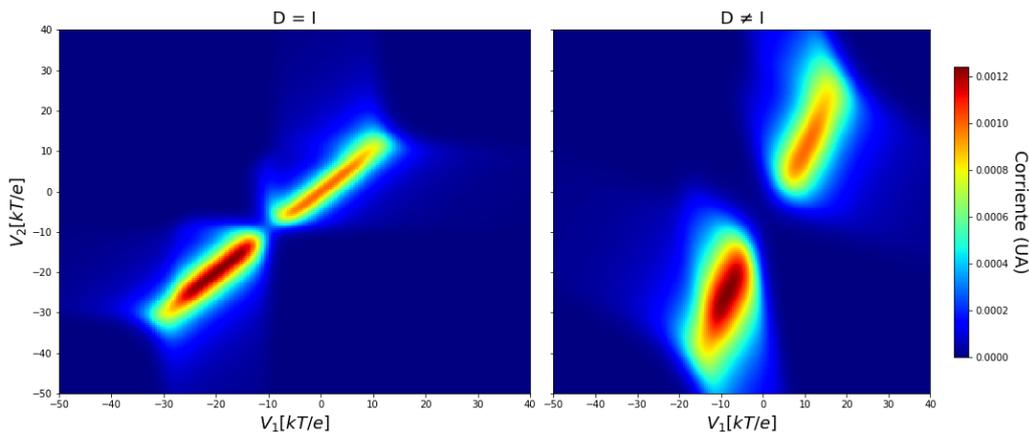
En primer lugar, el voltaje de compuerta que se aplica (o medido) no necesariamente es el voltaje que experimenta el electrón. Segundo, debido a las proximidades entre los puntos, el voltaje de compuerta de uno, puede afectar la energía del otro.

Esta situación se puede plasmar introduciendo una matriz de correlación  $\mathbf{D}$  entre los voltajes "medidos" y las energías de los puntos (Ec. 5.1). Por simplicidad, ya que el voltaje y la energía son proporcionales a la carga  $e$ , la matriz de correlación se aplica entre energía medida ( $e_i$ ) y energía real del punto ( $\epsilon_i$ ), es decir, la diferencia entre la medición del voltaje aplicado y el voltaje que siente el sistema.

$$\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \quad (5.1)$$

Es decir, las energías  $e_1$  y  $e_2$  son la medición directa del aparato y las energías  $\epsilon_1, \epsilon_2$  son las que van en el Hamiltoniano 3.23. Los coeficientes  $d_{12}$  y  $d_{21}$  representan la influencia de los voltajes en los puntos adyacentes y  $d_{11}, d_{22}$  representa la influencia sobre el mismo punto. Es de esperar entonces que los coeficientes  $d_{12}$  y  $d_{21}$  sean cercanos a cero y que los coeficientes  $d_{11}, d_{22}$  sean de orden uno en las unidades adecuadas al problema.

La influencia de estos coeficientes se puede ver en la figura 5.1 en donde se calcula bajo los mismos parámetros, dos mapas de corriente para un sistema en donde  $D = I$  y luego se repite el procedimiento para un sistema en donde  $D \neq I$ .



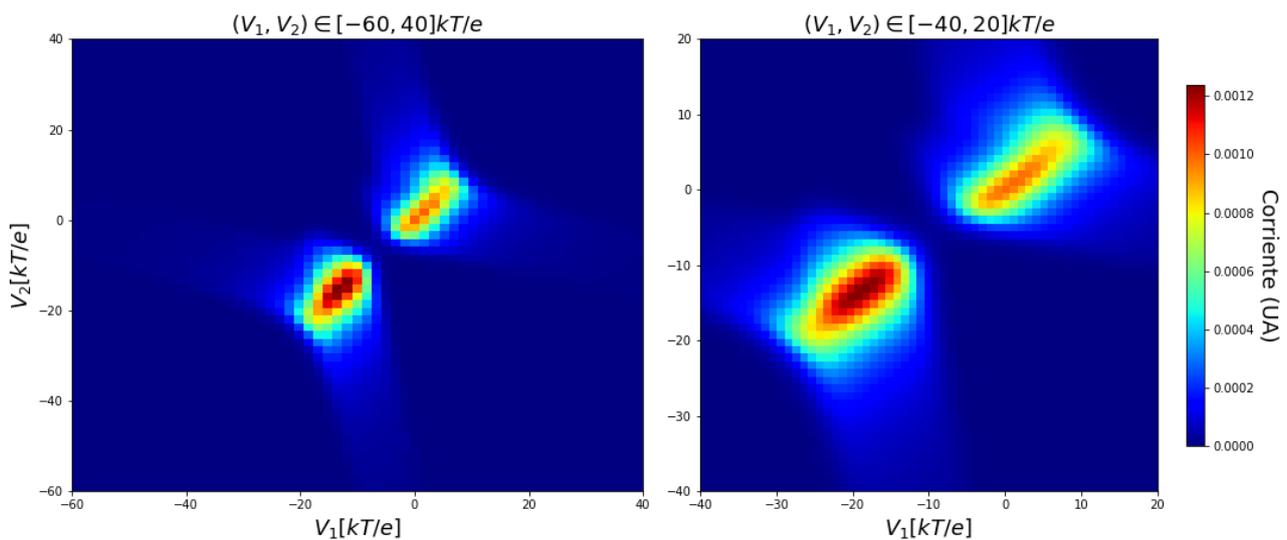
**Figura 5.1:** Comparación entre un sistema con  $D = I$  y  $D \neq I$

Es aquí en donde entra el objetivo de este capítulo. Se busca crear una base de datos, variando estos coeficientes utilizando una distribución normal y con dicha base entrenar una red neuronal. La idea es que al crear la base de datos uno conoce los valores para la matriz  $D$ , entonces, una vez entrenada la red, uno puede pasar los valores de corrientes medidos en cierto sistema y el modelo podrá predecir los coeficientes del mismo. En resumen, el mapa de corriente será el "vector de entrada" en la red neuronal y la matriz de correlación el "vector de salida".

## 5.2. Base de datos

El primer paso para la creación de una base de datos de entrenamiento es decidir que estructura es conveniente para los datos. Es decir, decidir que parámetros quedarán fijos y cual será el valor de estos. Ya que la idea es realizar un análisis exploratorio sobre la ejecución de aprendizaje automático aplicado a un sistema de puntos cuánticos, en principio se deja todo parámetro, más allá de los voltajes de compuerta y la matriz de correlación, fijos. Sin embargo, en futuros modelos se puede agregar una variación de estas sobre la base de datos, para luego predecirlos o no.

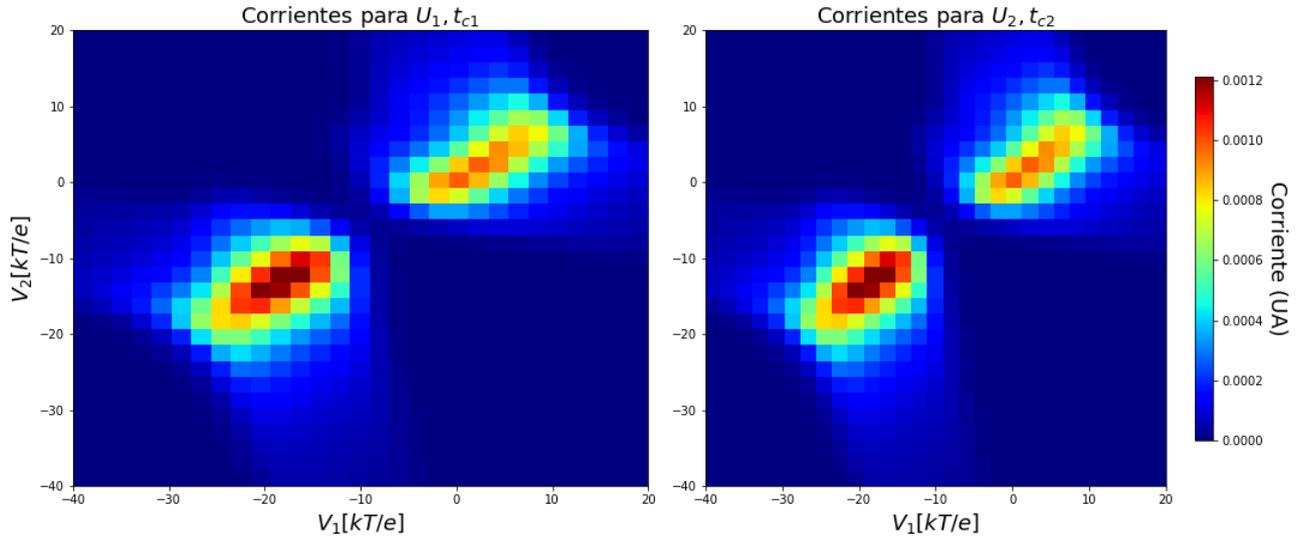
En primer lugar se explora el rango de voltajes que se debe utilizar y cuan grande es la zona de corriente no nula para diferentes valores de  $D$ . Para esto se calculan varios mapas de corrientes con diferentes parámetros, tal como se puede ver en la figura 5.2.



**Figura 5.2:** Cálculo de corrientes para diferentes valores de  $D$ ,  $T$ ,  $t_c$ ,  $U$ , etc. Ambos mapas con una definición de 60x60 píxeles.

Observando en la figura 5.2 la variación relativa entre los valores de corriente no nulos para el sistema, se decide bajar la resolución de las imágenes a 30x30 píxeles (figura 5.3). Esto también se debe a que el objetivo

de esta sección será predecir cuatro coeficientes, por lo que se considera que una imagen de 30x30 contiene la información necesaria para que el modelo pueda realizar esta predicción. Además, cuanto mas alta sea la resolución, mas alto será el tiempo de ejecución, tanto para la creación del dataset, como para el entrenamiento del modelo.



**Figura 5.3:** Cálculo de corrientes para diferentes valores de  $\mathbf{D}$  con una definición de 30x30 píxeles utilizando los valores de la tabla 5.4.

Finalmente, se decide establecer los parámetros en los valores presentados en la tabla 5.4. Se elije una temperatura tal que los estados del sistema estén bien definidos, ya que al influir en la distribución de Fermi, cuanto más lejana sea al cero absoluto, menos escalonada será la distribución. Por otro lado, el intervalo para los voltajes se elije tal que contenga los dos picos de corriente característicos de un sistema de doble punto cuántico.

$T$	$1K$
$t_c$	$5kT$
$U$	$30kT$
$V_{1,2}$	$[-40, 20][kT/e]$
$V_{bias}$	$15[kT/e]$

**Figura 5.4:** Parámetros utilizados en la simulación de un sistema de punto cuántico doble.

Una vez definidos los valores para estos parámetros, el siguiente paso es ejecutar un algoritmo para la creación de una base de datos (dataset).

### 5.2.1. Algoritmo para la creación del dataset

Para la creación de la base de datos se utilizan dos programas, ambos ejecutados en el lenguaje "Fortran 90". Esta elección se basa en que para cada valor de corriente, se debe calcular una cantidad considerable de coeficientes (elementos del kernel " $W_{a_i}^{a_j}$ ") y que cada para cada imagen se deben calcular 30x30 valores de corriente. Otros de los puntos a favor para utilizar este lenguaje es que cuanto más complejo sea el modelo, mejor es la decisión de utilizar un lenguaje diseñado para el cálculo numérico.

### 5.2.2. Ejecución del algoritmo

- **Módulo con las funciones necesarias:** Este módulo contiene todas la funciones y subrutinas necesarias para calcular el valor de una corriente, desde las funciones de Fermi, hasta subrutinas para el cálculo de los elementos del kernel. Además contiene las subrutinas que asignan en base a una distribución normal los valores para la matriz  $\mathbf{D}$ . Los coeficientes diagonales ( $d_{11}, d_{22}$ ) se generan con una media

$\mu = 1$  y una desviación  $\sigma = 0.1$ , para el caso de los coeficientes no diagonales ( $d_{12}, d_{21}$ ) se utiliza  $\mu = 0$  y  $\sigma = 0.1$  tomando el valor absoluto generado por la distribución ya que que el voltaje no puede afectar con un peso negativo al punto adyacente.

- Programa para la creación de la base de datos:** Este programa lee los parámetros necesarios para crear el data set, tal como el largo del vector de coeficientes (la matriz **D** se guarda en un vector) como el ancho y alto de la imagen y el tamaño del dataset y los guarda en un archivo 'double\_dot.h5'. Luego ejecuta una iteración  $N$  veces en donde  $N$  es el tamaño del dataset, la cual genera los coeficientes  $d_{ij}$  y luego en base a estos calcula la matriz de corriente utilizando las funciones del módulo óccup.f90 para luego grabar esto valores en el archivo.

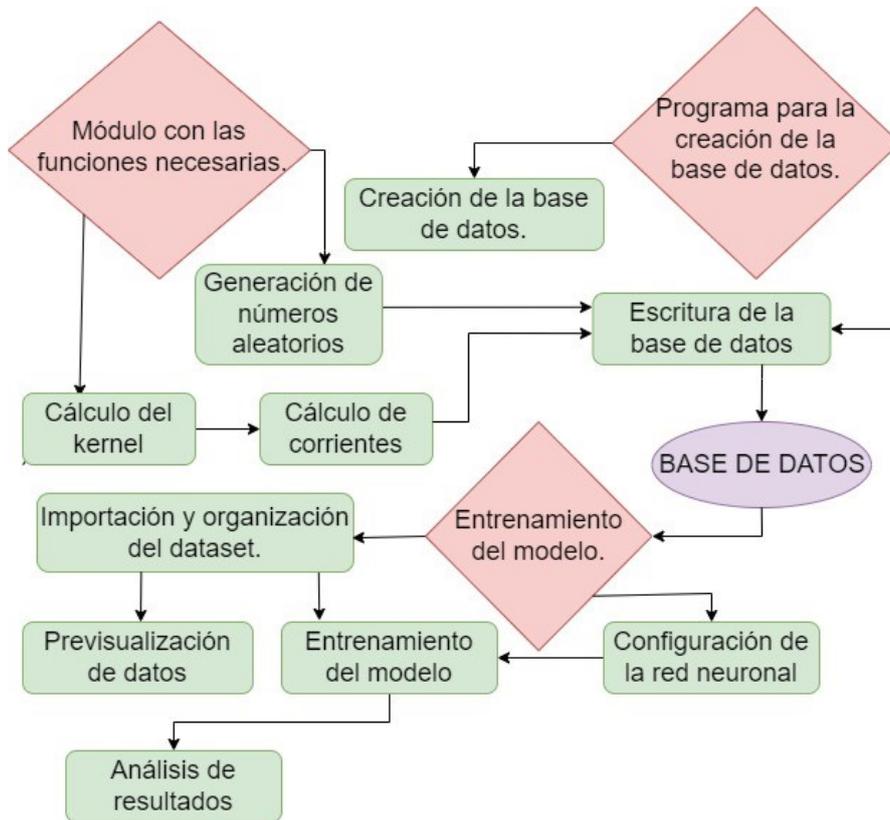


Figura 5.5: Descripción del proceso para la creación y análisis del dataset.

### 5.3. Entrenamiento del modelo

Una vez obtenido el dataset, se comienza con el entrenamiento de los modelos siguiendo el proceso mostrado en la figura 5.5. En un principio se visualizan los datos generados con el objetivo de corroborar la correcta generación e importación de los mismo. En la figura 5.6 se pueden observar dos corrientes generadas a partir de diferentes valores para  $d_{ij}$  utilizando los modelos planteados en el capítulo 3.4.

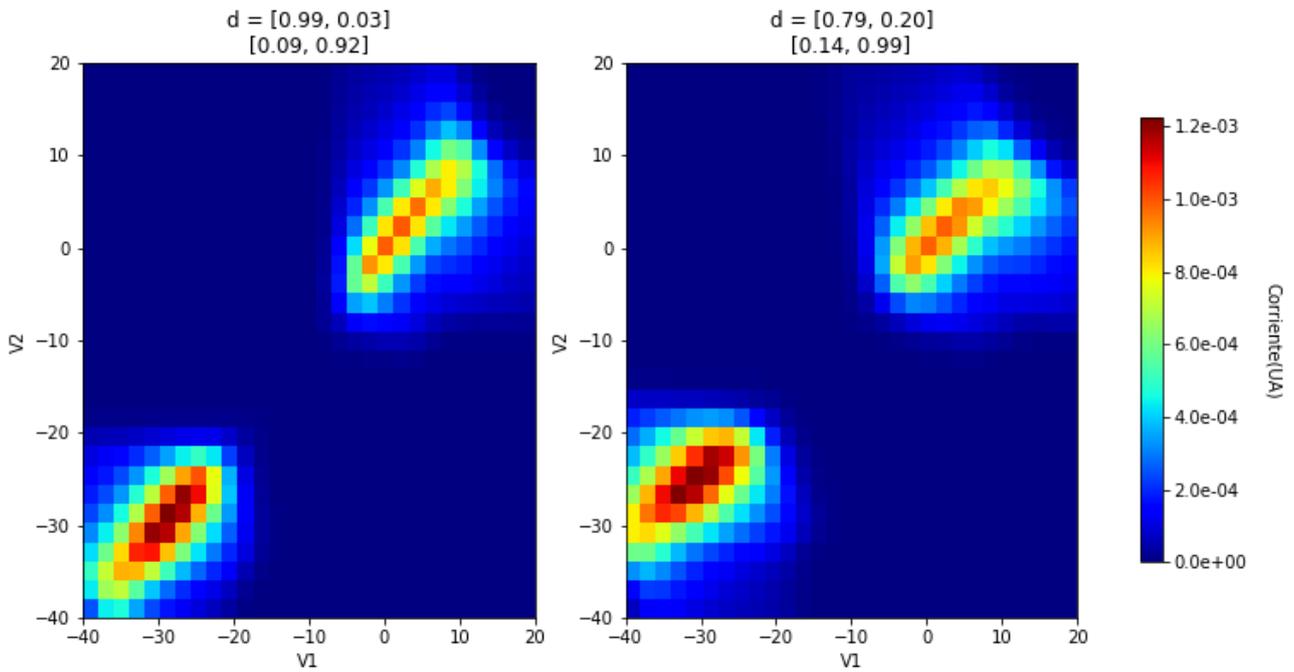


Figura 5.6: Corrientes simuladas por el programa con diferentes valores de  $D$

Para el entrenamiento de diferentes modelos a lo largo de este trabajo, ciertos parámetros del mismo no varían, por lo que se los indica una sola vez en la tabla 5.7.

Función activación	ReLU
Dropout	0.25
Radio de aprendizaje	0,001
Función pérdida	Entropía cruzada
Método de minimización	Descenso por el gradiente
Optimizador del gradiente	ADAM optimizer

Figura 5.7: Parámetros utilizados en los entrenamientos a lo largo del trabajo

### 5.3.1. Primer modelo (CNN1)

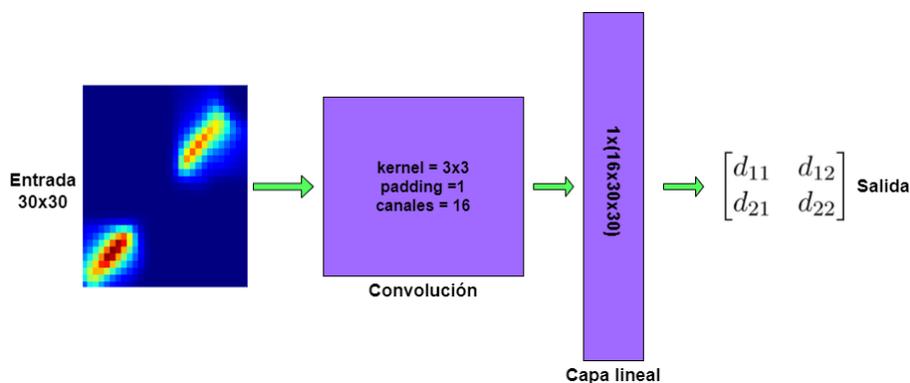
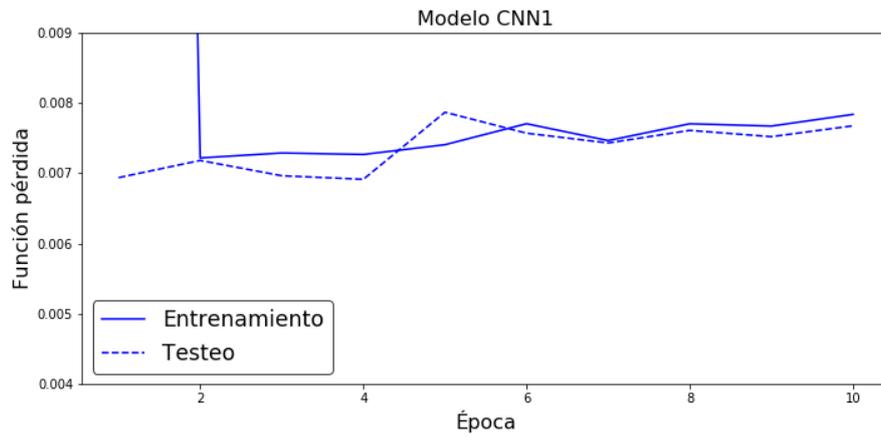


Figura 5.8: Arquitectura del primer modelo utilizado.

En primera instancia se utiliza una arquitectura sencilla de una capa convolucional con un kernel de 3x3 con un padding=1 y 16 canales de salida. Entonces, la dimensión del tensor de salida es de 16x30x30, el cual se achata a un tensor de rango 2 con dimensiones 1x(16x30x30). Se pasa este tensor por una capa lineal de neuronas. Se entrena el modelo durante diez épocas.



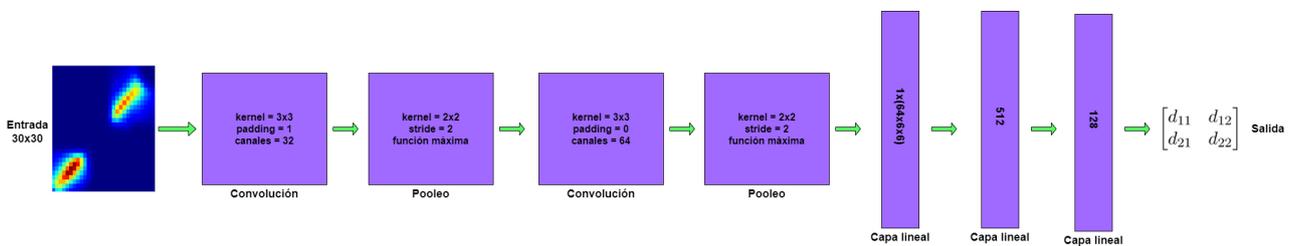
**Figura 5.9:** Entrenamiento del primer modelo durante diez épocas

Como se puede ver en la figura 5.9, el modelo obtiene la mayor precisión posible en las diez instancias de entrenamiento, logrando predecir los coeficientes hasta el segundo dígito. Aumentar las épocas de entrenamiento, solo haría que el modelo sobreajuste el conjunto de entrenamiento, bajando la precisión en la predicción sobre el conjunto de validación.

	$d_{1,1}$	$d_{1,2}$	$d_{2,1}$	$d_{2,2}$
Valor real	1.01	0.10	0.10	0.96
Valor predicho	1.09	0.07	0.14	1.08

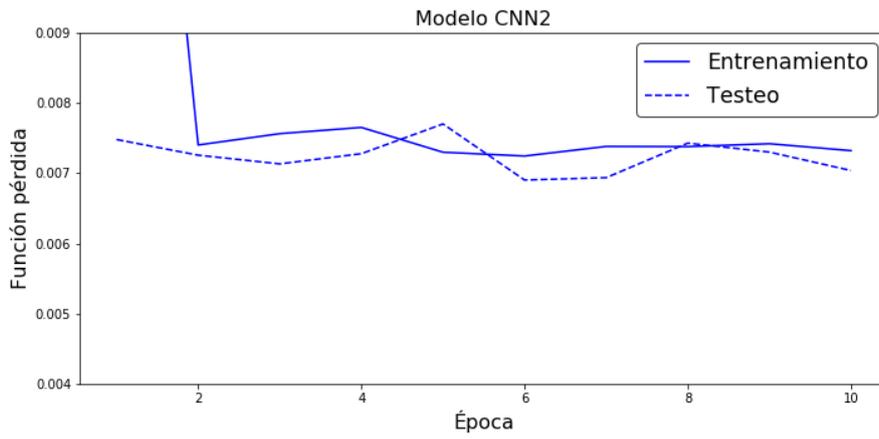
**Figura 5.10:** Valores reales vs valores predichos para cierto dato utilizando el modelo CNN1

### 5.3.2. Segundo modelo (CNN2)



**Figura 5.11:** Arquitectura del segundo modelo utilizado.

En esta instancia, se utiliza un modelo mas complejo en busca de aumentar el alcance del mismo. Se trabaja con una red convolucional de dos convoluciones y dos capas ocultas. En la primera convolución se utiliza un kernel de 3x3 con un padding=1 y 32 canales de salida, para el pooleo se usa un kernel 2x2 y un stride de 2 con la función máxima (toma el valor máximo del kernel). Para la segunda convolución, se utiliza un kernel de 3x3 con un padding nulo y 64 canales de salida, para el pooleo se usa un kernel 2x2 y un stride de 2 con la función máxima. Entonces, la dimensión del tensor luego de la convolución queda de 64x6x6, el cual se achata a un tensor de rango 2 con dimensiones 1x(64x6x6). Luego se pasa este vector por una capa oculta de 512 neuronas y una segunda capa de 128 neuronas.

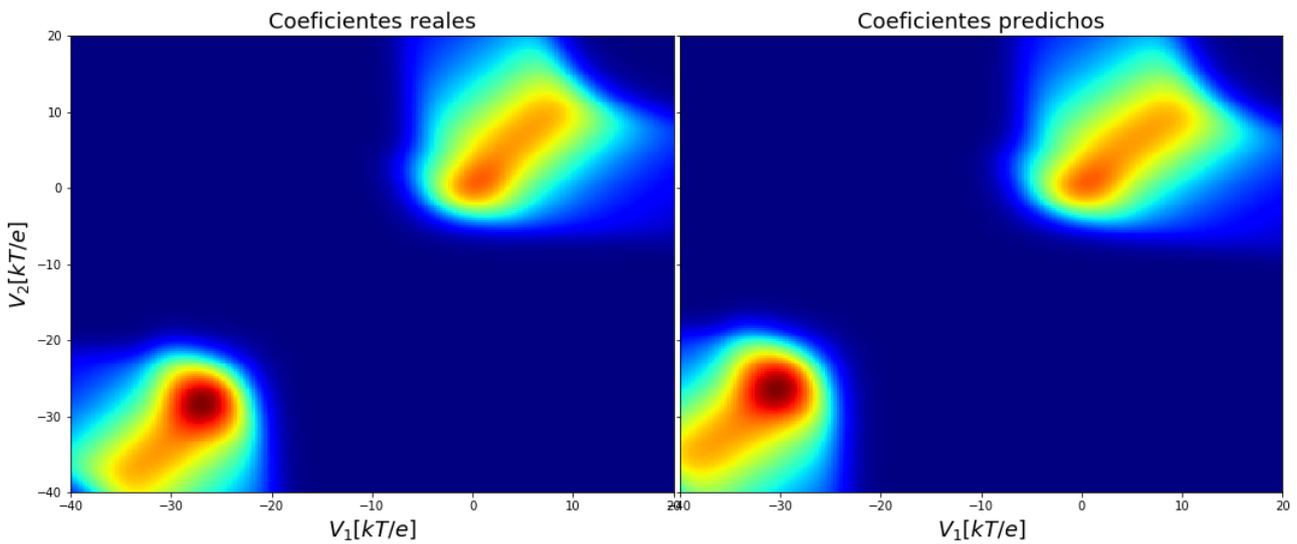


**Figura 5.12:** Entrenamiento del segundo modelo durante diez épocas

	$d_{1,1}$	$d_{1,2}$	$d_{2,1}$	$d_{2,2}$
Valor real	1.03	0.03	0.08	1.03
Valor predicho	1.01	0.11	0.02	0.97

**Figura 5.13:** Valores reales vs valores predichos para cierto dato utilizando el modelo CNN2

Como se puede ver en la figura 5.12, no solo el modelo obtiene su máxima precisión en las diez épocas de entrenamiento, si no que además, el alcance es el mismo al del primer modelo. Esto indica que no es necesario aumentar la complejidad de las CNN. Esto se debe principalmente a la manera en la que fueron simulados los datos. Para los cuatro coeficientes  $d_{ij}$  se utilizó un desviación estandar de  $\sigma = 0.1$  y es aquí donde llega la precisión del modelo. Es decir, para aumentar el alcance de estos modelos, se debe modificar la manera en la que se generan los datos de entrenamiento.



**Figura 5.14:** Cálculo de corrientes para los valores de  $\mathbf{D}$  vs el valor arrojado por la red.

En la figura 5.14 se puede observar como la predicción de estos coeficientes se da con la precisión necesaria como para replicar el mismo mapa de corriente. Es decir, si para cierto valor de  $\mathbf{D}$ , se calcula la corriente 30x30 y se pasa esta matriz por la red convolucional, el valor que devuelva esta permitirá replicar el mapa de corriente ya con la definición deseada, en el caso de la figura 5.14 es de 200x200 píxeles. Si bien los mapas se encuentran desplazados levemente, el comportamiento del sistema se logra replicar.

## 5.4. Ruido en las mediciones

Hasta el momento se trabajó con valores calculados de manera teórica. Sin embargo, como en cualquier medición uno debe lidiar con la incertidumbre de la misma, por lo que el siguiente paso lógico es agregar ruido a los corrientes calculadas.

### 5.4.1. Generación de ruido

Se agrega una nueva rutina al módulo que contiene las funciones necesarias, la cual devuelve una matriz de tamaño  $L \times L$  ( $L$  valor de entrada) con valores entre 1 y -1 generados por una distribución uniforme como se puede ver en la figura 5.15. Luego se modifica el programa en donde se genera la base de datos para que a la par de escribir el arreglo de corriente en el archivo, escriba un arreglo  $30 \times 30$  de ruido de manera paralela. Es decir, ahora cada ítem del dataset consiste de tres elementos: un arreglo  $30 \times 30$  con las corrientes calculadas utilizando los valores de  $\mathbf{D}$ , un arreglo  $2 \times 2$  con los valores de  $\mathbf{D}$  y un arreglo  $30 \times 30$  con el ruido generado.

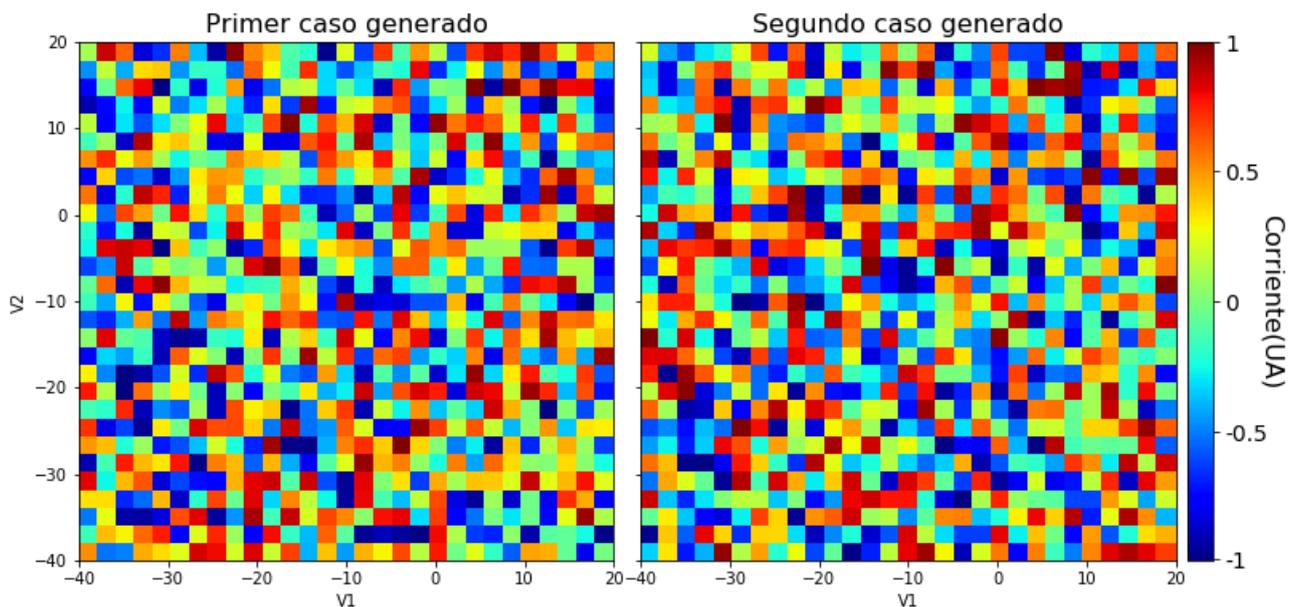


Figura 5.15: Ruido generado junto a los dos primeros mapas de corriente simulados.

### 5.4.2. Importación del ruido

La inclusión del ruido a las corrientes simuladas se realiza a la hora de importar la base de datos de la siguiente manera:

$$I_c = I_{c_0}(1 + \delta \times noise) \quad (5.2)$$

en donde  $I_{c_0}$  es la corriente calculada de manera teórica, *noise* es el ruido generado y  $\delta$  es el peso de la incertidumbre, es decir, si  $\delta = 0.01$  representa un error relativo del 1%. En la figura 5.16 se puede visualizar como a medida que aumenta el peso relativo del error, la imagen contiene mas ruido.

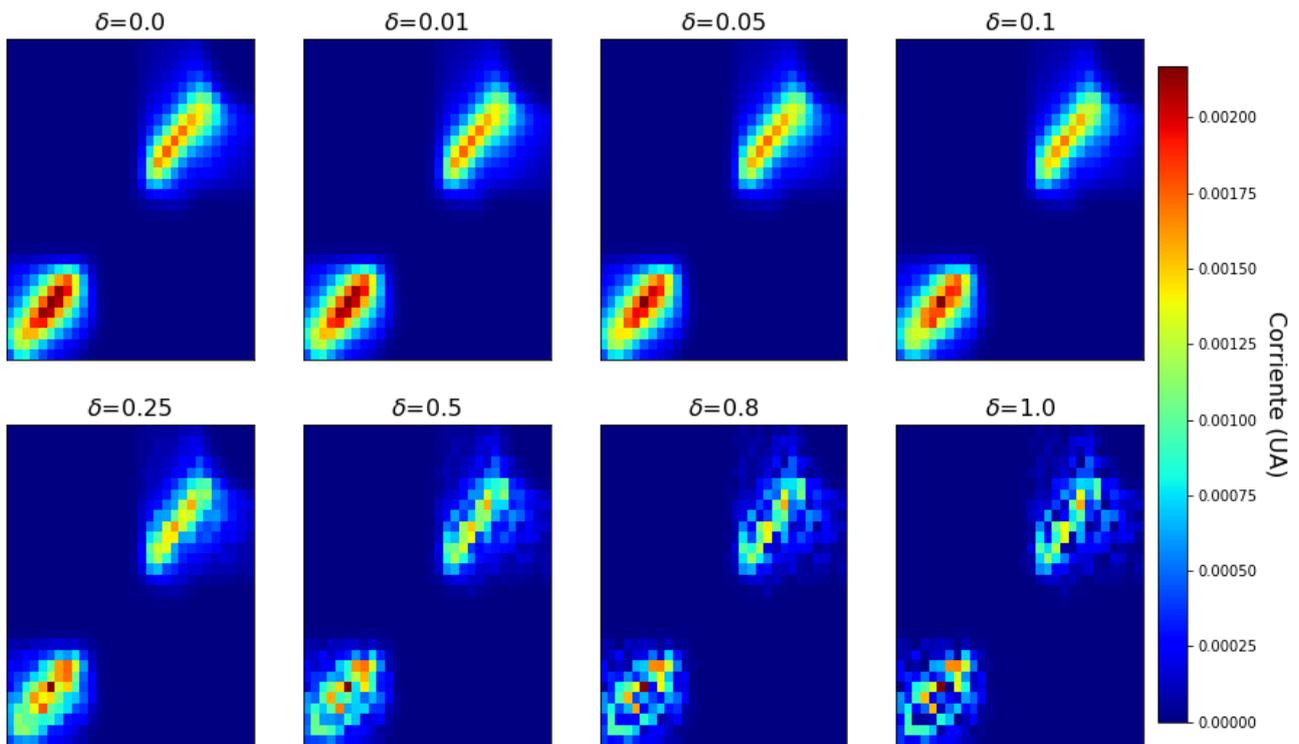


Figura 5.16: Variación del peso de la incertidumbre ( $\delta$ )

Debido a que el modelo CNN2 tiene el mismo alcance que el modelo CNN1, se utiliza este último ya que al ser menos complejo, requiere menor tiempo de ejecución. Se sigue el mismo procedimiento de entrenamiento utilizado anteriormente para los valores de  $\delta = (0.01, 0.05, 0.1, 0.5)$ .

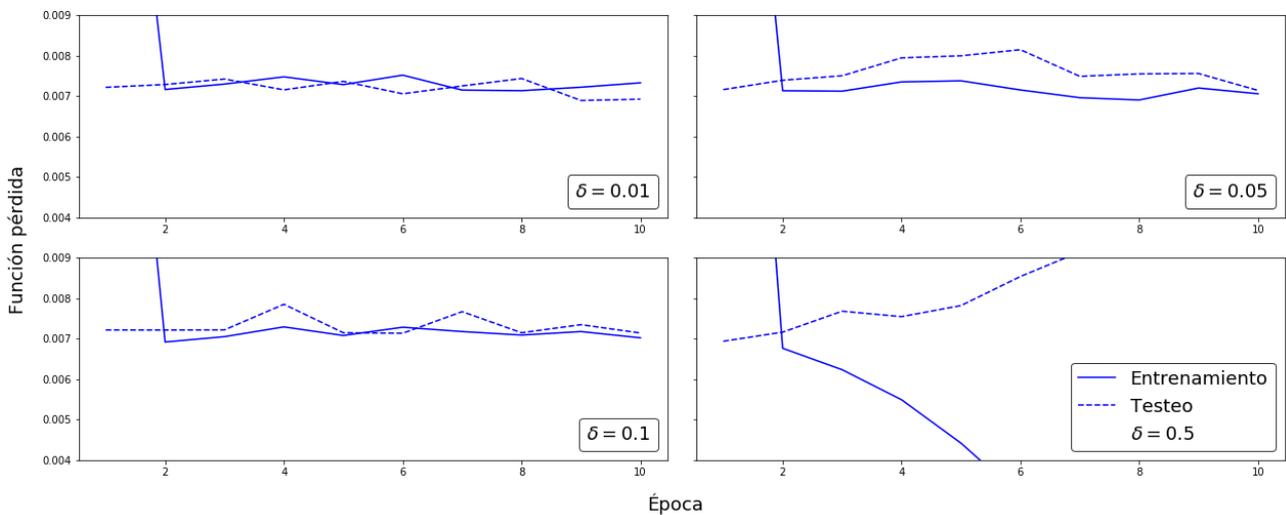


Figura 5.17: Valor de la función pérdida a través de las épocas de entrenamiento para diferentes valores de  $\delta$

Como se puede observar en la figura 5.17, los modelos neuronales logran el mismo alcance que el sistema sin la inclusión de ruido. A pesar de que para  $\delta = 0.5$  el modelo comienza a sobre-ajustar luego de la segunda época de entrenamiento, logra el mismo alcance que para el resto de valores.

## Capítulo 6

# Conclusiones

En este trabajo el objetivo principal fue desarrollar los códigos necesarios para la creación de una base de datos y el entrenamiento de modelos de aprendizaje automático para la parametrización de un sistema de punto cuántico doble.

Si bien para los coeficientes no diagonales de la matriz **D** no se obtuvieron buenos resultados, si se obtuvo buenos resultados para los términos diagonales. Sin embargo, en la recreación de mapas de corrientes utilizando los valores predichos, la imagen generada queda idéntica a la generada con los valores reales de la matriz **D**.

Se puede ver que la precisión en la predicción de los coeficientes fue la misma con la cual se generaron en la base de datos ( $\sigma = 0.1$ ), razón que explica la mala predicción para los coeficientes no diagonales. Si a esto le sumamos el hecho de que la media para los términos no diagonales es de un orden menor, se puede concluir que la corriente depende más de los coeficientes diagonales que de los no diagonales. Entonces los resultados son aceptables como una demostración de que la metodología funciona.

Lo interesante es que debido a la simplicidad del sistema planteado, en ningún caso hizo falta entrenar un modelo durante más de tres épocas para alcanzar el mínimo en la función pérdida. En particular los modelos más simples lograron el mismo alcance que modelos más complejos. Esto lleva a pensar que se podría variar aún más diferentes parámetros del sistema, ya sea para predecirlos o no. Es decir, se podría entrenar un algoritmo que sintonice los parámetros  $d_{ij}$  independientemente de los valores del resto de parámetros del sistema. Es decir, al poder replicar los mapas de corriente alcanzando el mínimo de la función pérdida con un número bajo de épocas de entrenamiento, nos indica que se puede complejizar el sistema.

Con pequeñas modificaciones en el módulo del cálculo de corriente se podría generar una base de datos en la cual varíen el resto de parámetros, como la energía de bias, la tasa de transporte, las energías de interacción y la temperatura. Esto haría que la red neuronal pueda calcular los coeficientes  $d_{ij}$  para un sistema genérico.

Otra modificación en el programa que escribe el dataset permitiría guardar los parámetros del sistema como vector de salida, es decir, se podría entrenar modelos para que sintonicen estos valores utilizando el mapa de corriente, lo cual permitiría de manera más sencilla conocer las propiedades de cada sistema en particular. También cabe la posibilidad de utilizar estos algoritmos en su manera más clásica, la clasificación. Es decir, utilizar métodos de sintonizado y ajuste para encontrar los estados deseados del sistema, por ejemplo, distinguir los diferentes bloques de carga.

Con respecto a la inclusión de ruido en el sistema, se vio como el modelo puede ajustar bien en todos los casos. Sin embargo, se debe tener cuidado a la hora de incluir ruido muy grande, ya que se puede observar como para el caso de  $\delta = 0.5$  la red comienza a ajustar el ruido, sobre ajustando la función a predecir.

La principal complicación que presentó este trabajo fue la diagonalización de matrices. Como se mostró en el desarrollo teórico, el proceso para el cálculo de corriente del sistema requiere los auto-estados del mismo, por lo cual se debe diagonalizar el Hamiltoniano. Si se trabaja en un régimen de no decoherencia, el cálculo de auto-valores se puede realizar por bloque, sin embargo, esto aumenta considerablemente el tiempo de cómputo. El cálculo de una matriz de corriente 30x30 cuyo proceso incluye la diagonalización tarda 3.5 minutos en una computadora promedio, para un sistema que tiene nueve estados posibles. Claramente, a medida que el sistema se vuelve más complejo, este problema escalará exponencialmente.

Para este trabajo, se utilizó la forma explícita de cada auto-valor, lo que permitió reducir el tiempo de eje-

cución lo necesario como para crear una base de datos. Sin embargo esto no resulta razonable en sistemas mas complejos para los cuales el número de estados posibles es mayor. Por ejemplo, si para el modelo de punto cuántico doble simplemente se permitiera la doble ocupación, el número de estados posibles es  $N = 16$ . En un futuro se deben explorar opciones para resolver este problema o simplemente recurrir a dispositivos de cómputo avanzando que permitan realizar estos cálculos en un menor tiempo.

## Apéndice A

# Desacople entre ocupaciones y coherencias en el régimen de acoplamiento fuerte

Partimos de la ecuación maestra generalizada para todos los elementos del operador densidad, es decir que se incluyen en el cálculo de  $\mathbf{p}$  tanto las ocupaciones como las coherencias:

$$(-i\mathbf{L} + \mathbf{W})\mathbf{p} = \mathbf{0}. \quad (\text{A.1})$$

En esta ecuación, tomamos a  $\mathbf{p}$  como un vector conformado por sus ocupaciones (elementos diagonales) y coherencias (elementos no-diagonales) según  $\mathbf{p} = (\mathbf{p}_d, \mathbf{p}_n)^\top$ , de modo que la ecuación anterior puede reescribirse como

$$\begin{pmatrix} \mathbf{W}_{dd} & \mathbf{W}_{dn} \\ \mathbf{W}_{nd} & -i\mathbf{L}_{nn} + \mathbf{W}_{nn} \end{pmatrix} \begin{pmatrix} \mathbf{p}_d \\ \mathbf{p}_n \end{pmatrix} = \begin{pmatrix} \mathbf{0}_d \\ \mathbf{0}_n \end{pmatrix}, \quad (\text{A.2})$$

donde puede verse, a partir de la ecuación (3.12), que el Liouvilliano solo entra en el bloque nn. Si expresamos a  $\mathbf{p}_n$  a partir del segundo sistema de ecuaciones, se tiene

$$\mathbf{p}_n = -\frac{1}{-i\mathbf{L}_{nn} + \mathbf{W}_{nn}} \mathbf{W}_{nd} \mathbf{p}_d. \quad (\text{A.3})$$

Ahora bien, para los elementos del kernel  $\mathbf{W}$ , se tiene que estos dependen linealmente de  $\Gamma$ , de modo que  $\mathbf{W}_{nn} \sim \Gamma$ , mientras que para los elementos del Liouvilliano  $\mathbf{L}_{nn} \sim (\epsilon, t_c)$ . Si suponemos un régimen de acoplamiento fuerte, dado por  $\epsilon, t_c \gg \Gamma$ , tenemos que en el denominador de la expresión anterior domina el Liouvilliano frente al kernel, de modo que podemos tomar

$$\mathbf{p}_n \simeq \frac{1}{i\mathbf{L}_{nn}} \mathbf{W}_{nd} \mathbf{p}_d. \quad (\text{A.4})$$

Volviendo entonces al primer sistema de ecuaciones, y reemplazando esta forma para las coherencias, obtenemos

$$\left( \mathbf{W}_{dd} + \mathbf{W}_{dn} \frac{1}{i\mathbf{L}_{nn}} \mathbf{W}_{nd} \right) \mathbf{p}_d = \mathbf{0}_d, \quad (\text{A.5})$$

por lo que podemos considerar un nuevo kernel efectivo para las ocupaciones, dado por

$$\mathbf{W}_{\text{eff}} = \mathbf{W}_{dd} + \mathbf{W}_{dn} \frac{1}{i\mathbf{L}_{nn}} \mathbf{W}_{nd}. \quad (\text{A.6})$$

Sin embargo, bajo la suposición de acople fuerte anterior, notamos que los términos diagonales  $\mathbf{W}_{dd}$  dominan sobre las correcciones no-diagonales, ya que

$$\mathbf{W}_{dd} \sim \Gamma, \quad \mathbf{W}_{dn} \frac{1}{i\mathbf{L}_{nn}} \mathbf{W}_{nd} \sim \frac{\Gamma^2}{t_c} \ll \Gamma, \quad (\text{A.7})$$

de modo que en este régimen podemos quedarnos únicamente con las contribuciones diagonales del kernel.

# Bibliografía

- [1] T. Ihn, *Semiconductor Nanostructures : Quantum States and Electronic Transport* (Oxford University Press, New York, 2015).
- [2] R. Landauer, "Electrical resistance of disordered one-dimensional lattices," *Philos. Mag.* **21**, 863 (1970).
- [3] M. Büttiker, "Four-terminal phase-coherent conductance," *Phys. Rev. Lett.* **57**, 1761 (1986).
- [4] Y. Meir y N. S. Wingreen, "Landauer formula for the current through an interacting electron region," *Phys. Rev. Lett.* **68**, 2512 (1992).
- [5] A.-P. Jauho, N. S. Wingreen, y Y. Meir, "Time-dependent transport in interacting and noninteracting resonant-tunneling systems," *Phys. Rev. B* **50**, 5528 (1994).
- [6] M. Leijnse, *Transport Spectroscopy and Control of Molecular Quantum Dots*, Phd thesis, RWTH Aachen University, Aachen (2009).
- [7] J. P. Zwolak, T. McJunkin, S. S. Kalantre, J. Dodson, E. MacQuarrie, D. Savage, M. Lagally, S. Coppersmith, M. A. Eriksson, y J. M. Taylor, "Autotuning of double-dot devices in situ with machine learning," *Phys. Rev. Appl.* **13**, 034075 (2020).
- [8] J. Ziegler, T. McJunkin, E. Joseph, S. S. Kalantre, B. Harpt, D. Savage, M. Lagally, M. Eriksson, J. M. Taylor, y J. P. Zwolak, "Toward robust autotuning of noisy quantum dot devices," *Phys. Rev. Appl.* **17**, 024069 (2022).
- [9] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (CRC Press, New York, 1994).
- [10] P. Dayan y L. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Massachusetts, 2001).
- [11] T. M. Mitchell, *Machine Learning* (McGraw-Hill, New York, 1997).
- [12] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, y L. M. K. Vandersypen, "Spins in few-electron quantum dots," *Rev. Mod. Phys.* **79**, 1217 (2007).
- [13] K. Ono, D. G. Austing, Y. Tokura, y S. Tarucha, "Current rectification by pauli exclusion in a weakly coupled double quantum dot system," *Science* **297**, 1313 (2002).
- [14] A. C. Johnson, J. R. Petta, C. M. Marcus, M. P. Hanson, y A. C. Gossard, "Singlet-triplet spin blockade and charge sensing in a few-electron double quantum dot," *Phys. Rev. B* **72**, 165308 (2005).
- [15] W. G. van der Wiel, S. De Franceschi, J. M. Elzerman, T. Fujisawa, S. Tarucha, y L. P. Kouwenhoven, "Electron transport through double quantum dots," *Rev. Mod. Phys.* **75**, 1 (2002).
- [16] D. V. Averin y K. K. Likharev, "Coulomb blockade of single-electron tunneling, and coherent oscillations in small tunnel junctions," *J. Low Temp. Phys.* **62**, 345 (1986).
- [17] H. Bruus y K. Flensberg, *Many-Body Quantum Theory in Condensed Matter Physics* (Oxford University Press, New York, 2004).

- [18] M. Braun, J. König, y J. Martinek, "Theory of transport through quantum-dot spin valves in the weak-coupling regime," *Phys. Rev. B* **70**, 195345 (2004).
- [19] A. Donarini, M. Grifoni, y K. Richter, "Dynamical symmetry breaking in transport through molecules," *Phys. Rev. Lett.* **97**, 166801 (2006).
- [20] H. Schoeller y G. Schön, "Mesoscopic quantum transport: Resonant tunneling in the presence of a strong coulomb interaction," *Phys. Rev. B* **50**, 18436 (1994).
- [21] J. König, H. Schoeller, y G. Schön, "Zero-bias anomalies and boson-assisted tunneling through quantum dots," *Phys. Rev. Lett.* **76**, 1715 (1996).
- [22] H. Schoeller, "A perturbative nonequilibrium renormalization group method for dissipative quantum mechanics," *Eur. Phys. J. Spec. Top.* **168**, 179 (2009).
- [23] J. König, H. Schoeller, y G. Schön, "Cotunneling at resonance for the single-electron transistor," *Phys. Rev. Lett.* **78**, 4482 (1997).
- [24] A. Thielmann, M. H. Hettler, J. König, y G. Schön, "Cotunneling current and shot noise in quantum dots," *Phys. Rev. Lett.* **95**, 146806 (2005).
- [25] I. Weymann, J. König, J. Martinek, J. Barnaś, y G. Schön, "Tunnel magnetoresistance of quantum dots coupled to ferromagnetic leads in the sequential and cotunneling regimes," *Phys. Rev. B* **72**, 115334 (2005).
- [26] B. Kubala y J. König, "Quantum-fluctuation effects on the thermopower of a single-electron transistor," *Phys. Rev. B* **73**, 195316 (2006).
- [27] T. Korb, F. Reininghaus, H. Schoeller, y J. König, "Real-time renormalization group and cutoff scales in nonequilibrium applied to an arbitrary quantum dot in the coulomb blockade regime," *Phys. Rev. B* **76**, 165316 (2007).
- [28] H. Schoeller y F. Reininghaus, "Real-time renormalization group in frequency space: A two-loop analysis of the nonequilibrium anisotropic kondo model at finite magnetic field," *Phys. Rev. B* **80**, 045117 (2009).
- [29] V. E. Tarasov, "Pure stationary states of open quantum systems," *Phys. Rev. E* **66**, 056116 (2002).
- [30] S. Mukamel, "Superoperator representation of nonlinear response: Unifying quantum field and mode coupling theories," *Phys. Rev. E* **68**, 021111 (2003).
- [31] M. Leijnse y M. R. Wegewijs, "Kinetic equations for transport through single-molecule transistors," *Phys. Rev. B* **78**, 235424 (2008).
- [32] S. Juergens, F. Haupt, M. Moskalets, y J. Splettstoesser, "Thermoelectric performance of a driven double quantum dot," *Phys. Rev. B* **87**, 245423 (2013).