



Instituto Carlos I de Física Teórica y Computacional
y Departamento de Física Moderna
Universidad de Granada

**FÍSICA ESTADÍSTICA DE PROCESOS
MARCOVIANOS: ESTUDIO DE REDES
DE NEURONAS Y SISTEMAS AFINES**

TESIS DOCTORAL
JOAQUÍN JAVIER TORRES AGUDO
GRANADA, 13 de Junio de 1997

FÍSICA ESTADÍSTICA DE PROCESOS MARCOVIANOS: ESTUDIO DE REDES DE NEURONAS Y SISTEMAS AFINES

TESIS DOCTORAL

presentada por

JOAQUÍN JAVIER TORRES AGUDO

para obtener el grado de

DOCTOR EN FÍSICA

Instituto Carlos I de Física Teórica y Computacional

Universidad de Granada

13 de Junio de 1997

Joaquín Marro Borau, Catedrático de Física de la Materia Condensada de la Universidad de Granada, CERTIFICA que este trabajo ha sido realizado bajo su dirección, y constituye la Tesis Doctoral de D. Joaquín Javier Torres Agudo.

Granada, a 14 de Abril de 1997.

Fdo: Joaquín Marro Borau

Quiero agradecer a mi Director de Tesis, D. Joaquín Marro Borau, su enorme interés en la realización de este trabajo, sus consejos y sugerencias, así como su dedicación para llevar a buen fin el mismo. Quiero agradecer la paciencia que ha mostrado conmigo y la confianza que ha depositado en mi, sin las cuales este trabajo no podría haberse realizado. Asimismo, quiero agradecer al profesor D. Pedro L. Garrido su ayuda y aquellos consejos que tanto me han servido para la realización de esta Tesis.

Agradezco también al Instituto Carlos I de Física Teórica y Computacional, al Departamento de Física Aplicada y a la Universidad de Granada los medios que ha puesto a mi disposición.

Quiero mostrar un sincero agradecimiento a mis amigos de departamento, José Ramón por todo lo que he aprendido de él y por las largas discusiones de física, José Antonio Vacas por sus consejos sobre ordenadores y por todo lo que me ha enseñado, Paco, Cristobal, Fattah, Juanjo, José Ignacio, Miguel, a toda la pandilla de los comedores: Africa, Eva, Ana Belén, Santi, Luis Alberto, Hanane, Hind, a la “élite” Intercampus, y sobre todo, a mis amigos Patxi, Bienvenido, Gonzalo, Ramón y Elvira, por todo lo que me han ayudado en los momentos más difíciles.

Por último quiero mostrar mi más especial y entrañable agradecimiento a Houda y a toda mi familia, en especial a mis padres y hermanos por todo lo que me han ayudado y comprendido, y porque sólo ellos saben lo que ha costado ésto.

Contenidos

Perspectiva	1
1 Algunas consideraciones previas	7
1.1 Introducción	7
1.2 Neurobiología	7
1.2.1 Neuronas	8
1.2.2 Sinapsis	11
1.2.3 Neurofisiología	12
1.3 Psicología experimental	18
1.4 Inteligencia artificial	18
1.5 Redes de neuronas	19
2 Modelo cinético de red de neuronas	25
2.1 Introducción	25
2.2 Definición del modelo	27
2.3 Resultados exactos para sinapsis simétricas	33
2.4 Sinapsis asimétricas	37
2.5 El formalismo de réplicas	40
2.6 Casos adicionales del modelo	48
2.7 Resumen y conclusiones	53
3 Red de neuronas fuera del equilibrio	57
3.1 Introducción y definición del modelo	57
3.2 Cambios de fase continuos	60
3.3 Fluctuaciones alrededor de la regla de Hebb	63

3.4	Estabilidad y cambios de fase de primer orden	68
3.5	Sinapsis asimétricas	80
3.6	Conclusiones	83
4	Simulación Monte Carlo	89
4.1	Introducción	89
4.2	Modelo	90
4.3	Resultados Monte Carlo	93
4.4	Conclusión	98
5	Difusión de iones magnéticos	101
5.1	Introducción	101
5.2	Definición del modelo	103
5.3	Algunos casos particulares del modelo	106
5.4	Resultados	108
	Conclusiones	117
	Apéndices	123
A	Modelo de Hopfield y técnica de réplicas	123
A.1	Función de partición en el límite $\alpha \rightarrow 0$	123
A.2	Red de Hopfield para α finito	125
B	Obtención de hamiltonianos efectivos	129
C	Diagonalización de matrices \mathcal{Q}	133
	Bibliografía	137

Incluso en el sueño el cerebro está pulsando, palpitando y destelleando con el complejo negocio de la vida humana: soñar, recordar, imaginar cosas. Nuestros pensamientos, visiones y fantasías poseen una realidad física. Si nos encogiéramos al nivel de las neuronas, podríamos presenciar formas elaboradas, intrincadas y evanescentes. Una podría ser la chispa de un recuerdo o el olor a lilas en un camino campestre de nuestra infancia. Otra podría ser un ansioso boletín enviado a todos los puntos: “¿Dónde he dejado mis llaves?”

Carl Sagan, “Cosmos”.

Perspectiva

Los mecanismos de la mente humana han fascinado a los científicos y filósofos durante siglos. Sin embargo, el conocimiento de la mente ha sido más bien pobre hasta hace poco más de un siglo. Sólo recientemente, el hombre ha sido capaz de responder a determinadas cuestiones muy simples, como la forma en que el cerebro procesa la información visual, pero sigue siendo incapaz de clarificar la naturaleza de la conciencia y la posible existencia de propiedades mentales que trascienden a la propia sustancia biológica de la que está constituido el cerebro.

Descartes fue el primero en dar una imagen aproximada de la mente, situando el pensamiento en el cerebro, pero errando al localizarlo en las partes internas del mismo¹. La información, según intuyó, es recogida por los órganos sensoriales y conducida a una unidad central mediante nervios, pero no respondió a la cuestión de cómo controla la información dicha unidad central.

Actualmente no conocemos todavía la naturaleza de la mente, en parte debido a la complejidad que presenta un sistema como el cerebro, con un gran número de componentes elementales, llamados *neuronas*, cada una interactuando con otras a través de las *sinapsis*, también éstas en un número muy elevado. Pero se han dado pasos importantes, como muestra el enorme desarrollo reciente de disciplinas como la neurobiología, la psicología experimental y la inteligencia artificial. Aunque estas disciplinas se han desarrollado de forma independiente, en los últimos años han aparecido importantes puntos de contacto entre ellas debido, sobre todo, al estudio de las llamadas *redes de neuronas*. Estos sistemas son modelos matemáticos bien definidos que

¹A diferencia de lo que pensaba Descartes, hoy se sabe que el sistema nervioso central de los mamíferos y, en particular, el del hombre muestra una gran organización, con una serie de partes bien diferenciadas como el córtex, tálamo, hipotálamo, cerebelo, etc, que tienen gran importancia en el funcionamiento del mismo. De entre ellas, la parte más externa o córtex juega un papel fundamental como procesador central.

tienen en cuenta, en cierta forma, diferentes aspectos de cada una de estas tres disciplinas. Las redes de neuronas constituyen un intento de formalizar, desde un punto de vista matemático, las principales propiedades que presenta la mente o, mejor, otros sistemas biológicos semejantes pero menos evolucionados. Entre las propiedades de interés merecen destacarse las tres siguientes: (i) *Memoria asociativa*, propiedad de la mente que permite recuperar una determinada idea mediante un conocimiento parcial de la misma. (ii) El *aprendizaje*, supervisado o no, de conocimientos, esto es, la capacidad de almacenar información que luego puede ser recuperada, y que está fuertemente ligada a la propiedad de memoria asociativa. (iii) *Auto-organización* del córtex de tal forma que partes adyacentes del cuerpo están controladas por regiones vecinas del córtex.

En la presente memoria describimos una serie de resultados de nuestro trabajo durante los últimos años para profundizar en el conocimiento de las redes de neuronas y de los sistemas llamados *vidrios de espines*, esto es, sistemas magnéticos desordenados que tiene una gran semejanza formal y conceptual con las redes de neuronas. Hemos intentado así comprender mejor muchos de los fenómenos que tienen lugar en los sistemas de neuronas biológicas y en ciertas sustancias desordenadas como son las aleaciones de un metal noble con impurezas magnéticas. Nos ha preocupado la elaboración de modelos originales de redes de neuronas que, en algún aspecto, muestren un comportamiento macroscópico similar al de los sistemas biológicos; por ejemplo, modelos que presenten la propiedad de memoria asociativa o que simulen sistemas inteligentes con capacidad de aprendizaje. La aparición de estas propiedades en nuestros modelos está íntimamente relacionada con la existencia de cambios de fase cuyo estudio permite determinar, por ejemplo, en qué condiciones (temperatura, número de patrones que el sistema es capaz de memorizar, etc.) los sistemas modelados tienen capacidad de aprender y almacenar información. En un sentido, nuestros modelos constituyen una generalización del modelo de Hopfield [1], que presenta una fase con memoria asociativa y otra fase sin esta propiedad, y puede definirse un parámetro *de orden* que sólo es no nulo en la primera de ellas.

Los modelos actuales de redes de neuronas no explican satisfactoriamente los mecanismos que intervienen en los procesos neuronales reales dadas las simplificaciones que

introducen. Por este motivo es conveniente construir nuevos modelos que incorporen y permitan analizar ciertos aspectos particulares y microscópicos de la generación y transmisión de los impulsos nerviosos hasta ahora no tenidos en cuenta. En concreto, y como veremos en el capítulo primero, un estudio neurofisiológico de la transmisión del impulso nervioso parece evidenciar la importancia de considerar diferentes mecanismos, con carácter competitivo, en la transmisión del impulso nervioso. Es interesante estudiar cómo estos mecanismos influyen en el comportamiento *macroscópico* del sistema, es decir, en propiedades tales como memoria asociativa, aprendizaje, etc.

Los sistemas presentados en esta memoria están, generalmente, fuera del equilibrio (como cuando existe un agente externo no-hamiltoniano, o una corriente a través del sistema), y van a ser estudiados en el marco de una teoría general basada en el estudio dinámico de una ecuación maestra característica de cada modelo particular. Esta circunstancia es coherente con el hecho de que la mayoría de los sistemas biológicos se encuentran fuera del equilibrio. En el caso de los sistemas desordenados reales, la separación del equilibrio puede ser debida a la difusión de impurezas. En algunos sistemas estudiados, la competición entre dinámicas induce estados estacionarios relativamente sencillos que admiten una descripción en términos de un *hamiltoniano efectivo*. Por otra parte, nuestros modelos de redes de neuronas presentan una alta conectividad entre sus elementos, lo que hace que los tratamientos de tipo campo medio o auto-consistentes sean especialmente adecuados. Además de resultados teóricos analíticos, presentamos algunos resultados numéricos obtenidos mediante la simulación en el ordenador de realizaciones Monte Carlo de los modelos. Así comprobamos a veces la bondad de las aproximaciones realizadas. Nuestros resultados teóricos van a ser comparados también con el comportamiento de sistemas reales y con el de otros modelos relacionados.

El capítulo primero es introductorio, analizando ciertos aspectos de la neurofisiología del sistema nervioso. Ésta es conveniente para comprender ciertas hipótesis y aproximaciones en nuestros modelos. En particular, hacemos un análisis detallado de los procesos dinámicos que tienen lugar en las neuronas y sinapsis referentes a la generación y transmisión del impulso nervioso. Los cuatro capítulos siguientes son originales; algunos resultados del capítulo quinto han sido obtenidos ya en la tesis de

licenciatura [2], aunque ahora se presentan más elaborados y se ha llegado a conclusiones más definitivas. El contenido de estos cuatro capítulos puede resumirse como sigue.

En el capítulo segundo presentamos un modelo de red de neuronas cinético, caracterizado por una ecuación maestra, en el que tanto las sinapsis como las neuronas evolucionan con el tiempo, pero con distinta velocidad. Para ciertos casos particulares, el modelo puede ser descrito en términos de un hamiltoniano efectivo que ha sido resuelto, en aproximación de campo medio, cuando las sinapsis evolucionan con una velocidad mucho mayor que las actividades neuronales. Se presenta y estudia un caso que se corresponde con el modelo de Hopfield con fluctuaciones aleatorias de las intensidades sinápticas, de tal forma que éstas se acoplan entre sí, en promedio, como establece la regla de aprendizaje de Hebb [3]. Las consecuencias de tales fluctuaciones en el sistema son analizadas con detalle para una elección específica de las probabilidades de transición y de la distribución de fluctuaciones, incluyendo el caso en el que las sinapsis son asimétricas.

En el capítulo tercero se estudia otro modelo de red de neuronas cinético (que podemos considerar como una versión del modelo general presentado en el capítulo segundo) que incluye también sinapsis fluctuantes. Está caracterizado por una distribución de fluctuaciones, diferente a la del modelo anterior, que no permite una descripción hamiltoniana del sistema. Sin embargo, se puede obtener información del estado estacionario en aproximación de campo medio. Los resultados obtenidos han sido comparados con los del modelo de Hopfield. A diferencia de lo que ocurre en éste, el error asociado al proceso de recuerdo de información puede ser reducido bajo determinadas condiciones para las fluctuaciones. Se presentan resultados analíticos tanto para el caso de acoplamientos sinápticos simétricos como asimétricos.

En el capítulo cuarto se presentan resultados de simulaciones Monte Carlo del modelo presentado en el capítulo tercero. Algunas simulaciones muestran que la teoría de campo medio desarrollada en el capítulo tercero es exacta. Otras simulaciones muestran una comparación entre el modelo propuesto en el capítulo tercero y el modelo de Hopfield. Los resultados de estas simulaciones muestran cómo nuestro modelo optimiza los procesos de aprendizaje y recuerdo comparado con el modelo de Hopfield.

En el capítulo quinto elaboramos un modelo cinético de sistema microscópicamente desordenado. Las energías de intercambio entre pares de espines cambian con el tiempo de forma aleatoria, lo que simula una difusión iónica (ideal) rápida. El modelo ha sido resuelto mediante una técnica original de campo medio, basada en agrupamientos o *dominios* de espines, en la que se puede trabajar, de manera natural, con diferentes órdenes de aproximación. En concreto, hemos resuelto los dos primeros órdenes correspondientes a agrupamientos con un sólo espín (para dimensión espacial arbitraria) y con cuatro espines (para dimensión dos). Los resultados muestran, entre otras cosas, cambios de fase con procesos reentrantes reminiscentes del comportamiento observado en ciertos materiales como mezclas binarias de metales nobles con impurezas magnéticas.

La memoria termina con una exposición de las conclusiones más importantes que se siguen de nuestro trabajo, resumiendo las contribuciones originales más importantes. También se plantean algunos problemas abiertos que quedan por resolver y que pueden ser objeto de futuros estudios.

Capítulo 1

Algunas consideraciones previas

1.1 Introducción

Las redes de neuronas surgen como un intento de modelar ciertos aspectos o propiedades de los sistemas neuronales biológicos. En este intento, se busca una posible utilidad inmediata en el campo de la computación. El estudio de estos sistemas, además de matemáticas, involucra a la neurobiología, a la psicología experimental y a la disciplina llamada inteligencia artificial. Aunque estas disciplinas son esencialmente diferentes entre sí, cada vez existen más nexos de unión entre ellas. En los últimos años, los físicos teóricos han mostrado un creciente interés por estos sistemas, dada la semejanza conceptual y formal que tienen con los sistemas magnéticos desordenados, particularmente con los llamados vidrios de espines. En este capítulo repasamos brevemente algunas contribuciones de todas estas disciplinas, recopilando información sobre los sistemas neuronales biológicos, motivando así el diseño de nuevos modelos que, aunque necesariamente artificiales, sean más realistas y mejoren los ya existentes.

1.2 Neurobiología

La neurobiología trata de los principios biofísicos del funcionamiento del sistema nervioso de los animales. Uno de sus hitos principales es el descubrimiento en 1791 por Galvani de la naturaleza eléctrica de las señales nerviosas. Fundamental es, también, la contribución de Ramón y Cajal, que demostró, a finales del siglo pasado, cómo el tejido nervioso está constituido por un conjunto de células, bien definidas, las

neuronas, en contacto entre sí a través de un determinado número de uniones, las sinapsis. Los mecanismos que intervienen en la creación y propagación de las señales eléctricas en las neuronas fueron explicados, cinco décadas más tarde, por Hodgkin y Huxley. En la actualidad, existe gran interés en el estudio de ciertas moléculas llamadas *neurotransmisores* y de proteínas complejas, los *neurorreceptores*, que intervienen en la transmisión de la señal eléctrica a través de las sinapsis. También interesan ahora las reacciones químicas en las que intervienen estas moléculas, así como los procesos iónicos responsables de la dinámica neuronal. Últimamente, hay un creciente interés en el estudio de los mecanismos microscópicos que intervienen en la transmisión sináptica y que se suponen responsables de la variabilidad observada en la respuesta neuronal ante dos estímulos consecutivos; cf., por ejemplo, Ref. [4].

1.2.1 Neuronas

Las neuronas, como constituyentes elementales del tejido nervioso, son células altamente especializadas, cuya función principal es la generación y transmisión de la señal eléctrica, llamada *impulso nervioso* o *potencial de acción*. El número de neuronas en el cerebro humano es del orden de 10^{11} .¹ A grandes rasgos, en una neurona típica podemos distinguir tres partes bien diferenciadas (cf. figura 1.1):

- **Soma.** Está constituido por el cuerpo celular, donde residen los *orgánulos* celulares y el *núcleo* de la célula. Su tamaño en una neurona típica es del orden de $10 - 100\mu m$. Es una parte muy importante para la actividad nerviosa pues, aparte de contener todos los orgánulos que regulan la actividad de la célula, en cierta zona de su membrana celular tiene lugar la creación del potencial de acción, responsable de la actividad nerviosa.
- **Axón.** Es una prolongación de la *membrana celular* del soma que hace de cable conductor por el que se transmite el potencial de acción hacia otras neuronas vecinas o hacia los músculos. El axón se desprende de una región cónica del soma

¹Estimación obtenida contando pequeñas regiones del cerebro, situadas en el córtex, donde se observa que la densidad de neuronas es prácticamente constante, del orden de 150000 neuronas por milímetro cuadrado. Este hecho, junto con una estimación para el área total del córtex del orden de $200000mm^2$, da lugar a ese número [5].

llamada *cono axónico* o *zona de Hillock* que llega a alcanzar un diámetro hasta de $100\mu m$ [6]. El axón tiene una longitud típica de $0.07cm$ [7]. En muchas neuronas está recubierto por unas cápsulas, de una sustancia llamada *mielina*, separadas por unos nódulos amielínicos llamados *nodos de Ranvier*. Estas cápsulas favorecen la conducción del impulso nervioso. La figura 1.2 muestra el enorme aumento en la velocidad de conducción del potencial de acción debido al efecto de la cápsula de mielina que hace que disminuya la capacitancia del axón y que los potenciales de acción se generen sólo a nivel de los nodos de Ranvier.

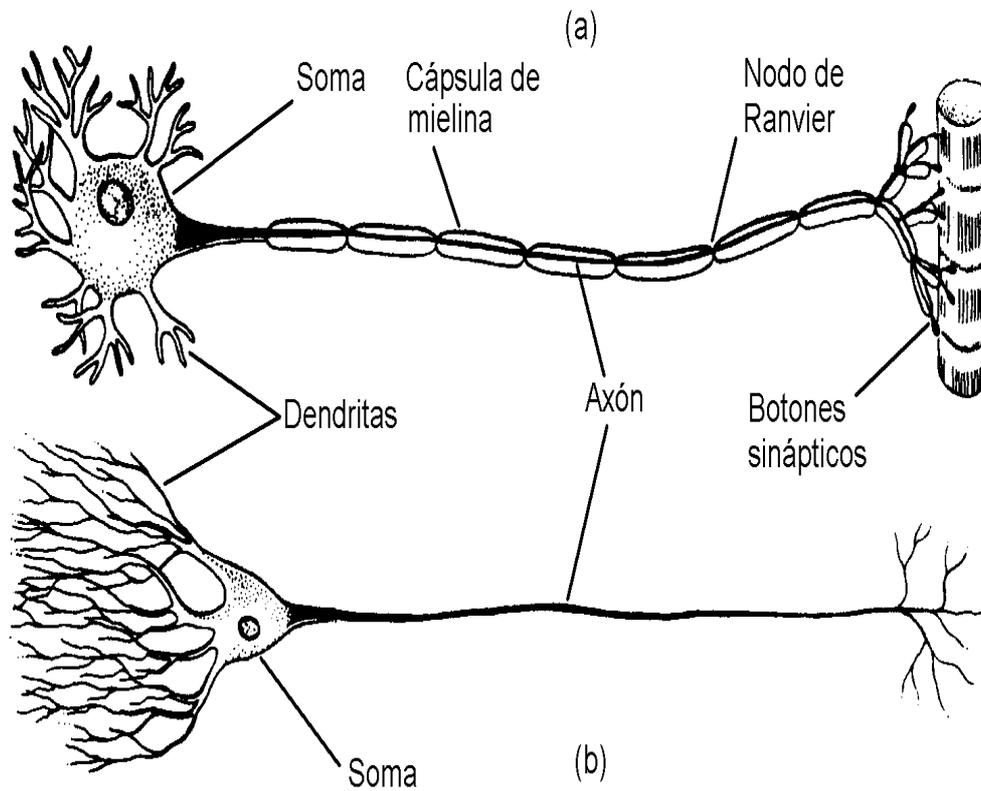


Figura 1.1: La figura muestra dos tipos de neuronas con sus partes fundamentales. (a) Neurona motora con un axón mielinizado para favorecer la conducción del impulso nervioso a los músculos y (b) neurona cerebral con dendritas muy ramificadas y un único axón amielínico. (Redibujado de R.M. Berne y M.N. Levy, *Fisiología*, Times Mirror de España, Madrid 1993.)

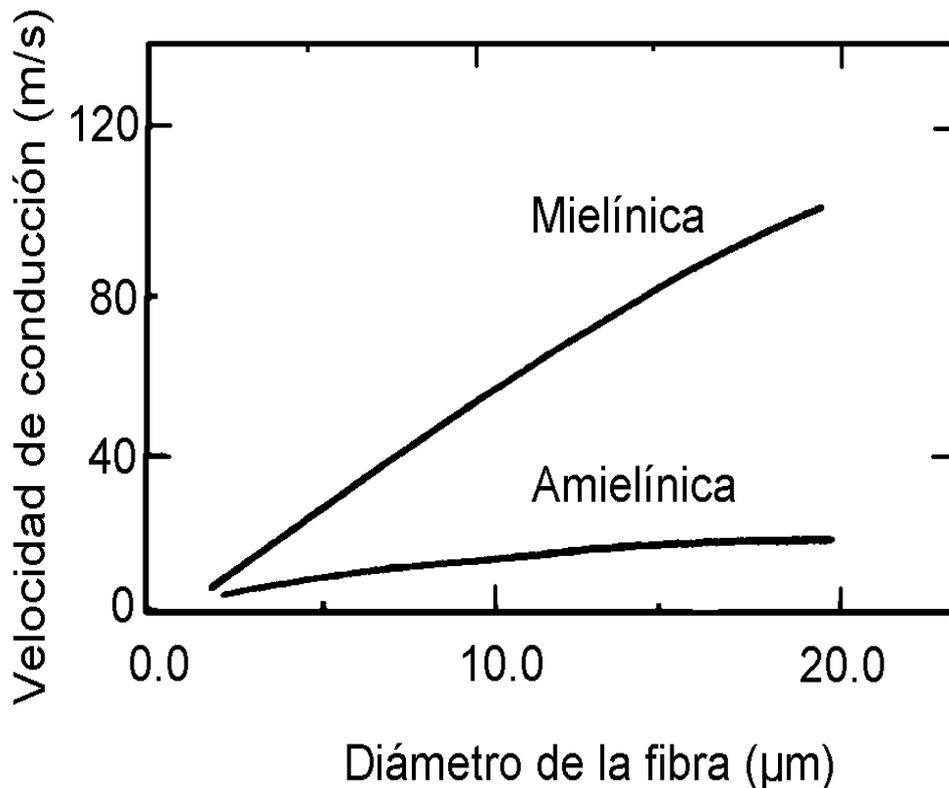


Figura 1.2: Velocidades de conducción en función del diámetro de la fibra para axones mielínicos y amielínicos de vertebrados. (Redibujado de C.J. Flickinger, y colaboradores, *Medical Cell Biology*, W.B. Saunders, Filadelfia 1979.)

- **Dendritas.** Son prolongaciones del soma, generalmente de menor tamaño que el axón, que conducen a la neurona el impulso nervioso procedente de otras neuronas. Las neuronas típicas tienen dos o más dendritas que se van ramificando en una estructura divergente. Constituyen la zona receptora que recibe y procesa impulsos nerviosos, los cuales producen una respuesta graduada decreciente sobre el soma. Son verdaderas extensiones *protoplasmáticas* del cuerpo celular. Rara vez alcanzan los $700\mu\text{m}$ de longitud. Tienen unas espinas llamadas *gémulas* o *botones sinápticos*, en las que tienen lugar las uniones o sinapsis entre dos neuronas. El número de gémulas varía de unas neuronas a otras, llegando a unas 4000 en algunas neuronas piramidales [6].

1.2.2 Sinapsis

Las sinapsis son regiones de contacto especializadas entre neuronas, entre éstas y órganos efectores, o entre dos fibras musculares. El cerebro humano tiene del orden de 10^{14} sinapsis [6]; la mayoría se desarrollan a los pocos meses del nacimiento del individuo. Las sinapsis pueden variar en intensidad y localización precisa. El extremo del axón puede tener desde unas cuantas hasta 5×10^4 sinapsis, siendo el promedio del orden de 10^3 botones sinápticos por neurona [6]. Además, una neurona puede recibir contactos sinápticos de muchas otras neuronas, desde varios miles hasta 10^5 terminaciones axónicas separadas [6]. Según la zona de las neuronas entre las que tiene lugar el contacto sináptico, existen tres tipos frecuentes de sinapsis: *axosomática*, *axodendrítica*, *axoaxónica*. Según cómo tiene lugar la transmisión del impulso nervioso a través de ellas, las sinapsis se clasifican en *químicas* y *eléctricas*.

- **Sinapsis química.** Cerca de la membrana presináptica existe una alta concentración de *vacuolas*, rellenas de unas sustancias, específicas de cada tipo de neurona, llamadas neurotransmisores, que nunca aparecen en la membrana postsináptica. Esto último implica que las sinapsis químicas tienen carácter unidireccional y asimétrico. Entre las dos membranas (pre y postsináptica) hay una anchura de unos 200Å [6] que, cuando tiene lugar la transmisión del impulso nervioso, produce lo que se conoce con el nombre de *retardo sináptico*; cf. figura 1.3.
- **Sinapsis eléctrica.** A diferencia de la sinapsis química, en este tipo no existen vacuolas con neurotransmisores. Ambas membranas, presináptica y postsináptica, son similares y la sinapsis es, por lo tanto, simétrica. Las dos neuronas que intervienen en la sinapsis están electrónicamente acopladas mediante puentes iónicos, o pasos naturales en la membrana, a través de los que hay un flujo de iones entre la célula y el medio que la rodea. No existe el retardo sináptico y la anchura entre las dos membranas es de unos 20Å [6]. Por lo tanto, son sinapsis más rápidas que las químicas; cf. figura 1.3.

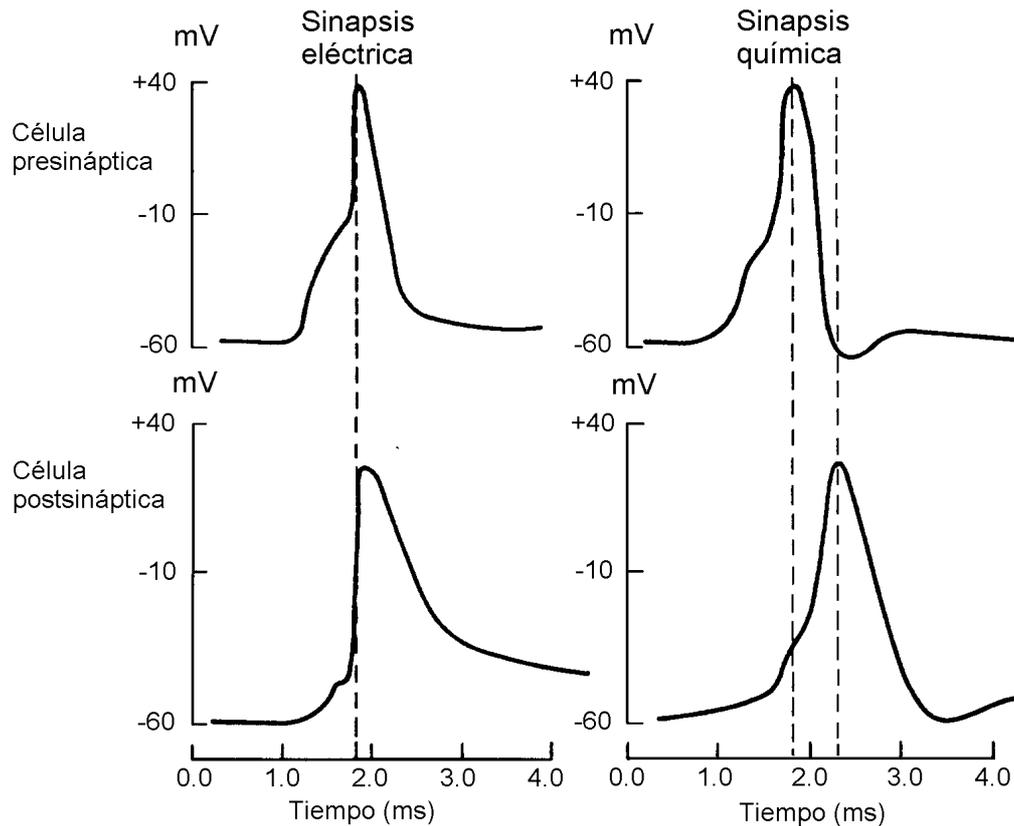


Figura 1.3: La transmisión sináptica en sinapsis eléctricas y químicas. En la sinapsis eléctrica los iones fluyen directamente desde la neurona presináptica a la postsináptica. En la sinapsis química hay un retraso del orden de 0.5ms debido a la secreción del neurotransmisor, a su difusión y a la respuesta postsináptica. (Redibujado de R.M. y M.N. Levy, *Fisiología*, Times Mirror de España, Madrid 1993.)

1.2.3 Neurofisiología

La generación de los impulsos nerviosos, y su transmisión a través de las neuronas y las sinapsis, tiene lugar en una serie de etapas que a continuación se detallan.

Potencial de reposo.

La observación de una neurona en reposo muestra que, a ambos lados de la membrana celular, existen concentraciones específicas de diferentes tipos de iones. En el medio acuoso que rodea a la neurona hay mayor concentración de iones Na^+ y Cl^- que en el interior de la célula, siendo en su conjunto eléctricamente neutro. Por otra parte,

en el medio intracelular hay mayor abundancia de iones K^+ y radicales orgánicos proteicos que en el exterior de la célula, siendo el conjunto también eléctricamente neutro. Los iones tienden a difundirse a partir de las regiones de elevada concentración hacia zonas en las que la concentración específica de dicho ión es menor. La membrana de la neurona es selectiva a la hora de dejar pasar iones, siendo más permeable a los iones Na^+ y K^+ . La diferencia de concentración de iones, a uno y otro lado de la membrana, provoca la aparición de potenciales eléctricos asociados a cada uno de los iones, de acuerdo con la Ley de Nernst [8]:

$$V_x = \frac{RT}{FZ_x} \log \left(\frac{n^x_{ext}}{n^x_{int}} \right), \quad (1.1)$$

donde R es la constante de los gases, T la temperatura absoluta, F la constante de Faraday y Z_x la valencia del ión x , y n^x_{int} y n^x_{ext} son las concentraciones del ión x dentro y fuera de la célula respectivamente. Por otra parte, como resultado de la diferencia de concentración de iones entre el medio intracelular y el medio extracelular, los iones K^+ tienden a salir de la neurona rápidamente, hasta que se establece un equilibrio en la concentración del mismo a los dos lados de la membrana celular. Así, el interior de la neurona queda depolarizado respecto del exterior. Este equilibrio es lo que se llama *potencial de reposo*, que suele tomar un valor $V = -70mV$ (a veces $-100mV$). El origen de la diferente concentración de iones entre el interior y el exterior de la neurona se cree que es debido a la actividad metabólica de la neurona, que ocasiona que el Na^+ sea bombeado hacia afuera de la célula y el K^+ hacia dentro. La energía necesaria para este proceso proviene de enlaces fosfato ricos en energía.

Generación del potencial de acción.

Una excitación local eléctrica causa la apertura de los canales o puentes específicos del Na^+ , produciéndose una depolarización del exterior de la membrana, incrementándose el potencial desde -70 a $+40mV$. Esto provoca la apertura de los canales de K^+ después de $0.5ms$ de haberse abierto los del Na^+ , repolarizándose la membrana hasta $-80mV$. Este pico constituye el potencial de acción. Después de esto transcurren entre 3 y 4 milisegundos, tiempo en el que se restablecen las concentraciones originales de iones. Este tiempo recibe el nombre de *período refractario* [7]; cf. figura 1.4.

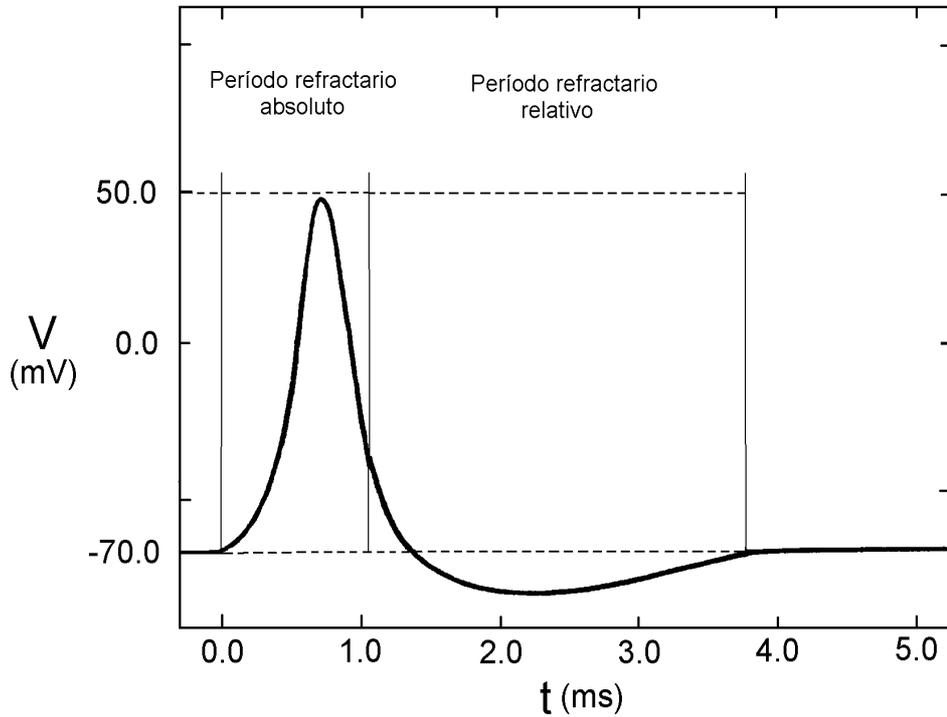


Figura 1.4: Potencial de acción generado en una neurona mostrando los períodos refractarios absoluto y relativo. (Redibujado de R.M. y M.N. Levy, *Fisiología*, Times Mirror de España, Madrid 1993.)

Propagación del potencial de acción

El potencial creado por la apertura de los canales iónicos difunde hacia otras regiones de la membrana donde produce la apertura de nuevos canales de Na^+ y K^+ . En el proceso se crean dos ondas que se propagan a lo largo de la membrana del axón en direcciones opuestas, a una velocidad que en general depende del axón [6]. Durante la propagación del potencial de acción, la neurona gana Na^+ y pierde K^+ , bombeando el Na^+ ganado posteriormente mediante procesos metabólicos. El bombeo de iones Na^+ hacia fuera de la neurona se acompaña de la entrada de una cantidad equivalente de iones K^+ .

La zona de la neurona donde ocurre la excitación inicial se llama zona de Hillock

y contiene una alta densidad de canales o puentes iónicos. Cuando el potencial de membrana de esta zona supera un umbral $\theta = -30mV$, los canales iónicos se abren y el potencial de acción se propaga a lo largo del axón. El otro potencial de acción no puede moverse hacia el soma y las dendritas, donde hay menos densidad de canales iónicos [7]. En este caso, el potencial de acción sólo puede difundir, lo que en cierta forma explica el carácter unidireccional de la transmisión del impulso nervioso. Este tipo de conducción del potencial de acción es el más normal y suele darse en neuronas conectadas entre si mediante axones amielínicos. Existe otro tipo de conducción, más rápida que la anterior, que se conoce con el nombre de *conducción saltatoria* y tiene lugar en los axones mielínicos. El impulso nervioso salta de un nodo de Ranvier a otro. En los nodos de Ranvier tiene lugar el intercambio iónico entre el líquido intersticial y el intracelular. Este movimiento iónico discontinuo en el nodo es eficiente desde el punto de vista de conservación de la energía. En la conducción saltatoria el impulso salta debido a que sólo se encuentran pequeñas regiones de membrana plasmática del nervio que pueden ser depolarizadas para propagarlo. Cada nodo es un sitio de autoregeneración del impulso. Las fibras mielínicas juegan un papel análogo al de un cable transoceánico, donde los nodos de Ranvier semejan estaciones repetidoras de relevo que, a determinados intervalos, amplifican los impulsos eléctricos debilitados.

Periodos refractarios

Se llama período refractario absoluto al tiempo que va desde que pasa un pulso hasta que puede generarse otro; suele ser del orden de $1ms$. Esto implica que una fibra no puede transmitir más de 1000 pulsos por segundo. Después del período refractario absoluto va el período refractario relativo durante el cual la fibra nerviosa puede propagar un potencial de acción siempre y cuando el estímulo sea mayor que el umbral. El período refractario completo supera unos cuantos milisegundos [6]; cf. figura 1.4.

Transmisión sináptica

Fisiológicamente las sinapsis son asimétricas: los neurotransmisores sólo se producen en la neurona presináptica. Cada neurona elabora y libera el mismo tipo de neurotransmisor en todas sus terminaciones sinápticas. El tiempo de transmisión de

los neurotransmisores químicos a través de la sinapsis química requiere de 0.3 a 1ms, período conocido con el nombre de *retardo sináptico*. Este período de tiempo es mucho más corto en el caso de sinapsis eléctricas debido a que la señal pasa directamente entre dos células adyacentes. Por otra parte, la sinapsis se fatiga, como consecuencia de la pérdida de reservas de neurotransmisores químicos [6].

Los neurotransmisores, en general, se liberan en forma de paquetes o *cuantos*, del orden de 1000 moléculas cada uno, de forma aleatoria. Esta liberación aumenta enormemente cuando llega un potencial de acción a la membrana presináptica, que causa la emisión de 100 – 300 cuantos en un período de tiempo muy corto. Los experimentos muestran que el modo en que se liberan las vacuolas y su efecto sobre la neurona postsináptica van a depender de la zona del cerebro en la que está localizada la sinapsis. Katz y colaboradores descubrieron, hace más de cuarenta años, que el vaciado de vacuolas de neurotransmisores ocurre, en determinadas posiciones individuales, de forma estocástica y en una fracción de tiempo pequeña [9,10]. Los experimentos, llevados a cabo en neuronas motoras con uniones sinápticas sobre músculos, mostraban que, ante un potencial de acción presináptico dado, se producía una respuesta postsináptica con mucha probabilidad —pese a la aleatoriedad en el vaciado— debido al hecho de que la respuesta del músculo está promediada sobre un número muy grande (> 1000) de posiciones de vaciado de neurotransmisores. En el cortex parece ser que el mecanismo es diferente; los experimentos parecen evidenciar que las comunicaciones entre neuronas corticales son, a menudo, altamente no fiables y ruidosas. En particular la probabilidad de vaciado es pequeña, 0.1 o menor [10,11], haciendo falta 9 o 10 estímulos (potenciales presinápticos) para producir un vaciado de las vacuolas. A diferencia de lo que ocurre en las neuronas motoras, en las sinapsis entre neuronas corticales hay muy pocas posiciones de vaciado de neurotransmisores [12,13]. La respuesta postsináptica a un estímulo presináptico es altamente variable, ya que está promediada sobre una población de posiciones de vaciado de neurotransmisores muy pequeña y poco efectiva. Esta variabilidad puede producir fluctuaciones en el tiempo y en el número de potenciales postsinápticos creados, dando respuestas postsinápticas diferentes a dos estímulos presinápticos consecutivos [4, 14]. En neuronas del hipocampo los experimentos han mostrado que el vaciado de neurotransmisores tiene dos componentes que

actúan a diferentes velocidades: una rápida y sincrónica y otra más lenta asincrónica [15].

La probabilidad de vaciado depende, también, de la historia de la actividad presináptica [16]. Así en algunas sinapsis, si un par de potenciales presinápticos ocurren en rápida sucesión, la probabilidad de vaciado aumenta de forma dramática para el segundo potencial, mientras que en otras sinapsis, la misma sucesión de eventos provoca una degradación de la respuesta en el segundo potencial. De esta forma la probabilidad de transmisión podría proporcionar el rango dinámico necesario para una rápida modulación de las intensidades sinápticas y dar cierta funcionalidad a la variabilidad sináptica. Recientes experimentos parecen evidenciar esto último [17], es decir, diferentes poblaciones de células presinápticas, con diferentes probabilidades de vaciado, sufren diferentes formas de modulación de corto periodo. Además, otros experimentos muestran que los cambios en la fiabilidad de la respuesta postsináptica parecen estar relacionados con ciertos mecanismos de plasticidad sináptica [18].

En resumen la transmisión sináptica parece tener una componente aleatoria muy importante y, para el caso de las neuronas corticales, dicha transmisión parece ser altamente ruidosa y poco fiable, produciendo una variabilidad estocástica en la respuesta postsináptica que podría tener cierta importancia en el buen funcionamiento del cerebro.

Recepción e integración somática

La interacción de un neurotransmisor con la membrana postsináptica puede provocar respuestas excitadoras o inhibitoras en ella. La respuesta excitatoria produce una depolarización por la entrada de iones Na^+ y la salida de K^+ a través de la membrana postsináptica. La inhibitora produce una hiperpolarización por la entrada de iones Na^+ y la salida de Cl^- . La forma en que un sitio de recepción responde a un neurotransmisor específico debe determinarse individualmente. Una vez que los neurotransmisores depolarizan parcialmente la membrana originan un potencial excitador postsináptico. La pequeña depolarización constituye un potencial graduado, decremental (analógico) que termina rápidamente en unos milisegundos o décimas de segundo. La respuesta gradual en una sinapsis no siempre alcanza el segmento inicial del axón (zona de Hillock), y la mayoría de los potenciales excitadores postsinápticos,

aunque lleguen a dicha zona, no generan potenciales de acción pues no alcanzan el umbral necesario. Así, las dendritas y el cuerpo celular están adaptadas no para la transmisión a larga distancia sino para la integración de la actividad sináptica. Todos los impulsos son acumulativos. El primer impulso que llega a la sinapsis produce un potencial excitador postsináptico que se suma, antes de desvanecerse, a un segundo impulso y así sucesivamente con un tercero, etc. La suma de sucesivos potenciales de la misma sinapsis recibe el nombre de *sumación o adición temporal*. La suma de potenciales excitadores postsinápticos, en diferentes sinapsis de la misma neurona recibe el nombre de *sumación o adición espacial*.

1.3 Psicología experimental

La psicología experimental tiene como uno de sus primeros y más importantes hitos los trabajos de Paulov (1906) sobre condicionamiento clásico. Recientemente se ha desarrollado de manera importante. Mencionamos los avances en la medida de los tiempos entre el estímulo de determinadas zonas del sistema nervioso y la respuesta que de dicho estímulo se deriva. En concreto, dichos tiempos se miden hoy en día con una precisión de un milisegundo. Estos conocimientos dan información sobre cómo han de ser las arquitecturas de los modelos a diseñar, así como la cantidad de datos cuantitativos que se tienen que tener en cuenta a la hora de crear un modelo concreto.

Son muchas las relaciones entre neurobiología y psicología experimental. Destacamos por su relevancia los trabajos de Kandel en los años setenta sobre la *aplysia*, un molusco marino. Kandel demostró que los parámetros neurofisiológicos son modificados por el condicionamiento. Esta observación demostraba la hipótesis de Hebb (1949) de que el condicionamiento podría modificar la intensidad de las sinapsis en una forma definida, dando una base fisiológica al proceso de aprendizaje [3, 7].

1.4 Inteligencia artificial

La inteligencia artificial (IA) se desarrolla como disciplina en los años sesenta en el MIT, particularmente por Minsky. Su aparición va emparejada al desarrollo de ordenadores cada vez más rápidos, en un intento por hacer que los ordenadores sean

capaces de imitar el comportamiento humano. Los éxitos de la IA han sido importantes en el campo de reconocimiento de patrones y solución de problemas complejos. Como principal logro destaca la creación de sistemas expertos, aunque dichos sistemas requieren participación humana. Uno de los puntos débiles de la IA está en la utilización de programas escritos para máquinas que trabajan en serie, de forma que la dinámica asociada es secuencial, muy lejos de la forma de operar tanto de las redes de neuronas biológicas como de las artificiales.

1.5 Redes de neuronas

Los modelos llamados redes neuronales artificiales o redes de neuronas son algoritmos matemáticos que simulan determinadas propiedades cognitivas de la mente, tales como el aprendizaje, la asociación de información y la optimización, que parecen ser fenómenos cooperativos. Están constituidas, en general, por los siguientes elementos:

1. Un conjunto de nodos o neuronas cuyo estado se representa mediante una variable de estado, $s_{\mathbf{x}}$, donde \mathbf{x} representa la posición espacial del nodo que sólo es relevante para distinguir los nodos entre sí.
2. Los nodos se relacionan mediante uniones, llamadas sinapsis, representadas por unas variables J_{xy} que, en general, toma valores reales.
3. Cada nodo lleva asociado un valor real, $\theta_{\mathbf{x}}$, que representa una energía umbral de excitación.
4. A cada nodo se le asocia una función de transferencia, $f_{\mathbf{x}}$, que determina su estado en función de las variables $\theta_{\mathbf{x}}$, J_{xy} y $s_{\mathbf{y}} \forall \mathbf{y} \neq \mathbf{x}$.²

El primer modelo de red de neuronas artificial fue propuesto por McCulloch y Pitts en 1943. En él se consideran las neuronas como puertas lógicas con dos posibles valores internos, activo o inactivo, representados por los valores $s_{\mathbf{x}} = 1$ o $s_{\mathbf{x}} = 0$, respectivamente. La neurona en \mathbf{x} tiene un conjunto de entradas constituidas por las salidas del resto de las neuronas ($s_{\mathbf{y}} \forall \mathbf{y} \neq \mathbf{x}$). Las entradas son sumadas y el estado de la neurona

²La terminología utilizada es la misma que la de las redes neuronales biológicas para resaltar la semejanza de las redes de neuronas artificiales con los sistemas biológicos.

es determinado por el valor de la señal resultante, llamado *campo local*, definido como $h_{\mathbf{x}} = \sum_{\mathbf{y} \neq \mathbf{x}} J_{xy} s_{\mathbf{y}} - \theta_{\mathbf{x}}$.³ Para un cierto umbral $\theta_{\mathbf{x}}$, la salida es $s_{\mathbf{x}} = 1$ si $h_{\mathbf{x}} > 0$ y $s_{\mathbf{x}} = 0$ si $h_{\mathbf{x}} < 0$. Para determinados valores de los parámetros J_{xy} y $\theta_{\mathbf{x}}$, este modelo tan sencillo permite obtener todas las operaciones lógicas sencillas (NOT, AND y OR) y, por combinación de ellas, cualquier otra operación lógica (como XOR; cf. figura 1.5) de forma que cualquier máquina universal de Turing⁴ puede ser diseñada usando neuronas como componentes fundamentales. El problema de este modelo, pese a su potencia de cálculo (asociada con el hecho de que hay un conjunto de procesadores, las neuronas, funcionando en paralelo) está en que no es adaptativo y no proporciona criterios para elegir el valor de las sinapsis a la hora de realizar una determinada tarea cognitiva. En 1954 Cragg y Temperley hicieron explícita la idea de que la actividad de una red de neuronas es el resultado del comportamiento cooperativo de las neuronas que la constituyen. En 1974 Little notó la similitud entre una red de neuronas del tipo propuesto por McCulloch y Pitts con ciertos sistemas magnéticos, introduciendo la posibilidad de ruido en la red, esto es, considerando la red de neuronas como una red autómatas probabilista. Estas ideas fueron desarrolladas por Hopfield en 1982 [1] que mostró cómo la memoria es un proceso cooperativo, resultado de la acción conjunta de un número elevado de unidades elementales (neuronas) y cómo la memorización puede tener lugar tomando como valor de las sinapsis el que se sigue de la prescripción realizada por Hebb. Según ésta, la intensidad de una determinada sinapsis puede ser ajustada si su nivel de actividad cambia, de forma que una sinapsis que activa continuamente a la neurona postsináptica crecerá en intensidad, mientras que las menos activas tenderán a debilitarse [3, 5]. Esta plasticidad sináptica ha sido comprobada en

³El campo local así definido da cuenta de la influencia de las señales (estímulos o potenciales presinápticos) provenientes del conjunto de neuronas que están en contacto con una neurona dada, aquí representada por \mathbf{x} , sobre dicha neurona, provocando o no en la misma la emisión de un potencial de acción. En ciertos modelos deterministas ($T=0$) la respuesta de la neurona en \mathbf{x} bajo la acción de estos estímulos es de la forma:

$$s_{\mathbf{x}} = \text{sign}(h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})), \quad (1.2)$$

expresión que es análoga a las que se obtienen en los modelos de condicionamiento clásico [7].

⁴Una máquina de Turing es una máquina ideal constituida por una cinta infinita, en la que están codificadas, en alguna forma, una serie de instrucciones algorítmicas, y un dispositivo, que tiene un número finito de estados internos, que puede utilizar de alguna forma las instrucciones de la cinta para llevar a cabo un determinado cálculo. Una máquina universal de Turing es una máquina de Turing en la que la parte inicial de la cinta contiene la lista de instrucciones correspondientes a cualquier máquina de Turing [19].

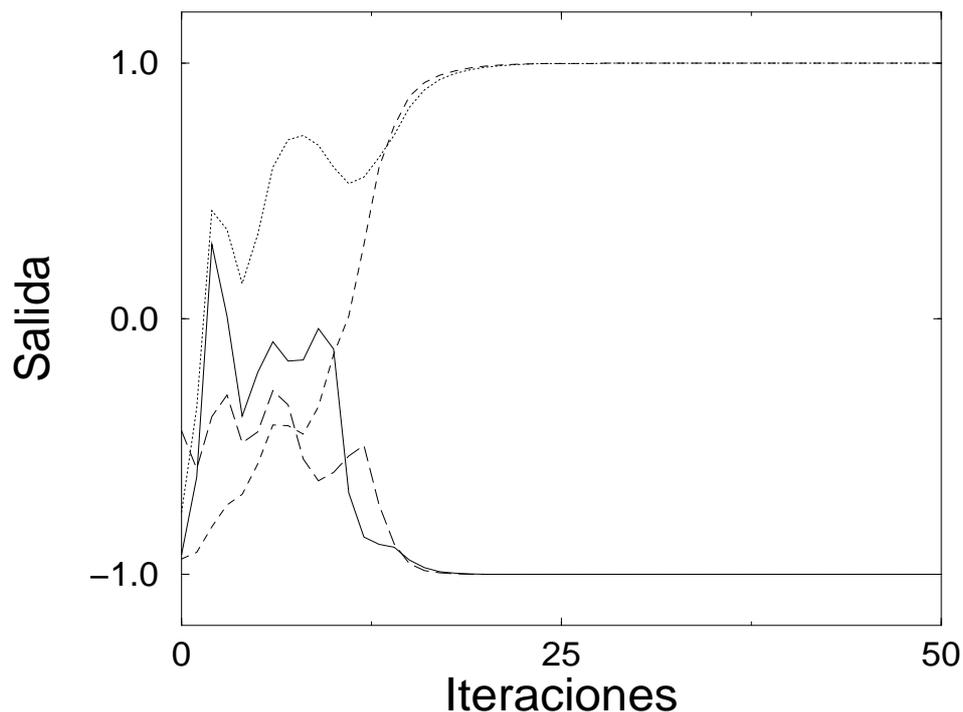


Figura 1.5: La figura muestra la salida de una red de neuronas utilizada para construir la operación lógica XOR. Esta operación lógica sólo puede realizarse con una red de neuronas que incluya capas centrales. Cada curva representa una de las cuatro posibles salidas que tiene dicha operación lógica y su evolución en el proceso de entrenamiento de la red.

observaciones experimentales, hasta el punto de que el modelo de Hopfield es considerado como el primer modelo matemáticamente bien definido que establece una relación útil entre neurofisiología y psicología experimental.

La idea central de todos los modelos de redes de neuronas es que la actividad neuronal es un fenómeno cooperativo que involucra una fuerte conectividad entre todas las unidades elementales que constituyen el sistema. Dicha conectividad puede ser completa, con todas las neuronas interconectadas entre sí, como ocurre en el modelo de Hopfield, o bien puede existir un flujo de información que va desde unas neuronas que funcionan como entrada, hacia neuronas que funcionan como salida, como ocurre con los modelos *feedforward*. En todos los modelos de redes de neuronas el aprendizaje es un proceso en el que se ajustan los parámetros del sistema, es decir, la intensidad de

las sinapsis, cuyo número es del orden N^2 , y los N umbrales $\theta_{\mathbf{x}}$, donde N es el número de neuronas, de forma que el sistema dé salidas deseadas a determinadas entradas. El problema del aprendizaje es más difícil de resolver en redes completamente conectadas que en redes *feedforward*. En todo caso, ha de notarse que es este carácter cooperativo, y la sencillez matemática de los sistemas de redes de neuronas, lo que permite aplicar aquí con éxito los potentes métodos de la física estadística, hecho éste que explica su popularidad desde la propuesta de Hopfield.

Son muchos los resultados que la física estadística ha proporcionado en este campo como, por ejemplo, la caracterización de los estados estacionarios, diagramas de fase que ponen de manifiesto en qué regiones de los mismos, esto es, para qué valores de los parámetros el sistema puede almacenar y recordar información, determinación del número máximo de memorias que dichos sistemas pueden almacenar, etc. Estos modelos funcionan bien, desde un punto de vista computacional, en ciertas regiones de sus correspondientes diagramas de fase. Sin embargo, estos modelos difieren bastante de las redes biológicas en muchos aspectos, esto es, resultan muy artificiales. Entre estos aspectos no realistas vamos a mencionar tres: En primer lugar, son modelos que consideran al sistema en equilibrio a una cierta temperatura T , estado que se supone caracterizado por un cierto funcional energía $\mathcal{H}(\mathbf{s}, \mathbf{J})^5$. En segundo lugar, una vez que ha finalizado el proceso de aprendizaje de la red, las sinapsis permanecen fijas, sin cambiar con el tiempo. En tercer lugar, en estos modelos las sinapsis son consideradas, en general, simétricas. Ninguno de estos tres aspectos es realista, puesto que los correspondientes procesos biológicos, son dinámicos y mantienen al sistema fuera del equilibrio, aunque pueden presentarse situaciones estacionarias. Por otra parte, como hemos discutido antes (sección §1.2), en las sinapsis tienen lugar multitud de procesos dinámicos —de carácter estocástico en algunos casos— que pueden transcurrir a diferentes velocidades (sinapsis químicas y eléctricas, diferentes grados de fiabilidad en la transmisión sináptica, etc), que hacen que pueda haber fluctuaciones en la intensidad de las mismas (variación local y rápida del número de vacuolas de neurotransmisores como consecuencia de la llegada de un potencial de acción, diferente respuesta postsináptica a dos estímulos presinápticos consecutivos, etc). Por último,

⁵Aquí $\mathbf{s} \equiv \{s_{\mathbf{x}}\}$ y $\mathbf{J} \equiv \{J_{xy}\}$ representan una *configuración* de estado de todas las neuronas y de todas las sinapsis, respectivamente

hemos visto (sección §1.2) que, para determinados tipos de sinapsis (sinapsis químicas), éstas pueden presentar un fuerte grado de asimetría (sólo la membrana presináptica genera vacuolas llenas de neurotransmisores). Estos argumentos sugieren la conveniencia de construir modelos de redes de neuronas más realistas que los actuales, que tengan en cuenta aspectos neurofisiológicos no considerados hasta ahora y mejoren su aplicación en el campo de la computación.

Capítulo 2

Modelo cinético de red de neuronas

2.1 Introducción

El modelo de Hopfield de memoria asociativa¹ [1], así como gran parte de los modelos basados en él, consiste de un conjunto de N neuronas binarias, cuya actividad puede tomar uno de los valores $s_{\mathbf{x}} = \pm 1$ en cada instante de tiempo. En estos modelos, en general, la actividad evoluciona con el tiempo mediante algún tipo de proceso que puede ser determinista o estocástico (correspondiendo cada uno, respectivamente, a una situación en la que tenemos temperatura cero o temperatura finita distinta de cero). Las neuronas, que podemos imaginar (pero no es esencial) localizadas en los nudos de una red d -dimensional, Λ_d , están interactuando entre sí de acuerdo con una determinada regla de aprendizaje como, por ejemplo, la regla de Hebb [3]. Esto es, la neurona que ocupa el nudo que está en la posición \mathbf{x} está conectada con la neurona que ocupa el nudo cuya posición es \mathbf{y} mediante una sinapsis cuya intensidad es

$$J_{xy} = \frac{1}{N} \sum_{\mu=1}^P \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}, \quad (2.1)$$

donde $\{\xi_{\mathbf{x}}^{\mu} = \pm 1; \mathbf{x} \in \Lambda_d\} \equiv \boldsymbol{\xi}^{\mu}$ representa $\mu = 1, \dots, P$ patrones almacenados. En estos modelos, los valores J_{xy} son independientes del tiempo. Esto suele ser interpretado de dos formas, bien considerando que las intensidades sinápticas han sido determinadas durante un proceso previo de aprendizaje (*plasticidad sináptica*), consistente en que,

¹El término memoria asociativa está ligado a la capacidad que tiene la red de neuronas de recuperar la información correspondiente a un determinado número de patrones, que previamente ha almacenado, a partir del conocimiento parcial o imperfecto de dicha información, sin conocer la localización precisa donde ha sido almacenada.

cada vez que el sistema memoriza un determinado patrón ξ^μ , las intensidades sinápticas se modifican en la forma

$$J_{xy}^{new} = J_{xy}^{old} + \frac{\xi_x^\mu \xi_y^\mu}{N}, \quad (2.2)$$

o bien considerando que existen dos escalas temporales, bien definidas y diferentes, para la evolución de las sinapsis y de las neuronas, respectivamente, la primera mucho más grande que la segunda.

En el modelo de Hopfield interesa evaluar la importancia que tienen los detalles microscópicos en las propiedades macroscópicas de la red, y así han surgido una serie de modelos, modificaciones del original, que tratan de explicar características esenciales de los sistemas biológicos; cf. [20–25]. Apuntando en esta dirección y teniendo en cuenta las consideraciones sobre neurofisiología en el capítulo anterior, podemos argumentar que no es realista el considerar sólo las variaciones temporales de las sinapsis que ocurren en el proceso de aprendizaje. Una situación típica en los sistemas biológicos [9–14, 16, 17, 26] consiste en que las intensidades sinápticas varían con el tiempo, también, en una escala temporal mucho menor que aquélla en la que tiene lugar la evolución de la actividad de las neuronas. Así, en un intervalo de tiempo Δt del orden del tiempo que transcurre entre la generación de dos potenciales de acción consecutivos, el estado de las sinapsis puede cambiar de forma sustancial, debido, por ejemplo, a las variaciones en el número de vacuolas en los botones sinápticos que inducen fluctuaciones locales en la concentración de neurotransmisores [5, 7]. La motivación para considerar variaciones locales de las intensidades sinápticas J_{xy} en los modelos de redes de neuronas, también tiene su origen en el hecho de que las neuronas biológicas están, en la práctica, conectadas por más de una sinapsis [6], cada una teniendo diferente naturaleza, es decir químicas o eléctricas, las cuales transmiten el potencial de acción a diferente velocidad; cf. §1.2.

En este capítulo se presenta un modelo de red de neuronas [27] que, además de tener en cuenta la *lenta* plasticidad sináptica debida al aprendizaje, tiene en cuenta fluctuaciones locales, relativamente rápidas, de las intensidades sinápticas. Motivado por lo que ocurre en los sistemas biológicos, las intensidades J_{xy} varían aleatoriamente con el tiempo de tal forma que su promedio sobre Δt , \bar{J}_{xy} , tiene el valor correspondiente a una determinada regla de aprendizaje. Por ejemplo, asumiendo que dicha regla de

aprendizaje es la regla de Hebb, J_{xy} constantemente *visita* al azar, durante toda la evolución, los patrones memorizados, tomando así valores instantáneos $J_{xy} = \frac{\xi_x^\mu \xi_y^\mu}{N}$ con una determinada probabilidad, de forma que \bar{J}_{xy} viene dado por (2.1).

Está claro, a la vista de la situación descrita en el capítulo anterior, que es interesante ver la influencia de tales fluctuaciones en propiedades emergentes del sistema. En este capítulo de la memoria nos vamos a centrar en el estudio de distribuciones específicas para las fluctuaciones que producen algunos resultados exactos, como hamiltonianos efectivos, tanto para la situación en la que las intensidades sinápticas son simétricas, $J_{xy} = J_{yx}$ (§2.3), como para el caso en el que dichas intensidades son asimétricas, $J_{xy} \neq J_{yx}$ (§2.4; ver también §2.6). Partiendo de estos hamiltonianos, usamos la técnica de *réplicas* (cf. §2.5 y apéndice A) para obtener resultados explícitos. La descripción más general que hacemos del modelo se consigue mediante una aproximación cinética de campo medio (§2.6), que revela un variado comportamiento del sistema. También se prueba explícitamente que las fluctuaciones introducen un ruido adicional que afecta significativamente la propiedad de memoria asociativa del sistema. En particular, comparando con la situación ordinaria, en la que no ocurren tales fluctuaciones, la región del espacio de las fases, a temperatura finita, en la que aparecen los estados vidrio de espines es menor en el modelo que aquí se presenta, y no aparece a temperatura cero (es decir, en ausencia de ruido térmico) por encima de un valor crítico para el número de patrones almacenados. Por otra parte, esta versión del modelo no es afectada de forma crítica por la *línea Almeida-Thouless* o límite de estabilidad de la solución con simetría de réplicas [28]. Es también interesante notar el hecho de que nuestro modelo admite distribuciones para las fluctuaciones más generales de las que se han estudiado explícitamente aquí, y que uno puede idear procesos temporales para las sinapsis más realistas que los considerados en este capítulo, algunos de los cuales se desarrollan en el próximo capítulo de esta memoria [29].

2.2 Definición del modelo

Consideremos un sistema constituido por N neuronas con actividad representamos por $s_{\mathbf{x}}$, conectadas entre sí mediante sinapsis cuya intensidad denotamos mediante la variable J_{xy} . Denotemos por $\mathbf{s} \equiv \{s_{\mathbf{x}} = \pm 1; \mathbf{x} \in \Lambda_d\}$ y $\mathbf{J} \equiv \{J_{xy} \in \mathfrak{R}; \mathbf{x}, \mathbf{y} \in \Lambda_d\}$ una

configuración de neuronas y de sinapsis, respectivamente. La probabilidad del estado (\mathbf{s}, \mathbf{J}) en el instante de tiempo t varía de forma estocástica de acuerdo con la ecuación maestra:

$$\partial_t P_t(\mathbf{s}, \mathbf{J}) = \sum_{\mathbf{s}', \mathbf{J}'} [P_t(\mathbf{s}', \mathbf{J}') c(\mathbf{s}', \mathbf{J}' \rightarrow \mathbf{s}, \mathbf{J}) - P_t(\mathbf{s}, \mathbf{J}) c(\mathbf{s}, \mathbf{J} \rightarrow \mathbf{s}', \mathbf{J}')], \quad (2.3)$$

donde la doble sumatoria está extendida a todas las configuraciones de neuronas y de sinapsis, respectivamente. Esto es, consideramos que determinados agentes, entre los que se incluye un baño térmico a temperatura T , inducen cambios $(\mathbf{s}, \mathbf{J}) \rightarrow (\mathbf{s}', \mathbf{J}')$ cuya probabilidad por unidad de tiempo (*rate*) es

$$c(\mathbf{s}, \mathbf{J} \rightarrow \mathbf{s}', \mathbf{J}') = p c_1(\mathbf{s} \rightarrow \mathbf{s}' | \mathbf{J}) \delta_{\mathbf{J}, \mathbf{J}'} + (1 - p) c_2(\mathbf{J} \rightarrow \mathbf{J}') \delta_{\mathbf{s}, \mathbf{s}'}, \quad (2.4)$$

donde $\delta_{X, X'}$ representa la función delta de Kronecker,

$$c_1(\mathbf{s} \rightarrow \mathbf{s}' | \mathbf{J}) = \sum_{\mathbf{y}} \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{y}) \delta_{s_{\mathbf{y}}, -s'_{\mathbf{y}}} \prod_{\mathbf{x} \neq \mathbf{y}} \delta_{s_{\mathbf{x}}, s'_{\mathbf{x}}} \quad (2.5)$$

y

$$c_2(\mathbf{J} \rightarrow \mathbf{J}') = \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \varpi(J_{xy} \rightarrow J'_{xy}) (1 - \delta_{J_{xy}, J'_{xy}}) \prod_{\substack{\mathbf{u}, \mathbf{v} \\ (\mathbf{u}, \mathbf{v}) \neq (\mathbf{x}, \mathbf{y})}} \delta_{J_{uv}, J'_{uv}}, \quad (2.6)$$

donde las sumas y productos se extienden a todo el sistema con las restricciones mostradas explícitamente en cada caso. Esto es, la evolución temporal tiene lugar mediante la superposición de dos clases diferentes de procesos elementales (sólo uno en la unidad de tiempo), que son cambios $s_{\mathbf{x}} \rightarrow -s_{\mathbf{x}}$ que ocurren con *velocidad* $p \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x})$ para una configuración de sinapsis dada, y cambios $J_{xy} \rightarrow J'_{xy}$ que ocurren con *velocidad* $(1 - p) \varpi(J_{xy} \rightarrow J'_{xy})$, independiente de la configuración de neuronas que en ese instante hay en el sistema. Aquí el parámetro p mide la probabilidad del proceso de evolución de las neuronas con respecto al de evolución de las sinapsis. En estas condiciones, haciendo uso de las expresiones (2.4)-(2.6) e introduciéndolas en (2.3), podemos escribir la ecuación maestra (2.3) en la forma

$$\begin{aligned} \partial_t P_t(\mathbf{s}, \mathbf{J}) = & p \sum_{\mathbf{x}} [\varpi_{\mathbf{J}}(\mathbf{s}^{\mathbf{x}}; \mathbf{x}) P_t(\mathbf{s}^{\mathbf{x}}, \mathbf{J}) - \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x}) P_t(\mathbf{s}, \mathbf{J})] \\ & + (1 - p) \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \sum_{J'_{xy}} [\varpi(J'_{xy} \rightarrow J_{xy}) P_t(\mathbf{s}, \mathbf{J}^{xy}) - \varpi(J_{xy} \rightarrow J'_{xy}) P_t(\mathbf{s}, \mathbf{J})], \end{aligned} \quad (2.7)$$

donde \mathbf{s}^x representa la configuración de neuronas obtenida de \mathbf{s} después de cambiar $s_x \rightarrow -s_x$, y \mathbf{J}^{xy} es la configuración de sinapsis que se obtiene a partir de \mathbf{J} después de que haya tenido lugar el cambio $J_{xy} \rightarrow J'_{xy}$. Por simplicidad, a partir de este momento nos restringimos a la situación particular en la que se tiene

$$\varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x}) = \Psi [2\beta s_x h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})], \quad (2.8)$$

donde $\beta = (k_B T)^{-1}$, k_B es la constante de Boltzmann, y

$$h_{\mathbf{x}}(\mathbf{s}, \mathbf{J}) = \sum_{y \neq x} J_{xy} s_y - \theta_{\mathbf{x}} \quad (2.9)$$

es el campo local asociado a la neurona que ocupa la posición \mathbf{x} con $\theta_{\mathbf{x}}$ la energía umbral necesaria para activar esa neurona. En (2.8) Ψ es una función arbitraria con la única condición de que verifique $\Psi(0) = 1$, $\lim_{X \rightarrow +\infty} \Psi(X) = 0$, y $\Psi(X) = \Psi(-X) \exp(-X)$; esta última condición se conoce con el nombre de condición de balance detallado. Ejemplos de este tipo de funciones son

$$\Psi(X) = \begin{cases} e^{-\frac{X}{2}}, & \text{rate } V \\ 2(1 + e^X)^{-1}, & \text{rate } K \\ \min\{1, e^{-X}\}, & \text{rate } M. \end{cases} \quad (2.10)$$

En el caso de ciertos límites bien definidos, el considerar estas elecciones simplificatorias, es decir (2.8), (2.9), y la independencia de ϖ ($J_{xy} \rightarrow J'_{xy}$) en \mathbf{s} , reduce el modelo a situaciones más familiares, haciendo que nuestro modelo sea una generalización de las mismas, válido para estudiar otras situaciones hasta ahora no tenidas en cuenta. Así, para $p \equiv 1$, $c(\mathbf{s}, \mathbf{J} \rightarrow \mathbf{s}', \mathbf{J}') = c_1(\mathbf{s} \rightarrow \mathbf{s}' | \mathbf{J}) \delta_{\mathbf{J}, \mathbf{J}'}$, y (2.7), la ecuación maestra dada en (2.3) queda transformada en

$$\partial_t P_t(\mathbf{s}, \mathbf{J}) = \sum_{\mathbf{y}} [\varpi_{\mathbf{J}}(\mathbf{s}^y; \mathbf{y}) P_t(\mathbf{s}^y, \mathbf{J}) - \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{y}) P_t(\mathbf{s}, \mathbf{J})]. \quad (2.11)$$

Esta ecuación describe una situación en el sistema en la que las sinapsis permanecen *congeladas* (*quenched*) durante el intervalo de tiempo que caracteriza la variación temporal de \mathbf{s} . Así, el valor congelado de \mathbf{J} estaría determinado, por ejemplo, mediante un proceso previo de memorización de P patrones ξ^μ , tras el cual no ocurre ninguna

modificación sináptica adicional (en la escala de tiempo de variación de \mathbf{s}). Varios casos particulares, asociados con diferentes simetrías y reglas de aprendizaje, han sido considerados antes por diferentes autores, por ejemplo:

(i) $J_{xy} = J_{yx}$. En general esta simetría no es realista si se trata de modelar sistemas biológicos [6], pero tiene la ventaja de que reduce el problema original a una forma más sencilla. Esto es así porque la condición de balance detallado implica entonces que el estado estacionario se corresponda con un estado de equilibrio termodinámico a temperatura T con energía

$$\mathcal{H}_{\mathbf{J}}(\mathbf{s}) = -\frac{1}{2} \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} J_{xy} s_{\mathbf{x}} s_{\mathbf{y}} + \sum_{\mathbf{x}} \theta_{\mathbf{x}} s_{\mathbf{x}}. \quad (2.12)$$

Este es el hamiltoniano de Ising si $J_{xy} = J$ para cualquier par de vecinos próximos y cero en cualquier otro caso. Si consideramos, por ejemplo, una distribución espacialmente aleatoria para las sinapsis J_{xy} , podemos obtener el modelo de Edwards-Anderson de vidrios de espines [30], o bien cualquiera de las familiares versiones de él. Alternativamente, tomando como regla de aprendizaje la regla de Hebb (2.1), da lugar al modelo de Hopfield [1]. Podemos considerar también una regla de aprendizaje no lineal [31–33] como, por ejemplo, $J_{xy} = N^{-1} P^{1/2} \phi(P^{1/2} \boldsymbol{\xi}_{\mathbf{x}} \cdot \boldsymbol{\xi}_{\mathbf{y}})$, donde $\boldsymbol{\xi}_{\mathbf{x}} \equiv \{\xi_{\mathbf{x}}^{\mu} = \pm 1; \mu = 1, \dots, P\}$; este caso se reduce al modelo de Hopfield para $\phi(X) = X$.

(ii) $J_{xy} \neq J_{yx}$. Para esta situación, no es posible, en general, hacer una descripción hamiltoniana, pero podemos obtener alguna información interesante a partir de la ecuación maestra, por ejemplo, multiplicando los dos miembros de (2.11) por $s_{\mathbf{x}}$ y sumando sobre todas las posibles configuraciones \mathbf{s} de actividades neuronales, se obtiene que

$$\partial_t \langle s_{\mathbf{x}} \rangle = -2 \langle s_{\mathbf{x}} \Psi [2\beta s_{\mathbf{x}} h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})] \rangle. \quad (2.13)$$

Después de introducir el campo local medio $m_{\mathbf{x}}(\mathbf{J}) \equiv \sum_{\mathbf{y}} J_{xy} \langle s_{\mathbf{y}} \rangle$ (por simplicidad, tomamos $\theta_{\mathbf{x}} = 0 \forall \mathbf{x}$ de aquí en adelante), y considerando la aproximación de campo medio $s_{\mathbf{x}} \approx \langle s_{\mathbf{x}} \rangle$, que implica $h_{\mathbf{x}}(\mathbf{s}, \mathbf{J}) \approx m_{\mathbf{x}}(\mathbf{J})$, se obtiene que la solución a (2.13) que caracteriza el estado estacionario del sistema verifica $\langle s_{\mathbf{x}} \rangle \approx \tanh \{\beta m_{\mathbf{x}}(\mathbf{J})\}$, donde el campo local medio se obtiene mediante la relación de autoconsistencia

$$m_{\mathbf{x}}(\mathbf{J}) = \sum_{\mathbf{y} \neq \mathbf{x}} J_{xy} \tanh \{\beta m_{\mathbf{x}}(\mathbf{J})\}, \quad J_{xy} \neq J_{yx}. \quad (2.14)$$

Ésta es la solución estacionaria obtenida antes por Crisanti y Sompolinsky para su modelo dinámico de red de Hopfield con acoplamientos asimétricos [34].

La discusión precedente indica que la ecuación (2.7), que caracteriza una clase muy interesante de modelos tipo Hopfield, puede permitirnos ilustrar, en cierta forma, la influencia, en el comportamiento emergente de los sistemas de redes de neuronas, de algunas características o propiedades de los mismos no consideradas hasta el momento y, como consecuencia, evaluar su papel en sistemas más realistas como son los sistemas neuronales biológicos. Como un primer paso, estamos particularmente interesados aquí en el límite $p \rightarrow 0$. Esto se corresponde con una situación en la que, una vez que el proceso de aprendizaje ha sido completado (o durante el mismo pero en una escala de tiempo muy inferior a la de la variación de las sinapsis consecuencia del aprendizaje), las sinapsis cambian muy rápidamente (fluctuaciones rápidas) comparado con los cambios en la actividad de las neuronas, de tal forma que se pueden distinguir dos escalas de tiempo bien definidas [35]. Por una parte hay una escala temporal microscópica, τ , para las fluctuaciones rápidas de las sinapsis, en la que la actividad de las neuronas no evoluciona de forma apreciable, y por otra parte hay una escala temporal, $t = p\tau$ para $p \rightarrow 0$ y $\tau \rightarrow \infty$, diferente de la anterior, en la que las neuronas evolucionan bajo una distribución estacionaria para las sinapsis. Bajo tal condición, para la que hay ciertas motivaciones biológicas (cf. §1.2), la ecuación (2.7) se transforma en

$$\partial_{\tau} f_{\tau}(\mathbf{J}) = \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \sum_{J'_{xy}} [\varpi(J'_{xy} \rightarrow J_{xy}) f_{\tau}(\mathbf{J}^{xy}) - \varpi(J_{xy} \rightarrow J'_{xy}) f_{\tau}(\mathbf{J})] \quad (2.15)$$

para las sinapsis, y

$$\partial_t P_t(\mathbf{s}) = \sum_{\mathbf{x}} [\bar{\varpi}(\mathbf{s}^{\mathbf{x}}; \mathbf{x}) P_t(\mathbf{s}^{\mathbf{x}}) - \bar{\varpi}(\mathbf{s}; \mathbf{x}) P_t(\mathbf{s})] \quad (2.16)$$

para las neuronas, donde

$$\bar{\varpi}(\mathbf{s}; \mathbf{x}) = \int d\mathbf{J} f(\mathbf{J}) \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x}) \quad (2.17)$$

con $f(\mathbf{J})$ la solución estacionaria a (2.15), es decir $f(\mathbf{J}) = \lim_{\tau \rightarrow \infty} f_{\tau}(\mathbf{J})$. En otras palabras, el sistema se reduce en el límite $p \rightarrow 0$ al proceso marcoviano (2.16) con velocidades efectivas dadas por (2.17).

Este es el sistema que hemos estudiado con detalle para diferentes elecciones de $f(\mathbf{J})$ y de $\varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x})$. Solamente vamos a considerar aquí velocidades elementales $\varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x})$ elegidas de forma que se verifiquen (2.8) y (2.9) (aunque pueden ser consideradas otro tipo de situaciones sencillas). La elección para $f(\mathbf{J})$ debe hacerse de forma que tenga en cuenta la naturaleza de simetría o asimetría de J_{xy} . A la hora de considerar variaciones del modelo, correspondientes a situaciones más familiares, suponemos que las fluctuaciones descritas por la distribución $f(\mathbf{J})$ son alrededor de valores medios correspondientes a una determinada regla de aprendizaje, esto es, por ejemplo,

$$\bar{J}_{xy} \equiv \int d\mathbf{J} f(\mathbf{J}) J_{xy} = \frac{1}{N} \sum_{\mu=1}^P \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}, \quad (2.18)$$

para el caso de la regla de Hebb. Una elección particular que es consistente con esta consideración consiste en tomar

$$f(\mathbf{J}) = \prod_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \delta \left(J_{xy} - \frac{1}{N} \sum_{\mu=1}^P \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} \right); \quad (2.19)$$

sin embargo esta elección impide en la práctica la posibilidad de cualquier tipo de fluctuación, y el modelo se reduce al modelo de Hopfield con $J_{xy} = J_{yx}$ dados por la regla de Hebb. Más interesante es la elección

$$f(\mathbf{J}) = \prod_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \sum_{\mu=1}^P a_{\mu} \delta \left(J_{xy} - \eta_{xy}^{\mu} \right), \quad (2.20)$$

que implica (2.18) para

$$\eta_{xy}^{\mu} = \lambda_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}, \quad \lambda_{\mu} = (a_{\mu} N)^{-1}. \quad (2.21)$$

La elección (2.20) se corresponde con la situación en la que los patrones memorizados contribuyen con una cantidad η_{xy}^{μ} con probabilidad a_{μ} a la intensidad sináptica J_{xy} , independientemente de las otras, de forma que no existe correlación espacial entre las diferentes sinapsis. Una variedad de resultados para esta situación serán mostrados en las próximas secciones de este capítulo. Otra situación de interés que implica (2.18) consiste en tomar

$$f(\mathbf{J}) = \sum_{\mu=1}^P a_{\mu} \prod_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \delta \left(J_{xy} - \eta_{xy}^{\mu} \right), \quad (2.22)$$

es decir, cada patrón memorizado ξ^μ contribuye con probabilidad a_μ a una configuración completa de sinapsis \mathbf{J} , permitiendo, a diferencia de la anterior elección, es decir (2.20), que existan correlaciones espaciales entre los diferentes patrones almacenados. El estudio detallado de nuestro sistema correspondiente a esta situación será el objeto de los dos próximos capítulos de esta memoria; cf §3, §4 y Ref. [29].

2.3 Resultados exactos para sinapsis simétricas

Consideramos fluctuaciones en las sinapsis de tal forma que, *en promedio*, las neuronas estén unidas unas con otras mediante acoplamientos sinápticos dados a partir de alguna regla de aprendizaje. A la hora de determinar las consecuencias de esto en las propiedades macroscópicas del modelo, en primer lugar vamos a tratar una de sus más simples realizaciones. El propósito de esto es obtener resultados exactos, que sirvan para describir el sistema en función de un hamiltoniano efectivo [36, 37], objetivo que va a restringir bastante las posibles elecciones de las funciones $\Psi(X)$ y de la distribución $f(\mathbf{J})$ que podemos utilizar. De hecho vamos a considerar sólo el caso de *rate V* —cf. ecuación (2.10)— y

$$f(\mathbf{J}) = \prod_{\substack{x,y \\ x \neq y}} f(J_{xy}). \quad (2.23)$$

De aquí se sigue que existe un hamiltoniano efectivo (cf. apéndice B) que, como en el caso de Hopfield, contiene la información relevante sobre los patrones almacenados. Vamos a considerar tanto la situación en la que los acoplamientos son simétricos como la situación en que son asimétricos; este último caso será estudiado con detalle en §2.4.

Consideremos pues la elección (2.20) en el contexto de sinapsis simétricas, es decir, $f(J_{xy}) = \sum_\mu a_\mu \delta(J_{xy} - \lambda_\mu \xi_x^\mu \xi_y^\mu)$, por lo que las fluctuaciones son alrededor de valores medios dados por la regla de Hebb, sin más que tomar $\lambda_\mu = (a_\mu N)^{-1}$; en el caso particular de que todos los patrones sean equiprobables y de la condición de normalización $\sum_\mu a_\mu = 1$, se tiene $\lambda_\mu = \alpha$ con $\alpha \equiv PN^{-1}$; ² cf. (2.21). Las fluctuaciones alrededor de la regla de Hebb entonces están caracterizadas por

²Aquí el parámetro α es una medida de la capacidad de almacenamiento de información de la red de neuronas.

$$\sigma_{xy}^2(\boldsymbol{\xi}) \equiv \overline{(J_{xy} - \bar{J}_{xy})^2} = \alpha^2 - \frac{1}{N^2} \sum_{\mu, \nu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} \xi_{\mathbf{x}}^{\nu} \xi_{\mathbf{y}}^{\nu}, \quad (2.24)$$

magnitud que varía entre 0 y $2\alpha^2$. Esto es, en el límite termodinámico, $N \rightarrow \infty$, con $\alpha \neq 0$ (es decir, el número P de patrones que el sistema almacena es comparable con el tamaño del sistema), el valor mayor de $\sigma_{xy}^2(\boldsymbol{\xi})$ ocurre cuando los patrones son tales que verifican $\xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} \xi_{\mathbf{x}}^{\nu} \xi_{\mathbf{y}}^{\nu} = -1 \quad \forall \mu, \nu, \mathbf{x}, \mathbf{y}$. En caso contrario, es decir $\alpha = 0$, el último término de (2.24), que mide la correlación entre patrones aprendidos, también se anula. En algunas situaciones tomamos los patrones memorizados de forma aleatoria con probabilidades

$$p(\xi_{\mathbf{x}}^{\mu}) = \frac{1}{2} [\delta(\xi_{\mathbf{x}}^{\mu} - 1) + \delta(\xi_{\mathbf{x}}^{\mu} + 1)]; \quad (2.25)$$

de donde para esta caso se obtiene $\sigma_{xy}^2(\boldsymbol{\xi}) \simeq \alpha^2$. Una primera conclusión que se obtiene es que el efecto de las fluctuaciones depende de forma muy importante del grado de correlación entre los patrones almacenados por el sistema, incrementándose conforme los patrones tienden a ser independientes los unos de los otros. Por otra parte, también las fluctuaciones se incrementan con el número α de patrones almacenados relativo al tamaño del sistema.

Vamos a estudiar con más detalle la influencia de las fluctuaciones en el estado estacionario del sistema. Para el caso particular de *rate* V y $f(\mathbf{J})$ dada por (2.23), se puede probar [38, 39] que la solución estacionaria de (2.16) es³

$$P_{st}(\mathbf{s}) \propto \exp(-\beta \mathcal{H}_{eff}) \quad (2.26)$$

con

$$\mathcal{H}_{eff}(\mathcal{J}, \mathbf{s}) = -\frac{1}{2} \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \mathcal{J}_{xy} s_{\mathbf{x}} s_{\mathbf{y}}, \quad (2.27)$$

donde $\mathcal{J} = \{\mathcal{J}_{xy}\}$ representa una configuración de intensidades sinápticas *efectivas* dadas por

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln \frac{\overline{\exp(\beta J_{xy})}}{\exp(-\beta J_{xy})}. \quad (2.28)$$

En el caso particular en que la distribución de sinapsis viene dada por la expresión (2.20) con $a_{\mu} = P^{-1}$, de forma que (2.18) se verifique, se tiene que

³Los detalles pueden verse en el apéndice B.

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln \left\{ \frac{1 + \rho_{xy} \tanh(\beta\alpha)}{1 - \rho_{xy} \tanh(\beta\alpha)} \right\} \quad (2.29)$$

con $\rho_{xy} \equiv \frac{1}{P} \sum_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}$. Si, además de que los patrones son equiprobables, los elementos de los patrones, es decir las variables $\xi_{\mathbf{x}}^{\mu}$, son variables aleatorias estadísticamente independientes con probabilidades dadas por (2.23), es de esperar que ρ_{xy} tome valores muy pequeños en el límite $N, P \rightarrow \infty$,⁴ de donde, a partir de (2.29) y utilizando esto último, se tiene

$$\mathcal{J}_{xy} \approx \frac{\mathcal{A}(\alpha)}{N} \sum_{\mu=1}^P \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}, \quad \mathcal{A}(\alpha) = (\beta\alpha)^{-1} \tanh(\beta\alpha). \quad (2.30)$$

Esto es, las intensidades sinápticas efectivas vienen dadas entonces por una regla de aprendizaje similar a la regla de Hebb salvo por el factor $\mathcal{A}(\alpha)$ que contiene toda la información relevante acerca de las fluctuaciones.

La función de partición resultante es

$$Z = \sum_{\mathbf{s}} \exp[-\beta \mathcal{H}_{eff}(\mathbf{s})] = \sum_{\mathbf{s}} \exp[-\beta \mathcal{A}(\alpha) \mathcal{H}_H(\mathbf{s})], \quad (2.31)$$

donde $\mathcal{H}_H(\mathbf{s})$ es el hamiltoniano del modelo de Hopfield. Por lo tanto, bajo las condiciones expuestas, el sistema puede ser caracterizado por dicho hamiltoniano a una *temperatura efectiva* definida como

$$T_{eff} \equiv \frac{T}{\mathcal{A}(\alpha)} = \frac{\alpha}{\tanh(\alpha T^{-1})}. \quad (2.32)$$

Propiedad 2.1 *En el límite $P \rightarrow \infty$ se tiene que $T_{eff} \geq T \forall T, \alpha$. La igualdad se da para $\alpha T^{-1} = 0$.*

Demostración: Consideremos la función $f(x) = x \tanh(x)^{-1}$. En virtud de la definición (2.32) se tiene $T_{eff} T^{-1} = f(\alpha T^{-1})$. La propiedad queda demostrada sin más que ver que $x \geq \tanh(x) \forall x \geq 0$. Esto es trivial pues la función $h(x) \equiv x - \tanh(x)$ es una función estrictamente creciente, y $h(x) = 0 \Leftrightarrow x = 0$.[†]

⁴En virtud del teorema de Chebyshev y bajo las consideraciones que estamos realizando, ρ_{xy} es una variable estocástica que con probabilidad uno tiende, para $P \rightarrow \infty$, a cero (este es el valor esperado de la variable $\xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}$ con la distribución dada por (2.23)). Por otra parte y en virtud del teorema del límite central las desviaciones de los valores de ρ_{xy} respecto de este valor medio son $\mathcal{O}(P^{1/2})$.

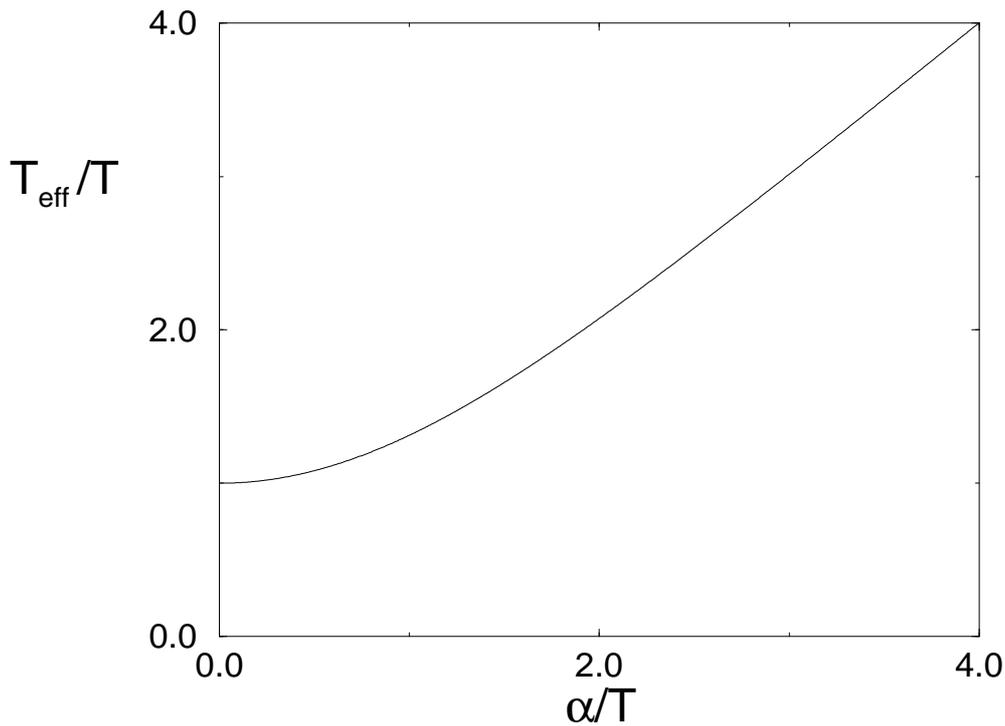


Figura 2.1: Representación del cociente T_{eff}/T (que es una medida de la función $\mathcal{A}(\alpha)$; cf. texto principal) en función del cociente α/T . La gráfica muestra que la temperatura efectiva es siempre mayor que la temperatura ordinaria para α y T finitos y no nulos, mostrando además que sólo para $\alpha/T = 0$ se tiene que $T_{\text{eff}} = T$, que ocurre para $T \rightarrow \infty$ y $\alpha \neq 0$ (el ruido debido a las fluctuaciones se hace irrelevante) y para $\alpha \rightarrow 0$ y $T \neq 0$ (el sistema almacena un número finito de patrones en $N \rightarrow \infty$).

Consecuencia de esta propiedad es que se tiene que $\mathcal{A}(\alpha) \leq 1$. A nivel físico la propiedad anterior nos dice que el tener en cuenta las fluctuaciones en las intensidades sinápticas implica, en la práctica, añadir un ruido adicional al ruido térmico ya existente que induce una temperatura mayor en el sistema que la debida sólo a efectos térmicos. Se tiene además que $\mathcal{A}(\alpha) \rightarrow 1$ cuando $\alpha T^{-1} \rightarrow 0$, de forma que la temperatura efectiva tiende a la temperatura ordinaria, $T_{\text{eff}} \rightarrow T$. Esto ocurre para $\alpha \neq 0$ cuando $T \rightarrow \infty$ (el exceso de ruido debido a las fluctuaciones es irrelevante) o para $T \neq 0$ cuando $\alpha \rightarrow 0$ (el número de patrones almacenados por la red es finito en el límite termodinámico); en ambas situaciones, el modelo, para el que el exceso de ruido es proporcional a α^2 , como hemos mostrado anteriormente, se reduce al modelo de

Hopfield. Es de destacar también el hecho de que $T_{eff} = \alpha$ para $T = 0$, lo que indica que en ausencia de ruido térmico el sistema, debido al efecto de las fluctuaciones, se ve afectado por un ruido que es proporcional al número de patrones que el sistema puede almacenar. En la figura 2.1 se muestra la relación entre T_{eff} y T en función del cociente α/T observándose todas las propiedades que acabamos de exponer.

2.4 Sinapsis asimétricas

Es conocido que el estado fisiológico de ciertos tipos de sinapsis biológicas es consistente con la simetría $J_{xy} = J_{yx}$, pues dichas sinapsis consisten en puentes iónicos entre neuronas a través de los cuales se establece una transmisión eléctrica del potencial de acción de la membrana presináptica a la postsináptica, siendo ambas membranas fisiológicamente iguales (cf. §1). Sin embargo este tipo de sinapsis no es el más común y es más realista en general el considerar $J_{xy} \neq J_{yx}$. Esto es así porque la transmisión del potencial de acción en la neurona y en la sinapsis tiene un carácter altamente asimétrico; en la neurona el potencial de acción generado se propaga, sin pérdida, en la dirección del axón debido a la abundancia de canales iónicos en la membrana del mismo, mientras que en sentido contrario acaba difundiéndose al ir disminuyendo la cantidad de canales iónicos en la membrana del soma (cf. §1.2); en la sinapsis la asimetría se manifiesta en el hecho de que el potencial de acción puede propagarse desde la membrana presináptica —con una alta concentración de vacuolas de neurotransmisores— hasta la membrana postsináptica —donde no hay neurotransmisores—, pero no al contrario. Por lo tanto, es interesante el investigar las implicaciones que tiene el considerar este tipo de simetría. Ciertos tipos de experimentos electrofisiológicos [7] ponen de manifiesto la relación entre la asimetría de las sinapsis y el aprendizaje, experimentos que llevan a suponer que la asimetría sináptica tiene su origen en el aprendizaje. Una descripción bastante general consistente con esto está en considerar

$$J_{xy} = \frac{1}{N} \sum_{\mu=1}^P [A_1 \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} + A_2 \xi_{\mathbf{x}}^{\mu} + A_3 \xi_{\mathbf{y}}^{\mu} + A_4], \quad (2.33)$$

para la que no existe una descripción hamiltoniana equivalente a la del modelo de Hopfield con la regla de Hebb. En esta sección discutimos la posible extensión de los

resultados previos, obtenidos para el caso de sinapsis simétricas, a la situación en la que tenemos acoplamientos sinápticos asimétricos (cf. §2.6 para un estudio posterior de asimetría).

Para sinapsis asimétricas, el estado estacionario tiene la estructura canónica dada por (2.26) para la *rate* V , algunas elecciones de $f(\mathbf{J})$ con la propiedad de factorización (2.23), y

$$\bar{\omega}(\mathbf{s}; \mathbf{x}) = \left[\prod_{\mathbf{y} \neq \mathbf{x}} \mathcal{N}_{xy} \right] \exp \left[-\beta \Delta \mathcal{H}_{eff}^{\mathbf{x}}(\mathcal{J}, \mathbf{s}) \right]. \quad (2.34)$$

El primer factor que aparece aquí es un factor de normalización, y $\Delta \mathcal{H}_{eff}^{\mathbf{x}}(\mathcal{J}, \mathbf{s}) = \mathcal{H}_{eff}(\mathcal{J}, \mathbf{s}^{\mathbf{x}}) - \mathcal{H}_{eff}(\mathcal{J}, \mathbf{s})$. Por sencillez, vamos a exigir que $\mathcal{H}_{eff}(\mathcal{J}, \mathbf{s})$ sea tipo Ising, ecuación (2.27).

La asimetría en la sinapsis implica que J_{xy} y J_{yx} son dos variables independientes. Bajo esta condición, se puede probar que (2.28) se sigue verificando para los coeficientes \mathcal{J}_{xy} , y que

$$\mathcal{N}_{xy} = \left\{ \overline{\exp(\beta J_{xy}) \exp(-\beta J_{xy})} \right\}^{\frac{1}{2}}. \quad (2.35)$$

Una diferencia con respecto a lo que ocurriría en el caso de simetría es que la elección de $f(J_{xy})$ en (2.23) debe restringirse a aquellas distribuciones que tengan en cuenta la simetría $J_{xy} \neq J_{yx}$. La situación más sencilla que se desea es que lo anterior se verifique y que además las sinapsis efectivas sean simétricas, $\mathcal{J}_{xy} = \mathcal{J}_{yx}$, lo cual garantiza que (2.27) es un hamiltoniano efectivo. Con esto en la mente elegimos

$$f(J_{xy}) = \sum_{\mu=1}^P a_{\mu} \delta(J_{xy} - \eta_{xy}^{\mu}), \quad (2.36)$$

con $\eta_{xy}^{\mu} \neq \eta_{yx}^{\mu}$. Suponiendo además aquí que $a_{\mu} = P^{-1} \quad \forall \mu$, y que

$$\eta_{xy}^{\mu} = \alpha \left(A_1 \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} + A_2 \xi_{\mathbf{x}}^{\mu} + A_3 \xi_{\mathbf{y}}^{\mu} + A_4 \right), \quad \alpha = PN^{-1}, \quad (2.37)$$

se obtiene que el modelo describe fluctuaciones en torno a valores medios de las intensidades sinápticas dadas por la regla de aprendizaje asimétrica (2.33).

Con las elecciones (2.36) y (2.37) y tras promediar sobre la distribución $f(\mathbf{J})$ dada por (2.23) se obtiene

$$\overline{\exp(\pm \beta J_{xy})} = \exp(\pm \mathcal{A}_4) \left[\prod_{i=1}^3 \cosh \mathcal{A}_i \right] \sum_{j=1}^4 \lambda_{\pm}^j \kappa_{xy}^j, \quad (2.38)$$

donde $\mathcal{A}_i \equiv \beta\alpha A_i$, $i = 1, 2, 3, 4$, y

$$\begin{aligned}\kappa_{xy}^1 &= \sum_{\mu=1}^P a_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}, & \kappa_{xy}^4 &= \sum_{\mu=1}^P a_{\mu} = 1, \\ \kappa_{xy}^2 &= \sum_{\mu=1}^P a_{\mu} \xi_{\mathbf{x}}^{\mu}, & \kappa_{xy}^3 &= \sum_{\mu=1}^P a_{\mu} \xi_{\mathbf{y}}^{\mu},\end{aligned}\tag{2.39}$$

y donde los constantes λ_{\pm}^i , $i = 1, 2, 3, 4$, vienen dadas mediante las expresiones

$$\begin{aligned}\lambda_{\pm}^1 &= \pm \tanh \mathcal{A}_1 + \tanh \mathcal{A}_2 \tanh \mathcal{A}_3, \\ \lambda_{\pm}^2 &= \pm \tanh \mathcal{A}_2 + \tanh \mathcal{A}_1 \tanh \mathcal{A}_3, \\ \lambda_{\pm}^3 &= \pm \tanh \mathcal{A}_3 + \tanh \mathcal{A}_2 \tanh \mathcal{A}_1, \\ \lambda_{\pm}^4 &= 1 \pm \prod_{i=1}^3 \tanh \mathcal{A}_i.\end{aligned}\tag{2.40}$$

Por lo tanto, introduciendo las expresiones (2.38)-(2.40) en la expresión (2.28), que nos da los acoplamientos sinápticos efectivos \mathcal{J}_{xy} que aparecen en el hamiltoniano (2.27), se tiene la siguiente forma explícita para dichos acoplamientos

$$\mathcal{J}_{xy} = \frac{\mathcal{A}_4}{\beta} + \frac{1}{2\beta} \ln \mathcal{K}_{xy},\tag{2.41}$$

donde \mathcal{K}_{xy} es

$$\mathcal{K}_{xy} = \left(\sum_{i=1}^4 \lambda_{+}^i \kappa_{xy}^i \right) \left(\sum_{i=1}^4 \lambda_{-}^i \kappa_{xy}^i \right)^{-1}.\tag{2.42}$$

A la hora de tener $\mathcal{J}_{xy} = \mathcal{J}_{yx}$, es necesario exigir que $\mathcal{K}_{xy} = \mathcal{K}_{yx}$, que va a depender de la simetría de los factores κ_{xy}^i $i = 1, 2, 3, 4$ al intercambiar \mathbf{x} por \mathbf{y} , a saber, $\kappa_{xy}^1 = \kappa_{yx}^1$; $\kappa_{xy}^4 = \kappa_{yx}^4$; $\kappa_{xy}^2 = \kappa_{yx}^3$; $\kappa_{yx}^2 = \kappa_{xy}^3$. En este punto, es conveniente distinguir los siguientes casos particulares (notar que la simetría en (2.33) está controlada por los parámetros A_2 y A_3):

(i) $\lambda_{+}^2 = \lambda_{+}^3$ y $\lambda_{-}^2 = \lambda_{-}^3$ y, por lo tanto, $\mathcal{A}_2 = \mathcal{A}_3$. En este caso hay que tener en cuenta dos posibilidades:(a) $\mathcal{A}_2 = \mathcal{A}_3 = 0$; esto ocurre para cualquier $\alpha, \beta \neq 0$ si $A_2 = A_3 = 0$ (es decir intensidades sinápticas inicialmente simétricas en (2.33)), o por el contrario para $\alpha = 0$, es decir P finito en el límite termodinámico. En el último caso se tiene $\mathcal{J}_{xy} \sim \mathcal{O}(1/N)$, mientras que en el primero, (2.41) se reduce a

$$\mathcal{J}_{xy} = \alpha A_4 + \frac{1}{2\beta} \ln \left[\frac{1 + \rho_{xy} \tanh(\beta\alpha A_1)}{1 - \rho_{xy} \tanh(\beta\alpha A_1)} \right],\tag{2.43}$$

que es similar a (2.29) para el caso de fluctuaciones en torno a la regla de Hebb.

(b) $\mathcal{A}_2 = \mathcal{A}_3 \neq 0$, que requiere $A_2 = A_3 \neq 0$, y $\alpha, \beta \neq 0$, es decir, η_{xy}^{μ} es simétrico,

concretamente, $\alpha^{-1}\eta_{xy}^\mu = A_1\xi_x^\mu\xi_y^\mu + A_2(\xi_x^\mu + \xi_y^\mu) + A_4$. Se obtiene para las sinapsis efectivas

$$\mathcal{J}_{xy} = \alpha A_4 + \frac{1}{2\beta} \ln \left[\frac{\lambda_+^4 + \rho_{xy}\lambda_+^1 + (\eta_x + \eta_y)\lambda_+^2}{\lambda_-^4 + \rho_{xy}\lambda_-^1 + (\eta_x + \eta_y)\lambda_-^2} \right], \quad (2.44)$$

$i = 1, 2, 3,$, con $\eta_x \equiv P^{-1} \sum_\mu \xi_x^\mu$.

(ii) $A_2 \neq A_3$, es decir, originalmente tenemos asimetría en (2.33), y además $\alpha \neq 0$, de forma que trabajamos bajo la idea de que el sistema puede almacenar un número macroscópico de patrones, es decir $P = \alpha N$. A la hora de tener $\mathcal{K}_{xy} = \mathcal{K}_{yx}$ y, por lo tanto, acoplamientos sinápticos efectivos simétricos, es necesario introducir ciertas condiciones sobre los patrones almacenados. Por ejemplo si los patrones almacenados son tales que $\eta_x = \eta_y = \eta \quad \forall \mathbf{x}, \mathbf{y}$, en el límite $P, N \rightarrow \infty$, es decir patrones que no presentan inhomogeneidades espaciales locales, es posible hacer una descripción hamiltoniana para el caso de sinapsis asimétricas, con acoplamientos sinápticos efectivos dados por

$$\mathcal{J}_{xy} = \alpha A_4 + \frac{1}{2\beta} \ln \left[\frac{\lambda_+^4 + \rho_{xy}\lambda_+^1 + \eta(\lambda_+^2 + \lambda_+^3)}{\lambda_-^4 + \rho_{xy}\lambda_-^1 + \eta(\lambda_-^2 + \lambda_-^3)} \right]. \quad (2.45)$$

Un caso particular del anterior es el caso en que el sistema almacena patrones que tiene el mismo número de variables ξ_x^μ con valor $+1$ y -1 de forma que $\eta = 0$. Estos ejemplos son los casos más sencillos con asimetría intrínseca en las sinapsis que puede ser descrito mediante un hamiltoniano efectivo tipo Hopfield. Otras situaciones más generales que las aquí presentadas pueden ser estudiadas bajo una teoría de campo medio, por ejemplo, como mostraremos más adelante; cf. §2.6.

2.5 El formalismo de réplicas

En esta sección vamos a aplicar el formalismo de réplicas (cf. apéndice A) para estudiar implicaciones adicionales de los resultados exactos obtenidos en las secciones previas. El caso particular de acoplamientos asimétricos es considerado en §2.6; en esta sección vamos a ilustrar el caso de acoplamientos simétricos; en particular, queremos resolver la función de partición (2.31). Esto ha sido ya calculado en el caso particular de $\mathcal{A}(\alpha) \equiv 1$ en el marco de una aproximación de campo medio usando el formalismo de réplicas y el método del punto de silla [40–42]. Después de sustituir β por

$\beta_{eff} \equiv \beta \mathcal{A}(\alpha)$ en la solución conocida, se obtiene para la energía libre por neurona la siguiente expresión:

$$f = \frac{\alpha}{2} + \frac{\alpha}{2\beta_{eff}n} \text{Tr} \ln [(1 - \beta_{eff}) \mathbf{I} - \beta_{eff} \mathbf{q}] + \frac{1}{2n} \sum_{\rho=1}^n \left[\sum_{\nu=1}^k (m_{\rho}^{\nu})^2 + \alpha \beta_{eff} \sum_{\substack{\sigma=1 \\ \sigma \neq \rho}}^n r_{\rho\sigma} q_{\rho\sigma} \right] - \frac{1}{n\beta_{eff}} \left\langle \ln \text{Tr}_{\vec{s}} e^{\beta_{eff} H_{\xi}(\vec{s})} \right\rangle_{\xi}, \quad (2.46)$$

en el límite $n \rightarrow 0$, donde ρ y σ son índices que describen n réplicas, $\vec{s} \equiv \{s^1, s^2, \dots, s^n\}$,

$$H_{\xi}(\vec{s}) \equiv \sum_{\rho=1}^n \left[\frac{\alpha \beta_{eff}}{2} \sum_{\substack{\sigma=1 \\ \sigma \neq \rho}}^n r_{\rho\sigma} s^{\rho} s^{\sigma} + \sum_{\nu=1}^k (m_{\rho}^{\nu} + h^{\nu}) \xi^{\nu} s^{\rho} \right], \quad (2.47)$$

y donde $\langle \dots \rangle_{\xi}$ es un promedio sobre la distribución de patrones aleatorios (2.25); \mathbf{I} es la matriz identidad en el espacio de réplicas con elementos de matriz $\delta_{\rho\sigma}$, y \mathbf{q} es la matriz, también en el espacio de réplicas, que tiene elementos $q_{\rho\sigma} \neq 0$, en general, para $\rho \neq \sigma$ y ceros para $\rho = \sigma$. El significado físico de los parámetros m_{ρ}^{ν} , $q_{\rho\sigma}$ y $r_{\rho\sigma}$ en (2.46) está dado en virtud de las ecuaciones del punto de silla. Esto es, se obtiene

$$m_{\rho}^{\nu} = \frac{1}{N} \left\langle \sum_{\mathbf{y}} \xi_{\mathbf{y}}^{\nu} \langle s_{\mathbf{y}}^{\rho} \rangle \right\rangle_{\xi}, \quad (2.48)$$

como una medida del solapamiento de un estado dado con un patrón almacenado dado, y

$$q_{\rho\sigma} = \frac{1}{N} \left\langle \sum_{\mathbf{y}} \langle s_{\mathbf{y}}^{\rho} \rangle \langle s_{\mathbf{y}}^{\sigma} \rangle \right\rangle_{\xi}, \quad r_{\rho\sigma} = \frac{1}{\alpha} \sum_{\mu=k+1}^P \left\langle m_{\rho}^{\mu} m_{\sigma}^{\mu} \right\rangle_{\xi}, \quad (2.49)$$

que son los parámetros de orden de Edwards-Anderson [30] y AGS [41], respectivamente. El índice ν en estas expresiones describe un número (finito), k , de *patrones condensados*, definidos como aquellos patrones con los que el estado del sistema tiene el solapamiento m^{ν} no nulo en el límite termodinámico. Los *patrones no condensados*, descritos mediante el índice μ por ejemplo en (2.49) y en la expresión para $p(\xi_{\mathbf{x}}^{\mu})$, tienen un solapamiento que es $\mathcal{O}(\frac{1}{\sqrt{N}})$. A continuación estudiamos algunas de las implicaciones de (2.46).

Consideremos la situación en que tenemos simetría de réplicas, es decir, $m_{\rho}^{\nu} = m^{\nu}, \forall \rho = 1, \dots, n$, y $q_{\rho\sigma} = q$ y $r_{\rho\sigma} = r$ para $\rho \neq \sigma$. En este caso las ecuaciones del punto de silla pueden ser dadas en una forma cerrada; así a partir de (2.46),

derivando f con respecto a los parámetros m^ν, r , y q se tiene para $n \rightarrow 0$ que

$$m^\nu = \left[\left[\xi^\nu \tanh \beta_{eff} \left\{ z\sqrt{\alpha r} + (\mathbf{m} + \mathbf{h}) \cdot \boldsymbol{\xi} \right\} \right] \right], \quad (2.50)$$

$$q = \left[\left[\tanh^2 \beta_{eff} \left\{ z\sqrt{\alpha r} + (\mathbf{m} + \mathbf{h}) \cdot \boldsymbol{\xi} \right\} \right] \right], \quad r = q(1 - \beta_{eff} + \beta_{eff}q)^{-2}. \quad (2.51)$$

Aquí $\mathbf{m} \equiv \{m^\nu; \nu = 1, \dots, k\}$, $\mathbf{h} \equiv \{h^\nu; \nu = 1, \dots, k\}$ son campos conjugados asociados a los patrones condensados, que aparecen como un término adicional, $-\sum_{\nu=1}^k h^\nu \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\nu s_{\mathbf{x}}$, en el hamiltoniano (2.27), y

$$[[\dots]] \equiv \frac{1}{\sqrt{2\pi}} \int dz e^{-z^2/2} \langle \dots \rangle_{\xi^\nu}. \quad (2.52)$$

El promedio $\langle \dots \rangle_{\xi^\nu}$, que aparece en el segundo término de la definición (2.52), es similar al promedio que aparece en (2.46) pero, ahora, es un promedio sólo sobre los patrones condensados; la integral gaussiana mostrada explícitamente corresponde a integrar sobre el ruido gaussiano cuyo origen está en el conjunto de patrones no-condensados.

Resolviendo numéricamente las ecuaciones (2.50) y (2.51) se encuentran las principales implicaciones, para $h^\nu = 0 \quad \forall \nu$, que a continuación detallamos. Para el caso en que $T = 0$, la temperatura efectiva es $T_{eff} = \alpha$. En esta situación hay una solución vidrio de espines, es decir, $m^\nu = 0$ con $q \neq 0$ para algún $\alpha < \alpha_q = 2.618$; la dependencia resultante de q en α se muestra en la figura 2.2. Ésta difiere esencialmente del resultado $q = 1$, independiente del número de patrones almacenados, que se obtiene en las mismas condiciones para el modelo de Hopfield sin fluctuaciones [41]. Esto es, las fluctuaciones en las sinapsis tienden a impedir la fase vidrio de espines, y el ruido asociado a las fluctuaciones es tan grande, cuando el número de patrones almacenados excede alrededor de $3N$ veces el número de neuronas, que la solución vidrio de espines desaparece. En cierto sentido, esto es una cualidad interesante de nuestro modelo, pues la fase vidrio de espines no es deseable para la realización óptima de la red, es decir, para la eficiencia de procesos tales como el almacenamiento y el recuerdo de patrones; la mejoría sería real si se pudiera simultáneamente extender la región en la que aparecen los estados de Mattis cerca de α_q (cf. más abajo). Como era de esperar, las fluctuaciones desaparecen para $\alpha \rightarrow 0$, de forma que en dicho límite recuperamos el resultado del modelo de Hopfield ($q = 1$).

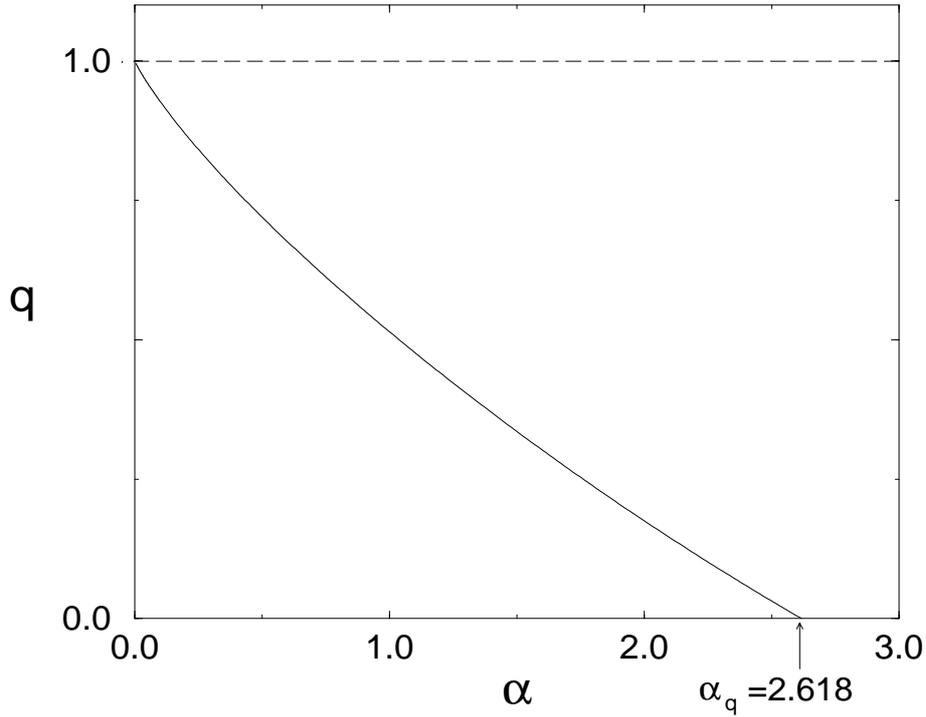


Figura 2.2: La línea continua representa la solución $q = q(\alpha)$, que se obtiene a partir de las ecuaciones (2.50) y (2.51) para $m^\nu = 0$ y $h^\nu = 0 \quad \forall \nu$, y en $T = 0$ ($T_{eff} = \alpha$). La fase es de naturaleza vidrio de espines y q decrece de forma continua, conforme se incrementa el número de patrones almacenados, hasta hacerse cero para $\alpha > \alpha_q = 2.618$ teniendo lugar un cambio de fase continuo. Esto es una diferencia con el modelo de Hopfield (línea a trazos) para el que $q = 1$ independientemente del valor de α .

Las soluciones tipo Mattis [43],

$$m^\nu = m\delta_{\nu\nu_0}, \quad (2.53)$$

es decir, aquellas soluciones que exhiben un único solapamiento no nulo con un patrón determinado, por ejemplo ν_0 , aparecen (dos por cada patrón) en $T = 0$ para $\alpha < \alpha_M \simeq 0.132$, valor que es ligeramente inferior al valor de 0.138 que aparece en el modelo de Hopfield. Esto es, el principal efecto de las fluctuaciones es el de hacer decrecer algo la eficiencia del proceso de memoria asociativa; en la figura 2.3 se muestra una comparación entre los estados de Mattis que aparecen a $T = 0$ en el modelo con fluctuaciones comparados con los estados de Mattis que aparecen en las mismas

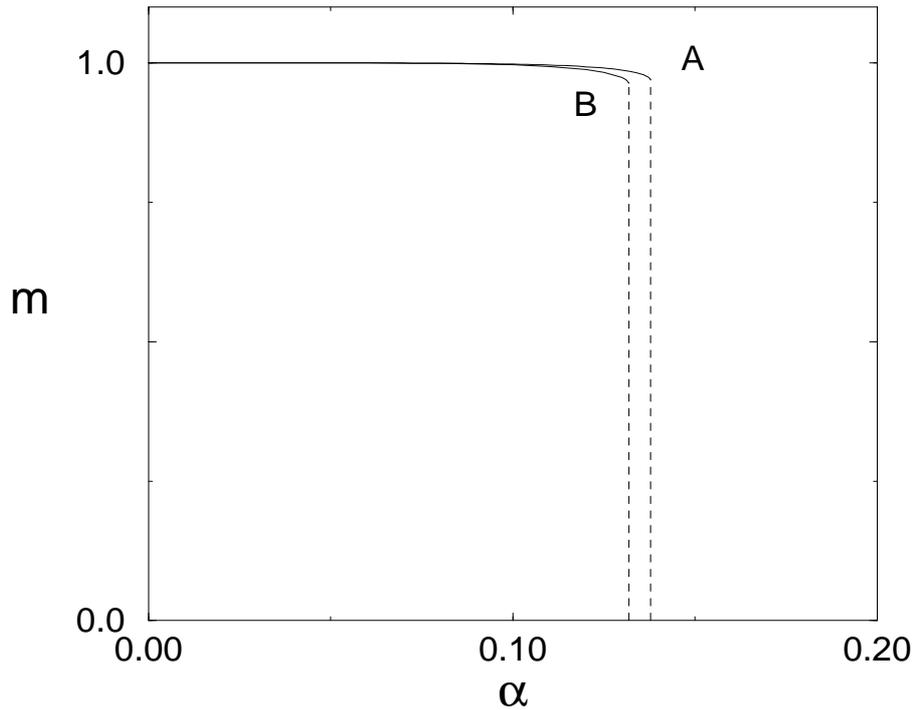


Figura 2.3: Soluciones tipo estados de Mattis para $T = 0$ en el modelo de Hopfield (A) y para el caso del modelo de fluctuaciones en las sinapsis (B). La figura muestra que los estados de Mattis, en ambos modelos, aparecen de forma discontinua, por debajo de sendos valores críticos de α . Para $\alpha \rightarrow 0$, aparece un comportamiento similar de las soluciones de ambos modelos. Las diferencias más significativas ocurren para valores de α próximos a aquellos valores críticos en los que tiene lugar el cambio de fase.

condiciones en el modelo de Hopfield. Cuando $\alpha \rightarrow 0$, en ambos modelos se recupera un patrón dado sin error, es decir, $|m| \rightarrow 1$. El error se incrementa con α , y esto ocurre ligeramente en mayor cuantía en presencia de fluctuaciones en las sinapsis. De los resultados expuestos se deduce que las diferencias entre los dos modelos, a saber, modelo de Hopfield sin fluctuaciones y modelo con fluctuaciones sinápticas, son más pequeñas que en el caso de las soluciones tipo vidrio de espines, debido sobre todo al hecho de que los estados de Mattis aparecen para valores muy pequeños de α y los dos modelos coinciden cuando $\alpha \rightarrow 0$. Cuando α se incrementa tiene lugar un cambio de fase discontinuo, como se aprecia en la figura 2.3. Se tiene además que $q \neq 0$ para las soluciones tipo Mattis, es decir, dichos estados involucran cierto orden vidrio de espines. La

estabilidad de las soluciones es similar a la del modelo de Hopfield, concretamente, el estado fundamental está constituido sólo por estados de Mattis para $\alpha < \alpha_C \simeq 0.052$, aunque como hemos visto más arriba pueden aparecer estados vidrios de espines pero con una energía libre f mayor. Para $\alpha_M > \alpha > \alpha_C$ los estados de Mattis dejan de ser

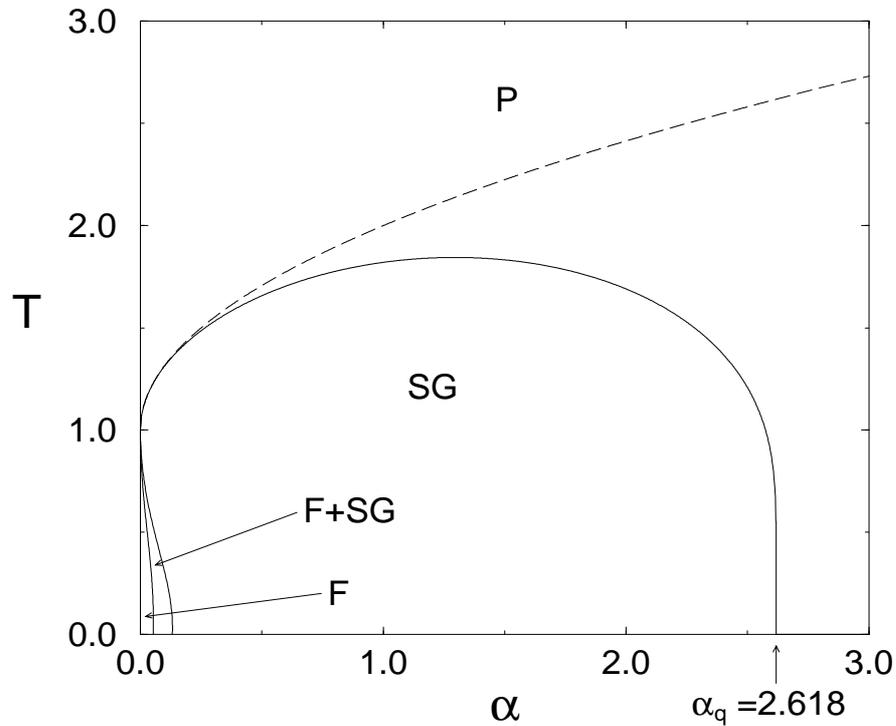


Figura 2.4: Diagrama de fases para la red de neuronas definida por (2.31), es decir, hay un hamiltoniano efectivo tipo Hopfield, (2.27) con (2.28), para el caso de simetría de réplicas. Usando terminología del magnetismo, las diferentes regiones corresponden a una fase paramagnética (P), vidrio de espines (SG), y ferromagnética (F). La línea continua, que se corresponde a la ecuación (2.54), separa los estados paramagnéticos de los estados vidrios de espines; la línea a trazos es el correspondiente resultado del modelo de Hopfield. A diferencia del modelo de Hopfield, en el modelo con fluctuaciones no aparecen soluciones vidrio de espines para $\alpha > \alpha_q$ (cf. figura 2.2)

el estado fundamental del sistema siendo ahora estados metaestables (mínimos locales de la energía libre); el estado fundamental está constituido ahora por la solución vidrio de espines. Por encima de α_C la única solución que aparece es la vidrio de espines. El valor de α_C es prácticamente el mismo para los dos modelos, pues las fluctuaciones

son relativamente pequeñas para tales valores pequeños de α .

Para $T \neq 0$, nuestro modelo se corresponde con el modelo de Hopfield a una temperatura efectiva dada por (2.32). Se obtiene que la fase vidrio de espines tiene lugar para valores de T y α tales que $T < T_{sg}(\alpha)$ con

$$T_{sg}^{-1}(\alpha) = \begin{cases} \frac{1}{2\alpha} \ln \left\{ \frac{1-\alpha\sqrt{\alpha}}{1-2\alpha+\alpha\sqrt{\alpha}} \right\}, & \text{for } \alpha \neq 1 \\ \tanh^{-1} \frac{1}{2}, & \text{for } \alpha = 1. \end{cases} \quad (2.54)$$

En el límite $\alpha \rightarrow 0$, esta expresión se reduce al conocido resultado del modelo de Hopfield, $T_{sg}(\alpha) = 1 + \sqrt{\alpha}$. Esta línea ha sido obtenida realizando un desarrollo en serie de potencias de las ecuaciones (2.50) y (2.51) en torno a $q = 0$ (para visualizar donde desaparece la fase vidrio de espines), de forma que dicha línea marca un cambio de fase de segundo orden, es decir, q es continuo en T_{sg} para algún α ; cf. figura 2.4. Aparte de este cambio de fase, encontramos dos líneas más que marcan la existencia de sendos cambios de fase (cf. figura 2.5): una de ellas es $T_M(\alpha)$ y es tal que los estados de Mattis no parecen para $T > T_M(\alpha)$; la otra, que denotamos por $T_C(\alpha)$, aparece siempre por debajo de $T_M(\alpha) \forall \alpha$ y es tal que para $T > T_C(\alpha)$ los estados de Mattis no se corresponden con el mínimo global de la energía libre. Los cambios de fase en $T_M(\alpha)$ y $T_C(\alpha)$ son discontinuos para cualquier α ; estas líneas son muy parecidas a las que aparecen en el modelo de Hopfield, dado que aparecen para valores de α pequeños, valores para los que las fluctuaciones son pequeñas.

En resumen, el diagrama de fases completo para nuestro modelo, en el caso en que se considera simetría de réplicas (cf. figura 2.4), es cualitativamente similar al correspondiente al modelo de Hopfield, más parecido cuanto menor es el valor de α , coincidiendo para el caso en que $\alpha \rightarrow 0$. Sin embargo, aparecen importantes diferencias cuando α se incrementa, es decir cuando las fluctuaciones en las sinapsis son significativas. En particular, se sigue que los estados vidrios de espines que aparecen en el modelo de Hopfield se ven fuertemente influidos por las fluctuaciones y no aparecen para valores grandes de α .

La solución vidrio de espines con simetría de réplicas es, en general, inestable. Esto se refleja, por ejemplo, en el modelo de vidrio de espines de Serrington-Kirpatrick [40] por el hecho de que la entropía asociada con estos estados es negativa para $T = 0$ [40]. El modelo de Hopfield tiene un comportamiento paralelo, es decir, la solución vidrio

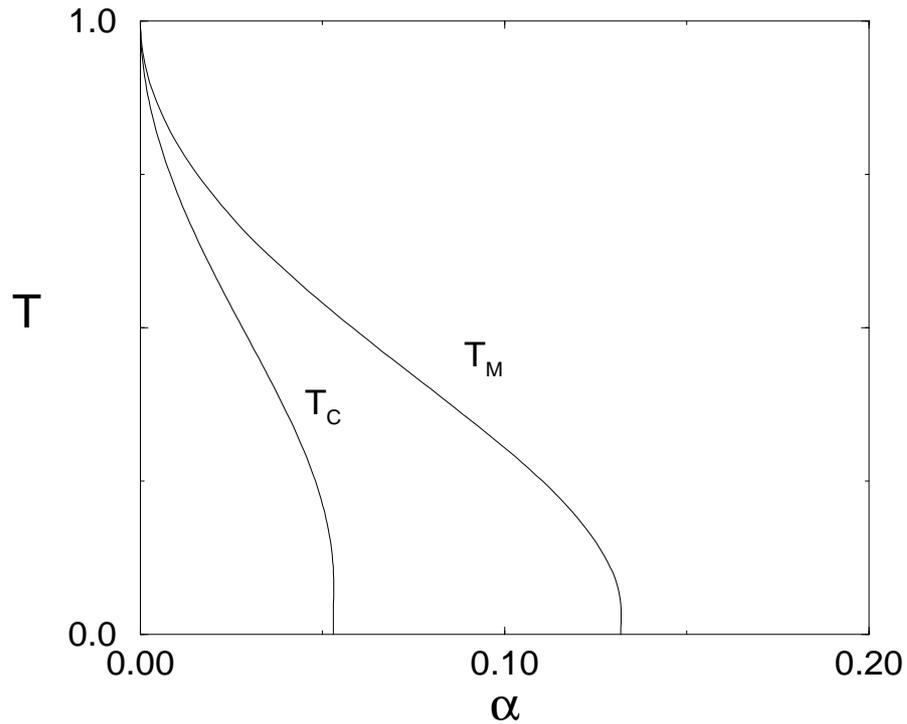


Figura 2.5: Detalle cerca del origen en la figura 2.4 para ilustrar las funciones $T_M(\alpha)$ y $T_C(\alpha)$, definidas en el texto principal, correspondientes a cambios de fase discontinuos.

de espines con $q_{\rho,\sigma} = q$ para $\rho \neq \sigma$ es inestable para $T < T_{sg} = 1 + \sqrt{\alpha}$, y la entropía a temperatura cero asociada con los estados vidrios de espines y estados de Mattis es negativa [41]. Una cuestión importante es ver cuando deja de ser estable la solución ferromagnética con simetría de réplicas, que es la solución más relevante desde el punto de vista de la memoria asociativa. El límite de estabilidad de la solución ferromagnética es conocido como línea Almeida-Thouless (AT) [28]. Para el modelo de Hopfield [41], la línea AT separa estados estables con simetría de réplicas de una región en la que la estabilidad se rompe. Esta región es relativamente pequeña y aparece a temperaturas próximas a cero cerca de $\alpha_M (= 0.138)$. Para este valor de α la línea AT tiene un valor máximo en la región ferromagnética, decreciendo rápidamente hasta cero cuando $\alpha \rightarrow 0$, de forma que no afecta significativamente a las propiedades de memoria asociativa. En presencia de fluctuaciones en las sinapsis, la línea AT no aparece, incluso para $T = 0$, debido al exceso de ruido introducido por los patrones almacenados, que

se incrementa con α . En esta situación se tiene que la entropía asociada a las soluciones de las ecuaciones (2.50) y (2.51) tipo estados de Mattis (estados ferromagnéticos) es

$$S(\beta, q, m) = \beta^2 (\beta_{eff}^{-2} - \alpha^2) S_H(\beta_{eff}, q, m), \quad (2.55)$$

donde $S_H(\beta_{eff}, q, m)$ es la entropía que resulta en el modelo de Hopfield con β sustituida por β_{eff} . Esta expresión tiende a cero cuando $\beta \rightarrow \infty$, pues $T_{eff} \rightarrow \alpha$ de forma exponencial cuando $T \rightarrow 0$, como se comprueba fácilmente a partir de la definición (2.32).

2.6 Casos adicionales del modelo

En esta sección mostramos los principales resultados que se obtienen de un análisis general de campo medio que incluye ecuaciones de evolución temporal. Obtenemos una descripción detallada de diferentes cambios de fase. El punto de interés en el que nos centramos en esta sección está en las versiones sencillas del modelo definido en §2.2, en el límite $p \rightarrow 0$, para el caso de funciones $\varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x})$ y $f(\mathbf{J})$ para las que, en general, no se obtiene una estructura canónica del tipo dado en (2.26) y (2.27). A partir de la expresión (2.14) se obtiene para la variación temporal de la actividad neuronal promedio, esto es $\partial_t \langle s_{\mathbf{x}} \rangle = -2 \langle s_{\mathbf{x}} \bar{\omega}(\mathbf{s}; \mathbf{x}) \rangle$, que

$$\partial_t \langle s_{\mathbf{x}} \rangle = -2 \langle s_{\mathbf{x}} \bar{a}_{\mathbf{x}}(\mathbf{s}) \rangle - 2 \langle \bar{b}_{\mathbf{x}}(\mathbf{s}) \rangle, \quad (2.56)$$

donde $\bar{a}_{\mathbf{x}}(\mathbf{s})$ y $\bar{b}_{\mathbf{x}}(\mathbf{s})$ se obtienen promediando con respecto a $f(\mathbf{J})$, como en (2.18), las funciones

$$a_{\mathbf{x}}(\mathbf{s}, \mathbf{J}) \equiv \frac{1}{2} \{ \Psi [2\beta h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})] + \Psi [-2\beta h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})] \} \quad (2.57)$$

y

$$b_{\mathbf{x}}(\mathbf{s}, \mathbf{J}) \equiv \frac{1}{2} \{ \Psi [2\beta h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})] - \Psi [-2\beta h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})] \}, \quad (2.58)$$

respectivamente. (Notar que hay aquí una dependencia explícita en \mathbf{x} que tiene su origen en la dependencia espacial que exhiben los patrones memorizados; cf. (2.1) y (2.9).) Bajo la aproximación de campo medio $s_{\mathbf{x}} \approx \langle s_{\mathbf{x}} \rangle$, la expresión (2.56) se transforma en

$$\partial_t \langle s_{\mathbf{x}} \rangle \approx -2\bar{a}'_{\mathbf{x}} \langle s_{\mathbf{x}} \rangle - 2\bar{b}'_{\mathbf{x}}, \quad (2.59)$$

donde $\bar{a}'_{\mathbf{x}}$ y $\bar{b}'_{\mathbf{x}}$ se obtienen, respectivamente, de $\bar{a}_{\mathbf{x}}(\mathbf{s})$ y $\bar{b}_{\mathbf{x}}(\mathbf{s})$ haciendo la sustitución $s_{\mathbf{x}} \approx \langle s_{\mathbf{x}} \rangle$. La correspondiente solución estacionaria a la expresión (2.59) es $m_{\mathbf{x}}(\mathbf{J}) = \sum_{\mathbf{y} \neq \mathbf{x}} J_{xy} \langle s_{\mathbf{y}} \rangle$, con

$$\langle s_{\mathbf{x}} \rangle \approx -\frac{\bar{b}'_{\mathbf{x}}}{\bar{a}'_{\mathbf{x}}} = \frac{\overline{\Psi[-2\beta m_{\mathbf{x}}(\mathbf{J})] - \Psi[2\beta m_{\mathbf{x}}(\mathbf{J})]}}{\overline{\Psi[-2\beta m_{\mathbf{x}}(\mathbf{J})] + \Psi[2\beta m_{\mathbf{x}}(\mathbf{J})]}}, \quad (2.60)$$

resultado que hemos obtenido usando las expresiones (2.56)-(2.58) y el hecho de que se tiene $h_{\mathbf{x}}(\mathbf{s}, \mathbf{J}) \approx m_{\mathbf{x}}(\mathbf{J})$ bajo la hipótesis de campo medio que estamos considerando.

Notemos algunas propiedades generales que se obtienen de la expresión (2.60), antes de especificar una forma concreta para las *rates*. La expresión (2.60) se verifica para $m_{\mathbf{x}}(\mathbf{J}) = 0$; por lo tanto, soluciones no triviales se pueden obtener desarrollando dicha expresión alrededor de esta solución trivial en potencias de $m_{\mathbf{x}}(\mathbf{J})$. De esta forma se tiene en primer orden que $m_{\mathbf{x}}(\mathbf{J}) \approx \beta \sum_{\mathbf{y} \neq \mathbf{x}} J_{xy} \overline{m_{\mathbf{y}}(\mathbf{J})}$, o equivalentemente

$$\overline{m_{\mathbf{x}}(\mathbf{J})} \approx \beta \sum_{\mathbf{y} \neq \mathbf{x}} \bar{J}_{xy} \overline{m_{\mathbf{y}}(\mathbf{J})}, \quad (2.61)$$

después de promediar con respecto a $f(\mathbf{J})$. Por lo tanto, hay un cambio de fase, es decir aparecen soluciones $\overline{m_{\mathbf{x}}(\mathbf{J})} \neq 0$, a una temperatura crítica que es solución a la ecuación

$$\det \{ \delta_{xy} - \beta \bar{J}_{xy} \} = 0 \quad (2.62)$$

A continuación presentamos algunos casos ilustrativos de las propiedades generales que acabamos de describir:

(i) $f(\mathbf{J})$ es tal que $\bar{J}_{xy} = J/N$ con J independiente de $N \quad \forall \mathbf{x}, \mathbf{y}$. Bajo estas hipótesis se tiene $\overline{m_{\mathbf{x}}(\mathbf{J})} = Jm \quad \forall \mathbf{x}$, donde $m \equiv N^{-1} \sum_{\mathbf{x}} \langle s_{\mathbf{x}} \rangle$ (actividad neuronal promedio en la red), que haciendo uso de la expresión (2.61) da como resultado $m \approx \beta Jm$. Por lo tanto, el sistema muestra un cambio de fase de una fase paramagnética a una fase ferromagnética a $T_c \approx J$.

(ii) $f(\mathbf{J})$ es tal que \bar{J}_{xy} viene dado por (2.18) con $\{\xi_{\mathbf{x}}^{\mu}; \mu = 1, \dots, P; \mathbf{x} \in \Lambda_d\}$ constituyendo un conjunto de $N \times P$ variables aleatorias estadísticamente independientes con distribuciones de probabilidad dadas por (2.25). Después de utilizar la condición

de balance detallado para la *rate* $\Psi(X)$ y desarrollando en serie dichas *rates* alrededor de $X = 0$, se tiene a partir de (2.60), en primer orden en X , que

$$\langle s_{\mathbf{x}} \rangle \approx \beta \alpha \sum_{\mathbf{y} \neq \mathbf{x}} \left(P^{-1} \sum_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} \right) \langle s_{\mathbf{y}} \rangle. \quad (2.63)$$

Para $N, P \gg 1$, utilizando la ley de los grandes números tenemos que con probabilidad tendiendo a 1, y salvo correcciones $\mathcal{O}(N^{-\frac{1}{2}})$, se tiene

$$\sum_{\mathbf{y} \neq \mathbf{x}} \langle s_{\mathbf{y}} \rangle \delta_{xy} (1 - \alpha \beta) \approx 0. \quad (2.64)$$

Esta ecuación se satisface para $\langle s_{\mathbf{x}} \rangle = 0 \quad \forall \mathbf{x}$, i.e., $m = 0$; soluciones no triviales aparecen para $T < T_c$, donde la temperatura crítica se obtiene como solución a la ecuación $\det [\delta_{xy} (1 - \alpha \beta)] = 0$ dando $T_c = \alpha$. Esta situación se corresponde por lo tanto con un cambio de fase de segundo orden entre una fase paramagnética y una fase ferromagnética. Para la misma forma de la distribución $f(\mathbf{J})$, hemos estudiado el parámetro de orden

$$m^{\mu} = \frac{1}{N} \sum_{\mathbf{y}} \xi_{\mathbf{y}}^{\mu} \langle s_{\mathbf{y}} \rangle, \quad (2.65)$$

definido ya, en otra situación, en §2.5, y que describe el solapamiento de una configuración del sistema con un patrón memorizado ξ^{μ} . Se tiene a partir de (2.60) y (2.65) que $m^{\mu} = -N^{-1} \sum_{\mathbf{y}} \xi_{\mathbf{y}}^{\mu} (\bar{b}'_{\mathbf{x}} / \bar{a}'_{\mathbf{x}})$, de donde en primer orden de aproximación se tiene que

$$\sum_{\mu=1}^P m^{\mu} \left[\delta_{\mu\nu} - \frac{\beta}{N} \sum_{\mathbf{y}} \xi_{\mathbf{y}}^{\mu} \xi_{\mathbf{y}}^{\nu} \right] \approx 0, \quad (2.66)$$

después de desarrollar $\Psi(X)$. La solución trivial, para la que el sistema no puede recuperar mediante asociación ninguno de los patrones memorizados, es $m^{\mu} = 0 \quad \forall \mu$. Soluciones no triviales, que marcan la presencia de un cambio a una fase cuyos estados presentan la propiedad de memoria asociativa, pueden aparecer sólo por debajo de una temperatura crítica que verifica $\det \left\{ \delta_{\mu\nu} - \beta N^{-1} \sum_{\mathbf{y}} \xi_{\mathbf{y}}^{\mu} \xi_{\mathbf{y}}^{\nu} \right\} = 0$. Para $N \gg 1$, salvo correcciones $\mathcal{O}(N^{-\frac{1}{2}})$, se tiene $N^{-1} \sum_{\mathbf{y}} \xi_{\mathbf{y}}^{\mu} \xi_{\mathbf{y}}^{\nu} \approx \delta_{\mu\nu}$, o condición de cuasi-ortogonalidad de los patrones almacenados⁵. En el caso particular de patrones ortogonales, la condición anterior se satisface exactamente sin necesidad de tomar el límite

⁵Esta condición puede ser tomada como la definición de patrones cuasi-ortogonales, es decir patrones cuyos elementos son elegidos aleatoriamente e independientes entre sí y que verifican dicha condición para N grande.

de $N \gg 1$. La condición de ortogonalidad o cuasi ortogonalidad aplicada a (2.66) da como resultado $T_c = 1$.

Consideramos a continuación la expresión (2.56) para diferentes elecciones de la *rate* $\Psi(X)$ cuando $f(\mathbf{J})$ está dado por (2.20) y (2.21). Para el caso particular de *rate* K se obtiene

$$\bar{a}_{\mathbf{x}}(\mathbf{s}) = 1, \quad \bar{b}_{\mathbf{x}}(\mathbf{s}) = -\overline{\tanh[\beta h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})]} \quad (2.67)$$

y consecuentemente para la actividad neuronal promedio, $\langle s_{\mathbf{x}} \rangle = \langle \overline{\tanh[\beta h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})]} \rangle$, que tras promediar queda en la forma

$$\langle s_{\mathbf{x}} \rangle = \left\langle \sum_{\mu_1=1}^P \cdots \sum_{\mu_{N-1}=1}^P a_{\mu_1} \cdots a_{\mu_{N-1}} \tanh[\beta E] \right\rangle, \quad (2.68)$$

donde $E \equiv \eta_{xy_1}^{\mu_1} s_{y_1} + \cdots + \eta_{xy_{N-1}}^{\mu_{N-1}} s_{y_{N-1}}$.

Para el caso de *rate* M se tiene

$$\bar{a}_{\mathbf{x}}(\mathbf{s}) = \frac{1}{2} \left[\overline{1 + \exp(-2\beta |h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})|)} \right] \quad (2.69)$$

y

$$\bar{b}_{\mathbf{x}}(\mathbf{s}) = \frac{1}{2} \text{sign}(h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})) \left(1 - e^{-2\beta |h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})|} \right),$$

y por lo tanto se sigue

$$\langle s_{\mathbf{x}} \rangle = -B/A, \quad (2.70)$$

donde

$$A = \frac{1}{2} \left[1 + \frac{1}{P^{N-1}} \sum_{\mu_1=1}^P \cdots \sum_{\mu_{N-1}=1}^P e^{-2\beta |\tilde{E}|} \right] \quad (2.71)$$

y

$$B = \frac{1}{2P^{N-1}} \sum_{\mu_1=1}^P \cdots \sum_{\mu_{N-1}=1}^P \left(1 - e^{-2\beta |E|} \right) \text{sign} \tilde{E} \quad (2.72)$$

con $\tilde{E} \equiv \eta_{xy_1}^{\mu_1} \langle s_{y_1} \rangle + \cdots + \eta_{xy_{N-1}}^{\mu_{N-1}} \langle s_{y_{N-1}} \rangle$. Las situaciones particulares dadas por las expresiones (2.68) y (2.70) son difíciles de tratar analíticamente y requerirían simplificaciones drásticas antes de obtener resultados explícitos.

Para el caso de *rate* V , podemos desarrollar las expresiones (2.57) y (2.58) de forma explícita. Consideremos en primer lugar la elección de sinapsis simétricas (2.20) y (2.21) de forma que se verifique (2.18). Se obtiene

$$\bar{a}_{\mathbf{x}}(\mathbf{s}) = K_{\mathbf{x}} \cosh \left(\beta \sum_{\mathbf{y} \neq \mathbf{x}} \mathcal{J}_{xy} s_{\mathbf{y}} \right), \quad \bar{b}_{\mathbf{x}}(\mathbf{s}) = -K_{\mathbf{x}} \sinh \left(\beta \sum_{\mathbf{y} \neq \mathbf{x}} \mathcal{J}_{xy} s_{\mathbf{y}} \right), \quad (2.73)$$

donde

$$K_{\mathbf{x}} = \prod_{\mathbf{y} \neq \mathbf{x}} \left[\cosh^2(\beta\alpha) - \rho_{xy}^2 \sinh^2(\beta\alpha) \right]^{\frac{1}{2}} \quad (2.74)$$

y

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln \left[\frac{1 + \rho_{xy} \tanh(\beta\alpha)}{1 - \rho_{xy} \tanh(\beta\alpha)} \right], \quad (2.75)$$

con $\rho_{xy} \equiv P^{-1} \sum_{\mu=1}^P \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}$, de forma que se tiene $\mathcal{J}_{xy} = \mathcal{J}_{yx}$. Bajo la aproximación de campo medio $s_{\mathbf{x}} \approx \langle s_{\mathbf{x}} \rangle$, la correspondiente solución estacionaria verifica

$$\langle s_{\mathbf{x}} \rangle = \tanh \left(\beta \sum_{\mathbf{y} \neq \mathbf{x}} \mathcal{J}_{xy} \langle s_{\mathbf{y}} \rangle \right), \quad (2.76)$$

y utilizando la definición del parámetro de orden m^{μ} se tiene

$$m^{\mu} = \frac{1}{N} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^{\mu} \tanh \left(\beta \sum_{\mathbf{y} \neq \mathbf{x}} \mathcal{J}_{xy} \langle s_{\mathbf{y}} \rangle \right). \quad (2.77)$$

Este es el resultado que se obtiene en el modelo de Hopfield para P finito con \mathcal{J}_{xy} sustituido por la regla de Hebb (expresión (2.1)). Para P proporcional al tamaño del sistema N , se puede obtener (2.50) y (2.51) de forma heurística a partir de (2.77) asumiendo que \mathcal{J}_{xy} viene dado por la expresión (2.30) para $N, P \rightarrow \infty$.

En el caso en que tenemos acoplamientos sinápticos asimétricos, elección (2.37), se puede obtener después de algo de álgebra que

$$\bar{a}_{\mathbf{x}}(\mathbf{s}) = \left[\prod_{i=1}^3 \cosh(\beta\alpha A_i) \right] \left(\prod_{\mathbf{y} \neq \mathbf{x}} c_{xy} \right) \cosh \left(\beta \sum_{\mathbf{y} \neq \mathbf{x}} \mathcal{J}_{xy} s_{\mathbf{y}} \right) \quad (2.78)$$

y

$$\bar{b}_{\mathbf{x}}(\mathbf{s}) = - \left[\prod_{i=1}^3 \cosh(\beta\alpha A_i) \right] \left(\prod_{\mathbf{y} \neq \mathbf{x}} c_{xy} \right) \sinh \left(\beta \sum_{\mathbf{y} \neq \mathbf{x}} \mathcal{J}_{xy} s_{\mathbf{y}} \right) \quad (2.79)$$

donde se tiene

$$c_{xy}^2 = \left[1 + \kappa_{xy}^1 \tanh \mathcal{A}_2 \tanh \mathcal{A}_3 + \kappa_{xy}^2 \tanh \mathcal{A}_1 \tanh \mathcal{A}_3 + \kappa_{xy}^3 \tanh \mathcal{A}_1 \tanh \mathcal{A}_2 \right]^2 - \left[\prod_{i=1}^3 \tanh \mathcal{A}_i + \kappa_{xy}^1 \tanh \mathcal{A}_1 + \kappa_{xy}^2 \tanh \mathcal{A}_2 + \kappa_{xy}^3 \tanh \mathcal{A}_3 \right]^2; \quad (2.80)$$

cf. §2.4. Utilizando la aproximación de campo medio $s_{\mathbf{x}} \approx \langle s_{\mathbf{x}} \rangle$, el estado estacionario del sistema vuelve a estar descrito por las mismas ecuaciones de campo medio que para el caso de sinapsis simétricas, es decir, (2.76) y (2.77) pero ahora con \mathcal{J}_{xy} dado por la expresión (2.38). En este punto hay que notar que estas ecuaciones generalizan los resultados que obteníamos al final de la sección §2.4 para sinapsis asimétricas. Ahora no necesitamos imponer a los patrones almacenados las condiciones (restrictivas) que entonces imponíamos para tener una descripción hamiltoniana que aquí no requerimos, concretamente, $\eta_{\mathbf{x}} = \eta_{\mathbf{y}} = \eta \quad \forall \mathbf{x}, \mathbf{y}$.

2.7 Resumen y conclusiones

Como una extensión de un trabajo previo en sistemas desordenados [35, 38, 44], se ha propuesto y estudiado en este capítulo un nuevo modelo de memoria asociativa. Dicho modelo es una generalización, versión cinética, del modelo de red de neuronas de Hopfield, en el sentido de que las intensidades sinápticas no permanecen constantes después del proceso de aprendizaje sino que varían en el transcurso de tiempo en que tiene lugar los cambios en la actividad de las neuronas, como ocurre, en general, en los sistemas biológicos. En este capítulo se ha considerado explícitamente la situación particular en la que se tienen fluctuaciones rápidas de las intensidades sinápticas alrededor de valores medios que vienen determinados por una regla de aprendizaje tipo Hebb. Como consecuencia, la solución estacionaria de la ecuación maestra que caracteriza el sistema está determinada por una función, más o menos compleja, que actúa como probabilidad de transición en dicha ecuación maestra. (De hecho, debido a las fluctuaciones, el estado estacionario del sistema es en general un estado fuera del equilibrio, como si sobre el sistema estuvieran actuando agentes no hamiltonianos.) Se ha demostrado en el desarrollo del capítulo que esto puede inducir efectos interesantes en el aprendizaje, recuperación de información y otros procesos de memoria asociativa.

Para ciertas elecciones de la probabilidad de transición (*rate*), el modelo se reduce a situaciones familiares en la literatura como son el modelo de Ising o el modelo de Hopfield, para los que las intensidades sinápticas no fluctúan, correspondiendo a $p \equiv 1$ en la ecuación (2.12). En el límite de fluctuaciones muy rápidas, es decir $p \rightarrow 0$, las intensidades sinápticas son consideradas como variables aleatorias con distribución de

probabilidad $f(\mathbf{J})$. Se pueden tener en consideración diferentes elecciones de dicha distribución de probabilidad tanto en el caso de acoplamientos simétricos, $J_{xy} = J_{yx}$, como asimétricos, $J_{xy} \neq J_{yx}$. En cualquier caso el estado estacionario depende fuertemente tanto de la distribución $f(\mathbf{J})$ como de la *rate*. Se han estudiado algunas propiedades generales del modelo, y se ha presentado aquí la solución analítica para algunas elecciones de los parámetros en el marco de una teoría de tipo campo medio. En particular, se han considerado situaciones en las que cada sinapsis está asociada a uno de los elementos de cada patrón memorizado, es decir (2.20) donde la matriz η_{xy}^μ puede ser simétrica o asimétrica. Se han presentado ecuaciones cerradas para los parámetros de orden para el caso de *rate V* —cf. ecuación (2.10)— si $\eta_{xy}^\mu = \eta_{yx}^\mu$. El sistema en este caso (pero no para el caso de *rates K y M*) puede ser descrito mediante un hamiltoniano efectivo tipo Hopfield, en el que se muestra que las fluctuaciones sinápticas modifican la temperatura original T , encontrándose el sistema a una temperatura efectiva $T_{eff} > T$. Se ha aplicado el formalismo de réplicas para obtener el siguiente comportamiento explícito.

Un efecto importante de las fluctuaciones de las sinapsis es el exceso de ruido que aparece comparado con el caso del modelo de Hopfield. Consecuentemente, el valor crítico para la razón entre el número de patrones almacenados y el tamaño del sistema, $\alpha \equiv P/N$, por debajo del cual existe la propiedad de memoria asociativa de manera eficiente está reducido en un 4%. Es remarcable el hecho de que la solución con simetría de réplicas permanece estable en nuestro modelo cerca de α_C , debido al hecho de que $T_{eff} > T$, contrariamente a lo que ocurre en el modelo de Hopfield donde la línea Almeida-Thouless excluye la estabilidad en una región pequeña del diagrama de fases suficientemente cercana a α_C y a $T = 0$. Por otra parte, el exceso de ruido reduce notablemente la región del diagrama de fases donde la propiedad de memoria asociativa se ve dificultada por la aparición de estados vidrios de espines. A diferencia de lo que ocurre en el modelo de Hopfield, esta región es finita en nuestro caso; sólo aparece fase vidrio de espines por debajo de la línea $T_{sg}(\alpha)$, que se hace cero para $\alpha > 2.618$ (de forma que los estados vidrio de espines no pueden aparecer por encima de ese valor crítico de α). Aparte de lo anterior hemos encontrado ecuaciones cerradas para los valores estacionarios de los parámetros de orden para el caso de *rate V* si

$J_{xy} \neq J_{yx}$. En esta situación se sigue una descripción mediante un hamiltoniano efectivo bajo ciertas condiciones sobre los patrones memorizados; sin embargo, no existe para patrones arbitrarios una descripción con hamiltonianos efectivos, en general, pues las sinapsis efectivas resultantes son asimétricas, que es la situación habitual en los sistemas biológicos. Además mostramos que a partir de una aproximación más general se sigue un formalismo cinético de campo medio para el caso de sinapsis asimétricas. En los capítulos siguientes de esta memoria se analizan casos adicionales de la realización de nuestro modelo (2.7) correspondientes a elecciones diferentes de la distribución $f(\mathbf{J})$; cf. §3 y §4.

En resumen, hemos mostrado explícitamente que las variaciones temporales, o mejor, las fluctuaciones de las intensidades sinápticas durante la actividad neuronal, que como es sabido ocurren en los sistemas biológicos, afecta, en general, de forma importante al comportamiento de la red de neuronas. Por todo lo dicho sugerimos el interés en estudiar, mediante otro tipo de técnicas, nuevas versiones de nuestro modelo que no son tratables mediante los desarrollos analíticos aquí mostrados.

Capítulo 3

Red de neuronas fuera del equilibrio

3.1 Introducción y definición del modelo

Como se ha mostrado en el capítulo precedente, el modelo de red de neuronas de Hopfield [1] puede ser reformulado como un sistema cinético-estocástico en el que las intensidades sinápticas evolucionan con el tiempo, después de que ha tenido lugar el proceso de aprendizaje, mediante una dinámica local propia consistente en rápidos cambios que compiten con la dinámica asociada a los cambios en la actividad neuronal. Argumentábamos entonces que el comportamiento de los sistemas biológicos sirve de motivación para intentar modelar tales efectos. Por ejemplo, vimos en el capítulo primero que las neuronas biológicas, en general, están acopladas a otras mediante sinapsis que pueden tener diferente naturaleza (eléctrica y química) transmitiendo el potencial de acción a diferente velocidad. Por otra parte, los experimentos más recientes muestran que el vaciado de vacuolas de neurotransmisores al espacio intersináptico y los mecanismos relacionados con él parecen que tienen una componente aleatoria esencial, reflejando ciertos procesos temporales subyacentes, desconocidos hasta ahora, que pueden ser los responsables de la variabilidad observada en la respuesta neuronal a un determinado estímulo; cf. Refs. [4–18].

Como un primer paso en esta dirección, se ha estudiado con detalle en el capítulo segundo el sistema constituido por un conjunto de N neuronas binarias con configuraciones $\mathbf{s} = \{s_x = \pm 1\}$ que evolucionan con el tiempo mediante una ecuación estocástica de la forma

$$\partial_t P_t(\mathbf{s}) = \sum_{\mathbf{x}} [\bar{\omega}(\mathbf{s}^{\mathbf{x}}; \mathbf{x}) P_t(\mathbf{s}^{\mathbf{x}}) - \bar{\omega}(\mathbf{s}; \mathbf{x}) P_t(\mathbf{s})], \quad (3.1)$$

donde las probabilidades de transición vienen dadas en la forma

$$\bar{\omega}(\mathbf{s}; \mathbf{x}) = \int d\mathbf{J} f(\mathbf{J}) \Psi [2\beta s_{\mathbf{x}} h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})]. \quad (3.2)$$

Aquí $\mathbf{s}^{\mathbf{x}}$ es la configuración obtenida a partir de \mathbf{s} después de hacer el cambio $s_{\mathbf{x}} \rightarrow -s_{\mathbf{x}}$ (*spin-flip*); $\mathbf{J} = \{J_{xy} \in \mathfrak{R}\}$ representa una configuración de sinapsis; $\beta = (k_B T)^{-1}$ con k_B la constante de Boltzmann y T la temperatura del baño térmico con el que está en contacto la red; $h_{\mathbf{x}}(\mathbf{s}, \mathbf{J}) = \sum_{\mathbf{y} \neq \mathbf{x}} J_{xy} s_{\mathbf{y}} - \theta_{\mathbf{x}}$ es un campo local con $\theta_{\mathbf{x}}$ la energía umbral necesaria para activar la neurona que se encuentra en la posición \mathbf{x} (por simplicidad, vamos a considerar $\theta_{\mathbf{x}} = 0 \forall \mathbf{x}$ de aquí en adelante); y Ψ es una función verificando la condición de balance detallado, es decir $\Psi(X) = \Psi(-X) \exp(-X)$, y $\Psi(0) = 1$ y verificando además $\lim_{X \rightarrow +\infty} \Psi(X) = 0$; por ejemplo, (2.10) Como vimos en el capítulo segundo las expresiones (3.1) y (3.2) se siguen a partir de ecuaciones dinámicas más generales en las que las sinapsis y neuronas evolucionaban, aquéllas independientemente de la configuración de éstas, con diferentes velocidades. El resultado dado por (3.1) y (3.2) puede ser interpretado asumiendo que cuando el proceso de aprendizaje (plasticidad sináptica) ha sido completado, el subsistema constituido por las neuronas varía en una escala de tiempo t relativamente grande, mientras que el subsistema constituido por las sinapsis varía con el tiempo pero en una escala mucho más fina, $\tau = tp^{-1}$, para p pequeño. Para $p \rightarrow 0$ y $\tau \rightarrow \infty$, las sinapsis evolucionan muy rápidamente comparado con el tiempo típico en que tienen lugar los cambios en las neuronas (como veíamos en el capítulo segundo dichos cambios rápidos pueden asociarse con a fluctuaciones en el estado de las intensidades sinápticas), de forma que se puede considerar a los acoplamientos J_{xy} como variables aleatorias con distribución de probabilidad $f(\mathbf{J})$.

La rápida evolución de las sinapsis, compitiendo con la evolución de las neuronas, puede describirse adoptando una *rate* efectiva para las transiciones $s_{\mathbf{x}} \rightarrow -s_{\mathbf{x}}$ como la (3.2) (cf. §2), que implica una superposición de diferentes mecanismos canónicos, $\Psi [2\beta s_{\mathbf{x}} h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})]$, cada uno de los cuales se corresponde con un campo local diferente $h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})$. La superposición (3.2) no satisface la condición de balance detallado, en general. Por lo tanto, el sistema evoluciona asintóticamente hacia un estado esta-

cionario que, en general, está fuera del equilibrio y que depende fuertemente de la distribución $f(\mathbf{J})$ y de la *rate* $\Psi(X)$. Las consecuencias de tal competición han sido estudiadas explícitamente en el capítulo segundo para la distribución

$$f(\mathbf{J}) = \prod_{\substack{x,y \\ x \neq y}} \sum_{\mu=1}^P a_\mu \delta(J_{xy} - \eta_{xy}^\mu). \quad (3.3)$$

Esta elección implica, dado su carácter multiplicativo, ausencia de correlaciones espaciales entre diferentes sinapsis. Para $\eta_{xy}^\mu \propto \xi_x^\mu \xi_y^\mu$, esta falta de correlación espacial puede ser interpretada suponiendo que cada intensidad sináptica está asociada, con probabilidad a_μ , a dos de los elementos de un patrón almacenado, $\xi^\mu \equiv \{\xi_x^\mu = \pm 1\}$ con $\mu = 1, \dots, P$. La principal consecuencia que se tiene para esta situación, como vimos en el capítulo segundo, es que las fluctuaciones sinápticas añaden un exceso de ruido comparado con la situación correspondiente al modelo de Hopfield. Consecuentemente, el valor α_C , para la razón $\alpha \equiv P/N$, por debajo del cual existe memoria asociativa efectiva, decrece (ligeramente) y la región del diagrama de fases, donde la aparición de la fase vidrio de espines dificulta la propiedad de memoria asociativa, se reduce notablemente, y no aparece a temperatura cero por encima de cierto valor de α .

Pero la elección de (3.3) para la distribución de las fluctuaciones es sólo un caso particular del modelo más general descrito por (3.1)-(3.2); de hecho, el tratar con una descripción hamiltoniana sencilla ha sido la principal motivación para estudiar (3.3), pues esta elección da lugar a resultados exactos. Con objeto de entender mejor la influencia de la dinámica de las sinapsis en la realización de modelos tipo Hopfield, en el presente capítulo generalizamos la situación representada por (3.3) a un estudio de fluctuaciones en las sinapsis que pueden exhibir correlaciones espaciales. En particular, vamos a mostrar el comportamiento que se sigue de (3.1), (3.2) y (2.10) para

$$f(\mathbf{J}) = \sum_{\mu=1}^P a_\mu \prod_{\substack{x,y \\ x \neq y}} \delta(J_{xy} - \eta_{xy}^\mu). \quad (3.4)$$

Si $\eta_{xy}^\mu = a_\mu^{-1} \xi_x^\mu \xi_y^\mu$, la elección (3.4) implica que cada memoria ξ^μ contribuye con probabilidad a_μ a la configuración de sinapsis \mathbf{J} , de forma que puede haber correlaciones espaciales debido a las correlaciones dentro de los patrones ξ^μ . La elección (3.4) impide, en la práctica, una descripción hamiltoniana similar a la encontrada en el capítulo segundo para (3.3). Estamos interesados en una situación en la que cada intensidad J_{xy}

varía aleatoriamente con el tiempo de forma que su promedio durante el intervalo de tiempo que caracteriza la actividad neuronal, esto es \bar{J}_{xy} , tiene un valor correspondiente a una determinada regla de aprendizaje. Por ejemplo, suponiendo $\eta_{xy}^\mu \propto \xi_x^\mu \xi_y^\mu$ se tiene

$$\bar{J}_{xy} \equiv \int d\mathbf{J} f(\mathbf{J}) J_{xy} = \frac{1}{N} \sum_{\mu=1}^P \xi_x^\mu \xi_y^\mu, \quad (3.5)$$

es decir las sinapsis fluctúan alrededor de valores medios correspondientes a la regla de Hebb [3]. Queremos hacer notar que los resultados que a continuación describimos se refieren al límite termodinámico, es decir, al caso $N \rightarrow \infty$, y el comportamiento explícito que presentamos aquí es para el caso de P finito. Sin embargo, creemos, y presentamos cierta evidencia en ese sentido, que nuestros resultados para P grande pertenecen al régimen asintótico, de modo que también se cumplen para $P \rightarrow \infty$.

3.2 Cambios de fase continuos

La actividad neuronal promedio verifica, a partir de (3.1) que

$$\partial_t \langle s_{\mathbf{x}} \rangle = -2 \langle s_{\mathbf{x}} \bar{a}_{\mathbf{x}}^+(\mathbf{s}) \rangle - 2 \langle \bar{a}_{\mathbf{x}}^-(\mathbf{s}) \rangle, \quad (3.6)$$

donde

$$\bar{a}_{\mathbf{x}}^\pm(\mathbf{s}) \equiv \frac{1}{2} \int d\mathbf{J} f(\mathbf{J}) \{ \Psi [2\beta s_{\mathbf{x}} h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})] \pm \Psi [-2\beta s_{\mathbf{x}} h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})] \}. \quad (3.7)$$

Más explícitamente utilizando (3.4) se tiene

$$\partial_t \langle s_{\mathbf{x}} \rangle = -2 \left\langle s_{\mathbf{x}} \sum_{\mu=1}^P a_{\mu} \mathcal{A}_{\mu, \mathbf{x}}^+(\mathbf{s}) \right\rangle - 2 \left\langle \sum_{\mu=1}^P a_{\mu} \mathcal{A}_{\mu, \mathbf{x}}^-(\mathbf{s}) \right\rangle, \quad (3.8)$$

donde

$$\mathcal{A}_{\mu, \mathbf{x}}^\pm(\mathbf{s}) = \frac{1}{2} \left\{ \Psi \left[2\beta \sum_{\mathbf{y}} \eta_{xy}^\mu s_{\mathbf{y}} \right] \pm \Psi \left[-2\beta \sum_{\mathbf{y}} \eta_{xy}^\mu s_{\mathbf{y}} \right] \right\}. \quad (3.9)$$

El campo local $h_{\mathbf{x}}(\mathbf{s}, \mathbf{J}) = \sum_{\mathbf{y} \neq \mathbf{x}} J_{xy} s_{\mathbf{y}}$ implica considerar interacciones entre una neurona dada y el resto. Una hipótesis de prueba usual, consistente con el hecho de que haya tales interacciones de largo alcance y con el límite $N \rightarrow \infty$, consiste en la condición de campo medio $s_{\mathbf{x}} = \langle s_{\mathbf{x}} \rangle$; cf. §4. Esta condición implica que la solución estacionaria de (3.8) es

$$\langle s_{\mathbf{x}} \rangle = \frac{\sum_{\mu=1}^P a_{\mu} \{ \Psi(-2\beta m_{\mathbf{x}}^{\mu}) - \Psi(2\beta m_{\mathbf{x}}^{\mu}) \}}{\sum_{\mu=1}^P a_{\mu} \{ \Psi(-2\beta m_{\mathbf{x}}^{\mu}) + \Psi(2\beta m_{\mathbf{x}}^{\mu}) \}}, \quad (3.10)$$

donde $m_{\mathbf{x}}^{\mu} \equiv \sum_{\mathbf{y}} \eta_{xy}^{\mu} \langle s_{\mathbf{y}} \rangle$, $\forall \mu = 1, \dots, P$. Posteriormente remarcaremos algunas consecuencias generales de (3.10). La expresión (3.10) se verifica para $\langle s_{\mathbf{x}} \rangle = 0 \quad \forall \mathbf{x}$. Por lo tanto, las soluciones no triviales pueden ser obtenidas después de desarrollar (3.10), en potencias de $m_{\mathbf{x}}^{\mu}$, alrededor de $m_{\mathbf{x}}^{\mu} = 0 \quad \forall \mu = 1, \dots, P$. Se puede ver fácilmente que en primer orden (3.10) queda de la forma

$$\langle s_{\mathbf{x}} \rangle \approx \beta \sum_{\mathbf{y}} \sum_{\mu=1}^P a_{\mu} \eta_{xy}^{\mu} \langle s_{\mathbf{y}} \rangle. \quad (3.11)$$

Esta ecuación muestra que soluciones no nulas para los parámetros de orden $\langle s_{\mathbf{x}} \rangle$ y $m_{\mathbf{x}}^{\mu}$ aparecen por debajo de una temperatura crítica, T_c , que es solución a la ecuación

$$\det \left\{ \beta \sum_{\mu=1}^P a_{\mu} \eta_{xy}^{\mu} - \delta_{xy} \right\} = 0. \quad (3.12)$$

Es preciso remarcar que, a pesar de las muchas similitudes con el familiar cambio de fase de segundo orden de equilibrio, el presente cambio de fase tiene una naturaleza fuera del equilibrio [45]. Esto se refleja, por ejemplo, en el hecho de que (a diferencia del equilibrio) T_c depende de la elección particular para η_{xy}^{μ} y de las probabilidades a_{μ} , es decir, de la distribución (3.3) y por lo tanto de la *rate* (3.2). También es verdad que, en el caso presente, T_c es independiente de la función $\Psi(X)$. Sin embargo la validez de la expresión (3.11) se pierde cuando $T \rightarrow 0$, límite para el que las correspondientes soluciones exhiben una dependencia en las propiedades asintóticas de $\Psi(X)$ cuando $X \rightarrow \pm\infty$; en estos límites, debido a la relación de balance detallado, $\Psi(X) = \Psi(-X) e^{-X}$, el único comportamiento que hay que considerar a temperaturas bajas es $\lim_{X \rightarrow -\infty} \Psi(X) = \rho \exp(-\Delta X)$, donde $\rho = 1, \Delta = 1/2$ para *rate V*, $\rho = 2, \Delta = 0$ para *rate K*, y $\rho = 1, \Delta = 0$ para *rate M*.

Por ejemplo, consideremos el caso $\eta_{xy}^{\mu} = N^{-1} \eta^{\mu}$, es decir, cada patrón contribuye, con probabilidad a_{μ} , con la misma intensidad en todas las sinapsis, independientemente de \mathbf{x} y \mathbf{y} . Las ecuaciones (3.11) y (3.12) predicen un cambio de fase, de una fase

paramagnética a una fase ferromagnética a $T_c = \sum_{\mu=1}^P a_\mu \eta^\mu$. Cuando $T \rightarrow 0$, se tiene a partir de (3.10) que

$$\langle s_{\mathbf{x}} \rangle = \begin{cases} \pm \sum_{\mu} a_{\mu} \text{sign}(\eta^{\mu}) & \text{para } \sum_{\mu} a_{\mu} \text{sign}(\eta^{\mu}) > 0 \\ 0 & \text{en caso contrario} \end{cases} \quad (3.13)$$

para el caso de las *rates* K y M ($\Delta = 0$), y

$$\langle s_{\mathbf{x}} \rangle = \begin{cases} \pm 1 & \text{para } \text{sign}(\eta^{\alpha}) > 0 \\ 0 & \text{en caso contrario,} \end{cases} \quad (3.14)$$

para la *rate* V ($\Delta \neq 0$). El índice α indica aquel patrón para el que $|\eta^{\alpha}| = \max_{\mu} |\eta^{\mu}|$.

Como otro ejemplo ilustrativo, consideremos $\eta_{xy}^{\mu} = \lambda_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}$ con $\lambda_{\mu} > 0$. En este caso el interés está en ver en qué condiciones el sistema tiene, como soluciones estacionarias a (3.8), estados que tengan un solapamiento $m^{\mu}(\mathbf{s}) \equiv N^{-1} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^{\mu} s_{\mathbf{x}}$ no nulo con un determinado patrón ξ^{μ} . Obtenemos a partir de (3.11) que tales soluciones aparecen cuando

$$\sum_{\mu=1}^P m^{\mu} \left(\delta_{\mu\nu} - \beta \sum_{\mathbf{x}} a_{\mu} \lambda_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{x}}^{\nu} \right) = 0, \quad (3.15)$$

donde m^{μ} es igual a $m^{\mu}(\mathbf{s})$ con $s_{\mathbf{x}} = \langle s_{\mathbf{x}} \rangle \quad \forall \mathbf{x}$. La expresión (3.15) es una condición de criticalidad para la aparición de soluciones no triviales para el parámetro de orden m^{μ} , y por lo tanto indica el valor crítico de temperatura por debajo del cual el sistema puede tener la capacidad de memoria asociativa. En el caso de patrones ortogonales o cuasiortogonales (cf. §2.6), en el límite $N \rightarrow \infty$, la condición (3.15) revela un cambio de fase de segundo orden fuera del equilibrio [45], de una fase paramagnética a una fase ferromagnética, a una temperatura crítica, independiente de la función $\Psi(X)$, que es $T_c = \max_{\mu} a_{\mu}$ si $\lambda_{\mu} = N^{-1} \quad \forall \mu$, y $T_c = 1$ si $\lambda_{\mu} = (Na_{\mu})^{-1}$. Este último caso, que corresponde a considerar distribuciones $f(\mathbf{J})$ que describen fluctuaciones alrededor de la regla de Hebb, va a ser estudiado con detalle en las próximas secciones. Por otra parte, el comportamiento en $T \rightarrow 0$ de las soluciones no triviales, que aparecen en el cambio de fase descrito por la ecuación (3.15), depende de la elección particular de la función $\Psi(X)$ y no depende de la elección particular del parámetro λ_{μ} . Así a partir de (3.10) se obtiene que en $T = 0$ y para patrones ortogonales o cuasi-ortogonales es $m^{\mu} = \pm a_{\mu}$ para las *rates* K y M , y $m^{\mu} = \delta_{\mu\gamma} \text{sign}(m^{\gamma}) \quad \forall \mu$ para la *rate* V , donde γ indica aquel patrón tal que $|m^{\gamma}| = \max_{\mu} |m^{\mu}|$. Este comportamiento peculiar del

sistema a bajas temperaturas hace necesario un análisis más detallado del sistema cerca de $T = 0$, estudio que se muestra de forma precisa y detallada en las próximas secciones.

3.3 Fluctuaciones alrededor de la regla de Hebb

Los resultados anteriores sugieren el estudiar de una forma más detallada el caso en que las sinapsis fluctúan alrededor de valores medios determinados por una regla de aprendizaje dada; consideramos aquí explícitamente el caso de la regla de Hebb. Como vimos antes el considerar fluctuaciones alrededor de la regla de Hebb corresponde en nuestro formalismo a considerar la elección $\eta_{xy}^\mu = (a_\mu N)^{-1} \xi_x^\mu \xi_y^\mu$, que implica fluctuaciones de valor medio dado por (3.5) y variancia

$$\sigma_{xy}^2(\boldsymbol{\xi}) = \overline{(J_{xy} - \bar{J}_{xy})^2} = \left(\frac{P}{N}\right)^2 - \frac{1}{N^2} \sum_{\mu, \nu} \xi_x^\mu \xi_y^\mu \xi_x^\nu \xi_y^\nu, \quad (3.16)$$

con $\boldsymbol{\xi} \equiv \{\xi^\mu; \mu = 1, \dots, P\}$. En este caso, las ecuaciones (3.8) y (3.9) muestran explícitamente la dependencia en los patrones ξ^μ y en los solapamientos $m^\mu(\mathbf{s})$ en la forma

$$\partial_t \langle s_{\mathbf{x}} \rangle = -2 \left\langle s_{\mathbf{x}} \sum_{\mu=1}^P a_\mu \mathcal{B}_\mu^+(\mathbf{s}) \right\rangle - 2 \left\langle \sum_{\mu=1}^P a_\mu \xi_{\mathbf{x}}^\mu \mathcal{B}_\mu^-(\mathbf{s}) \right\rangle, \quad (3.17)$$

donde

$$\mathcal{B}_\mu^\pm(\mathbf{s}) = \frac{1}{2} \left\{ \Psi \left[2\beta a_\mu^{-1} m^\mu(\mathbf{s}) \right] \pm \Psi \left[-2\beta a_\mu^{-1} m^\mu(\mathbf{s}) \right] \right\}. \quad (3.18)$$

Estas ecuaciones tienen una importancia crucial pues para $N \rightarrow \infty$ con P permaneciendo finito y con la condición de campo medio $s_{\mathbf{x}} = \langle s_{\mathbf{x}} \rangle$, es posible encontrar una ecuación dinámica cerrada para los solapamientos; concretamente se tiene

$$\partial_t m^\mu = -2m^\mu \sum_{\nu=1}^P a_\nu \mathcal{B}_\nu^+ - 2a_\mu \mathcal{B}_\mu^-, \quad (3.19)$$

donde \mathcal{B}_μ^\pm se obtiene a partir de la definición dada en (3.18) con $s_{\mathbf{x}}$ reemplazado por $\langle s_{\mathbf{x}} \rangle$. Este resultado se verifica para patrones ξ^μ que son ortogonales entre sí (o por lo menos quasi-ortogonales entre sí para el caso en que $\{\xi_{\mathbf{x}}^\mu\}$ es un conjunto de $P \times N$ variables aleatorias estadísticamente independientes; cf. §2.6) de forma que se tiene

$\frac{1}{N} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{x}}^{\nu} = \delta_{\mu\nu}$ (o bien $\rightarrow \delta_{\mu\nu}$ cuando $N \rightarrow \infty$ con P finito). La solución estacionaria a (3.19) es entonces $\mathbf{m} = (m^1, \dots, m^P)$ con

$$m^{\mu} = a_{\mu} \mathcal{B}_{\mu}^{-} \left(\sum_{\nu=1}^P a_{\nu} \mathcal{B}_{\nu}^{+} \right)^{-1}. \quad (3.20)$$

Solución trivial a (3.20) es $\mathbf{m} = (0, 0, \dots, 0)$, y soluciones no triviales de la forma $\mathbf{m} = (m^1, \dots, m^n, 0, \dots, 0)$, con $n \leq P$ y $1 \geq |m^{\mu}| \geq 0 \quad \forall \mu = 1, \dots, P$, aparecen a temperaturas suficientemente bajas por debajo de $T_c = 1$. En cuanto a su estabilidad, que en general depende de la función Ψ , estas soluciones son *puntos de silla* con un determinado número de direcciones de inestabilidad. (Los detalles referentes a la estabilidad de las soluciones a (3.20) se estudian en §3.4.) De entre este tipo de soluciones aquellas con $n = 1$, es decir los estados de Mattis [43], van a ser los importantes para la propiedad de memoria asociativa; para el caso de P patrones almacenados hay $2P$ estados de Mattis correspondientes a los P patrones ξ^{μ} y sus negativos $-\xi^{\mu}$ (que pueden aparecer debido a la simetría ± 1 de los elementos $\xi_{\mathbf{x}}^{\mu}$). Los estados mezcla ($n > 1$) pueden ser de varios tipos. Para patrones que contribuyen con probabilidad a_{μ} a las intensidades sinápticas se tiene $|m^{\mu}| \neq |m^{\nu}|$ para $a_{\mu} \neq a_{\nu}$ y $|m^{\mu}| = |m^{\nu}|$ para $a_{\mu} = a_{\nu}$. En el caso en que todos los patrones contribuyen con la misma probabilidad a la intensidad de las sinapsis, es decir $a_{\mu} = \frac{1}{P}$, se tiene que las soluciones a (3.20) son independientes de μ . Como caso particular tenemos las soluciones $\mathbf{m} = m_n(1, \dots, 1, -1, \dots, -1, 0, \dots, 0)$, con k componentes $+1$ y s componentes -1 , siendo $n = k + s$; éstos son conocidos como *estados mezcla simétricos* para el caso en que se tiene $k = 0$ o $s = 0$, y *estados mezcla asimétricos* para $k, s \neq 0$.

Estos resultados generales son independientes de la función Ψ . Resultados más explícitos van a depender de la forma particular de la *rate* Ψ . En particular, para la *rate* K , la expresión (3.20) queda de la forma $m^{\nu} = a_{\mu} \tanh(\beta m^{\nu} / a_{\mu})$, es decir las soluciones son $m^{\nu} = a_{\mu} x_0(T)$, donde $x_0 = \tanh(\beta x_0)$. Por lo tanto, a una temperatura dada, se tiene $m^{\mu} \propto a_{\mu}$. El comportamiento de las soluciones cerca de $T_c = 1$ es de la forma

$$m^{\mu} \approx \pm \sqrt{3} a_{\mu} T_c (T_c - T)^{\frac{1}{2}}, \quad T \rightarrow T_c^{-}, \quad (3.21)$$

y se tiene $m^{\mu} = \pm a_{\mu}$ cuando $T \rightarrow 0$. Para $a_{\mu} = P^{-1}$, los únicos posibles valores de m^{μ} para una temperatura T dada son 0 y $\pm P^{-1} |x_0(T)|$, de forma que el solapamiento

decrece conforme aumenta el número de patrones almacenados. Como veremos en la sección §3.4, estos estados van a ser, localmente, puntos de silla con direcciones de inestabilidad asociadas a solapamientos nulos con cierto número de patrones; por lo tanto, el único estado localmente estable va a ser aquel que tiene solapamientos no nulos con todos los patrones (en particular los estados de Mattis van a ser localmente inestables).

Para la *rate* M , (3.20) queda de la forma

$$|m^\mu| = \frac{a_\mu [1 - \exp(-2\beta a_\mu^{-1} |m^\mu|)]}{\sum_{\nu=1}^P a_\nu [1 + \exp(-2\beta a_\nu^{-1} |m^\nu|)]}. \quad (3.22)$$

A diferencia de la *rate* K , ahora m^μ depende ahora de los $(P - 1)$ solapamientos restantes. El comportamiento del sistema a $T \rightarrow 0$ va a depender del tipo de solución $\mathbf{m} = (m^1, \dots, m^n, \dots, 0)$. Por ejemplo, para el caso en que los estados \mathbf{m} son tales que $n = P$, es decir todos los solapamientos son distintos de cero, se tiene $|m^\mu| = a_\mu$ en $T = 0$ (como para el caso de la *rate* K). Sin embargo, si el estado solución \mathbf{m} es tal que tiene $P - n$ (con $n < P$) solapamientos nulos, con otros tantos patrones, entonces el comportamiento del resto de los n solapamientos no nulos en $T = 0$ es

$$|m^\mu| = a_\mu \left[1 + \sum_{\nu=n+1}^P a_\nu \right]^{-1}. \quad (3.23)$$

En el caso que todos los patrones memorizados son equiprobables, es decir $a_\mu = \frac{1}{P}$, m^μ puede tomar sólo los valores 0 y $\pm|m_n|$, para una temperatura T dada, y se sigue que por debajo de $T_c = 1$ estos solapamientos son de la forma

$$|m_n| = \begin{cases} (P - n)^{-1} T_c (T_c - T) & \text{Si } n < P, \\ \sqrt{3} P^{-1} T_c (T_c - T)^{\frac{1}{2}} & \text{Si } n = P. \end{cases} \quad T \rightarrow T_c^-. \quad (3.24)$$

Esta solución cambia de forma continua hacia $|m_n| = (2P - n)^{-1}$, con $n \leq P$, cuando $T \rightarrow 0$. Por otra parte como veremos en la sección §3.4, al igual que para la *rate* K , sólo los estados mezcla con solapamientos no nulos con todos los patrones son localmente estables (cf. §3.4).

En el caso de *rate* V la ecuación (3.20) queda de la forma

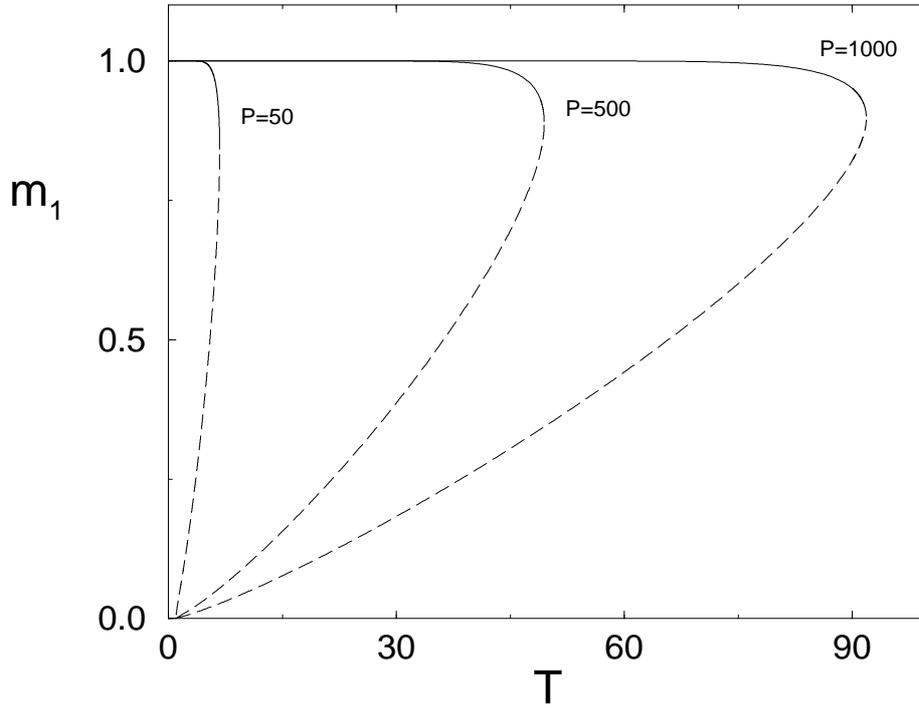


Figura 3.1: La figura muestra las dos soluciones que tiene la ecuación (3.20) para $n = 1$ y *rate* V , representadas en función de T (en unidades k_B^{-1}) y para diferentes valores de P . La línea continua representa la solución localmente estable mientras que la línea discontinua representa la solución inestable. Hay que hacer notar que las soluciones estables aparecen de forma discontinua por debajo de una temperatura $\tilde{T}(1, P)$, lo que ocurre siempre para $P \geq 3$.

$$m^\mu = a_\mu \sinh(\beta a_\mu^{-1} m^\mu) \left[\sum_{\nu=1}^P a_\nu \cosh(\beta a_\nu^{-1} m^\nu) \right]^{-1}. \quad (3.25)$$

Al igual que para el caso de *rate* M , m^μ depende ahora de los $(P - 1)$ solapamientos restantes y de a_μ , para una temperatura T dada. El comportamiento de la solución (3.25) para $T \rightarrow 0$ es

$$m^\mu = a_\mu \left(\sum_{\nu=1}^{n_\gamma} a_\nu \right)^{-1} \text{sign}(m^\mu) \delta_{\gamma\mu}, \quad (3.26)$$

donde γ indica aquel patrón tal que $|m^\gamma| = \max_\mu |m^\mu|$, y n_γ es el número de solapamientos tales que $|m^\mu| = |m^\gamma|$. Para $n_\gamma = 1$ se tiene $m^\mu = \text{sign}(m^\mu) \delta_{\gamma\mu}$; por lo tanto, en $T = 0$ el sistema es capaz de recuperar un patrón sin error. Por simplicidad, para

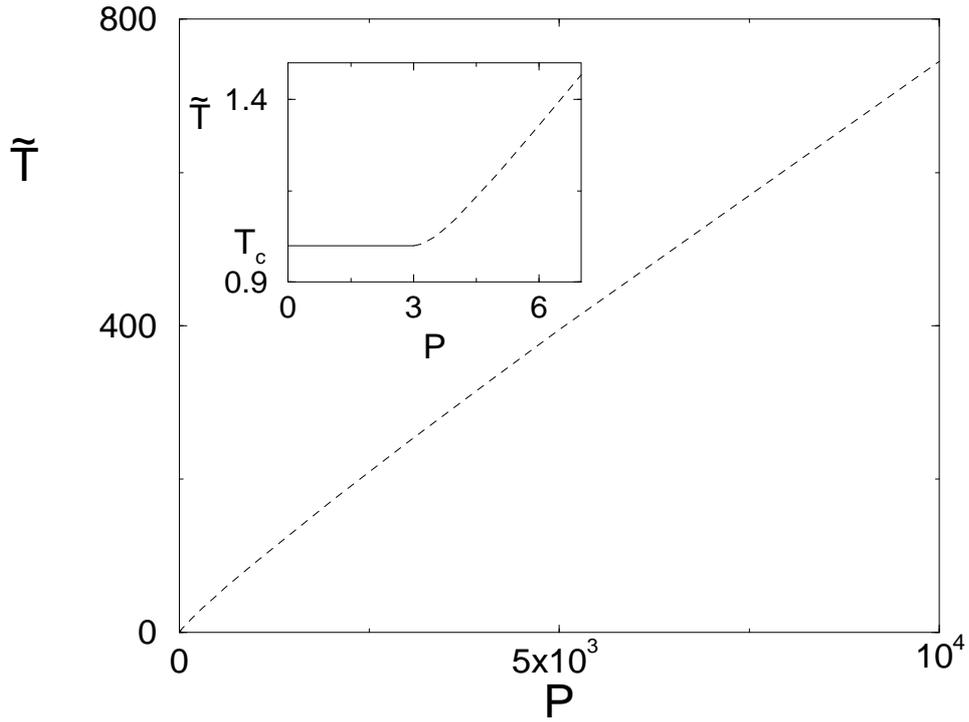


Figura 3.2: La figura muestra el comportamiento de la temperatura $\tilde{T}(1, P)$ en función del número de patrones que la red ha almacenado. La línea discontinua corresponde a cambios de fase de primer orden, mientras que la línea continua se corresponde con cambios de fase de segundo orden. El detalle cerca del origen, muestra como los cambios de fase se hacen discontinuos por encima del punto tricrítico $(T_c, P = 3)$.

$T > 0$ hemos estudiado con detalle el caso en que todos los patrones son equiprobables, esto es $a_\mu = \frac{1}{P}$, situación para la que se tiene $m^\mu = 0, \pm|m_n|$, donde m_n es función de T , n , y P . A diferencia de lo que ocurre para las *rates* K y M , ahora aparecen dos tipos de estados mezcla: Para $n > \frac{1}{3}P$, se tienen estados mezcla tales que $m_n = 0$ para $T > T_c = 1$ y

$$|m_n| = \frac{\sqrt{6}T_c}{\sqrt{P(3n - P)}}(T_c - T)^{\frac{1}{2}} \quad T \rightarrow T_c^-. \quad (3.27)$$

Sin embargo, para $n \leq \frac{P}{3}$, aparecen estados mezcla tales que m_n se comporta de forma discontinua a una temperatura $\tilde{T}(n, P) \equiv \tilde{m}_n P \theta_n^{-1}$, donde θ_n es solución de

$$n\theta_n + (P - n)(\theta_n \cosh \theta_n - \sinh \theta_n) - n \sinh \theta_n \cosh \theta_n = 0 \quad (3.28)$$

y

$$\tilde{m}_n = \sinh \theta_n (n \cosh \theta_n + P - n)^{-1} \quad (3.29)$$

es el valor de m_n para $T = \tilde{T}(n, P)$; se obtiene que $\tilde{T}(n, P < 3n) = T_c$. El cambio de fase que tiene lugar a $\tilde{T}(n, P)$ es de primer orden como se sigue de la existencia de solapamientos \tilde{m}_n no necesariamente pequeños (salvo para estados mezcla con n grande) justo por debajo de $\tilde{T}(n, P)$, que implica propiedades interesantes en el sistema. En cualquier caso se tiene que $|m_n| = n^{-1}$ cuando $T \rightarrow 0$.

Este comportamiento que caracteriza la *rate V* se ilustra en las figuras 3.1 y 3.2 para el caso de estados de Mattis $n = 1$. Las figuras muestran, en particular la existencia de un *punto tricrítico de no equilibrio* en $(T = T_c, P = 3)$. De hecho, las soluciones a (3.20) corresponden, en general, a puntos de silla con un número de direcciones de estabilidad que dependen de la función particular $\Psi(X)$. Como mostraremos explícitamente en la próxima sección, se sigue que para el caso de la *rate V* sólo los estados de Mattis ($n=1$) son localmente estables, mientras que para el caso de las *rates K* y *M* los únicos estados con estabilidad local van a ser estados mezcla con $n = P$.

3.4 Estabilidad y cambios de fase de primer orden

La estabilidad local de cualquier solución particular del sistema (3.19), digamos m_0^μ , puede ser estudiada linealizando dicho sistema en la forma (cf. Ref. [46], por ejemplo)

$$\partial_t n^\mu = \sum_{\nu=1}^P \mathcal{Q}_{\mu\nu} n^\nu + \mathcal{O}(\mathbf{n}^2), \quad (3.30)$$

donde $\mathbf{n} = (n^1, \dots, n^P)$ con $n^\mu \equiv m^\mu - m_0^\mu \quad \forall \mu$. La estabilidad local de la solución $\mathbf{m} = \mathbf{m}_0$ viene dada por la estabilidad de la solución $\mathbf{n} = \mathbf{0}$ del sistema (3.30).¹ La estabilidad local de dicha solución viene dada por el signo de los autovalores de la matriz \mathcal{Q} , que tienen que ser negativos para que la solución sea localmente estable. Los elementos de matriz de \mathcal{Q} dependen de la función *rate* $\Psi(X)$. Por lo tanto, los autovalores de \mathcal{Q} dependen también de la forma particular de Ψ (cf. apéndice C). Una primera conclusión se obtiene es que la estabilidad de (3.19) depende incluso cualitativamente de $\Psi(X)$. En particular, en el caso en que la solución en estudio es la solución

¹Los detalles pueden verse en la Ref. [46].

estacionaria a (3.19), es decir, la solución $\mathbf{m}_0 \equiv (m^1, \dots, m^P)$ con m^μ $\mu = 1, \dots, P$ verificando (3.20) se tiene el comportamiento que a continuación detallamos.

Para la *rate* K , los elementos de matriz de \mathcal{Q} tienen la forma

$$\mathcal{Q}_{\mu\nu} = \left\{ \frac{2\beta}{\cosh^2(\beta a_\mu^{-1} m^\mu)} - 2 \right\} \delta_{\mu\nu}, \quad (3.31)$$

donde los solapamientos m^μ verifican (3.20). Aunque los elementos de matriz dependen de las probabilidades a_μ , la forma diagonal que presenta en este caso la matriz \mathcal{Q} hace que podamos obtener de forma sencilla sus autovalores. Estos son

$$\tilde{\lambda}_\mu = \frac{2\beta}{\cosh^2(\beta a_\mu^{-1} m^\mu)} - 2 \quad \forall \mu. \quad (3.32)$$

Como vimos antes (cf. §3.3), los estados solución, con solapamientos no nulos con un determinado número n de patrones, aparecen de forma continua justo por debajo de $T_c = 1$. Por otra parte, a partir de (3.32) y (3.20) se tiene que, para $T < T_c$ (independientemente del valor de las probabilidades a_μ), es $\tilde{\lambda}_\mu < 0$ para $m^\mu \neq 0$ y $\tilde{\lambda}_\mu > 0$ para $m^\mu = 0$; cf. propiedad 3.1. De esta forma, el estado $\mathbf{m} = (m^1, \dots, m^n, 0, \dots, 0)$ es, localmente, un punto de silla con $(P - n)$ direcciones de inestabilidad; por lo tanto cualquier estado mezcla con $n < P$ —en particular los estados de Mattis ($n=1$)— son inestables localmente; sólo los estados mezcla con $n = P$, esto es estados que tienen solapamientos no nulos con los P patrones almacenados, son localmente estables. El hecho de que los estados de Mattis ($n = 1$) sean localmente inestables hace que el estado estacionario del sistema definido por la distribución (3.4) no presente la propiedad de memoria asociativa cuando el modelo está construido con la *rate* K .

Propiedad 3.1 *Sea \mathcal{Q} la matriz asociada al sistema dinámico descrito por el sistema de ecuaciones (3.19), particularizado al caso de *rate* K , cuyos elementos vienen dados por (3.31) y cuyos autovalores $\tilde{\lambda}_\mu$ vienen dados en (3.32). Entonces para $T < T_c$ se tiene $\tilde{\lambda}_\mu < 0$ si $m^\mu \neq 0$ y $\tilde{\lambda}_\mu > 0$ si $m^\mu = 0$.*

Demostración:

Si $m^\mu = 0$ entonces a partir de (3.32) y del hecho de que $T < T_c$ se tiene $\tilde{\lambda}_\mu > 0$. En caso contrario, es decir $m^\mu \neq 0$, consideremos la función $f(x) = (2x - 2 \sinh x \cosh x) (\cosh^2 x)^{-1}$. Trivialmente se ve que $f(x)$ es

continua y derivable $\forall x \in \mathfrak{R}$ y además verifica que $f(0) = 0$. Por otra parte, $f'(x) = -(4x \sinh x \cosh x) (\cosh^4 x)^{-1} < 0 \forall x \in \mathfrak{R}$, por lo que $f(x)$ es estrictamente decreciente. Por lo tanto, si $x < 0$ entonces $f(x) > 0$ y si $x > 0$ entonces $f(x) < 0$. Tomando $x = \beta m^\mu a_\mu^{-1}$ se tiene, a partir de (3.32) y del hecho de que las soluciones a (3.20), para el caso de *rate* K , verifican la relación $m^\mu = a_\mu \tanh(\beta m^\mu a_\mu^{-1})$, que $f(\beta m^\mu a_\mu^{-1}) = \tilde{\lambda}_\mu m^\mu a_\mu^{-1}$. Por lo tanto, si $m_\mu \neq 0$ se tiene que $\tilde{\lambda}_\mu < 0$ en virtud de la propiedad de decrecimiento de la función $f(x)$ y del hecho de que los coeficientes a_μ son definidos positivos. †

En el caso de *rate* M , los elementos de matriz de \mathcal{Q} tienen la forma

$$\mathcal{Q}_{\mu\nu} = 2\beta |m^\mu| \Xi_\nu + \delta_{\mu\nu} \left\{ 2\beta \Xi_\nu - \sum_{\alpha=1}^P a_\alpha (1 + \Xi_\alpha) \right\}, \quad (3.33)$$

con

$$\Xi_\nu \equiv \exp(-2\beta a_\nu^{-1} |m^\nu|) \quad (3.34)$$

donde $|m^\mu|$ viene dado en (3.22). En este caso, la matriz de estabilidad \mathcal{Q} no tiene una forma diagonal sencilla y su diagonalización tiene la dificultad de que los elementos de matriz dependen de las probabilidades a_μ . En el caso particular en que $a_\mu = \frac{1}{P} \forall \mu$, para una temperatura $T < T_c$, y estados solución tipo mezcla con $n \leq P$ solapamientos no nulos (cf. §3.3), la matriz puede ser diagonalizada² obteniendo los autovalores

$$\begin{aligned} \tilde{\lambda}_1 &= 2\beta(n|m_n| + 1)\tilde{\Xi}_n + \frac{n}{P} (1 - \tilde{\Xi}_n) - 2, \\ \tilde{\lambda}_2 &= 2\beta\tilde{\Xi}_n + \frac{n}{P} (1 - \tilde{\Xi}_n) - 2, \\ \tilde{\lambda}_3 &= 2\beta + \frac{n}{P} (1 - \tilde{\Xi}_n) - 2 \end{aligned} \quad (3.35)$$

donde

$$\tilde{\Xi}_n \equiv \exp(-2\beta P |m_n|) \quad (3.36)$$

y los solapamientos m_n son solución a la ecuación (3.22); $\tilde{\lambda}_2$ y $\tilde{\lambda}_3$ aparecen con multiplicidad $(n-1)$ y $(P-n)$, respectivamente, mientras que $\tilde{\lambda}_1$ aparece con multiplicidad uno. Se puede demostrar fácilmente que $\tilde{\lambda}_3 > 0$ (cf. propiedad 3.2), de forma que todas aquellas soluciones que lleven asociado este autovalor van a ser localmente inestables.

²Los detalles pueden verse en el apéndice C.

Dado que dicho autovalor sólo aparece cuando $n < P$, todos los estados mezcla con $n < P$ son localmente inestables, en particular los estados de Mattis ($n = 1$). En el caso de estados mezcla con solapamientos no nulos con todos los patrones, es decir $n = P$, el autovalor $\tilde{\lambda}_3$ no aparece y los otros dos, a saber $\tilde{\lambda}_1$ y $\tilde{\lambda}_2$, son negativos (cf. propiedad 3.2) de forma que dicho estado es localmente estable, de igual forma a lo que ocurriría para la *rate* K . Como consecuencia de esto, el sistema para el caso de *rate* M no presenta la propiedad de memoria asociativa.

Propiedad 3.2 Sea \mathcal{Q} la matriz asociada al sistema dinámico descrito por el sistema de ecuaciones (3.19), particularizado al caso de *rate* M , cuyos elementos vienen dados por (3.33) y cuyos autovalores vienen dados por (3.35) para el caso en que $a_\mu = \frac{1}{P} \quad \forall \mu$. Entonces, para una temperatura $T < T_c$ se tiene:

- (i) $\tilde{\lambda}_1 < 0$ y $\tilde{\lambda}_2 < 0$.
- (ii) $\tilde{\lambda}_3 > 0$.

Demostración:

Demostremos primero (ii). Consideremos la función $f(x) = x - \{1 - \exp(-x)\}$. Dicha función es continua y derivable $\forall x \in \mathfrak{R}$ y se tiene $f(0) = 0$. La función derivada es $f'(x) = 1 - \exp(-x)$. Trivialmente se tiene que $f'(x) > 0$ si $x > 0$ y $f'(x) < 0$ si $x < 0$, de forma que en $x = 0$ tiene un mínimo absoluto al ser $f''(x) > 0 \quad \forall x$. Tomando $x = 2\beta P|m_n| > 0$ y haciendo uso de las expresiones (3.22) y (3.35) se tiene que $f(2\beta P|m_n|) = \tilde{\lambda}_3 P|m_n|$ de donde fácilmente se tiene $\tilde{\lambda}_3 > 0$ sin más que tener en cuenta las propiedades de la función $f(x)$.

Demostremos a continuación (i). En primer lugar vamos a demostrar que $\tilde{\lambda}_2 < 0$. Al igual que antes consideramos una función auxiliar, concretamente $g(x) = x \exp(-x) - \{1 - \exp(-x)\}$. Como antes $g(x)$ es continua $\forall x \in \mathfrak{R}$ con $g(0) = 0$ y derivable $\forall x \in \mathfrak{R}$ con $g'(x) = -x \exp(-x) \quad \forall x \in \mathfrak{R}$. De lo anterior se deduce que la función $g(x)$ es estrictamente creciente si $x < 0$ y estrictamente decreciente si $x > 0$. Por lo tanto en $x = 0$ tiene un máximo absoluto. Tomando $x = 2\beta P|m_n| > 0$ se obtiene fácilmente que $g(2\beta P|m_n|) = \tilde{\lambda}_2 P|m_n|$, de donde se deduce trivialmente que $\tilde{\lambda}_2 < 0$, teniendo en cuenta las propiedades de la función $g(x)$ y el hecho de que en $x = 0$ dicha función tiene un máximo absoluto. Vamos a demostrar a continuación que $\tilde{\lambda}_1 < 0$. Utilizando la expresión para $\tilde{\lambda}_1$, ecuación (3.35), y utilizando (3.22) se llega a

$$\tilde{\lambda}_1 P |m_n| = h(2\beta P |m_n|) - (2\beta P |m_n|)^2 \exp(-2\beta P |m_n|) \left(\frac{1}{2} - \frac{Tn}{2P} \right), \quad (3.37)$$

donde $h(x) = \frac{1}{2}x^2 \exp(-x) + x \exp(-x) - \{1 - \exp(-x)\}$. Dado que las soluciones a (3.22) aparecen para $T < T_c = 1$ (cf. §3.3) y debido al hecho de que $n \leq P$, se tiene que $\tilde{\lambda}_1 P |m_n| \leq h(2\beta P |m_n|)$. Por otra parte, la función $h(x)$ es una función continua $\forall x \in \mathfrak{R}$ verificando que $h(0) = 0$, y derivable $\forall x \in \mathfrak{R}$ con $h'(x) = -\frac{1}{2}x^2 \exp(-x) \leq 0 \forall x \in \mathfrak{R}$. Por lo tanto, $h(x)$ es una función decreciente. Teniendo en cuenta que $h(0) = 0$ se tiene que $h(x) < 0$ si $x > 0$. Por lo tanto, se tiene que $\tilde{\lambda}_1 < 0$.†

En el caso de la *rate V*, la matriz \mathcal{Q} tiene los siguientes elementos de matriz

$$\mathcal{Q}_{\mu\nu} = \left\{ \beta \Xi_\mu^c - \sum_{\alpha=1}^P a_\alpha \Xi_\alpha^c \right\} \delta_{\mu\nu} - \beta m^\mu \Xi_\nu^s, \quad (3.38)$$

con

$$\Xi_\nu^{\left\{ \begin{smallmatrix} s \\ c \end{smallmatrix} \right\}} \equiv 2 \left\{ \begin{array}{c} \sinh \\ \cosh \end{array} \right\} (\beta a_\nu^{-1} m^\nu) \quad (3.39)$$

donde los solapamientos m^μ son solución a la ecuación (3.25). Como ocurre con el caso de *rate M*, la matriz \mathcal{Q} no tiene una forma diagonal sencilla y la dependencia de los elementos de matriz en las probabilidades a_μ dificulta en gran medida su diagonalización. Sin embargo, para $a_\mu = \frac{1}{P}$ se tiene que, tanto para el caso de estados mezcla simétricos (con $n \leq P$ solapamientos m_n no nulos) como para el caso de estados mezcla asimétricos (con k solapamientos tomando el valor $+m_n$ y $(n-k)$ solapamientos tomando el valor $-m_n$, con $m_n > 0$; cf. §3.3), la matriz \mathcal{Q} puede ser diagonalizada encontrándose los siguientes autovalores³

$$\begin{aligned} \tilde{\lambda}_1 &= -\frac{2}{P} \left[P - n + (n - \beta P) \tilde{\Xi}_n^c + n\beta P m_n \tilde{\Xi}_n^s \right] \\ \tilde{\lambda}_2 &= -\frac{2}{P} \left[P - n + (n - \beta P) \tilde{\Xi}_n^c \right] \\ \tilde{\lambda}_3 &= -\frac{2}{P} \left[P(1 - \beta) - n + n \tilde{\Xi}_n^c \right], \end{aligned} \quad (3.40)$$

con

$$\tilde{\Xi}_\nu^{\left\{ \begin{smallmatrix} s \\ c \end{smallmatrix} \right\}} \equiv \left\{ \begin{array}{c} \sinh \\ \cosh \end{array} \right\} (\beta P m_n) \quad (3.41)$$

³Para ver los detalles en la obtención de los autovalores $\tilde{\lambda}_i$, tanto para el caso de estados mezcla simétricos como asimétricos, puede verse el apéndice C.

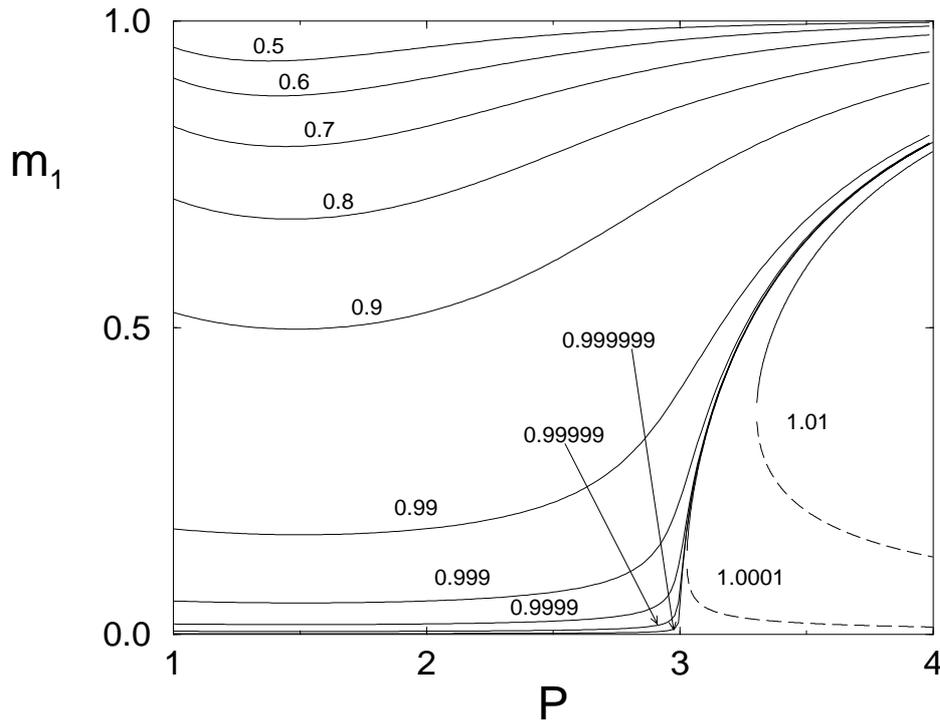


Figura 3.3: Isotermas $m_1 = m_1(P, T = \text{cte.})$ para $n = 1$ y diferentes valores de T , mostrando la presencia del punto tricrítico de no equilibrio ($T = T_c, P = 3$). Las líneas continuas representan estados estacionarios localmente estables, mientras que las líneas discontinuas representan estados estacionarios localmente inestables. La figura muestra resultados equivalentes a los que muestra la figura 3.1.

y donde los solapamientos m_n son solución a la ecuación (3.25), que aparecen con multiplicidad uno, $(n - 1)$ y $(P - n)$, respectivamente. Se puede demostrar que $\tilde{\lambda}_2 > 0$ (cf. propiedad 3.3). Por lo tanto, todos los estados mezcla con $n > 1$ son puntos de silla con un determinado número de direcciones de inestabilidad, independientemente del signo que tengan los otros dos autovalores. En particular, y a diferencia de lo que pasa en el caso de las *rates* K y M , los estados mezcla con $n = P$ solapamientos no nulos son localmente inestables. Sin embargo, para el caso de estados de Mattis, el que sea $\tilde{\lambda}_2 > 0$ no tiene relevancia alguna pues para esta situación es $n = 1$ y $\tilde{\lambda}_2$ no aparece al diagonalizar la matriz \mathcal{Q} . En cambio, sí tiene importancia, en el caso de estados de Mattis, el signo de los otros dos autovalores, es decir $\tilde{\lambda}_1$ y $\tilde{\lambda}_3$. En

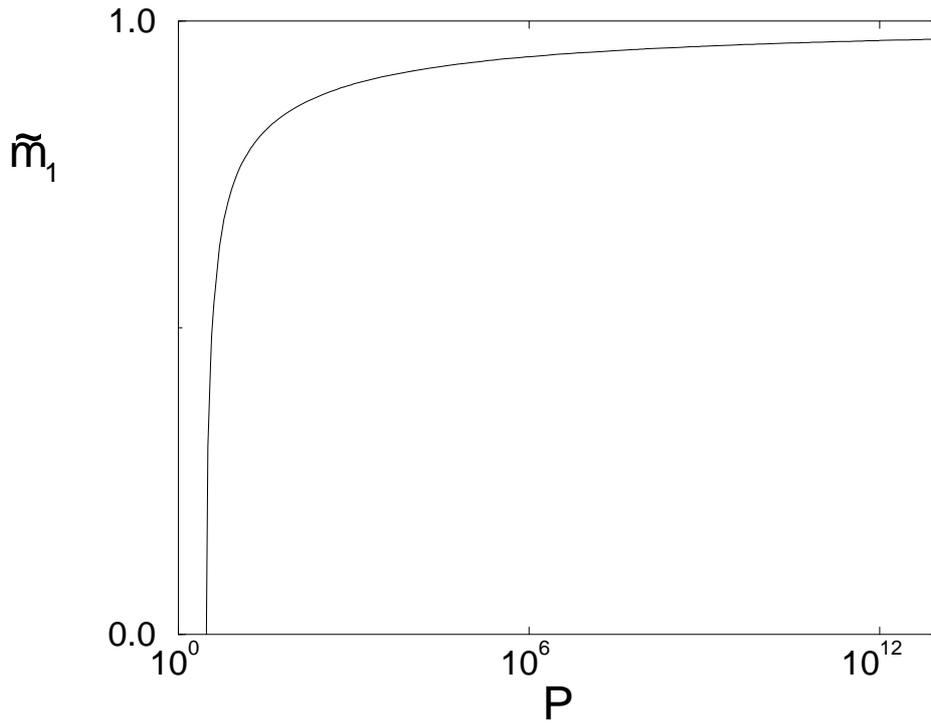


Figura 3.4: La figura muestra el comportamiento, en función del número de patrones almacenados P , del solapamiento \tilde{m}_1 . Se observa la tendencia hacia $\tilde{m}_1 = 1$ cuando $P \rightarrow \infty$.

general, como se demuestra en la propiedad 3.3, se tiene $\tilde{\lambda}_3 < 0$; en cuanto al signo de $\tilde{\lambda}_1$, se tiene que $\tilde{\lambda}_1 < 0$ para estados mezcla tales que $n > \frac{1}{3}P$ (cf. propiedad 3.3), que son estados que aparecen de forma continua a una temperatura justo por debajo de $T_c = 1$; en particular, los estados de Mattis con $P < 3$ son localmente estables y aparecen de forma continua, como en un cambio de fase de segundo orden, justo por debajo de $T_c = 1$; para estados mezcla tales que $n < \frac{1}{3}P$, $\tilde{\lambda}_1$ cambia de signo a la temperatura $\tilde{T}(n, P) = \tilde{m}_n P \theta_n^{-1}$, donde θ_n^{-1} y \tilde{m}_n están definidos en (3.28) y (3.29) respectivamente (cf. §3.3), temperatura que da cuenta de un cambio de fase de primer orden en el que el parámetro m_n se comporta de forma discontinua; así, $\tilde{\lambda}_1 < 0$ para $T < \tilde{T}(n, P)$ y $\tilde{\lambda}_1 > 0$ para $T > \tilde{T}(n, P)$ (cf. propiedad 3.3). Por lo tanto, sólo los estados de Mattis con $P < 3$ y $T < T_c$ y los estados de Mattis con $P > 3$ y $T < \tilde{T}(1, P)$ son localmente estables. El punto $(T = T_c, P = 3)$ se corresponde a

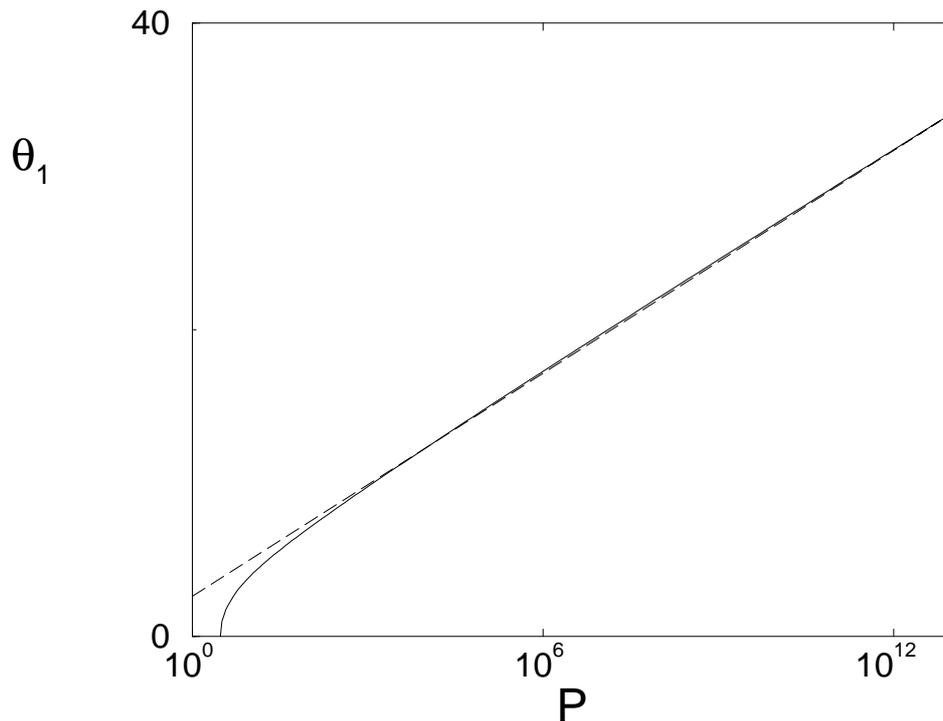


Figura 3.5: Comportamiento de θ_1 , es decir la solución a la ecuación (3.28) para el caso de estados de Mattis, en función de número total de patrones almacenados P (línea continua). El comportamiento para P grande ($P \in [10^4, 10^{13}]$) ha sido ajustado obteniendo $\theta_1 \approx 2.663 + 1.051 \ln P$ (línea discontinua).

un punto tricrítico fuera del equilibrio, que separa un régimen en el que los cambios de fase que tienen lugar son de segundo orden o continuos, de un régimen en el que dichos cambios de fase son discontinuos; cf. figuras 3.2 y 3.3. El significado físico de \tilde{m}_1 es el valor del solapamiento, que tiene el sistema con un patrón dado, justo en el punto donde tiene lugar el cambio de fase de primer orden. Su comportamiento en función del número de patrones almacenados se muestra en la figura 3.4, observándose la tendencia de que $\tilde{m}_1 \rightarrow 1$ cuando $P \rightarrow \infty$. En la figura 3.5 se muestra, asimismo, el comportamiento de θ_1 en función de P y un ajuste, en la forma $2.663 + 1.051 \ln P$, de dicho comportamiento para valores de P comprendidos en el intervalo $[10^4, 10^{13}]$. La gráfica muestra el comportamiento —cuasi-lineal— de θ_1 con $\ln P$. La situación descrita por las figuras 3.4 y 3.5 parece confirmar nuestra creencia de que la mayoría

de los resultados analíticos que se presentan, se cumplen en el régimen asintótico para $P \rightarrow \infty$.

Propiedad 3.3 Sea \mathcal{Q} la matriz asociada al sistema dinámico descrito por el sistema de ecuaciones (3.19) particularizado al caso de rate V , cuyos elementos vienen dados por (3.38) y cuyos autovalores vienen dados por (3.40) para el caso en que $a_\mu = \frac{1}{P} \quad \forall \mu = 1, \dots, P$. Entonces, los estados estacionarios (no triviales) solución a (3.20) tienen asociados autovalores (que dan cuenta de la estabilidad dinámica de dichas soluciones) que verifican:

(i) (a) $\tilde{\lambda}_1 < 0$ para estados mezcla tales que $3n > P$ y $T < T_c = 1$. (b) $\tilde{\lambda}_1 < 0$ para estados mezcla verificando $3n < P$ y $T < \tilde{T}(n, P) \equiv \tilde{m}_n P \theta_n^{-1}$ (con θ_n y \tilde{m}_n verificando (3.28), y (3.29) respectivamente) y $\tilde{\lambda}_1 > 0$ para estados mezcla con $3n < P$ y $T > \tilde{T}(n, P)$.

(ii) $\tilde{\lambda}_2 > 0$.

(iii) $\tilde{\lambda}_3 < 0$.

Demostración:

Demostremos la primera parte de (i): Consideremos la función auxiliar

$$h(x) \equiv x \cosh x - \sinh x - x \sinh^2 x (2 + \cosh x)^{-1}. \quad (3.42)$$

La función $h(x)$ es continua $\forall x \in \Re$ con $h(0) = 0$; también es derivable $\forall x \in \Re$ con

$$h'(x) = -\frac{1}{2} \sinh x (4 \sinh x + \sinh 2x - 6x) (2 + \cosh x)^{-2}. \quad (3.43)$$

Trivialmente se demuestra que $h'(x) < 0 \quad \forall x \in \Re$. Por lo tanto $h(x)$ es estrictamente decreciente y entonces se tiene que $h(x) < 0$ si $x > 0$. Por otra parte, utilizando las expresiones (3.40) y (3.20) y haciendo uso de que $3n > P$ obtenemos que $\tilde{\lambda}_1 < 2(P|m_n|)^{-1}h(\beta P|m_n|)$, que implica $\tilde{\lambda}_1 < 0$ para $m_n \neq 0$ al ser $h(x)$ decreciente.

De igual forma se demuestra la segunda parte de (i): Consideremos la función

$$g(x, \gamma) \equiv x - \sinh x \cosh x + \gamma(x \cosh x - \sinh x), \quad (3.44)$$

con $\gamma > 0$. Trivialmente se demuestra que $g(x, \gamma)$ es continua y derivable $\forall x \in \Re$, que $g(0, \gamma) = 0$, que $\lim_{x \rightarrow +\infty} g(x, \gamma) = -\infty$, y que $g(x, \gamma) = \frac{1}{3}(\gamma - 2)x^3 + \mathcal{O}(x^5)$

cuando $x \rightarrow 0$. Consideremos el intervalo abierto $I = (0, \infty)$; si $\gamma > 2$ y utilizando las propiedades anteriores de la función $g(x, \gamma)$ entonces siempre es posible encontrar dos números reales, $0 < a < b$ tales que $g(b, \gamma) < 0 < g(a, \gamma)$; utilizando el Teorema de los Ceros de Bolzano [47] se tiene que $\exists \theta \in [a, b]$ tal que $g(\theta, \gamma) = 0$. Vamos a ver que esta solución es única. Supongamos que hay dos soluciones $\theta_1 \neq \theta_2 \in [a, b]$ tales que $g(\theta_1, \gamma) = g(\theta_2, \gamma) = 0$. Utilizando esta última condición se llega a que $\gamma = f(\theta_1) = f(\theta_2)$ con $f(x) = (\sinh x \cosh x - x)(x \cosh x - \sinh x)^{-1}$, función que es continua, derivable y estrictamente creciente en I como se puede demostrar trivialmente. Esto último está en contradicción con la hipótesis de que $\theta_1 \neq \theta_2$, por lo tanto concluimos que $\theta_1 = \theta_2$ y queda demostrada la unicidad de θ en I ; como consecuencia de esto se tiene que si $x < \theta$ es $g(x, \gamma) > 0$ y si $x > \theta$ es $g(x, \gamma) < 0$. Por otra parte, utilizando (3.40) y (3.20) se obtiene que

$$\tilde{\lambda}_1 = 2n(P|m_n|)^{-1}\{n \cosh(\beta P|m_n|) + P - n\}^{-1}g(\beta P|m_n|, Pn^{-1} - 1), \quad (3.45)$$

con $m_n = \sinh(\beta P m_n)\{n \cosh(\beta P m_n) + P - n\}^{-1}$, de forma que aplicando las propiedades de la función $g(x, \gamma)$ demostradas antes y el hecho de que $m_n \neq 0$ (pues estamos considerando soluciones no triviales) queda demostrada la segunda parte de (i). Hay que hacer notar dos cosas: en primer lugar que la condición de que $\gamma \equiv Pn^{-1} - 1 > 2$ queda garantizada por el hecho de que $P > 3n$, y en segundo lugar que la aplicación del teorema de Bolzano a la función $g(x, \gamma)$ implica la existencia y unicidad de $\tilde{T}(n, P) \equiv \tilde{m}_n P \theta_n^{-1} > 0$ (temperatura dada en unidades k_B^{-1}), donde $\tilde{m}_n = \sinh \theta_n (n \cosh \theta_n + P - n)^{-1}$, con θ_n verificando $g(\theta_n, Pn^{-1} - 1) = 0$ (expresión (3.28)).

Vamos a demostrar a continuación (ii). Consideremos la función auxiliar $u(x) = \sinh x - x \cosh x$, que es continua $\forall x \in \mathfrak{R}$ y que verifica $u(0) = 0$. Además $u(x)$ es derivable $\forall x \in \mathfrak{R}$ con $u'(x) = -x \sinh x < 0 \forall x \in \mathfrak{R}$ por lo que $u(x)$ es decreciente. Utilizando (3.40) y (3.20) se tiene que $\tilde{\lambda}_2 = -2(P|m_n|)^{-1}u(\beta P|m_n|)$, con $m_n = \sinh(\beta P m_n)\{n \cosh(\beta P m_n) + P - n\}^{-1}$, de forma que (ii) queda demostrado sin más que utilizar la propiedad de decrecimiento de $u(x)$ y el hecho de que $m_n \neq 0$.

Por último demostramos (iii). Consideremos la función $v(x) = \sinh x - x$. Trivialmente se tiene que $v(x)$ es continua $\forall x \in \mathfrak{R}$ con $v(0) = 0$; también $v(x)$ es derivable

$\forall x \in \mathfrak{R}$, con $v'(x) = \cosh x - 1 > 0 \quad \forall x \neq 0$ por lo que es estrictamente creciente $\forall x \neq 0$. Utilizando (3.40) y (3.20) se tiene que $\tilde{\lambda}_3 = -2(P|m_n|)^{-1}v(\beta P|m_n|)$, con $m_n = \sinh(\beta P m_n) \{n \cosh(\beta P m_n) + P - n\}^{-1}$, de forma que (iii) queda demostrado utilizando la propiedad de crecimiento de $v(x)$ y el hecho de que $m_n \neq 0$. †

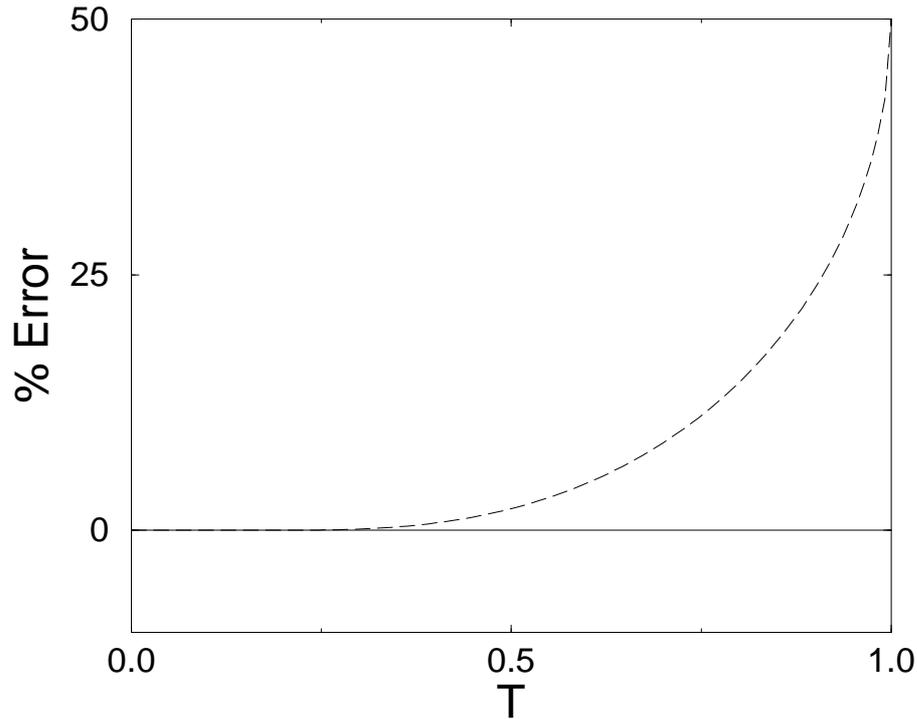


Figura 3.6: La figura muestra el porcentaje de error, definido como $\%Error \equiv (1 - m_1)/2$, en el proceso de recuperación de los estados de Mattis (*rate V*), en función de la temperatura (en unidades k_B^{-1}), con $P = 100$ (línea continua), comparado con el correspondiente resultado del modelo de Hopfield (línea a trazos).

En resumen, y para finalizar esta sección se concluye que sólo en el caso de que el sistema esté realizado con la *rate V* existen estados de Mattis localmente estables y, por lo tanto, el sistema exhibe la propiedad de memoria asociativa. En este caso, a diferencia de lo que ocurre en el modelo de Hopfield, y debido a la aparición de los estados de Mattis de forma discontinua justo por debajo de $\tilde{T}(1, P)$, los solapamientos con el correspondiente patrón pueden ser grandes (ceranos a +1 o -1) incluso para

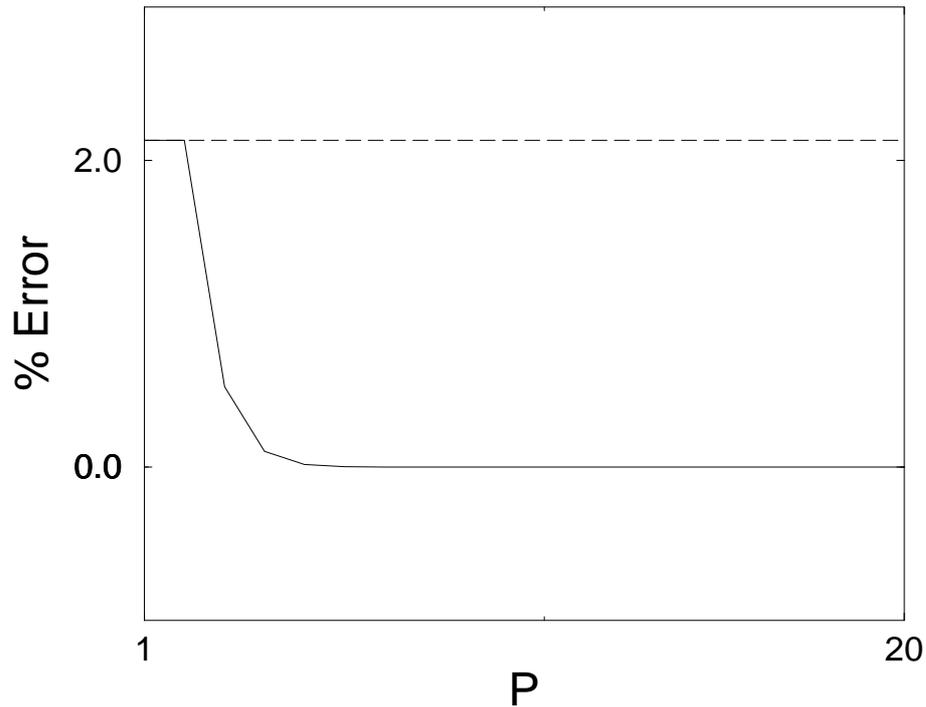


Figura 3.7: Porcentaje de error en el proceso de recuperación de los estados de Mattis (*rate V*) en función de P , para $T = 0.5$ (línea continua), comparado con el correspondiente resultado del modelo de Hopfield (línea a trazos).

valores de T y P grandes, por lo que los errores a la hora de recuperar ese patrón son sistemáticamente más pequeños que los correspondientes al modelo de Hopfield, como se puede apreciar en las figuras 3.6 y 3.7. Además los estados mezcla son localmente inestables lo que favorece la propiedad de memoria asociativa. Por el contrario, para el caso en que el sistema está realizado con las *rates K* y *M* no existen estados de Mattis localmente estables y sólo aquellos estados mezcla con solapamientos no nulos con todos los patrones son localmente estables. Por lo tanto, en estos dos últimos casos el sistema no posee la propiedad de memoria asociativa. Una conclusión definitiva que se obtiene es que el tipo de ruido sináptico considerado aquí optimiza significativamente, para ciertos casos particulares, el comportamiento del sistema como un mecanismo para la memoria asociativa.

3.5 Sinapsis asimétricas

La teoría desarrollada en las secciones precedentes para el caso de fluctuaciones alrededor de la regla de Hebb puede ser generalizada al estudio de fluctuaciones en torno a reglas de aprendizaje más realistas, como por ejemplo reglas de aprendizaje asimétricas. Como ilustración de esto consideramos a continuación distribuciones $f(\mathbf{J})$ como la dada en (3.4) con $\eta_{xy}^\mu \neq \eta_{yx}^\mu$. Una situación relativamente general consistente con esta simetría consiste en considerar $\eta_{xy}^\mu = (a_\mu N)^{-1} (A_1 \xi_x^\mu \xi_y^\mu + A_2 \xi_x^\mu + A_3 \xi_y^\mu + A_4)$, que induce valores medios:

$$\bar{J}_{x,y} = \frac{1}{N} \sum_{\mu=1}^P (A_1 \xi_x^\mu \xi_y^\mu + A_2 \xi_x^\mu + A_3 \xi_y^\mu + A_4). \quad (3.46)$$

Los parámetros A_2 y A_3 determinan el grado de asimetría. A partir de (3.8) y (3.9) se obtiene que

$$\partial_t \langle s_{\mathbf{x}} \rangle = -2 \left\langle s_{\mathbf{x}} \sum_{\mu=1}^P a_\mu \mathcal{C}^\mu(\mathbf{s}) \right\rangle - 2 \left\langle \sum_{\mu=1}^P a_\mu \{ \xi_{\mathbf{x}}^\mu + \mathcal{A}^\mu(\mathbf{s}) \} \mathcal{B}^\mu(\mathbf{s}) \right\rangle, \quad (3.47)$$

donde

$$\begin{aligned} \mathcal{A}^\mu(\mathbf{s}) &= [A_3 m^\mu(\mathbf{s}) + A_4 m(\mathbf{s})] [A_1 m^\mu(\mathbf{s}) + A_2 m(\mathbf{s})]^{-1}, \\ \mathcal{B}^\mu(\mathbf{s}) &= \frac{1}{2} [\Psi(X_\mu^+(\mathbf{s})) - \Psi(-X_\mu^-(\mathbf{s}))], \\ \mathcal{C}^\mu(\mathbf{s}) &= \frac{1}{2} [\{1 - \mathcal{A}^\mu(\mathbf{s})\} \Psi(X_\mu^+(\mathbf{s})) + \{1 + \mathcal{A}^\mu(\mathbf{s})\} \Psi(-X_\mu^-(\mathbf{s}))], \end{aligned} \quad (3.48)$$

con

$$X_\mu^\pm(\mathbf{s}) \equiv 2\beta a_\mu^{-1} [(A_1 \pm A_3) m^\mu(\mathbf{s}) + (A_2 \pm A_4) m(\mathbf{s})]. \quad (3.49)$$

Aquí $m^\mu(\mathbf{s})$, como ya hemos definido antes, es el solapamiento del sistema con un patrón dado ξ^μ , y $m(\mathbf{s}) \equiv N^{-1} \sum_{\mathbf{x}} s_{\mathbf{x}}$ es la actividad promedio en la red. Bajo la condición de campo medio $s_{\mathbf{x}} = \langle s_{\mathbf{x}} \rangle$ la ecuación (3.47) se transforma en el sistema de ecuaciones diferenciales acopladas siguiente

$$\begin{aligned} \partial_t m^\mu &= -2m^\mu \sum_{\nu=1}^P a_\nu \mathcal{C}^\nu - 2 \sum_{\nu=1}^P a_\nu \mathcal{B}^\nu \frac{1}{N} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\mu \xi_{\mathbf{x}}^\nu \\ &\quad - 2 \sum_{\nu=1}^P a_\nu \mathcal{A}^\nu \mathcal{B}^\nu \frac{1}{N} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\mu, \\ \partial_t m &= -2m \sum_{\nu=1}^P a_\nu \mathcal{C}^\nu - 2 \sum_{\nu=1}^P a_\nu \mathcal{B}^\nu \frac{1}{N} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\nu - 2 \sum_{\nu=1}^P a_\nu \mathcal{A}^\nu \mathcal{B}^\nu, \end{aligned} \quad (3.50)$$

donde m , m^μ , \mathcal{A}^μ , \mathcal{B}^μ y \mathcal{C}^μ se obtienen a partir de $m(\mathbf{s})$, $m^\mu(\mathbf{s})$, $\mathcal{A}^\mu(\mathbf{s})$, $\mathcal{B}^\mu(\mathbf{s})$ y $\mathcal{C}^\mu(\mathbf{s})$, con $s_{\mathbf{x}} = \langle s_{\mathbf{x}} \rangle$. Si los patrones son tales que $N^{-1} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\mu \xi_{\mathbf{x}}^\nu = \delta_{\mu\nu}$ y $N^{-1} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\nu = 0$, es decir patrones ortogonales que tienen el mismo número de componentes $+1$ que -1 , o patrones tales que $N^{-1} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\mu \xi_{\mathbf{x}}^\nu = \delta_{\mu\nu} + \mathcal{O}(N^{-\frac{1}{2}})$ y $N^{-1} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\nu = \mathcal{O}(N^{-\frac{1}{2}})$, en $N \rightarrow \infty$, es decir patrones cuasi-ortogonales, el sistema de ecuaciones (3.50) queda de la forma:

$$\begin{aligned} \partial_t m^\mu &= -2m^\mu \sum_{\nu=1}^P a_\nu \mathcal{C}^\nu - 2a_\mu \mathcal{B}^\mu \\ \partial_t m &= -2m \sum_{\nu=1}^P a_\nu \mathcal{C}^\nu - 2 \sum_{\nu=1}^P a_\nu \mathcal{A}^\nu \mathcal{B}^\nu, \end{aligned} \quad (3.51)$$

cuya solución estacionaria es

$$m^\mu = -\frac{a_\mu \mathcal{B}^\mu}{\sum_{\nu=1}^P a_\nu \mathcal{C}^\nu}; \quad m = -\frac{\sum_{\nu=1}^P a_\nu \mathcal{A}^\nu \mathcal{B}^\nu}{\sum_{\nu=1}^P a_\nu \mathcal{C}^\nu}. \quad (3.52)$$

sistema de ecuaciones que puede ser resuelto de forma numérica y cuya solución nos da, para una elección particular de los parámetros A_i $i = 1, \dots, 4$, el comportamiento estacionario de m^μ y m en función de T y P .

El sistema (3.51) se puede poner de forma más compacta definiendo el vector de dimensión $(P+1)$ $\mathbf{M} \equiv (M_0, \dots, M_P)$ con $M_0 = m$, $M_i = m^i$, $i = 1, \dots, P$. Así (3.51) queda de la forma

$$\partial_t \mathbf{M} = -2\mathbf{M} \sum_{\nu=1}^P a_\nu \mathcal{C}^\nu - 2\mathbf{N}, \quad (3.53)$$

donde \mathbf{N} es un vector de dimensión $(P+1)$ y componentes $N_0 = \sum_{\nu=1}^P a_\nu \mathcal{C}^\nu \mathcal{B}^\nu$ y $N_i = a_i \mathcal{B}^i$, $i = 1, \dots, P$.

Haciendo un desarrollo semejante al de la sección precedente, es posible hacer un estudio de la estabilidad de la solución estacionaria al sistema (3.53). Linealizando dicho sistema como antes (cf. §3.4), obtenemos la matriz cuyos autovalores nos dan la información sobre la estabilidad local del sistema y cuya forma es

$$\mathcal{Q}_{ij} = -2 \sum_{k=1}^P a_k \left\{ \delta_{ij} \mathcal{C}^k + M_i \mathcal{D}_{kj}^+ + \delta_{ik} \mathcal{D}_{kj}^- + (\delta_{ik} - M_i) (\mathcal{A}^k \mathcal{D}_{kj}^- + \mathcal{B}^k \mathcal{G}_{kj}) \right\}, \quad (3.54)$$

donde

$$\begin{aligned} \mathcal{D}_{ij}^\pm &= \beta a_i^{-1} \left[\dot{\Psi}(X_i^+) \{ (A_1 + A_3) \delta_{ij} + (A_2 + A_4) \delta_{0j} \} \right. \\ &\quad \left. \pm \dot{\Psi}(-X_i^-) \{ (A_3 - A_1) \delta_{ij} + (A_4 - A_2) \delta_{0j} \} \right] \end{aligned} \quad (3.55)$$

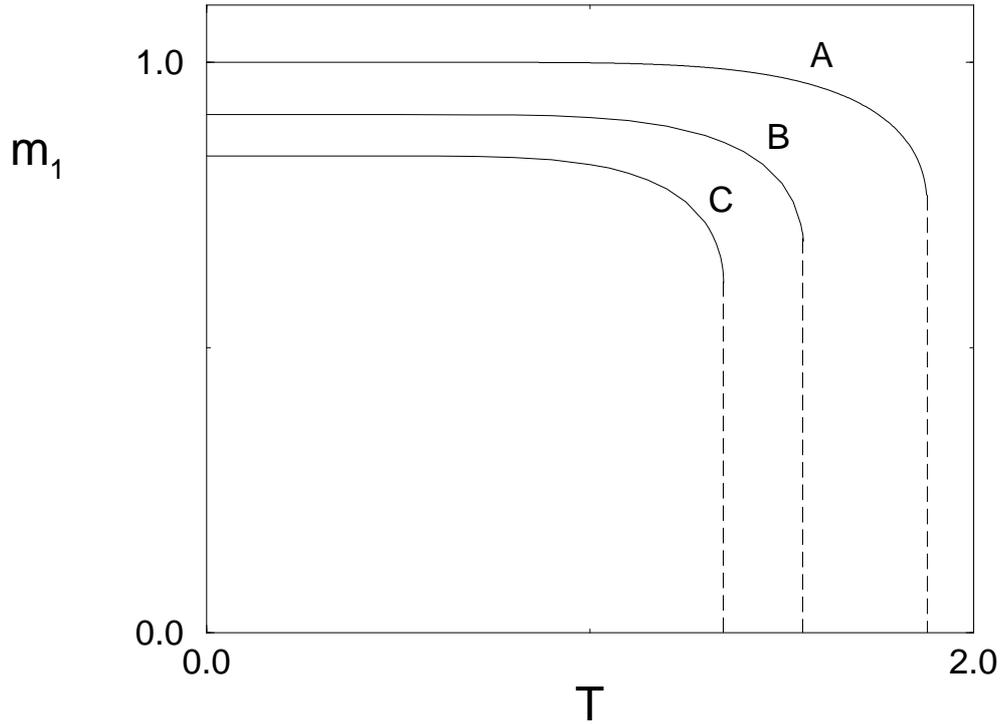


Figura 3.8: La figura muestra diferentes estados de Mattis (con un solapamiento no nulo m_1 con un determinado patrón y el resto de los solapamientos todos nulos) localmente estables en función de la temperatura T (unidades k_B^{-1}), que se obtienen al resolver el sistema de ecuaciones (3.52) para diferentes elecciones de los parámetros A_i y $P = 10$. (A) Corresponde a tomar $A_1 = 1$, $A_2 = A_3 = A_4 = 0$. (B) Corresponde a elegir $A_1 = 1$, $A_2 = -0.1$, $A_3 = 0.1$, $A_4 = 0$. (C) Corresponde a elegir $A_1 = 1$, $A_2 = 0.1$, $A_3 = 0.2$, $A_4 = 0$.

y

$$\mathcal{G}_{ij} = [A_1 A_4 - A_2 A_3] [M_i \delta_{0j} - M_0 \delta_{ij}] [A_1 M_i + A_2 M_0]^{-2}, \quad (3.56)$$

con $\dot{\Psi}(\pm X_i^\pm) \equiv [\partial \Psi(X) / \partial X]_{X=\pm X_i^\pm}$, y donde X_i^\pm se obtiene a partir (3.49) con $s_{\mathbf{x}} = \langle s_{\mathbf{x}} \rangle$.

Las ecuaciones (3.52) han sido resueltas de forma numérica y se ha estudiado la estabilidad de las soluciones mediante el análisis de los autovalores de la matriz \mathcal{Q} , obteniéndose los resultados que a continuación detallamos: Para ciertas elecciones particulares de los parámetros A_i $i = 1, \dots, 4$, y al igual que ocurre en el caso de fluctua-

ciones alrededor de reglas de aprendizaje simétricas como las situaciones estudiadas en las secciones precedentes, sólo para el caso de la *rate* V se encuentran estados de Mattis localmente estables, siendo los estados mezcla inestables localmente. Para el caso de las *rates* K y M , al igual que ocurre para el caso de simetría, no existen estados de Mattis localmente estables y sólo aquellos estados mezcla con $n = P$, es decir estados que tienen un solapamiento no nulo con todos los patrones, son localmente estables. De estos resultados se concluye que la propiedad de memoria asociativa, y por lo tanto la capacidad para recordar información almacenada por la red, sólo ocurre cuando el sistema es construido con la *rate* V , tanto para los casos de simetría como los casos de asimetría considerados aquí. En el caso particular de fluctuaciones alrededor de reglas asimétricas, como las dadas por (3.46), en las que $A_1 \neq 0$, $A_2 \neq 0$ y $A_3 = A_4 = 0$, se tiene que $\mathcal{A}^\mu = 0$ y, por lo tanto (3.51) se reduce a (3.19) (para patrones ortogonales o quasi-ortogonales). Así para esta situación, el sistema tiene un comportamiento equivalente a cuando consideramos fluctuaciones alrededor de la regla de Hebb. Para cualquier otra elección el sistema se comporta de forma sensiblemente diferente a la hora de recuperar un determinado patrón. En la figura 3.8 mostramos los estados de Mattis, que se obtienen para diferentes elecciones particulares de los parámetros A_i , en función de T y para el caso particular de $P = 10$ y la *rate* V . Los resultados están comparados con aquellos que resultan en el caso de fluctuaciones entorno a la regla de Hebb. Se observa que el principal efecto de la asimetría consiste en la pérdida de efectividad a la hora de recuperar un determinado patrón. Como en el caso de simetría los estados aparecen de forma discontinua justo por debajo de una temperatura que da cuenta de un cambio de fase de primer orden.

3.6 Conclusiones

La idea de que la redundancia y la robusted son esenciales para el buen funcionamiento de las redes de neuronas biológicas (a diferencia de lo que ocurre en los ordenadores digitales) data de von Neumann [48]. Sin embargo, no se comprende bien todavía como dichas propiedades son desarrolladas por la red y cual es la relación precisa entre las mismas y los diferentes tipos de *ruido* que parecen tener lugar en las redes biológicas, así como su influencia en las propiedades emergentes del sistema.

En el capítulo primero hemos descrito diferentes causas que hacen que las sinapsis individuales sean extremadamente poco fiables, pero que en conjunto parecen aumentar la potencia de procesamiento del sistema; cf. por ejemplo Refs. [4–18]. El modelo de Hopfield constituye un excelente marco en el que podemos comprobar las consecuencias que se obtienen de considerar diferentes tipos de ruido y fluctuaciones en la red. Una característica principal de este modelo es que la transmisión cooperativa compitiendo con el ruido térmico tiene como consecuencia, bajo determinadas circunstancias, la propiedad de memoria asociativa. En este capítulo hemos mostrado como el modelo de Hopfield puede ser adaptado a una especie de filtro, que mejora sensiblemente la propiedad de memoria asociativa, cuando se considera un tipo de ruido, diferente del ruido térmico, que parece tener lugar en los sistemas biológicos. Este ruido, cuyos efectos hemos investigado, consiste en fluctuaciones correlacionadas de las intensidades sinápticas que ocurren en una escala de tiempo menor que la de evolución de las actividades neuronales.

El modelo estudiado en este capítulo es una versión del modelo cinético, más general, de red de neuronas tipo Hopfield introducido en el capítulo segundo, en el que las sinapsis sufren fluctuaciones aleatorias rápidas además del proceso (relativamente lento) de plasticidad sináptica debido al aprendizaje. El modelo ha sido construido con una competición de dinámicas, entre variaciones de las neuronas y de las sinapsis, que lleva al sistema, asintóticamente, a un estado estacionario fuera del equilibrio. Las fluctuaciones en las intensidades sinápticas tienen una distribución $f(\mathbf{J})$ dada por (3.4), que respeta las correlaciones espaciales que caracterizan los patrones memorizados. Esta distribución induce en el sistema un comportamiento fuera del equilibrio completo, como realizado por un agente no hamiltoniano (a diferencia de las versiones del mismo modelo consideradas en el capítulo segundo en las que es posible hacer una descripción hamiltoniana).

El modelo ha sido resuelto analíticamente considerando una condición de campo medio correspondiente a interacciones de largo alcance, con $N \rightarrow \infty$ y P finito. Sin embargo, los resultados obtenidos para valores de P grande indican con gran evidencia que las conclusiones que se presentan no se ven afectadas por tales restricciones, y muestran, claramente, cual va a ser el comportamiento del sistema para $P \rightarrow \infty$. El

estado estacionario resultante depende fuertemente, es decir cualitativamente, de los detalles de la dinámica microscópica *efectiva* que hace evolucionar las actividades neuronales, ecuación (3.2). En el caso en el que hemos considerado fluctuaciones alrededor de la regla de Hebb (cf. §3.3) se tienen las principales propiedades que a continuación se detallan.

Si la dinámica efectiva (3.2) está construida con la *rate* K entonces el único estado localmente estable es aquel estado mezcla que tiene solapamientos no nulos con cada uno de los P patrones que el sistema puede almacenar, es decir $n = P$. Todos los demás estados mezcla, con $n < P$ son localmente puntos de silla, con tantas direcciones de inestabilidad como solapamientos nulos tenga dicho estado. Así estados con $n = 1$, es decir estados de Mattis, son puntos de silla con $(P - 1)$ direcciones de inestabilidad. Como consecuencia de lo anterior, el sistema no exhibe la propiedad de memoria asociativa, y el estado estacionario del sistema es un estado mezcla que tiene solapamientos no nulos con cada uno de los P patrones que puede memorizar.

Para la *rate* M ocurre algo similar a lo que ocurre para la *rate* K , encontrando que, para $a_\mu = \frac{1}{P} \quad \forall \mu = 1, \dots, P$, los únicos estados localmente estables son aquellos que tienen un solapamiento no nulo con todos los patrones; todos los demás estados mezcla con $n < P$ son puntos de silla con tantas direcciones de inestabilidad como solapamientos nulos tenga dicho estado. Por lo tanto, al igual que ocurre para el caso de la *rate* K , el sistema no exhibe tampoco la propiedad de memoria asociativa, ya que los estados de Mattis, $n = 1$, son puntos de silla con $(P - 1)$ direcciones de inestabilidad.

Finalmente, en el caso de *rate* V se obtienen resultados más interesantes que en las anteriores situaciones. Así, al igual que ocurre en el modelo de Hopfield, aparecen estados de Mattis localmente estables, por lo que el sistema exhibe la propiedad de memoria asociativa. Los estados de Mattis aparecen para $T < \tilde{T}(1, P)$ Para $P < 3$ aparecen de forma continua justo por debajo de la temperatura $\tilde{T}(1, P < 3) = T_c$; cf. figura 3.9. Para $P \geq 3$ los estados de Mattis aparecen de forma discontinua justo por debajo de la temperatura de transición $\tilde{T}(1, P \geq 3)$. Para valores de T suficientemente grandes, la única solución que existe es la trivial de forma que T_c y $\tilde{T}(1, P \geq 3)$ corresponden a un cambio de fase (fuera del equilibrio) de segundo y primer orden,

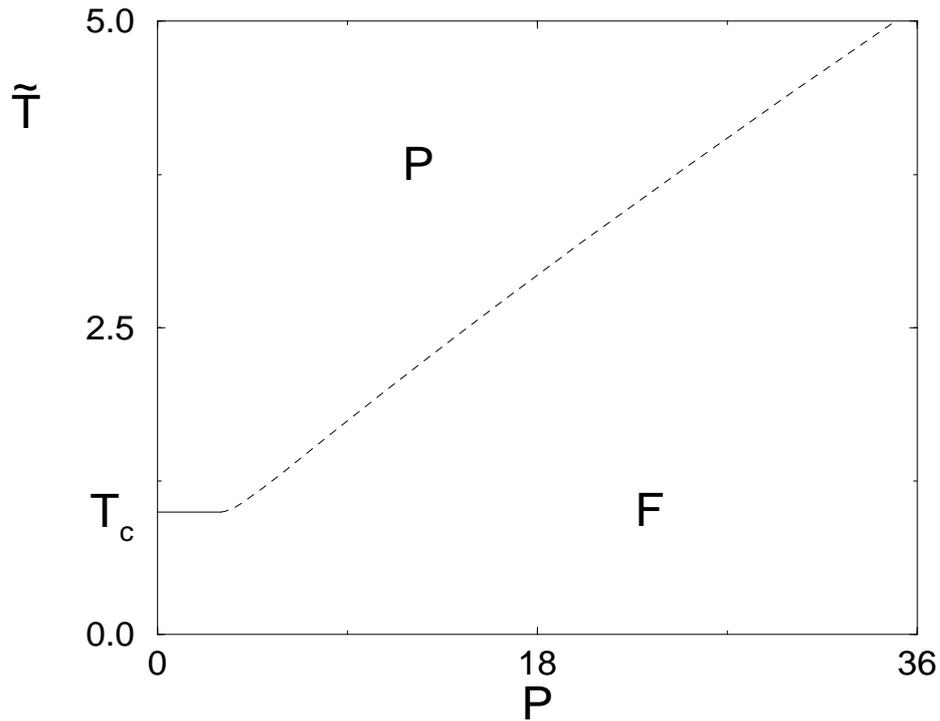


Figura 3.9: Diagrama de fases para $n = 1$ que caracteriza la ecuación (3.20) para *rate* V y $f(\mathbf{J})$ dada por (3.4). La línea $\tilde{T}(1, P)$ indica cambios de fase (fuera del equilibrio) de segundo (línea continua) y de primer orden (línea a trazos).

respectivamente; estos cambios de fase ocurren entre una fase paramagnética sin la propiedad de memoria asociativa, a temperaturas relativamente altas, y una fase ferromagnética a temperaturas relativamente bajas que sí exhibe dicha propiedad. Otra propiedad interesante es que la discontinuidad en $\tilde{T}(1, P \geq 3)$ ocurre para valores grandes (cerca de $+1$ o -1) del solapamiento del sistema con el patrón deseado para el caso, por otra parte más interesante, de valores de P grandes. A diferencia de lo que ocurre en el modelo de Hopfield ordinario, nuestro modelo no muestra por debajo de $T = 0.46$ estados mezcla localmente estables, que afectan de forma importante a la realización de la red de Hopfield a bajas temperaturas, impidiendo la propiedad de memoria asociativa. En nuestro caso todos los estados mezcla son, localmente, puntos de silla con $(n - 1)$ direcciones de inestabilidad. Otra diferencia importante que presenta nuestro modelo, comparado con el modelo de Hopfield, es que el error

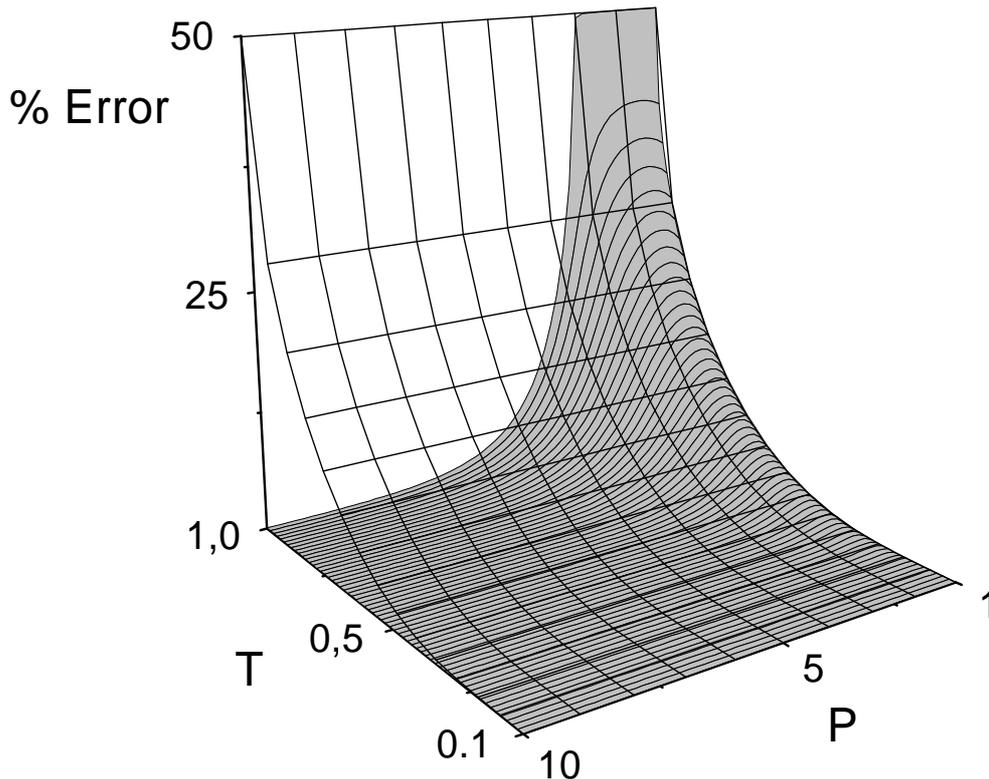


Figura 3.10: El porcentaje de error —para el sistema definido por (3.4) y *rate* V —, que es una medida de $(1 - m_1)/2$, durante el proceso de recuerdo para los estados de Mattis para $T < \bar{T}$ (superficie sombreada), mostrando que el error sólo es grande para valores de P pequeños cerca de T_c , y el correspondiente resultado para el modelo de Hopfield (entramado).

cometido cuando la red intenta recuperar un determinado patrón, por ejemplo el ξ^1 , definido como $\%Error \equiv \frac{1-m_1}{2}$, disminuye rápidamente hacia cero, en nuestro modelo, cuando el número de patrones almacenados P aumenta, para una determinada T , y permanece pequeño, para P dado, aun cuando $T (< T_c)$ se incrementa hasta $T = T_c$, tendiendo a cero para valores de P suficientemente grandes. Este comportamiento se ilustra en las figuras 3.10 y 3.11; en la figura 3.10 se muestra una comparación con el modelo de Hopfield donde el mismo error, definido más arriba, es independiente de P para una temperatura T dada, y se hace muy grande ($\sim 50\%$) cuando $T \rightarrow 1$.

Fluctuaciones alrededor de reglas de aprendizaje más generales han sido también consideradas. Concretamente se ha hecho un estudio analítico de campo medio para

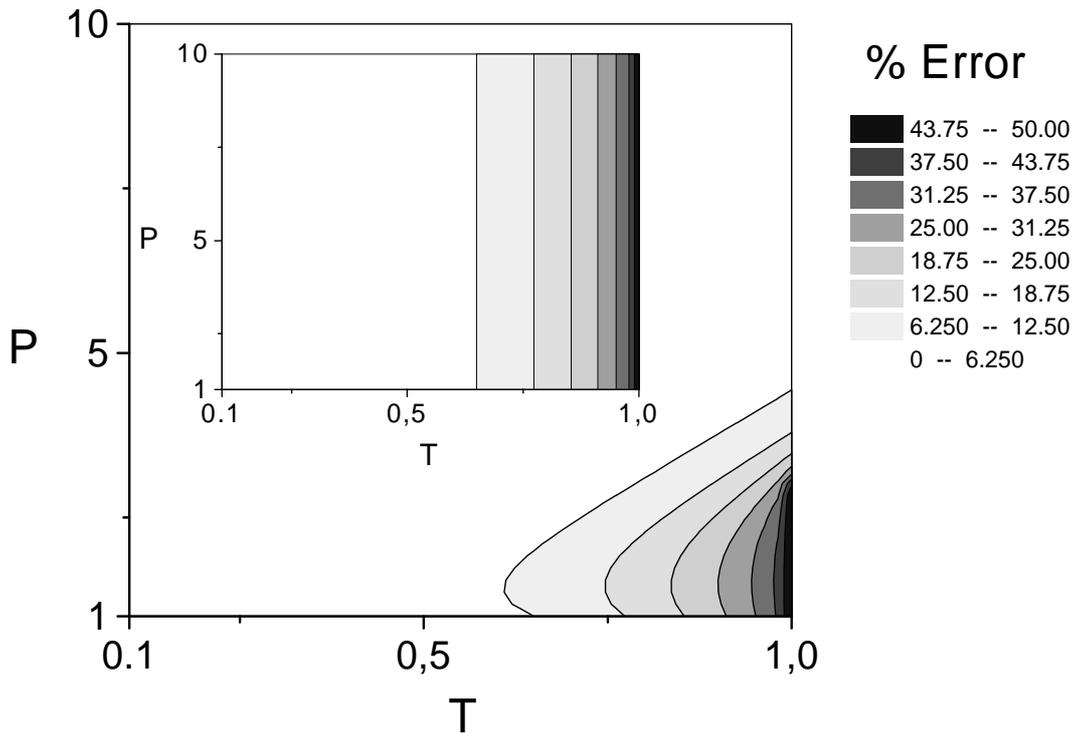


Figura 3.11: Curvas de nivel correspondientes a la superficie que se representa en la figura 3.10, mostrando la región en el espacio (T, P) donde el porcentaje de error en el proceso de recuerdo de los estados de Mattis es grande. El *inset* muestra el correspondiente resultado para el modelo de Hopfield.

reglas de aprendizaje asimétricas como la dada por (3.46). En este caso, se obtiene de forma analítica ecuaciones cerradas para los parámetros de orden m^μ y m , que una vez resueltas muestran que, al igual que para el caso de simetría, la única situación en la que se obtienen como solución estados de Mattis localmente estables es para el caso de la *rate* V ; asimismo se obtiene que el principal efecto debido a la asimetría es el dificultar la propiedad de memoria asociativa. Esto es así debido a que los correspondientes estados de Mattis tienen solapamientos, con el correspondiente patrón, que son en general más pequeños que si la situación fuera la de simetría. Los resultados muestran también que la propiedad de memoria asociativa depende fuertemente del valor de los parámetros de anisotropía dados en (3.46).

Capítulo 4

Simulación Monte Carlo

4.1 Introducción

La simulación en el ordenador es, actualmente, una herramienta bien establecida y útil en muchos campos de la ciencia. Existen múltiples motivaciones para simular sistemas físicos en el ordenador. Una de las principales es que pueden así eliminarse aproximaciones. Usualmente, al estudiar un problema analíticamente, hay que recurrir a aproximaciones que uno querría eliminar o, al menos, conocer bien su influencia en los resultados. Ambas cosas pueden conseguirse introduciendo simulación en el análisis. Típicamente, la simulación permite estudiar sistemas muy complejos, y estudiar con gran detalle y flexibilidad su comportamiento. Así, los resultados de la simulación proporcionan un patrón o modelo con el que las teorías aproximadas pueden ser comparadas, y probar la validez del modelo propuesto. Por último, mencionamos que algunas magnitudes o comportamientos de los sistemas físicos pueden ser difíciles o imposibles de medir u observar en el laboratorio, y la simulación de un modelo puede jugar entonces el papel del experimento. En todo caso, hay que notar que la simulación en el ordenador tiene algunas limitaciones importantes como, por ejemplo, el tiempo finito de observación, y el hecho de que los sistemas simulados tienen un tamaño finito. Esto último se soluciona en la práctica (para algunos casos) usando condiciones de contorno adecuadas y teorías de tamaño finito. En todo caso, ambos problemas, tiempo finito de observación y tamaño finito, vienen siendo minimizados cada vez más por los avances en el desarrollo de los ordenadores.

Las simulaciones que describimos en este capítulo están basadas en el método

Monte Carlo, que es un método de simulación estocástico que sólo requiere considerar la parte configuracional del problema físico de interés. Sólo hay que saber cómo inducir cambios temporales de una configuración del sistema a otra distinta, cambios que son consecuencia de una evolución probabilística realizada mediante un proceso de Markov [49]. Los resultados que mostramos corresponden a simulaciones del modelo que presentado en el capítulo tercero, que es una versión del modelo general introducido en el capítulo segundo. Este modelo introduce la novedad de considerar fluctuaciones en las intensidades sinápticas alrededor de valores medios dados por una determinada regla de aprendizaje, la regla de Hebb. El interés en realizar simulaciones de este modelo radica en que, como vimos antes (cf. §3), la teoría de campo medio, utilizada para resolverlo analíticamente, muestra resultados muy interesantes en las propiedades del sistema en el caso en que existen correlaciones espaciales entre fluctuaciones que reflejan las correlaciones espaciales entre los patrones memorizados. Esto es, el modelo mejora la propiedad de memoria asociativa y la robusted en el proceso de aprendizaje, comparado con el modelo de Hopfield ordinario que no tiene en cuenta dichas fluctuaciones. Las simulaciones han servido para comprobar la validez de la aproximación de campo medio utilizada en el capítulo tercero, y para hacer evidentes otras de sus propiedades.

4.2 Modelo

El modelo simulado consiste en un conjunto de N neuronas binarias, $s_{\mathbf{x}} = +1$ o -1 , cada uno de estos valores correspondiendo a un estado de la neurona en la que ésta emite un potencial de acción y a un estado de no emisión, respectivamente, acopladas entre sí mediante intensidades sinápticas J_{xy} . Una configuración cualquiera, $\mathbf{s} = \{s_{\mathbf{x}}\}$, cambia compitiendo con variaciones simultáneas de las configuraciones de sinapsis, $\mathbf{J} = \{J_{xy}\}$, de tal forma que el sistema alcanza un estado estacionario que en general está fuera del equilibrio. Esto es, la dinámica particular asociada a neuronas y sinapsis involucra un conflicto entre ellas cuyo efecto es el mismo que el que tendría un hipotético e indeterminado agente actuando sobre \mathbf{s} ; este agente, junto con la acción de un baño térmico a temperatura T , puede inducir sobre este sistema una evolución temporal estocástica hacia un estado estacionario que, en general, no está determinado

sólo por el valor de T y por la energía asociada a \mathbf{s} . En los modelos teóricos que hemos desarrollado antes (cf. capítulos §2 y §3), esta evolución se describe en términos de la ecuación maestra

$$\partial_t P_t(\mathbf{s}) = \sum_{\mathbf{x}} [\varpi(\mathbf{s}^{\mathbf{x}}; \mathbf{x}) P_t(\mathbf{s}^{\mathbf{x}}) - \varpi(\mathbf{s}; \mathbf{x}) P_t(\mathbf{s})], \quad (4.1)$$

donde $P_t(\mathbf{s})$ es la probabilidad de tener una determinada configuración \mathbf{s} en un instante de tiempo t , y $\varpi(\mathbf{s}; \mathbf{x})$ es la correspondiente probabilidad de transición o *rate*. En la práctica, tal evolución puede ser simulada en un ordenador de la siguiente forma: Se parte de una configuración inicial aleatoria \mathbf{s}_0 , en la que a cada variable $s_{\mathbf{x}}$ se le ha asignado, al azar, uno de los dos posibles valores $+1$ o -1 ; a continuación se elige una neurona al azar (por ejemplo la neurona etiquetada con \mathbf{x}), y se hace el cambio (*spin-flip*) $s_{\mathbf{x}} \rightarrow -s_{\mathbf{x}}$, con una cierta probabilidad (probabilidad de transición), mediante el método Monte Carlo, esto es, generamos un número aleatorio uniforme $\eta \in [0, 1]$, que comparamos con la probabilidad de transición de forma que, si ésta es mayor que η se acepta el cambio (*spin-flip*) y la configuración resultante se toma ahora como configuración inicial y en caso contrario, no se acepta el cambio y se vuelve a elegir una nueva neurona al azar en la configuración de partida. La unidad de tiempo, el paso Monte Carlo, corresponde a la realización de N sorteos, es decir el elegir N neuronas al azar. En este modelo general que se ha simulado se han estudiado con detalle la evolución del sistema con las siguientes probabilidades de transición particulares (*rates*)

$$\begin{aligned} \varpi_1 &= \min \left\{ 1, \exp \left[\frac{2P}{T} \left(\frac{1}{N} - s_{\mathbf{x}} \frac{1}{P} \sum_{\nu=1}^P \xi_{\mathbf{x}}^{\nu} m_{\nu}(\mathbf{s}) \right) \right] \right\}, \\ \varpi_2 &= \exp \left[-\frac{P}{T} \left(1 + s_{\mathbf{x}} \frac{1}{P} \sum_{\nu=1}^P \xi_{\mathbf{x}}^{\nu} m_{\nu}(\mathbf{s}) \right) \right], \\ \varpi_3 &= \frac{1}{P} \sum_{\mu=1}^P \exp \left[-\frac{P}{T} (1 + s_{\mathbf{x}} \xi_{\mathbf{x}}^{\mu} m_{\mu}(\mathbf{s})) \right], \end{aligned} \quad (4.2)$$

donde hemos asumido que el sistema es capaz de almacenar P patrones, $\xi^{\mu} \equiv \{\xi_{\mathbf{x}}^{\mu} = \pm 1\}$ con $\mu = 1, \dots, P$, y donde $m_{\mu}(\mathbf{s}) = N^{-1} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^{\mu} s_{\mathbf{x}}$ es una medida del solapamiento del estado actual del sistema con el patrón memorizado μ . En (4.2) la temperatura está dada en unidades de k_B^{-1} .

Las elecciones ϖ_1 y ϖ_2 para $\varpi(\mathbf{s}, \mathbf{x})$ corresponden al modelo de Hopfield con

la regla de aprendizaje de Hebb $J_{xy} = N^{-1} \sum_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}$, sin la posibilidad de fluctuaciones en las sinápsis. La probabilidad de transición ϖ_1 se obtiene usando la conocida probabilidad de transición de Metropolis, $\Psi(\beta\Delta) = \min\{1, e^{-\Delta/T}\}$ (que en los capítulos precedentes hemos denotado por simplicidad como *rate* M; cf. §2, ecuación (2.10)), donde Δ es el cambio en la energía que se produce cuando tiene lugar el cambio $s_{\mathbf{x}} \rightarrow -s_{\mathbf{x}}$. Dicha energía viene dada por el hamiltoniano de Hopfield $\mathcal{H}_{\mathbf{J}}(\mathbf{s}) = -\frac{1}{2} \sum_{\mathbf{x} \neq \mathbf{y}} J_{xy} s_{\mathbf{x}} s_{\mathbf{y}} + \sum_{\mathbf{x}} \theta_{\mathbf{x}} s_{\mathbf{x}}$ (a partir de aquí tomaremos $\theta_{\mathbf{x}} = 0$). La probabilidad de transición ϖ_2 se obtiene utilizando la *rate* V (cf. §2, ecuación (2.10)), es decir $\Psi(\beta\Delta) = e^{-\beta\Delta/2}$, en vez de la *rate* de Metropolis y normalizando a la unidad ¹. El caso representado por la *rate* ϖ_3 se corresponde con nuestro modelo, desarrollado en el capítulo tercero, para la situación particular en que se considera la *rate* V ². En efecto, como vimos en la sección §3.1, el modelo corresponde a una versión más simplificada de un proceso general de Markov, introducido en el capítulo §2, consistente en que el sistema cambia $s_{\mathbf{x}} \rightarrow -s_{\mathbf{x}}$ con probabilidad $p\Psi(\beta\Delta)$ y con probabilidad $(1-p)\Phi$, donde —por simplicidad— Φ es independiente de \mathbf{s} , tiene lugar el cambio $J_{xy} \rightarrow J'_{xy}$; $p \in [0, 1]$. Junto al caso bien conocido de equilibrio que se obtiene para $p = 1$, que ha sido ya considerado con la elección de las *rates* ϖ_1 y ϖ_2 , en el límite $p \rightarrow 0$ puede ser considerada una condición, simplificada, para la competición entre diferentes tendencias, cada una de ellas correspondiendo a un determinado valor de $\mathcal{H}_{\mathbf{J}}$. Este límite introduce dos escalas de tiempo relevantes en el sistema (después de un proceso previo de plasticidad sináptica debido al aprendizaje): una escala de tiempo fina, τ , en la que las neuronas no evolucionan apreciablemente pero en la que las intensidades sinápticas fluctúan localmente; y una escala más grande, $t = p\tau$ para $p \rightarrow 0$ y $\tau \rightarrow \infty$, en la que las neuronas evolucionan con una distribución estacionaria para las sinapsis; la existencia de estas dos escalas de tiempo bien definidas es consistente —como vimos antes— con ciertas observaciones realizadas en los sistemas biológicos [4–18]. Esta

¹Dicha normalización, que consiste en dividir la *rate* por el máximo valor que puede tener, no afecta al estado estacionario que el sistema alcanza sino sólo al tiempo de relajación del mismo y sólo implica una escala temporal nueva, como se obtiene fácilmente a partir de (4.1).

²Los casos particulares en los que se consideran las *rates* K y M han sido también estudiados, coincidiendo el estado estacionario de las simulaciones realizadas con el que predice la teoría de campo medio de nuestro modelo. Dicho estado estacionario no presenta la propiedad de memoria asociativa, por lo que hemos considerado más interesante mostrar en detalle sólo aquella situación en la que nuestro modelo sí presenta dicha propiedad, es decir para el caso de la *rate* V .

condición límite equivale a hacer el cambio $s_x \rightarrow -s_x$ con una probabilidad *efectiva* $\varpi(\mathbf{s}, \mathbf{x}) = \int d\mathbf{J} f(\mathbf{J}) \Psi(\beta\Delta)$, donde $f(\mathbf{J})$ es la distribución estacionaria para las fluctuaciones en las sinapsis. La probabilidad de transición ϖ_3 se obtiene cuando se considera

$$f(\mathbf{J}) = \frac{1}{P} \sum_{\mu=1}^P \prod_{\substack{x,y \\ x \neq y}} \delta \left(J_{xy} - \frac{P}{N} \xi_x^\mu \xi_y^\mu \right) \quad (4.3)$$

y normalizando a la unidad. Esto es, las intensidades sinápticas fluctúan aleatoriamente en nuestro modelo de acuerdo a la distribución (4.3) alrededor de valores medios dados por la regla de Hebb, es decir $\bar{J}_{xy} \equiv \int d\mathbf{J} f(\mathbf{J}) J_{xy} = N^{-1} \sum_{\mu} \xi_x^\mu \xi_y^\mu$. Esta elección para $f(\mathbf{J})$ es interesante porque implica que cada patrón ξ^μ contribuye con una cierta probabilidad a \mathbf{J} , y tiene en cuenta correlaciones entre fluctuaciones en diferentes sinapsis, que reflejan las correlaciones espaciales dentro de los patrones ξ^μ .

4.3 Resultados Monte Carlo

Con las *rates* dadas en (4.2), se han realizado diferentes simulaciones con tamaños de red N variando entre $N = 400$ y $N = 3600$. El número de patrones almacenados P que se han considerado ha sido variable, aunque los resultados que aquí se muestran son para $P \in [1, 10]$. Sin embargo, es muy probable que se verifiquen los mismos resultados en el límite $N \rightarrow \infty$ para una valor de P arbitrariamente grande. Otra restricción que se ha tomado en las simulaciones es que se ha considerado sólo la situación $J_{xy} = J_{yx}$, como se corresponde la elección dada por (4.3), aunque las simulaciones son fácilmente extensibles a elecciones de $f(\mathbf{J})$ que involucren asimetría en las sinapsis. Los P patrones utilizados en la simulación han sido generados de forma aleatoria, tomando la variable ξ_x^μ , al azar, uno de los dos posibles valores $+1$ o -1 .

El primer tipo de simulaciones realizadas ha consistido en ver cómo es el estado estacionario del sistema y su estabilidad cuando se utiliza la *rate* ϖ_3 . La simulación muestra que la condición de campo medio utilizada en los desarrollos analíticos en el capítulo tercero es exacta y, salvo los efectos debidos a tamaño finito, los datos Monte Carlo se ajustan, de una forma muy significativa, a las curvas de los solapamientos, $m_n(T)$, que da dicha teoría de campo medio. Esto se aprecia claramente en la figura 4.1. La simulación ha sido realizada para $N = 3600$ y $P = 1, 5, 10$. Los datos experimentales

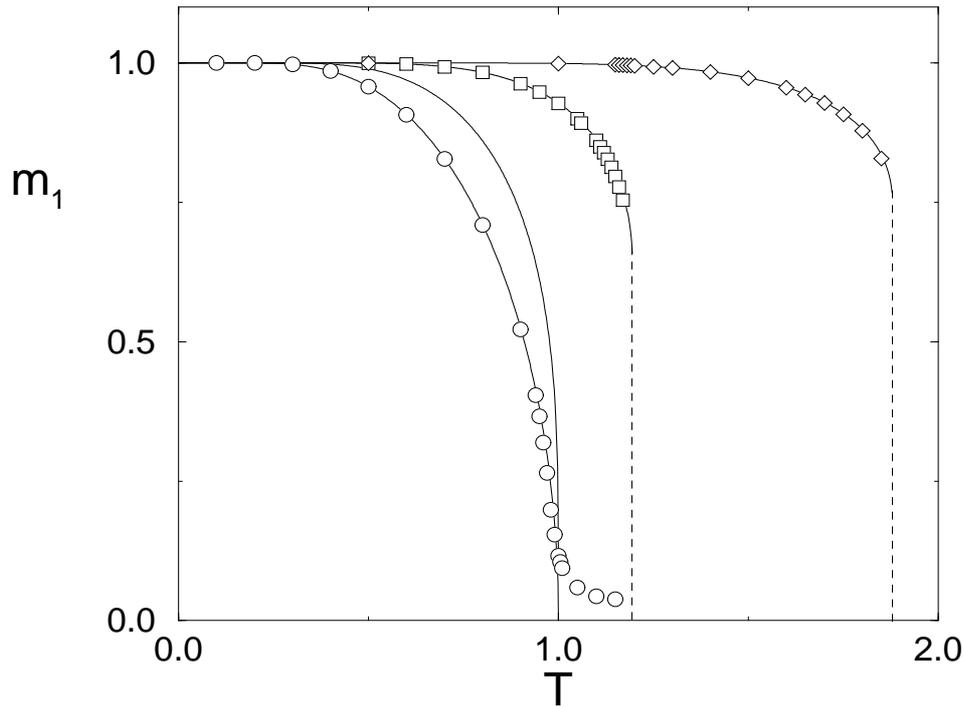


Figura 4.1: Las líneas continuas ilustran la variación con P de las soluciones estables $m_n(T)$ para $n = 1$ (estados puros de Mattis) que se obtienen a partir de (3.25) (*rate V*), para $P = 1, 3, 5$, y 10 , de izquierda a derecha. Las líneas discontinuas señalan el valor de $\tilde{T}(n, P)$ al que tiene lugar el cambio de fase de primer orden para cualquier $P > 3n$. Los símbolos son los correspondientes datos Monte Carlo para $N = 3600$ y $P = 1$ (\circ), 5 (\square), y 10 (\diamond). (La temperatura está dada en unidades k_B^{-1} .)

—representados en la figura por símbolos— se superponen con mucha exactitud con las curvas teóricas —representadas en líneas continuas—. El acuerdo tan sorprendente entre la teoría de campo medio y los resultados Monte Carlo se debe a que nuestro modelo es semejante a un modelo de Ising con interacciones de largo alcance, donde todos los espines interactúan con todos los demás. En nuestro caso, cada neurona está interactuando con las $N - 1$ restantes a través de las sinapsis. Como es sabido, para estos modelos tipo Ising con interacciones de largo alcance la teoría de campo medio es exacta.

En segundo lugar se han realizado una serie de simulaciones para $N = 2500$ y $N = 400$, con $P = 10$, comparando el modelo de Hopfield con el modelo que se ha

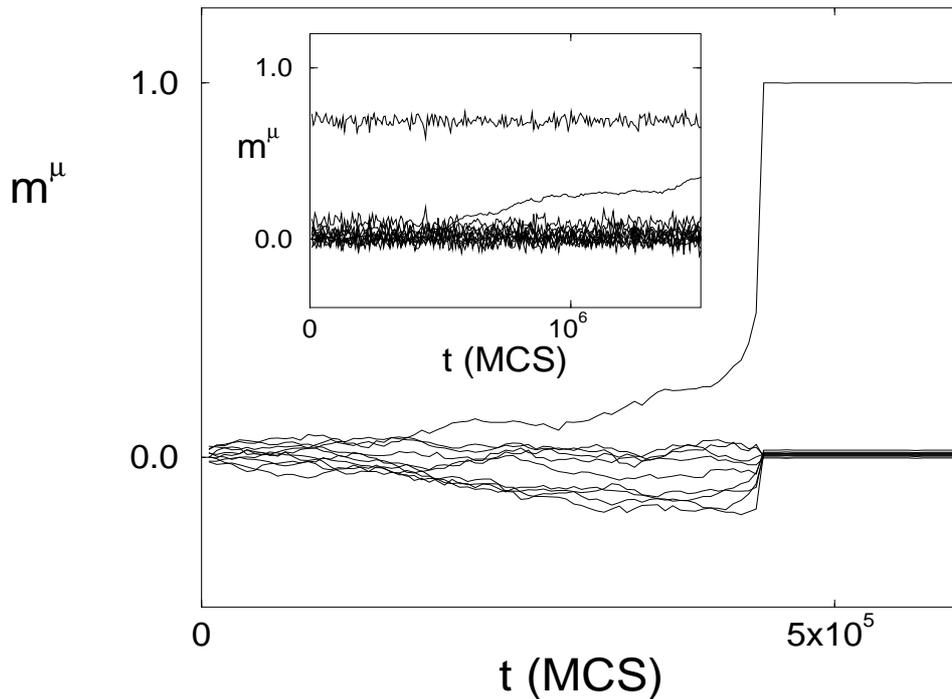


Figura 4.2: Evolución con el tiempo (en pasos Monte Carlo) del solapamiento $m^\mu(\mathbf{s}) = N^{-1} \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\mu s_{\mathbf{x}}$ para $N = 2500$ a la temperatura $T = 0.8$ (en unidades k_B^{-1}); Las diferentes líneas son para $\mu = 1, \dots, 10$. La gráfica principal ilustra el caso en el que hay fluctuaciones sinápticas, ϖ_3 . El *inset* contiene el resultado correspondiente al modelo de Hopfield; los comportamientos horizontales con ruido son para ϖ_1 (probabilidad de transición de Metropolis); la línea que se incrementa con el tiempo es para ϖ_2 . La figura muestra como fluctuaciones correlacionadas de sinapsis pueden hacer más estable y eficaz el proceso de recuerdo de información en una red de neuronas.

desarrollado analíticamente en el capítulo tercero. Las figuras 4.2-4.4 muestran las principales características que presentan ambos sistemas. El comportamiento horizontal con ruido en el *inset* de la figura 4.2 es el comportamiento familiar del modelo de Hopfield debido a la *rate* ϖ_1 . Esto es, partiendo de una configuración inicial aleatoria, \mathbf{s}_0 , hay una rápida evolución (no visible en la escala de esta figura) en la que el solapamiento con uno de los P patrones memorizados se hace relativamente grande (alrededor de 0.7 en esta simulación) mientras que los otros decaen prácticamente a cero. La situación permanece estacionaria excepto por el ruido térmico, que está añadido a la señal; incrementando T (que es 0.8 en esta simulación) se podría final-

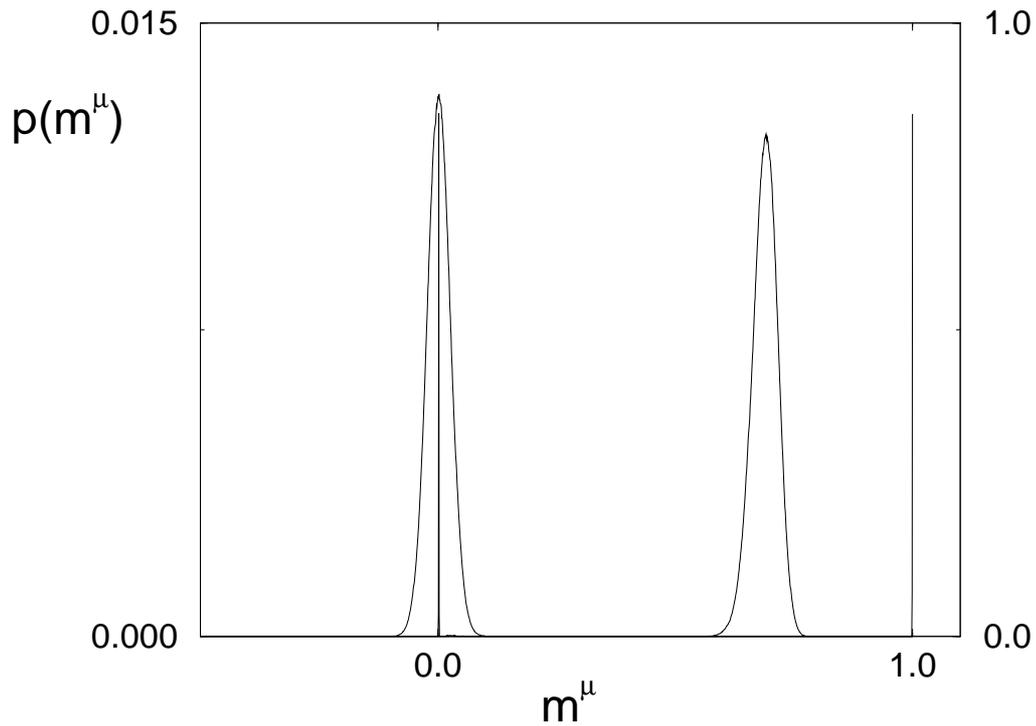


Figura 4.3: Distribución —normalizada a la unidad— de las fluctuaciones térmicas con el tiempo exhibidas por el sistema en la figura 4.2, durante el régimen estacionario, para dos valores típicos de μ . La escala en el lado izquierdo corresponde a las distribuciones con forma suave (centradas entorno a 0.0 y 0.7, respectivamente) que ocurren para ϖ_1 . La escala en el lado derecho corresponde a distribuciones (centradas entorno a 0.0 y muy cerca a 1, respectivamente) para ϖ_3 que se aproximan mucho (en la escala de esta figura) a funciones delta de Dirac.

mente impedir la propiedad de memoria asociativa. A efectos de comparación, el *inset* de la figura 4.2 también ilustra el comportamiento para ϖ_2 (línea que se incrementa con el tiempo), es decir, una elección diferente para la probabilidad de transición elemental $\Psi(\beta\Delta)$. El proceso de recuerdo en esta versión del modelo de Hopfield es muy lento pero más robusto frente al ruido térmico; esto era de esperar dado que $e^{-\beta\Delta/2}$ favorece los estados de baja energía más que el caso de la *rate* de Metrópolis. Como se ilustra en la gráfica principal de la figura 4.2, la realización del proceso de recuerdo es más efectivo si la evolución tiene lugar de acuerdo a ϖ_3 , o lo que es lo mismo, en presencia de las fluctuaciones sinápticas dadas por (4.3). Esto es, aunque la convergen-

cia al resultado deseado es más lenta con ϖ_3 que con ϖ_1 —aunque no es más lenta con ϖ_3 que con ϖ_2 —, la evolución y el estado estacionario son muy robustos y el error, en el proceso de recuerdo de un patrón dado, es despreciable con los parámetros de red y temperatura con los que ha tenido lugar la simulación. El mismo comportamiento cualitativo se observa cuando variamos N , P , y T dentro de un rango amplio. Es importante señalar que la naturaleza aditiva de (4.3), que implica la existencia de correlaciones entre las fluctuaciones en diferentes sinapsis inducidas por las correlaciones espaciales contenidas en los propios patrones ξ^μ , es esencial para este resultado. En particular, asumiendo que los efectos de diferentes patrones son multiplicativos en $f(\mathbf{J})$, se induce un simple ruido en el sistema, que se añade al ruido térmico no produciendo diferencias demasiado dramáticas respecto al modelo de Hopfield; cf. §2. La figura 4.3 confirma el resultado de la figura 4.2. Esta gráfica muestra claramente que con ϖ_3 , a diferencia de ϖ_1 , se alcanza la saturación y se induce un alto nivel de robustez al sistema. La figura 4.4 ilustra otra interesante característica de ϖ_3 . Las evoluciones aquí comienzan con el mismo estado inicial obtenido tras perturbar fuertemente uno de los patrones (aleatorios) memorizados. La gráfica principal muestra que ϖ_3 siempre lleva al sistema al patrón de partida permaneciendo estable una vez alcanzado. Las simulaciones han sido realizadas a $T = 0.6$. Si disminuimos la temperatura, el tiempo de relajación de las simulaciones aumenta pero no aparece ningún otro efecto cualitativo interesante. El *inset* muestra el resultado correspondiente para el modelo de Hopfield con ϖ_1 . En este caso el sistema es mucho menos eficiente a la hora de reconocer el patrón de partida y, eventualmente, se desestabiliza, debido sobre todo a las fluctuaciones térmicas, después de algún tiempo relativamente corto (hay que notar que el tiempo en la figura 4.4 está en escala logarítmica con el objeto de mostrar los detalles de la evolución al principio de la misma).

En resumen, mediante una serie de simulaciones, hemos extraído las principales propiedades que presenta el modelo de red de neuronas (definido en su forma general en §2.1 y que hemos desarrollado, para una determinada situación particular, en el capítulo tercero); estas simulaciones indican que el considerar fluctuaciones en las intensidades sinápticas en la forma dada por (4.3) induce una mayor robustez y una mejora significativa en los procesos de memoria asociativa y reconocimiento de pa-

trones. Asimismo, dichas simulaciones han servido para comprobar la teoría de campo medio desarrollada en §3.3 y §3.4, mostrando que dicha teoría es exacta.

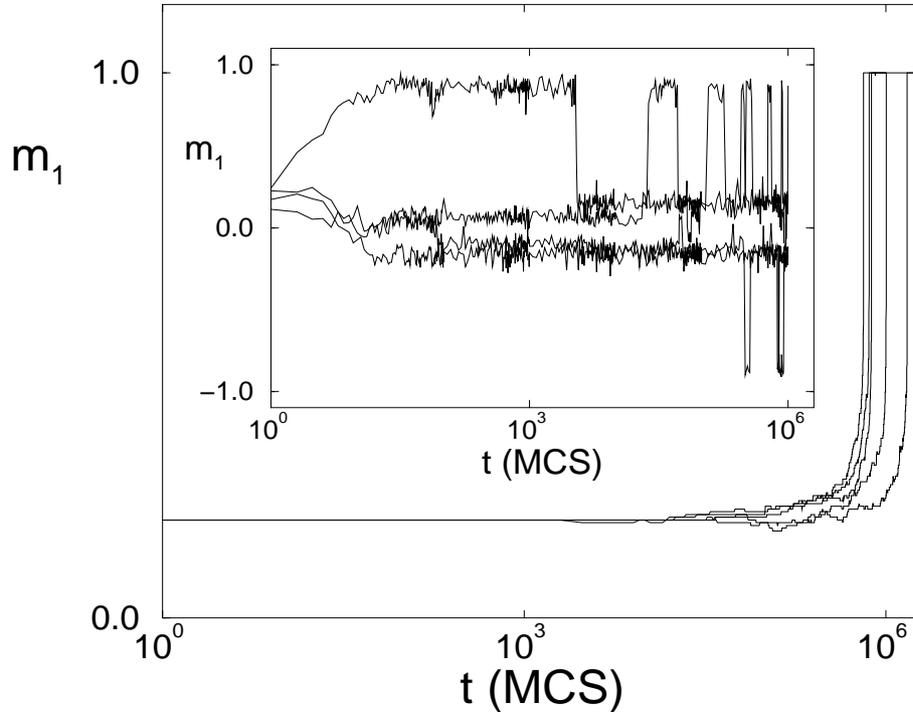


Figura 4.4: Representación semilogarítmica mostrando evoluciones independientes, para $T = 0.6$ (en unidades k_B^{-1}), $N = 400$, y $P = 10$ patrones aleatorios memorizados, comenzando con el mismo estado inicial. Este último se obtiene perturbando uno de los patrones de tal forma que en dicho estado inicial 236 neuronas (del total de 400), aleatoriamente situadas, están en el mismo estado que las correspondientes neuronas en ese patrón y el resto no. La gráfica principal muestra seis evoluciones típicas independientes con ϖ_3 . El *inset* muestra cuatro evoluciones típicas independientes con ϖ_1 . Esta figura ilustra que el proceso de reconocimiento de patrones en una red de neuronas se ve notablemente optimizado con la presencia de fluctuaciones correlacionadas en las sinapsis.

4.4 Conclusión

Mediante una serie de simulaciones Monte Carlo, hemos puesto de manifiesto la importancia, en la realización de determinadas redes de neuronas artificiales, de cierto

tipo de fluctuaciones (aleatorias) en las sinapsis, correlacionadas espacialmente y que reflejan correlaciones espaciales entre diferentes patrones memorizados. El principal efecto de tales fluctuaciones es una mejora sensible de la capacidad de memoria asociativa y una gran robusted en el proceso de aprendizaje de la red, frente a otros modelos que no consideran dichas fluctuaciones (por ejemplo, el modelo de Hopfield). Concluimos que este tipo de aparente ruido incrementa notablemente la eficiencia en la transmisión de la señal. Los resultados que obtenemos en las simulaciones están de acuerdo, por lo tanto, con las características que muestran las redes de neuronas biológicas, en las que las sinapsis individuales son extremadamente poco fiables y ruidosas, pero que en conjunto parecen aumentar la potencia de procesamiento del sistema; cf., por ejemplo, Refs. [4–18].

Los resultados Monte Carlo nos han permitido, asimismo, comprobar los principales resultados del estudio analítico realizado en el capítulo precedente, y comprobar el rango de validez de la aproximación de campo medio realizada en dicho estudio.

Capítulo 5

Difusión de iones magnéticos

5.1 Introducción

El modelo Ising de vidrio de espines [30, 40] es un ejemplo conceptualmente simple de sistema que presenta desorden microscópico. Consiste en un red regular d -dimensional —por ejemplo, cúbica simple Z^d — que tiene asociado el hamiltoniano

$$H_{\mathbf{J}}(\mathbf{s}) = - \sum_{\mathbf{x}, \mathbf{y} \in Z^d} J_{xy} s_{\mathbf{x}} s_{\mathbf{y}} - h_e \sum_{\mathbf{x} \in Z^d} s_{\mathbf{x}}, \quad (5.1)$$

donde J_{xy} es una variable aleatoria caracterizada por una cierta distribución de probabilidad, y h_e es una constante que representa un campo magnético sobre los espines. Cada configuración de las variables binarias de espín —definidas en cada nudo de la red—, $\mathbf{s} = \{s_{\mathbf{x}} = \pm 1; \mathbf{x} \in Z^d\}$, tiene una energía potencial (5.1) para cada posible distribución de desorden, $\mathbf{J} = \{J_{xy} \in \mathfrak{R}; \mathbf{x}, \mathbf{y} \in Z^d\}$. El hecho de que los acoplamientos entre espines, J_{xy} , cambien al azar de una pareja de espines a otra, trata de simular la situación en sistemas reales, donde los iones magnéticamente activos están distribuidos espacialmente al azar, y las energías de intercambio tienen un comportamiento oscilatorio con la distancia entre los iones¹. Una característica esencial de este modelo es que el conjunto \mathbf{J} es constante en el tiempo para cada realización particular del sistema;

¹Esta interacción de intercambio entre espines se conoce con el nombre de interacción RKKY, pues fue estudiada en detalle en mezclas magnéticas por Ruderman y Kittel en 1954, y posteriormente por Kasuya en 1956 y Yosida en 1957. La forma efectiva de dicha interacción en función de la distancia entre espines r es

$$J_{xy}(r) = J \frac{\cos(2k_F r + \varphi)}{r^3}, \quad r \rightarrow \infty, \quad (5.2)$$

donde k_F es el momento de Fermi y J y φ son parámetros característicos de la mezcla.

corresponde a un modelo con *desorden congelado* (*quenched*), es decir, las impurezas están *congeladas*, fijas en el tiempo. Por lo tanto, dado que la información física que encierra el modelo ha de ser independiente de la realización particular de desorden, interesa entonces promediar sobre la distribución estacionaria de las configuraciones de espines, $P_{st}(\mathbf{s})$, y sobre el peso, $f(\mathbf{J})$, de diferentes realizaciones de desorden [50–53].

Esta situación ha sido recientemente generalizada al caso en que la distribución espacial de desorden, \mathbf{J} , varía con el tiempo [35]. En este caso, se argumenta que, como un paso adicional, los modelos más realistas deberían considerar la posibilidad de variaciones temporales de los acoplamientos J_{xy} para tener en cuenta los efectos debidos a la difusión iónica. Ésta constantemente modifica la distancia, $\mathbf{x} - \mathbf{y}$, entre cada par específico de espines y , por lo tanto, hace que J_{xy} varíe en la práctica también con el tiempo, dado el carácter oscilatorio de la interacción RKKY con la distancia. Como una primera aproximación para entender el posible efecto debido a la difusión iónica, se ha estudiado la difusión rápida y aleatoria de impurezas [54]. En este caso, se asume que la probabilidad de una configuración de espines \mathbf{s} en el tiempo t , $P_t(\mathbf{s})$, satisface la ecuación maestra

$$\frac{\partial P_t(\mathbf{s})}{\partial t} = \sum_{\mathbf{x} \in Z^d} [c(\mathbf{s}^{\mathbf{x}}; \mathbf{x})P_t(\mathbf{s}^{\mathbf{x}}) - c(\mathbf{s}; \mathbf{x})P_t(\mathbf{s})], \quad (5.3)$$

donde la probabilidad por unidad de tiempo (*rate*) para el cambio (*spin-flip*) de la configuración $\mathbf{s} = \{\dots, s_{\mathbf{x}-1}, s_{\mathbf{x}}, s_{\mathbf{x}+1}, \dots\}$ a la configuración $\mathbf{s}^{\mathbf{x}} = \{\dots, s_{\mathbf{x}-1}, -s_{\mathbf{x}}, s_{\mathbf{x}+1}, \dots\}$ es una superposición de procesos más elementales de la forma

$$c(\mathbf{s}; \mathbf{x}) = \int_{-\infty}^{+\infty} d\mathbf{J} f(\mathbf{J}) c_{\mathbf{J}}(\mathbf{s}; \mathbf{x}). \quad (5.4)$$

La probabilidad de transición elemental $c_{\mathbf{J}}(\mathbf{s}; \mathbf{x})$ está asociada a una distribución particular de desorden \mathbf{J} y, por lo tanto, a un valor particular de la energía $H_{\mathbf{J}}(\mathbf{s})$. En la práctica, resulta conveniente tratar con funciones $c_{\mathbf{J}}(\mathbf{s}; \mathbf{x})$ que satisfagan la condición de balance detallado, es decir, $c_{\mathbf{J}}(\mathbf{s}; \mathbf{x}) = c_{\mathbf{J}}(\mathbf{s}^{\mathbf{x}}; \mathbf{x}) \exp[\beta H_{\mathbf{J}}(\mathbf{s}^{\mathbf{x}}) - \beta H_{\mathbf{J}}(\mathbf{s})]$. Para $f(\mathbf{J}) = \prod_{\mathbf{x}, \mathbf{y}} \delta(J_{xy} - J)$, es decir, cuando la dinámica consiste sólo en la acción de una única $c_{\mathbf{J}}(\mathbf{s}; \mathbf{x})$ con todas las J_{xy} s iguales a J , la condición de balance detallado implica que el sistema evoluciona con el tiempo hacia un estado de equilibrio correspondiente al inverso de la temperatura $\beta = (k_B T)^{-1}$ y energía $H_{\mathbf{J}}(\mathbf{s})$. En general, la

expresión (5.4) describe una competición entre diferentes tendencias cuyo resultado neto en el límite asintótico $t \rightarrow \infty$ es en general un estado estacionario fuera del equilibrio. Por lo tanto, el estudio de este modelo tiene interés no sólo para el estudio de sistemas desordenados, sino también para el estudio de fenómenos fuera del equilibrio.

En este capítulo presentamos un estudio de versiones sencillas y resolubles del sistema representado por (5.3)-(5.4), incluyendo también comparaciones con otras aproximaciones. La situación que hemos investigado difiere conceptualmente del modelo de vidrio de espines con *desorden recocido* (*annealed*) [55] (en el que las impurezas han alcanzado el estado de equilibrio junto con los espines, situación que usualmente ha sido reconocida como poco realista). De hecho, el modelo exhibe un comportamiento nuevo; en particular, describimos la aparición de una aleatoriedad adicional que implica cierta *frustración dinámica* —o conflicto durante la evolución temporal— que induce interesantes fenómenos, muchos de ellos semejantes a los que aparecen en materiales reales constituidos por mezclas binarias de impurezas magnéticamente activas con metales nobles. Entre estos fenómenos están el efecto de la frustración en la no saturación de la magnetización a bajas temperaturas y la existencia de fases *reentrantes* [56–58]

5.2 Definición del modelo

El estudio del sistema representado por las expresiones (5.3)-(5.4) es, en general, difícil. Por lo tanto, hemos trabajado con versiones sencillas del mismo motivadas por la suposición, generalmente admitida, de que un conjunto grande de grados de libertad puede ser reemplazado por la acción de un campo. Esto es, sólo vamos a tener en cuenta los detalles, incluyendo correlaciones entre espines, en un dominio compacto de nudos de la red relativamente pequeño, Λ_C . Este dominio consta de dos subconjuntos, Λ_I y Λ_F , tales que $\Lambda_C = \Lambda_I \cup \Lambda_F$, $\Lambda_I \cap \Lambda_F = \emptyset$, donde Λ_I es el dominio interior y Λ_F es la *frontera*, definida como el subconjunto de todos los nudos de Λ_C que tienen al menos un *vecino próximo* fuera de Λ_C ; cf. figura 5.1. Las probabilidades asociadas con las configuraciones, $\mathbf{s}_C = \{s_{\mathbf{x}}; \mathbf{x} \in \Lambda_C\}$, de $\mathbf{s}_I = \{s_{\mathbf{x}}; \mathbf{x} \in \Lambda_I\}$, y de $\mathbf{s}_F = \{s_{\mathbf{x}}; \mathbf{x} \in \Lambda_F\}$ pueden ser relacionadas mediante la ecuación [59]

$$P_t(\mathbf{s}_C) = P_t(\mathbf{s}_I/\mathbf{s}_F)P_t(\mathbf{s}_F), \quad (5.5)$$

donde $P_t(\mathbf{s}_I/\mathbf{s}_F)$ es la probabilidad condicional de tener la configuración \mathbf{s}_I en el interior supuesto que se tiene la configuración \mathbf{s}_F en la frontera, y donde asumimos las normalizaciones $\sum_{\mathbf{s}_I} P_t(\mathbf{s}_I/\mathbf{s}_F) = 1$, y $\sum_{\mathbf{s}_F} P_t(\mathbf{s}_F) = 1$, donde las sumatorias están extendidas a todas las posibles configuraciones del interior del dominio y de su frontera, respectivamente. Una característica principal de nuestro modelo es que, para cualquier configuración particular de la frontera \mathbf{s}_F , la probabilidad $P_t(\mathbf{s}_I/\mathbf{s}_F)$ satisface la ecuación maestra:

$$\frac{\partial P_t(\mathbf{s}_I/\mathbf{s}_F)}{\partial t} = \sum_{\mathbf{x} \in \Lambda_I} [c(\mathbf{s}_I^{\mathbf{x}}; \mathbf{x}) P_t(\mathbf{s}_I^{\mathbf{x}}/\mathbf{s}_F) - c(\mathbf{s}_I; \mathbf{x}) P_t(\mathbf{s}_I/\mathbf{s}_F)], \quad (5.6)$$

donde $c(\mathbf{s}_I; \mathbf{x})$ es una superposición similar a (5.4), es decir,

$$c(\mathbf{s}_I; \mathbf{x}) = [[c_{\mathbf{J}}(\mathbf{s}_I; \mathbf{x})]] \equiv \int_{-\infty}^{+\infty} d\mathbf{J} f(\mathbf{J}) c_{\mathbf{J}}(\mathbf{s}_I; \mathbf{x}). \quad (5.7)$$

Las características del modelo representadas por las expresiones (5.6)-(5.7) han sido completadas con una elección para $P_t(\mathbf{s}_F)$, que necesita incluir un campo coherente apropiado para tener en cuenta la relación entre espines en Λ_F y en el resto de la red Z^d . Esto es, la probabilidad total de tener en el sistema original una determinada configuración de espines $P_t(\mathbf{s})$ está sustituida aquí por la probabilidad asociada al dominio $P_t(\mathbf{s}_C)$, probabilidad que incluye un campo que debe ser calculado de forma autoconsistente. Este hecho implica que tenemos sólo un conjunto de 2^{N_I} ecuaciones acopladas a resolver, ecuación (5.6), (en vez de las 2^N , $N \gg N_I$, que tendría el problema original) cuya solución se puede encontrar para valores de $N_I = \text{card}(\Lambda_I)$ no demasiado grandes.

El estado estacionario del sistema descrito por la ecuación (5.6) está definido como la solución de $\partial P_{st}(\mathbf{s}_I/\mathbf{s}_F)/\partial t = 0$. Se puede escribir de forma bastante general que $P_{st}(\mathbf{s}_I/\mathbf{s}_F) \equiv \exp\{-\beta \mathcal{H}(\mathbf{s}_I, \mathbf{s}_F) - \Gamma(\mathbf{s}_F)\}$, que define $\mathcal{H}(\mathbf{s}_I, \mathbf{s}_F)$ y donde $\Gamma(\mathbf{s}_F)$ se determina a partir de la normalización de $P_{st}(\mathbf{s}_I/\mathbf{s}_F)$. De forma sencilla se sigue

$$\mathcal{H}(\mathbf{s}_I, \mathbf{s}_F) = \frac{1}{2^{N_I}} \ln \left[\prod_{\mathbf{s}_I} P_{st}(\mathbf{s}_I/\mathbf{s}_F) \right] - \ln P_{st}(\mathbf{s}_I/\mathbf{s}_F), \quad (5.8)$$

donde el productorio es sobre todas las configuraciones del interior del dominio \mathbf{s}_I . Para obtener (5.8) hemos asumido que $\mathcal{H}(\mathbf{s}_I, \mathbf{s}_F)$ puede ser escrito como una combinación de productos de espines dentro del dominio, es decir,

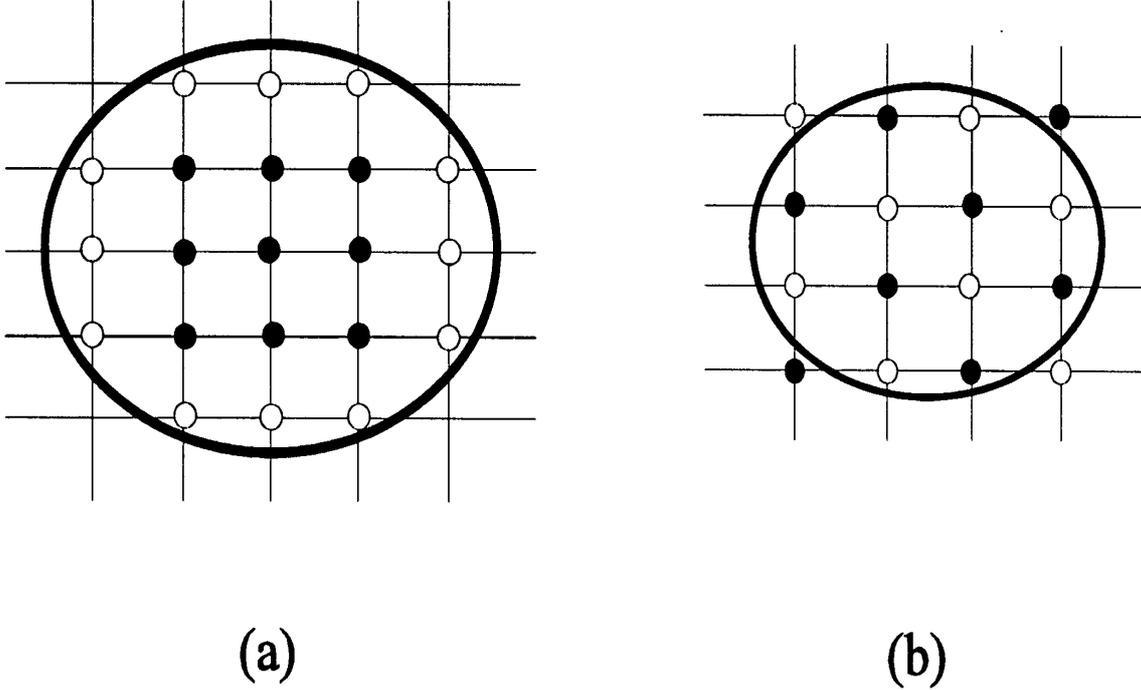


Figura 5.1: (a) La figura ilustra un dominio genérico de nudos en la red Z^d , $\Lambda_C \equiv \{\bullet, \circ\}$; el dominio interior, $\Lambda_I \equiv \{\bullet\}$; y la frontera, $\Lambda_F \equiv \{\circ\}$. (b) Se ilustra un dominio específico para $d = 2$, que consta de doce espines, como se considera en el texto; aquí se hace una distinción entre dos subredes, $\Lambda_A \equiv \{\bullet\}$, y $\Lambda_B \equiv \{\circ\}$.

$$\mathcal{H}(\mathbf{s}_I, \mathbf{s}_F) = \sum_{k=1}^{N_I} \sum_{(\mathbf{j}_1 \dots \mathbf{j}_k)} J_{\mathbf{j}_1 \dots \mathbf{j}_k}^{(k)} s_{\mathbf{j}_1} \dots s_{\mathbf{j}_k}, \quad (5.9)$$

donde $J_{\mathbf{j}_1 \dots \mathbf{j}_k}^{(k)}$ son números reales. Definimos la energía configuracional del dominio como

$$H_{\mathbf{J}}(\mathbf{s}_C) = - \sum_{\mathbf{x}, \mathbf{y} \in \Lambda_C} J_{xy} s_{\mathbf{x}} s_{\mathbf{y}} - h_e \sum_{\mathbf{x} \in \Lambda_C} s_{\mathbf{x}}. \quad (5.10)$$

Si escribimos la probabilidad en la frontera como

$$P_{st}(\mathbf{s}_F) \propto \exp \left\{ \Gamma(\mathbf{s}_F) + \beta h \sum_{\mathbf{y} \in \Lambda_F} s_{\mathbf{y}} \right\}, \quad (5.11)$$

donde h es el campo medio coherente, se obtiene después de usar (5.5) que

$$P_{st}(\mathbf{s}_C) = \frac{1}{\mathcal{N}} \exp \left\{ -\beta \mathcal{H}(\mathbf{s}_I, \mathbf{s}_F) + \beta h \sum_{\mathbf{y} \in \Lambda_F} s_{\mathbf{y}} \right\}, \quad (5.12)$$

con

$$\mathcal{N} = \sum_{\mathbf{s}_C} \exp \left\{ -\beta \mathcal{H}(\mathbf{s}_I, \mathbf{s}_F) + \beta h \sum_{\mathbf{y} \in \Lambda_F} s_{\mathbf{y}} \right\}. \quad (5.13)$$

El campo h en la frontera se calcula de forma autoconsistente, imponiendo el requisito de que la magnetización por espín en Λ_I debe ser la misma que en Λ_F . Este formalismo se puede generalizar de forma sencilla, en el sentido de que podemos incluir varios campos diferentes, que son necesarios para estudiar ciertos tipos de orden, como se hace de forma explícita en la próxima sección.

5.3 Algunos casos particulares del modelo

El formalismo anterior es válido para cualquier dimensión d y tamaño de Λ_I arbitrario. Vamos a considerar algunos casos para los que el modelo tiene solución analítica sencilla con un comportamiento no trivial. Casos más complicados pueden ser fácilmente estudiados resolviendo ecuaciones implícitas en el ordenador. Consideremos en primer lugar el caso más simple, concretamente $\Lambda_I = \{\mathbf{x}_0\}$ (para cualquier dimensión d). Es decir, el dominio está constituido por un espín central $s_{\mathbf{x}_0}$ y una frontera constituida por los q vecinos próximos de éste. Por simplicidad, nos vamos a restringir en todo el desarrollo que sigue a *rates* elementales tales que $c_{\mathbf{J}}(\mathbf{s}_I; \mathbf{x}) = \phi(\beta \Delta H_{\mathbf{J}}^{\mathbf{x}}(\mathbf{s}_C))$, con $\Delta H_{\mathbf{J}}^{\mathbf{x}}(\mathbf{s}_C) \equiv H_{\mathbf{J}}(\mathbf{s}_I^{\mathbf{x}}, \mathbf{s}_F) - H_{\mathbf{J}}(\mathbf{s}_I, \mathbf{s}_F)$ (usando la notación de la sección anterior), $\phi(X) = \phi(-X) \exp(-X)$ (correspondiente a la condición de balance detallado), $\phi(0) = 1$, y $\phi(\infty) = 0$. Tres situaciones familiares consistentes con estas propiedades son $\phi(X) = \exp(-X/2)$ (que denotamos como *rate V*), $\phi(X) = 2(1 + e^X)$ (*rate K*), y $\phi(X) = \min\{1, e^{-X}\}$ (*rate M*). La probabilidad estacionaria condicional se sigue entonces de forma sencilla a partir de $\partial P_{st}(\mathbf{s}_I/\mathbf{s}_F)/\partial t = 0$ como

$$P_{st}(s_{\mathbf{x}_0}/\mathbf{s}_F) = \frac{c(s_{\mathbf{x}_0}; \mathbf{x}_0)}{c(s_{\mathbf{x}_0}; \mathbf{x}_0) + c(-s_{\mathbf{x}_0}; \mathbf{x}_0)}, \quad (5.14)$$

después de usar (5.5) y (5.12). Se puede comprobar que

$$\frac{P_{st}(-s_{\mathbf{x}_0}/\mathbf{s}_F)}{P_{st}(s_{\mathbf{x}_0}/\mathbf{s}_F)} = \frac{c(-s_{\mathbf{x}_0}; \mathbf{x}_0)}{c(s_{\mathbf{x}_0}; \mathbf{x}_0)}, \quad (5.15)$$

es decir, éste es un caso muy simple en el que la competición de dinámicas —no sólo la *rate* elemental— también satisface la condición de balance detallado. Se obtiene a

partir de (5.8) y (5.14) que

$$\beta\mathcal{H}(s_{\mathbf{x}_0}, \mathbf{s}_F) = -\frac{1}{2}s_{\mathbf{x}_0} \ln \frac{[[\phi \{\beta\Delta_{\mathbf{J}}^{\mathbf{x}_0}(\mathbf{s}_F)\}]]}{[[\phi \{-\beta\Delta_{\mathbf{J}}^{\mathbf{x}_0}(\mathbf{s}_F)\}]]} \quad (5.16)$$

con $\Delta_{\mathbf{J}}^{\mathbf{x}_0}(\mathbf{s}_F) = 2 \left\{ \sum_{\mathbf{y} \in \Lambda_F} J_{\mathbf{x}_0\mathbf{y}} s_{\mathbf{y}} + h_e \right\}$. Por lo tanto,

$$\mathcal{H}(s_{\mathbf{x}_0}, \mathbf{s}_F) = - \sum_{\mathbf{y} \in \Lambda_F} J_{\mathbf{x}_0\mathbf{y}} s_{\mathbf{y}} s_{\mathbf{x}_0}, \quad (5.17)$$

donde

$$\beta \sum_{\mathbf{y} \in \Lambda_F} J_{\mathbf{x}_0\mathbf{y}} s_{\mathbf{y}} = -\frac{1}{2} \ln \frac{[[\phi \{\beta\Delta_{\mathbf{J}}^{\mathbf{x}_0}(\mathbf{s}_F)\}]]}{[[\phi \{-\beta\Delta_{\mathbf{J}}^{\mathbf{x}_0}(\mathbf{s}_F)\}]]}, \quad (5.18)$$

es un *hamiltoniano efectivo* tipo Ising [36] asociado al dominio.

La restricción de estas ecuaciones a $\Lambda_I = \{\mathbf{x}_0\}$ (dominio insuficiente para describir el orden antiferromagnético) aconseja considerar un único campo coherente, h , en (5.11). La solución final es entonces (5.12) con (5.16), donde h se obtiene mediante la relación de autoconsistencia

$$\langle s_{\mathbf{x}_0} \rangle = \left\langle \frac{1}{q} \sum_{\mathbf{y} \in \Lambda_F} s_{\mathbf{y}} \right\rangle \quad (5.19)$$

con $\langle \dots \rangle$ representando un promedio con respecto a $P_{st}(\mathbf{s}_C)$. Esta relación de autoconsistencia puede ser transformada en

$$\sum_{n=0}^q \binom{q}{n} y^n \Theta(n) = 0, \quad (5.20)$$

donde $y \equiv e^{2\beta h}$, y

$$\Theta(n) \equiv 2n \left[\left[\frac{A(n)}{B(n)} \right]^{\frac{1}{2}} + \left[\frac{B(n)}{A(n)} \right]^{\frac{1}{2}} \right] - 2q \left[\frac{B(n)}{A(n)} \right]^{\frac{1}{2}}, \quad (5.21)$$

con $A(n) = [[\phi(2\beta J(2n - q) + 2\beta h_e)]]$ y $B(n) = [[\phi(-2\beta J(2n - q) - 2\beta h_e)]]$.

Antes de analizar las consecuencias de (5.20), vamos a especificar el formalismo para un dominio de espines más grande. Ilustramos el caso de dimensión $d = 2$ con Λ_C definido como aparece en la figura 5.1 (b), es decir, nos restringimos al estudio de un dominio que contiene cuatro espines en el interior rodeados por ocho espines en la frontera. En dicho dominio definimos dos subredes, Λ_A y Λ_B , con $\Lambda_C = \Lambda_A \cup \Lambda_B$, $\Lambda_A \cap \Lambda_B = \emptyset$, como se representa en la figura 5.1 (b), y

$$m \equiv \frac{1}{N} \sum_{\mathbf{x} \in \Lambda} s_{\mathbf{x}}, \quad \mu \equiv \frac{1}{N_A} \sum_{\mathbf{x} \in \Lambda_A} s_{\mathbf{x}} - \frac{1}{N_B} \sum_{\mathbf{x} \in \Lambda_B} s_{\mathbf{x}}. \quad (5.22)$$

Resulta necesario el considerar ahora dos campos coherentes, h_A y h_B , actuando sobre los espines de la frontera en Λ_A y Λ_B , respectivamente. Esto es, la probabilidad asociada al dominio es ahora

$$P_{st}(\mathbf{s}_C) = \frac{1}{\mathcal{N}} \exp \left\{ -\beta \mathcal{H}(\mathbf{s}_I, \mathbf{s}_F) + \beta h_A \sum_{\mathbf{y} \in \Lambda_{F_A}} s_{\mathbf{y}} + \beta h_B \sum_{\mathbf{y} \in \Lambda_{F_B}} s_{\mathbf{y}} \right\} \quad (5.23)$$

con

$$\mathcal{N} = \sum_{\{\mathbf{s}_C\}} \exp \left\{ -\beta \mathcal{H}(\mathbf{s}_I, \mathbf{s}_F) + \beta h_A \sum_{\mathbf{y} \in \Lambda_{F_A}} s_{\mathbf{y}} + \beta h_B \sum_{\mathbf{y} \in \Lambda_{F_B}} s_{\mathbf{y}} \right\}. \quad (5.24)$$

Las correspondientes relaciones de autoconsistencia son:

$$2 \left\langle \sum_{\mathbf{x} \in \Lambda_I} s_{\mathbf{x}} \right\rangle = \left\langle \sum_{\mathbf{x} \in \Lambda_F} s_{\mathbf{x}} \right\rangle \quad (5.25)$$

$$2 \left\langle \sum_{\mathbf{x} \in \Lambda_{I_A}} s_{\mathbf{x}} - \sum_{\mathbf{x} \in \Lambda_{I_B}} s_{\mathbf{x}} \right\rangle = \left\langle \sum_{\mathbf{x} \in \Lambda_{F_A}} s_{\mathbf{x}} - \sum_{\mathbf{x} \in \Lambda_{F_B}} s_{\mathbf{x}} \right\rangle.$$

Se puede escribir esto de forma esquemática como

$$g_1(x, y) = 0, \quad g_2(x, y) = 0, \quad (5.26)$$

donde $x \equiv e^{2\beta h_A}$, $y \equiv e^{2\beta h_B}$, y las funciones g_i dependen de la elección particular de $f(\mathbf{J})$.

5.4 Resultados

Veamos explícitamente los resultados que se siguen de las ecuaciones obtenidas en la sección previa.

En lo referente a la situación en que se tiene $\Lambda_I = \{\mathbf{x}_0\}$, la solución viene dada por (5.12) con (5.16) y (5.20). Para $h_e = 0$ se tiene la propiedad de simetría $\Theta(n) = -\Theta(q - n)$. Se sigue entonces que $y = 1$ es siempre solución de (5.20), que corresponde a la solución trivial $m = 0$. Nos interesan las soluciones no triviales correspondientes a magnetización espontánea no nula.

En ausencia de competición entre diferentes tendencias dinámicas, como ocurre para $f(\mathbf{J}) = \prod_{\mathbf{x} \neq \mathbf{y}} \delta(J_{xy} - J_0)$, se tiene que $P_{st}(\mathbf{s}_C) = Z_{BP}^{-1} \exp\{-\beta H_{BP}(\mathbf{s}_C)\}$, donde

$H_{BP}(\mathbf{s}_C) = -J_0 \sum_{\mathbf{y} \in \Lambda_F} s_{\mathbf{x}_0} s_{\mathbf{y}} - h \sum_{\mathbf{y} \in \Lambda_F} s_{\mathbf{y}}$ es el hamiltoniano de Bethe-Peierls. Esto es, reproducimos la teoría —bien conocida— de campo medio para un sistema en equilibrio. De hecho, (5.20) en estas condiciones se reduce a

$$y = \left\{ \frac{1 + ye^{2\beta J_0}}{y + e^{2\beta J_0}} \right\}^{q-1}, \quad (5.27)$$

que tiene soluciones distintas de cero para $T < T_c$, donde $J_0/k_B T_c = \frac{1}{2} \ln(q/(q-2))$ señala un cambio de fase de una fase paramagnética a otra fase ferromagnética con magnetización espontánea no nula.

En presencia de competición, para dimensión $d = 2$, la expresión (5.20) se reduce a la condición $\Theta(0)y^2 + 4\Theta(1)y + \Theta(0) = 0$, que tiene dos soluciones distintas de cero:

$$y_{\pm} = -\frac{2\Theta(1)}{\Theta(0)} \pm \sqrt{\left\{ \frac{2\Theta(1)}{\Theta(0)} \right\}^2 - 1}. \quad (5.28)$$

Estas dos soluciones se transforman en una única solución para $\Theta(1)/\Theta(0) = \frac{1}{2}$, condición que define el valor de la temperatura crítica para el cambio de fase de segundo orden. El valor de T_c depende de $f(\mathbf{J})$ y de $\phi(X)$ como se deduce de forma sencilla a partir de (5.21). Nosotros hemos trabajado explícitamente con dos distribuciones diferentes para el desorden: una distribución gaussiana y otra bimodal.

Consideremos en primer lugar la distribución gaussiana

$$f(\mathbf{J}) = \prod_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \frac{1}{\sqrt{2\pi\tilde{J}}} \exp \left\{ -\frac{(J_{xy} - J_0)^2}{2\tilde{J}^2} \right\}. \quad (5.29)$$

Para la *rate* V , el cambio de fase es similar al de equilibrio descrito anteriormente; en particular, $k_B T_c/J_0 = 2.88$ for $d = 2$, independientemente de los parámetros que aparecen en (5.29). Esto es, la competición de dinámicas no induce efectos que sean remarcables. Por contraste, para la *rate* K y para la *rate* M , T_c depende de la anchura \tilde{J} de la distribución. Cuando $\tilde{J}/J_0 \rightarrow 0$, se observa que la magnetización a temperatura cero, $m(T = 0)$, tiende a la saturación; esto es, el efecto de las impurezas decrece cuando la distribución gaussiana tiende a una función delta y, eventualmente, se recupera la situación de equilibrio (en la que no hay competición de dinámicas). Por otra parte, $m(T) \rightarrow 0$, para cualquier T , cuando $\tilde{J}/J_0 \rightarrow 1.0336$; este valor último

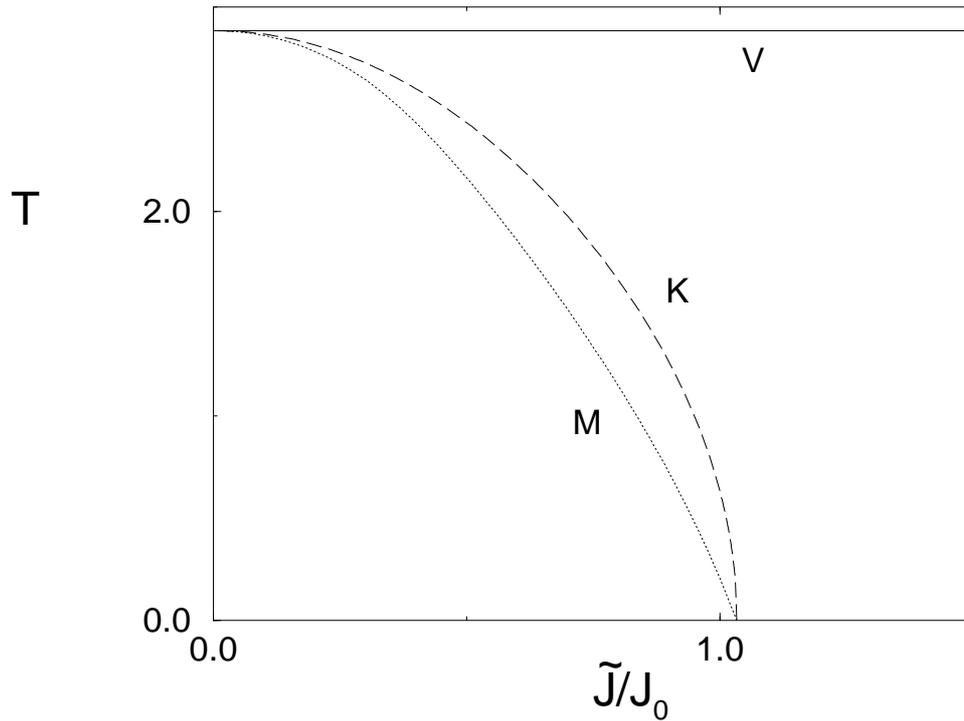


Figura 5.2: Diagrama de fases para dimensión $d = 2$, para un dominio que tiene un único espín en el interior, y evoluciona con una competición de dinámicas gaussiana, (5.29), y (de arriba a abajo, como se indica en la figura) *rates* V , K , y M , definidas en la sección 5.3. (La temperatura está dada en unidades J_0/k_B .)

se obtiene como solución a la ecuación $\text{erf}\left(J_0/\sqrt{2\tilde{J}}\right) = \frac{2}{3}$. Esto es, una distribución ancha para el desorden hace tan fuerte la aleatoriedad que cualquier tipo de orden de largo alcance desaparece; es interesante el que esto ocurra de forma repentina para un valor de \tilde{J}/J_0 bien definido. En la figura 5.2 se representa el diagrama de fases para varias elecciones de las *rates* cuando la competición de dinámicas es gaussiana, (5.29). El hecho de que la *rate* influya tan fuertemente en el estado estacionario del sistema es una característica distintiva del sistema cinético (5.3)-(5.4) [35].

Consideremos a continuación la distribución bimodal

$$f(\mathbf{J}) = \prod_{\substack{x,y \\ x \neq y}} \{p\delta(J_{xy} - J_0) + (1-p)\delta(J_{xy} + J_0)\}, \quad (5.30)$$

es decir, J_{xy} es igual a J_0 ($J_0 > 0$) con probabilidad p (tanto durante el tiempo en

que el sistema evoluciona como espacialmente en cualquier configuración del estado estacionario) y J_{xy} es igual a $-J_0$ con probabilidad $(1-p)$. El estado estacionario en esta situación resulta ser independiente de la *rate* (para los casos considerados), a diferencia de lo que ocurría para la distribución (5.29). De cualquier forma, la probabilidad p de que el acoplamiento J_{xy} sea positivo juega un papel similar al papel que juega la anchura de la gaussiana \tilde{J} que hemos introducido anteriormente. En particular, el sistema exhibe un cambio de fase de segundo orden a una temperatura crítica $T_c(p)$, de tal forma que, para que la magnetización alcance la saturación, se requiere que $p \rightarrow 1$, y cuando $p < p_0$ dicho cambio de fase no tiene lugar. Se tiene que $p_0 = \frac{5}{6}$ para $d = 2$. La figura 5.3 ilustra la situación que acabamos de describir y muestra una comparación con datos correspondientes a una simulación por ordenador del sistema (5.3)-(5.4) con la *rate* M [60].

La solución para el dominio bidimensional que viene representado en la figura 5.1 (b) es (5.23) con los campos medios coherentes verificando (5.26). Consideremos ahora explícitamente la situación en que la evolución temporal ocurre de acuerdo a la distribución bimodal, es decir, con probabilidades de transición dadas por (5.7) con $f(\mathbf{J})$ dada por (5.30). Para esta situación particular, las funciones que aparecen en (5.26) tienen la forma

$$\begin{aligned} g_1(x, y) &= a_1(1 - x^4y^4) + a_2(y + x - x^4y^3 - y^4x^3) \\ &+ a_3(y^2 + x^2 - x^4y^2 - x^2y^4) + a_4(y^3 + x^3 - x^4y - xy^4) \\ &+ a_5(xy - x^3y^3) + a_6(xy^2 + x^2y - x^3y^2 - x^2y^3), \end{aligned} \quad (5.31)$$

y

$$\begin{aligned} g_2(x, y) &= b_1(y - x - x^4y^3 + x^3y^4) + b_2(y^2 - x^2 + x^2y^4 - x^4y^2) \\ &+ b_3(y^3 - x^3 + xy^4 - x^4y) + b_4(y^4 - x^4) \\ &+ b_5(xy^2 - x^2y - x^3y^2 + x^2y^3) + b_6(xy^3 - x^3y), \end{aligned} \quad (5.32)$$

donde los coeficientes, a_i y b_i , $i = 1, \dots, 6$, dependen de T , p , y de la función $\phi(X)$. Estas ecuaciones han sido resueltas numéricamente obteniendo como principal resultado el siguiente: Para $p = 1$ y 0 , —situación en la que no hay competición de dinámicas y por lo tanto el estado estacionario es de equilibrio— la temperatura crítica (para cambios de fase de una fase paramagnética a una fase ferromagnética y a una fase antiferromagnética, respectivamente) es $k_B T_c / J_0 = 2.8309$, ligeramente más pequeña que el correspondiente resultado de la teoría de campo medio de Bethe-Peierls, descrita

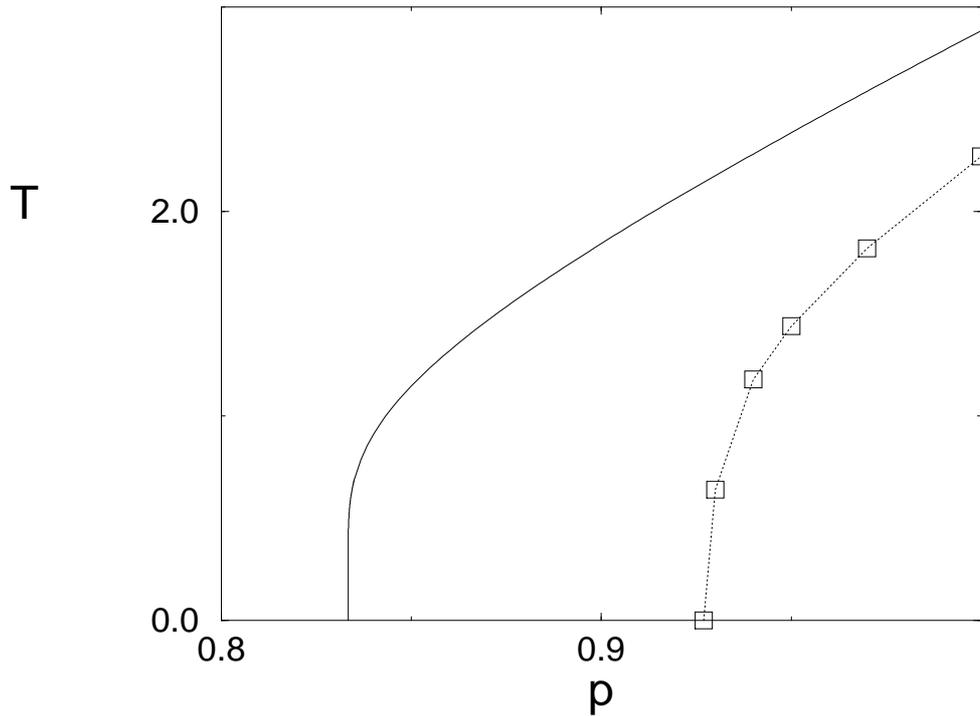


Figura 5.3: Diagrama de fases para dimensión $d = 2$, para un dominio que tiene un único espín central en su interior, y evoluciona mediante una competición de dinámicas bimodal, (5.30); el resultado es independiente de la *rate* elemental para las *rates* V , K , y M . Los símbolos corresponden a resultados correspondientes a una simulación por ordenador en Ref. [60] para el sistema (5.3)-(5.4) con la *rate* M . (La temperatura está dada en unidades J_0/k_B .)

anteriormente, que da para dicha temperatura crítica un valor de 2.88; además, esta versión del modelo da como resultado curvas de magnetización $m(T)$ que son muy parecidas a las que da la solución de Bethe-Peierls. Para el caso en que $p \in (0, 1)$, el comportamiento del sistema depende fuertemente de la elección de la función $\phi(X)$. En particular, con las *rates* K y M se obtiene un valor de $T_c(p)$, que es diferente para cada *rate* y que se hace cero para $p < p_0$, con $p_0 = 0.8636$, y 0.8664 , respectivamente. Este último valor puede ser comparado con el valor de p_0 que dan otras aproximaciones de campo medio; en concreto, con el valor de $p_0 = 0.8649$ que se obtiene con la aproximación de pares [61, 62], y con el valor de $p_0 = 0.928$ que se obtiene mediante simulación por ordenador [60], ambos resultados correspondientes al mismo sistema

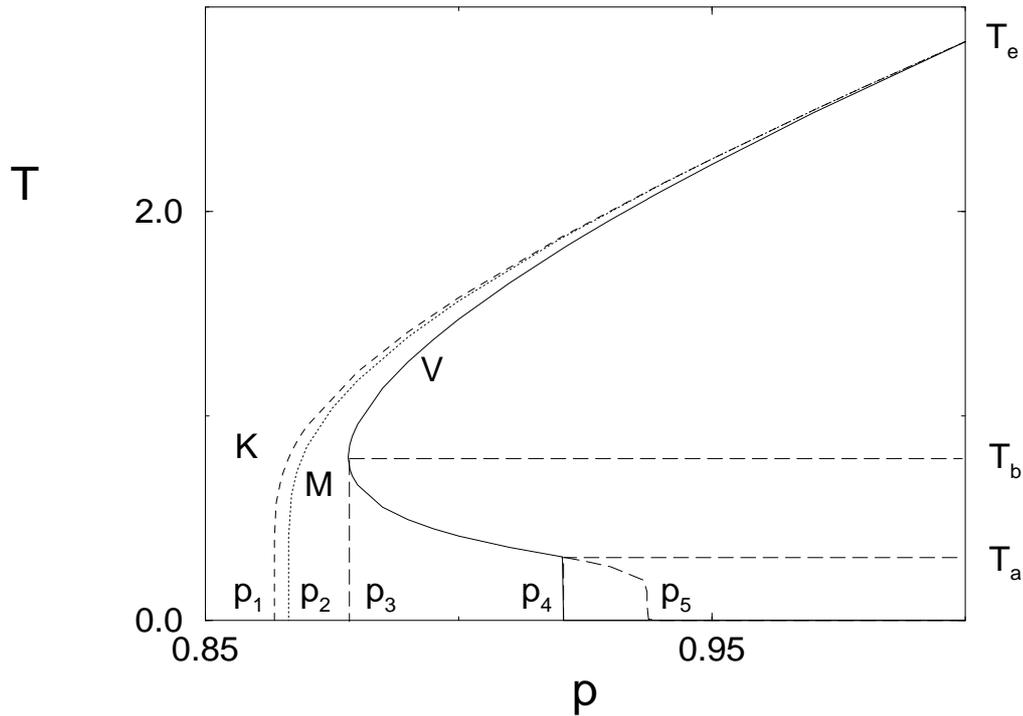


Figura 5.4: Diagrama de fases para el dominio bidimensional representado en la figura 5.1 (b), y para la competición de dinámicas bimodal, (5.30). Las diferentes curvas que aparecen son, de arriba a abajo, respectivamente, para las *rates* K , M , y V . Obsérvese para el caso V la aparición de una fase reentrante y la existencia de un punto tricrítico en (p_4, T_a) . (Aquí las unidades de T son J_0/k_B .)

original. La figura 5.4 ilustra el diagrama de fases correspondiente. Aquí, p_1 y p_2 son los valores anteriores de p_0 para el caso de las rates K y M , respectivamente; con la excepción de que ahora la forma de la *rate* determina el estado estacionario del sistema, las diferencias con la situación descrita en la figura 5.3 son pequeñas. La figura 5.4 muestra el comportamiento del sistema para $p > \frac{1}{2}$, valores para los que el cambio de fase es entre una fase paramagnética y una fase ferromagnética; la situación es similar para el caso en la que la fase ordenada es antiferromagnética, $p < \frac{1}{2}$. La figura 5.5 ilustra la variación con p del campo coherente h a tres temperaturas diferentes para el caso particular de rate K .

En ella aparecen claramente definidas las regiones en la que hay fase ferromagnética

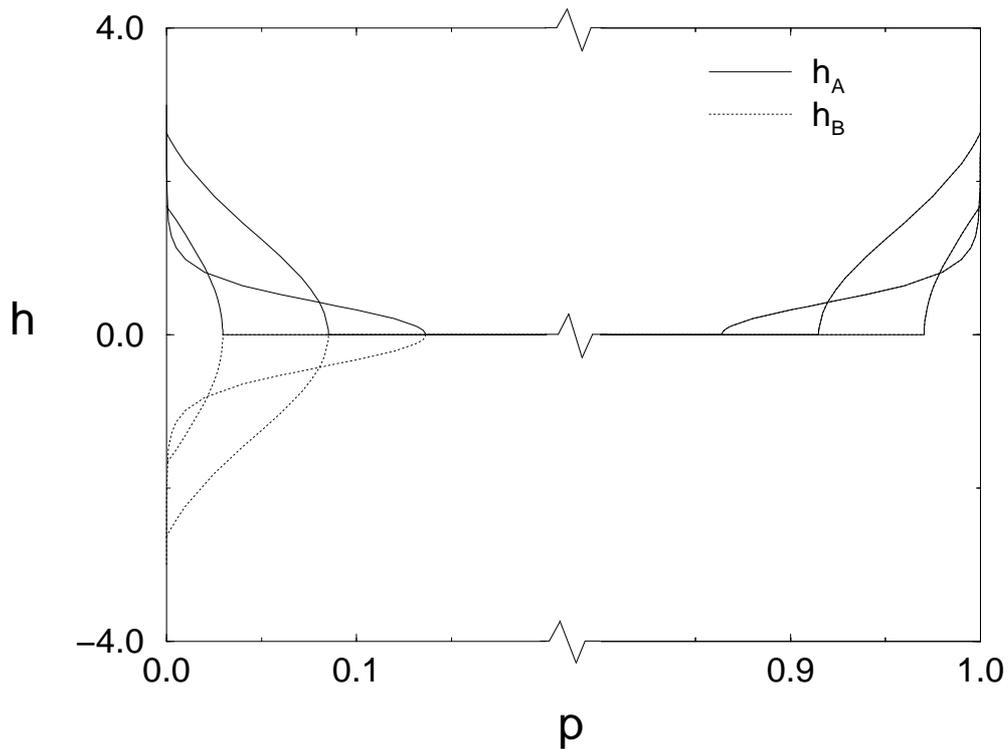


Figura 5.5: Tres isothermas ilustrativas mostrando la variación del campo coherente h con p para el sistema representado en la figura 5.4 con la *rate* K .

y fase antiferromagnética; esto último se pone de manifiesto porque se observa la simetría esperada para los dos campos involucrados, es decir $h_A = h_B$ en la fase ferromagnética y $h_A = -h_B$ en la fase antiferromagnética.

La situación para la *rate* V resulta bastante diferente, como se muestra en la figura 5.4. Como antes, hay un cambio de fase de segundo orden de una fase paramagnética a una fase ferromagnética y a una fase antiferromagnética, a una temperatura crítica $T_c = T_1(p) > T_b$; cf. figura 5.4. Cuando la temperatura es baja, a diferencia de lo que ocurría con las otras dos *rates*, aparece un segundo cambio de fase a $T_c = T_2(p) < T_b$, entre la fase ferromagnética o antiferromagnética y una fase en la que no existe ningún tipo de orden de largo alcance. Esto puede ser interpretado suponiendo que el sistema, a temperaturas inferiores a $T_2(p)$, presenta un tipo de orden diferente del típico orden de largo alcance, lo que podría sugerir la existencia de una fase vidrio de espines. Este segundo cambio de fase aparece de forma continua para $p > p_3 = 0.8782$ y $T_2(p) < T_b$.

Para $p > p_4 = 0.9206$ y $T_2(p) < T_a$ (cf. figura 5.4), este segundo cambio de fase se hace discontinuo como en un cambio de fase de primer orden, correspondiendo el punto (p_4, T_a) del diagrama de fases (p, T) a un punto tricrítico. El cambio de fase de primer orden tiene lugar para valores de $p < p_5 = 0.9375$. Para valores de $p > p_5$, el único cambio de fase que tiene lugar es el correspondiente a $T_1(p)$. La fase que acabamos de describir se llama reentrante y es típica de sistemas desordenados como los vidrios de espines. El comportamiento reentrante se aprecia claramente en la figura 5.6, en la que se muestran diferentes curvas de magnetización en función de la temperatura, para diferentes valores de p y para la rate V . La curva en la que la magnetización espontánea satura corresponde a $p = 1$, y es equivalente al comportamiento de equilibrio que da la teoría de Bethe, pero a una temperatura de equilibrio $k_B T_e / J_0 = 2.8309$, ligeramente inferior a la temperatura de Bethe. Para valores de p más pequeños se observa el comportamiento descrito más arriba, apareciendo para determinados valores de p curvas de magnetización reentrantes, semejantes a las que aparecen en ciertos materiales reales; cf. Refs. [56–58].

En cuanto a la estabilidad de las soluciones, se tiene que para cualquier configuración de la frontera \mathbf{s}_F , ya sea para el dominio con un único espín central, como para el caso del dominio en dimensión $d = 2$ que se representa en la figura 5.1 (b), las soluciones estacionarias a la ecuación (5.6) son estables. Esto se puede ver de forma sencilla en el caso del dominio con un sólo espín, donde los coeficientes de Lyapunov son uno negativo y el otro nulo —correspondiente a la solución estacionaria—. En $d = 2$, para el dominio representado en 5.1 (b), hay que diagonalizar la matriz dinámica asociada a dicha ecuación maestra, obteniendo que todos los autovalores son negativos, excepto el asociado a la solución estacionaria que es nulo, y todo esto para cualquier configuración de la frontera. Por lo tanto, los estados solución a las ecuaciones de autoconsistencia son estables. Lo anterior es consecuencia de la existencia del hamiltoniano efectivo (5.8), que implica la condición de balance detallado para las *rates* definidas en (5.7), condición que hace que las soluciones a la ecuación maestra sean estables; cf. Ref. [63].

En resumen, hemos estudiado una versión original de modelo que tiene en cuenta la posibilidad de difusión iónica, pues considera al desorden variando aleatoriamente en el tiempo. Dicho modelo ha sido estudiado mediante técnicas originales de campo

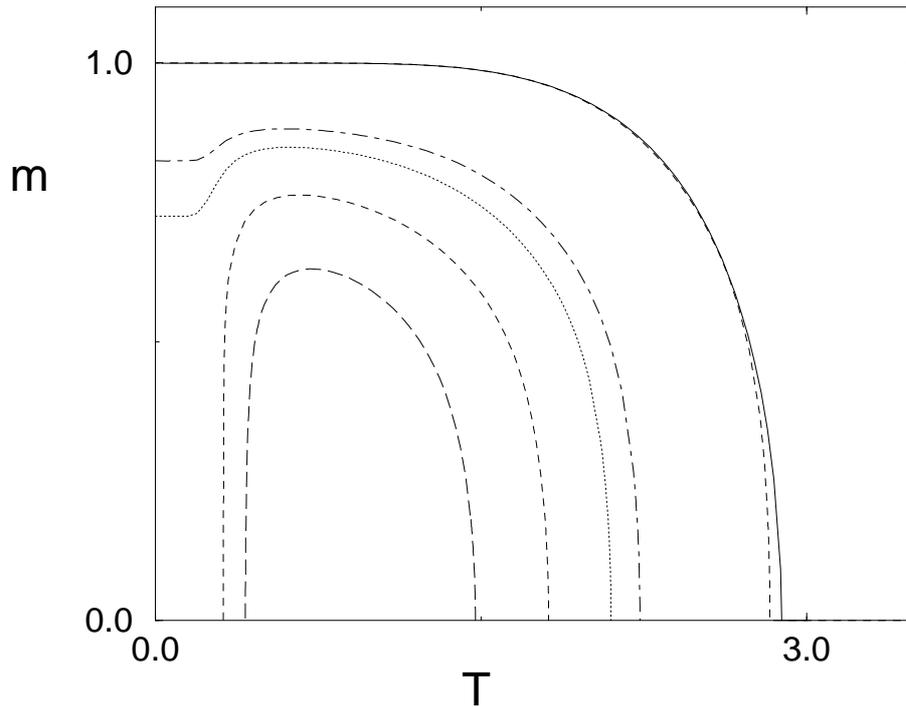


Figura 5.6: Variación con la temperatura de la magnetización para diferentes situaciones, para el dominio bidimensional representado en la figura 5.1 (b) y *rate* V . Las curvas son para la solución de equilibrio de Bethe (línea continua), y para el sistema cinético con $p = 1$ (equilibrio), 0.95, 0.94, 0.92, y 0.90, respectivamente, de arriba a abajo de la figura. La situación es similar para $p < \frac{1}{2}$ en la fase antiferromagnética y para el parámetro de orden μ ; cf. (5.3). (Aquí T está dada en unidades de J_0/k_B .)

medio, que implican considerar correlaciones entre espines sólo en un determinado dominio finito del sistema. Entre las principales características que presenta el sistema, están la existencia, en el estado estacionario, de hamiltonianos efectivos. El estado estacionario es altamente dependiente de la dinámica microscópica asociada al desorden \mathbf{J} . Asimismo, muchas de las propiedades emergentes que se obtienen recuerdan las que aparecen en ciertos materiales reales, por ejemplo, aparecen las fases reentrantes, típicas de algunos materiales.

Conclusiones

Hemos presentado y estudiado una serie de modelos reticulares fuera del equilibrio que representan diversos sistemas con desorden evolucionando con el tiempo. Todos ellos están basados en ecuaciones maestras en las que las probabilidades de transición por unidad de tiempo son superposición de diferentes tendencias, cada una de ellas intentando llevar al sistema a un determinado estado de equilibrio. Aquí, la aplicación más directa de estos modelos ha sido el estudio de sistemas de redes de neuronas y de sistemas magnéticos en los que hay difusión iónica. Cada capítulo describe con detalle los resultados más importantes de nuestro estudio; sólo a continuación enumeramos algunas contribuciones generales que constituyen hitos importantes originales en el trabajo realizado.

El primer modelo —que abarca los capítulos segundo, tercero y cuarto— consiste en una red en la que las neuronas y las intensidades sinápticas (desorden) evolucionan de forma estocástica, compitiendo entre sí. El modelo constituye una generalización de otros modelos de redes de neuronas en los que las intensidades sinápticas permanecen fijas en el tiempo o bien evolucionan en una escala de tiempo mayor que la que caracteriza la evolución de las neuronas —como es el caso del modelo de Hopfield—. Motivados por el comportamiento de las redes biológicas, hemos desarrollado ciertos límites del modelo general no estudiados hasta ahora. Nos hemos interesado en casos en los que las sinapsis evolucionan en una escala de tiempo menor que las neuronas, como cuando, aparte de las variaciones temporales lentas o *plasticidad sináptica* debida al aprendizaje, hay fluctuaciones rápidas de las intensidades sinápticas de carácter aleatorio. Comparado con modelos que no tienen en cuenta estas fluctuaciones, o con modelos que involucran *ruido* producido por un baño térmico a temperatura T , nuestro modelo es coherente con ciertos estudios experimentales. Esto es, se ha descrito

la existencia en sistemas biológicos de un comportamiento aleatorio esencial de las sinapsis, tanto en la transmisión del potencial de acción a través de ellas, como en la emisión de los cuantos de neurotransmisores al espacio intersináptico entre dos neuronas; estos estudios revelan, además, que este ruido sináptico puede tener un papel fundamental en el procesamiento de información y en el aprendizaje. En nuestro modelo, las intensidades sinápticas alcanzan rápidamente una distribución estacionaria, y las neuronas evolucionan en el tiempo mediante una ecuación maestra con transiciones (*rates*) efectivas que involucran diferentes mecanismos elementales, cada uno con una probabilidad, lo que hace que el sistema se encuentre en una situación fuera del equilibrio. Motivados por la idea de que el ruido existente en las sinapsis biológicas influye, en cierta forma, en el procesamiento de información, se han considerado distribuciones estacionarias de sinapsis donde éstas fluctúan alrededor de valores medios dados por una regla de aprendizaje, simétrica o asimétrica, dada.

En una primera versión del modelo general introducido en el capítulo segundo, se ha considerado una distribución de sinapsis que tiene cierta propiedad de factorización. Esta elección implica ausencia de correlaciones en las fluctuaciones sinápticas que, en la práctica, impide reflejar las correlaciones espaciales dentro de los patrones memorizados. Una primera consecuencia es que el estado estacionario es altamente dependiente de la forma de la *rate* elemental. Para ciertas elecciones de la misma es posible hacer una descripción del sistema mediante un *hamiltoniano efectivo* tipo Hopfield. El modelo ha sido resuelto entonces utilizando la técnica de réplicas. Algunas propiedades interesantes de esta versión del modelo son: cambios de fase de primer y segundo orden, existencia de fases ferromagnética, paramagnética, fase *vidrio de espines*. La principal conclusión es que las fluctuaciones tienen como principal efecto la aparición de un ruido extra, que se añade a la señal, que tiende a impedir, un poco, el proceso de memoria asociativa; para un valor crítico en el número de patrones que el sistema puede almacenar, este ruido puede impedir la aparición de estados mezcla o vidrios de espines. Otra propiedad interesante es que no aparece la línea Almeida-Thouless o límite de estabilidad de la solución ferromagnética con simetría de réplicas. Se ha realizado también una descripción cuando los patrones memorizados no tienen inhomogeneidades espaciales locales. Finalmente, se ha hecho un estudio dinámico de campo

medio de una versión más general del modelo, describiendo ciertos cambios de fase, distintos de los anteriores.

En una segunda versión (capítulo tercero y cuarto), se ha considerado un tipo de distribución para las fluctuaciones de las sinapsis que no tiene la propiedad de factorización del caso anterior. Esta elección permite correlaciones en las fluctuaciones de las sinapsis que, en cierto modo, reflejan la correlaciones espaciales dentro de los patrones memorizados. Sin embargo, no es posible, en ninguno de los casos estudiados, hacer una descripción del sistema mediante hamiltonianos efectivos. Aun así, podemos obtener información acerca del estado estacionario imponiendo una determinada condición de campo autoconsistente, apropiada en nuestro caso dado que las interacciones son entre una neurona y todas las demás, es decir, de largo alcance. El estado estacionario depende de la forma de la *rate* elemental, encontrándose que entre las tres elecciones consideradas, sólo una (*rate* V) permite la propiedad de memoria asociativa, que se manifiesta por la existencia de estados de Mattis localmente estables y la ausencia de estados mezcla estables que pudieran impedir dicha propiedad. Para esta elección, hay cambios de fase de primer orden a una temperatura $\tilde{T}(n, P)$ (para $n = 1$ y $P > 3$) y de segundo orden a $T_c = 1$ (para $n = 1$ y $P < 3$), entre una fase paramagnética a temperaturas relativamente altas, en la que no hay propiedad de memoria asociativa, y una fase ferromagnética a temperaturas relativamente bajas, en la que sí la hay. El diagrama de fases (T, P) refleja, pues, la existencia de un punto tricrítico fuera del equilibrio en $(T_c, P = 3)$. En la otras dos situaciones consideradas, es decir para las *rates* K y M , hemos demostrado que los únicos estados localmente estables son estados mezcla, que involucran un solapamiento no nulo con todos los patrones. Al comparar con el modelo de Hopfield, mostramos que las fluctuaciones en las intensidades sinápticas hacen más robusta la transmisión de la señal, y mejora sensiblemente la recuperación y recuerdo de información. En concreto, el error asociado con la recuperación de un determinado patrón es notablemente menor en nuestro modelo que en el modelo de Hopfield, y todavía disminuye más cuando aumenta el número de patrones memorizados por la red. El modelo también ha sido estudiado para fluctuaciones alrededor de reglas de aprendizaje asimétricas, mostrando que la asimetría tiende a impedir la propiedad de memoria asociativa.

En el capítulo cuarto se han presentado los resultados correspondientes a una serie de simulaciones Monte Carlo de la versión del modelo general estudiada en el capítulo tercero, es decir, aquella en la que la distribución de sinapsis incluye correlaciones en sus fluctuaciones. Los resultados de estas simulaciones han sido comparados con los que se obtienen del modelo de Hopfield. Mostramos así que, cuando se tienen en cuenta fluctuaciones sinápticas en la red, el sistema es más eficaz a la hora de recuperar un determinado patrón, en el sentido de que el estado estacionario se ve menos alterado por las fluctuaciones térmicas; éstas, en el caso de Hopfield, pueden desestabilizar el correspondiente estado de Mattis. Por otra parte, para los mismos valores de los parámetros P y T , el estado estacionario del modelo con fluctuaciones tiene un solapamiento mayor con un determinado patrón que en el caso de Hopfield. Las simulaciones han servido también para comprobar que la condición de campo autoconsistente utilizada para el desarrollo analítico del modelo es exacta, coincidiendo los datos Monte Carlo con las predicciones analíticas.

En el último capítulo de esta memoria se ha presentado otro modelo cinético de desorden fuera del equilibrio, que se ha estudiado usando un método de campo medio original que tiene en cuenta correlaciones entre espines pertenecientes a un dominio finito del sistema. Este método puede ser aplicado en diferentes órdenes de aproximación y, aunque sólo hemos ilustrado algunas situaciones sencillas, es fácilmente generalizable a órdenes superiores de aproximación. El modelo está basado en una ecuación maestra, en la que las *rates* son una superposición que trata de simular difusión de impurezas, que podría estar entre las causas del comportamiento de determinados sistemas reales conocidos. Al igual que en los modelos anteriores, el estado estacionario del sistema depende de la forma de la *rate* elemental. Los resultados muestran que, para determinadas elecciones de la misma, aparecen ciertas propiedades semejantes a las que se observan en ciertos sistemas reales, como efectos de no equilibrio y cambios de fase reentrantes. Para ciertas elecciones de las *rates* elementales (*rates* K y M) sólo hay cambios de fase de segundo orden entre una fase paramagnética y una fase ferromagnética o antiferromagnética. En otros casos (*rate* V) aparecen cambios de fase de primer orden y puntos tricríticos.

Los resultados obtenidos a lo largo de nuestro trabajo están siendo sometidos a

publicación en diferentes revistas internacionales:

- **Neural networks with fast time-variations of synapses**, J.J. Torres, P.L. Garrido, and J. Marro, *J. Phys. A: Mathematics and General*, por aparecer.
- **Nonequilibrium Hopfield neural network with competing dynamics**, P.L. Garrido, J. Marro, and J.J. Torres, preprint.
- **Effect of correlated fluctuations of synapses in the performance of Hopfield's Neural Network**, J. Marro, P.L. Garrido, and J.J. Torres, preprint.
- **Hopfield neural network with fluctuating synapses**, J.J. Torres, J. Marro, and P.L. Garrido, preprint (1997)
- **On ionic diffusion in spin glasses**, J.J. Torres, P.L. Garrido, and J. Marro, preprint.

Apéndice A

Modelo de Hopfield y técnica de réplicas

El modelo de Hopfield es un modelo mecánico-estadístico constituido por una red Λ_d d -dimensional, en cuyos nudos hay N neuronas representadas por variables de estado $s_{\mathbf{x}} = \pm 1$. Este sistema es capaz de aprender un conjunto de P patrones $\xi^\mu \equiv \{\xi_{\mathbf{x}}^\mu = \pm 1, \mathbf{x} \in \Lambda_d\}$, $\nu = 1, \dots, P$. Los patrones son elegidos de forma aleatoria e independientes entre sí tomando sus elementos, es decir, $\xi_{\mathbf{x}}^\mu$ el valor $+1$ y -1 con igual probabilidad. El sistema viene descrito por el hamiltoniano

$$\mathcal{H}_{\mathbf{J}}(\mathbf{s}) = -\frac{1}{2} \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} J_{xy} s_{\mathbf{x}} s_{\mathbf{y}} + \sum_{\mathbf{x}} \theta_{\mathbf{x}} s_{\mathbf{x}}. \quad (\text{A.1})$$

con $J_{xy} = N^{-1} \sum_{\mu=1}^P \xi_{\mathbf{x}}^\mu \xi_{\mathbf{y}}^\mu$ (Regla de Hebb) y $\theta_{\mathbf{x}} \equiv -\sum_{\mu=1}^P h^\mu \xi_{\mathbf{x}}^\mu$ representando el umbral de excitación asociado a cada neurona. Aquí $\mathbf{h} \equiv (h^1, \dots, h^P)$ es un vector constante de P componentes.

El estudio del modelo de Hopfield depende sustancialmente del comportamiento de la relación $\alpha \equiv P/N$ en el límite $N \rightarrow \infty$ como veremos a continuación.

A.1 Función de partición en el límite $\alpha \rightarrow 0$

Dado que las variables J_{xy} son fijas, el problema de evaluar la función de partición se reduce a calcular

$$Z = \sum_{\mathbf{s}} e^{-\beta \mathcal{H}_{\mathbf{J}}(\mathbf{s})} = e^{-\frac{1}{2}\beta P} \sum_{\mathbf{s}} \exp \left[\frac{\beta}{2N} \sum_{\mu=1}^P \left\{ 2N h^\mu \sum_{\mathbf{x}} \xi_{\mathbf{x}}^\mu s_{\mathbf{x}} + \left(\sum_{\mathbf{x}} \xi_{\mathbf{x}}^\mu s_{\mathbf{x}} \right)^2 \right\} \right], \quad (\text{A.2})$$

donde $\sum_{\mathbf{s}} \equiv \sum_{s_{\mathbf{x}_1=\pm 1}} \cdots \sum_{s_{\mathbf{x}_N=\pm 1}}$. Utilizando la integración gaussiana $\int_{-\infty}^{+\infty} dz \exp(-az^2 + bz) = \left(\frac{\pi}{a}\right)^{1/2} \exp\left(\frac{b^2}{4a}\right)$, con $a = \frac{\beta N}{2}$ y $b = \beta \sum_{\mathbf{x}} \xi_{\mathbf{x}}^{\mu} s_{\mathbf{x}}$, se tiene

$$Z = e^{-\frac{1}{2}\beta P} \left(\frac{\beta N}{2\pi}\right)^{P/2} \int d\mathbf{m} e^{-\frac{1}{2}\beta N \mathbf{m}^2} \prod_{\mathbf{x}} \left[\sum_{s_{\mathbf{x}}=\pm 1} \exp\{\beta(\mathbf{m} + \mathbf{h}) \cdot \boldsymbol{\xi}_{\mathbf{x}} s_{\mathbf{x}}\} \right] \quad (\text{A.3})$$

donde $\mathbf{m} \equiv (m^1, \dots, m^P)$, con m^{μ} $\mu = 1, \dots, P$ variables mudas de integración y $\mathbf{m}^2 \equiv \sum_{\mu=1}^P (m^{\mu})^2$. La sumatoria, dentro del productorio, puede evaluarse de forma sencilla obteniendo

$$Z = \left(\frac{\beta N}{2\pi}\right)^{P/2} \int d\mathbf{m} \exp\{-\beta N f(\mathbf{m})\} \quad (\text{A.4})$$

con $f(\mathbf{m}) = \frac{\mathbf{m}^2}{2} - \frac{1}{\beta N} \sum_{\mathbf{x}} \ln [2 \cosh \{\beta(\mathbf{m} + \mathbf{h}) \cdot \boldsymbol{\xi}_{\mathbf{x}}\}] + \frac{P}{2N}$.

En el límite $\alpha \rightarrow 0$, con P finito y $N \rightarrow \infty$, la integral dada por (A.4) puede ser evaluada mediante el método del punto de silla, que consiste en desarrollar la función $f(\mathbf{m})$ en serie de Taylor alrededor del mínimo y, dado que dicha función aparece multiplicada por N en el argumento de una exponencial, sólo van a contribuir al valor de la integral de forma importante los términos dominantes del desarrollo. Así se obtiene para la energía libre del sistema la expresión

$$\frac{F}{N} = -\frac{1}{N\beta} \ln Z = f(\mathbf{m}_0) + \mathcal{O}\left(\frac{\ln N}{N}\right), \quad (\text{A.5})$$

donde \mathbf{m}_0 es solución a la ecuación

$$\left. \frac{\partial f(\mathbf{m})}{\partial \mathbf{m}} \right|_{\mathbf{m}=\mathbf{m}_0} = 0, \quad (\text{A.6})$$

que da

$$\mathbf{m}_0 = \frac{1}{N} \sum_{\mathbf{x}} \boldsymbol{\xi}_{\mathbf{x}} \tanh\{\beta(\mathbf{m}_0 + \mathbf{h}) \cdot \boldsymbol{\xi}_{\mathbf{x}}\}, \quad (\text{A.7})$$

que es la solución de campo medio del modelo para $\alpha \rightarrow 0$. El resultado dado por (A.5) permite definir la energía libre del sistema como $F = Nf(\mathbf{m})$. Por otra parte se tiene que

$$\sum_{\mathbf{x}} \boldsymbol{\xi}_{\mathbf{x}} \langle s_{\mathbf{x}} \rangle = \lim_{\mathbf{h} \rightarrow 0} \frac{1}{\beta} \frac{\partial}{\partial \mathbf{h}} \ln Z, \quad (\text{A.8})$$

que, utilizando (A.5) y (A.7), implica $\mathbf{m}_0 = N^{-1} \sum_{\mathbf{x}} \boldsymbol{\xi}_{\mathbf{x}} \langle s_{\mathbf{x}} \rangle$, ecuación que constituye la solución estacionaria de campo medio del modelo. Se puede demostrar que la sumatoria sobre todas las neuronas que aparece en (A.7) se puede sustituir por un promedio sobre la distribución de patrones $\boldsymbol{\xi}^{\mu}$, quedando

$$\mathbf{m}_0 = \langle \boldsymbol{\xi}_{\mathbf{x}} \langle s_{\mathbf{x}} \rangle \rangle_{\xi} \quad (\text{A.9})$$

A.2 Red de Hopfield para α finito

En este caso $P = \alpha N$ con $N \rightarrow \infty$ y $\alpha \neq 0$, por lo que el método seguido en la sección anterior no puede llevarse a cabo, dado que el número de integraciones que deberíamos hacer no está acotado; cf. (A.4). Si evaluamos la energía por neurona a partir de (A.1), entonces se debe hacer la hipótesis de que sólo un determinado número k (finito) de patrones tiene solapamientos $\mathcal{O}(1)$ en $N \rightarrow \infty$, mientras que el resto de los solapamientos han de ser $\mathcal{O}(1/\sqrt{N})$ para que dicha energía no diverja; estos k patrones van a ser lo que van a contribuir al umbral $\theta_{\mathbf{x}}$ en la forma dada por (A.1). Ahora hay que poner más cuidado a la hora de promediar sobre la distribución de patrones. En este caso la energía libre del sistema, dado que las sinapsis son fijas, se obtiene en la forma

$$\frac{F}{N} = -\frac{1}{N\beta} \langle \ln Z \rangle_{\xi}. \quad (\text{A.10})$$

El obtener ahora la energía libre tiene la dificultad matemática de evaluar el promedio del logaritmo. Para llevarlo a cabo se utiliza la técnica de réplicas que consiste en utilizar la relación:

$$\ln Z = \lim_{n \rightarrow 0} \frac{Z^n - 1}{n}, \quad (\text{A.11})$$

por lo que el problema se reduce a evaluar $\langle Z^n \rangle_{\xi}$. Como antes (P finito), tras introducir las integraciones gaussianas, obtenemos

$$\langle Z^n \rangle_{\xi} = e^{-\frac{1}{2}\beta n P} \left(\frac{\beta N}{2\pi} \right)^{nP/2} \left\langle \sum_{\mathbf{s}^1 \dots \mathbf{s}^n} \int \left(\prod_{\rho=1}^n \prod_{\mu=1}^P dm_{\rho}^{\mu} \right) X(\mathbf{m}, \vec{s}, \boldsymbol{\xi}) \right\rangle_{\xi}, \quad (\text{A.12})$$

donde

$$\begin{aligned}
X(\mathbf{m}, \vec{s}, \xi) = & \exp \left\{ -\frac{1}{2} \beta N \sum_{\rho=1}^n \sum_{\mu=1}^P (m_{\rho}^{\mu})^2 \right\} \\
& \times \exp \left\{ \beta \sum_{\rho=1}^n \sum_{\mathbf{x}} \left[\sum_{\mu > k} m_{\rho}^{\mu} \xi_{\mathbf{x}}^{\mu} s_{\mathbf{x}}^{\rho} + \sum_{\nu=1}^k (m_{\rho}^{\nu} + h^{\nu}) \xi_{\mathbf{x}}^{\nu} s_{\mathbf{x}}^{\rho} \right] \right\}, \quad (\text{A.13})
\end{aligned}$$

los índices ρ y σ denotan n réplicas del sistema y $\vec{s} \equiv (s^1, \dots, s^n)$. Introduciendo la representación integral $1 = \int_{\mathbb{R}} dq_{ab} \delta(q_{ab} - N^{-1} \sum_{\mathbf{x}} s_{\mathbf{x}}^a s_{\mathbf{x}}^b)$, y la representación integral de la delta de Dirac, se pueden evaluar las integrales sobre las variables m_{ρ}^{μ} , $\rho = 1, \dots, n$ y $\mu > k$, ya que para estas variables la integral factoriza en forma de integrales gaussianas que no llevan ninguna dependencia en las variables neuronales. Así (A.12) queda de la forma

$$\begin{aligned}
\langle Z^n \rangle_{\xi} &= e^{-\beta P n / 2} \int \left(\prod_{\rho=1}^n \prod_{\nu=1}^k dm_{\rho}^{\nu} \right) \int \left(\prod_{\rho \neq \sigma} dq_{\rho\sigma} dr_{\rho\sigma} \right) e^{-\beta N \Phi} \\
\Phi &= \frac{1}{2} \sum_{\rho=1}^n \sum_{\nu=1}^k (m_{\rho}^{\nu})^2 + \frac{\alpha}{2\beta} \text{Tr} \ln \{ (1 - \beta) \mathbf{I} - \beta \mathbf{q} \} + \frac{\alpha\beta}{2} \sum_{\substack{\rho, \sigma \\ \rho \neq \sigma}} r_{\rho\sigma} q_{\rho\sigma}, \quad (\text{A.14}) \\
&\quad - \frac{1}{\beta} \left\langle \ln \text{Tr}_{\vec{s}} e^{\beta H_{\xi}(\vec{s})} \right\rangle_{\xi}
\end{aligned}$$

con

$$H_{\xi}(\vec{s}) \equiv \sum_{\rho=1}^n \left[\frac{\alpha\beta}{2} \sum_{\substack{\sigma=1 \\ \sigma \neq \rho}}^k r_{\rho\sigma} s^{\rho} s^{\sigma} + \sum_{\nu=1}^k (m_{\rho}^{\nu} + h^{\nu}) \xi^{\nu} s^{\rho} \right]. \quad (\text{A.15})$$

En la expresión de Φ que aparece en (A.14), el promedio $\langle \dots \rangle_{\xi}$ denota promedio sobre la distribución de los $\nu = 1, \dots, k$ patrones condensados. \mathbf{I} es la matriz identidad con elementos $\delta_{\rho\sigma}$ y \mathbf{q} es la matriz con elementos $q_{\rho\sigma}$ si $\rho \neq \sigma$ y ceros para $\rho = \sigma$. La energía libre por espín es entonces

$$\begin{aligned}
\frac{F}{N} &= \frac{\alpha}{2} + \frac{\alpha}{2\beta n} \text{Tr} \ln [(1 - \beta) \mathbf{I} - \beta \mathbf{q}] + \\
&\quad + \frac{1}{2n} \sum_{\rho=1}^n \left[\sum_{\nu=1}^k (m_{\rho}^{\nu})^2 + \alpha\beta \sum_{\substack{\sigma=1 \\ \sigma \neq \rho}}^k r_{\rho\sigma} q_{\rho\sigma} \right] - \frac{1}{n\beta} \left\langle \ln \text{Tr}_{\vec{s}} e^{\beta H_{\xi}(\vec{s})} \right\rangle_{\xi}. \quad (\text{A.16})
\end{aligned}$$

Como en el caso de P finito, la energía interna puede ser evaluada en $N \rightarrow \infty$ utilizando la aproximación del punto de silla, que este caso está definido por las condiciones

$$\frac{\partial \Phi}{\partial m_{\rho}^{\nu}} = 0, \quad \frac{\partial \Phi}{\partial r_{\rho\sigma}} = 0, \quad \frac{\partial \Phi}{\partial q_{\rho\sigma}} = 0, \quad (\text{A.17})$$

ecuaciones que dan significado físico a los parámetros de orden m_ρ^ν , $q_{\rho\sigma}$ y $r_{\rho\sigma}$, esto es

$$m_\rho^\nu = \frac{1}{N} \left\langle \sum_{\mathbf{y}} \xi_{\mathbf{y}}^\nu \langle s_{\mathbf{y}}^\rho \rangle \right\rangle_\xi \quad (\text{A.18})$$

como una medida del solapamiento del estado del sistema en la réplica ρ con un patrón almacenado dado, y

$$q_{\rho\sigma} = \frac{1}{N} \left\langle \sum_{\mathbf{y}} \langle s_{\mathbf{y}}^\rho \rangle \langle s_{\mathbf{y}}^\sigma \rangle \right\rangle_\xi, \quad r_{\rho\sigma} = \frac{1}{\alpha} \sum_{\mu=k+1}^P \langle m_\rho^\mu m_\sigma^\mu \rangle_\xi, \quad (\text{A.19})$$

que en el caso de simetría de réplicas ($q_{\rho\sigma} = q$ $\rho \neq \sigma$ y $q_{\rho\rho} = 0$) reducen a los conocidos parámetros de orden de EA y AGS, respectivamente.

Apéndice B

Obtención de hamiltonianos efectivos

En este apéndice desarrollamos las técnicas que hemos utilizado para la obtención de hamiltonianos efectivos tipo Hopfield¹ en el capítulo 2, tanto para el caso de sinapsis simétricas como asimétricas y dimensión d arbitraria.

Estas técnicas están basadas en técnicas desarrolladas para modelos reticulares fuera del equilibrio [38], constituidos por una red Λ_d en cuyos nudos hay definidas unas variables que pueden tomar dos valores $s_{\mathbf{x}} = \pm 1$. En estos modelos, la probabilidad de encontrar al sistema en una determinada configuración de *espines*² $\mathbf{s} = \{s_{\mathbf{x}}, \mathbf{x} \in \Lambda_d\}$ en el tiempo t satisface la ecuación maestra

$$\partial_t P_t(\mathbf{s}) = \sum_{\mathbf{x}} [\bar{\varpi}(\mathbf{s}^{\mathbf{x}}; \mathbf{x}) P_t(\mathbf{s}^{\mathbf{x}}) - \bar{\varpi}(\mathbf{s}; \mathbf{x}) P_t(\mathbf{s})], \quad (\text{B.1})$$

con

$$\bar{\varpi}(\mathbf{s}; \mathbf{x}) \equiv \int \left[\prod_{A \subset \Lambda_d} dJ_A f_A(J_A) \right] \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x}), \quad (\text{B.2})$$

donde $\varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x}) = \Psi(\beta \Delta H^{\mathbf{x}}(\mathbf{s}, \mathbf{J}))$ con $\Delta H^{\mathbf{x}}(\mathbf{s}, \mathbf{J}) = H(\mathbf{s}^{\mathbf{x}}, \mathbf{J}) - H(\mathbf{s}, \mathbf{J})$, siendo $H(\mathbf{s}, \mathbf{J}) = -\sum_{A \subset \Lambda_d} J_A s_A$, $s_A \equiv \prod_{\mathbf{x} \in A} s_{\mathbf{x}}$ un hamiltoniano tipo Ising y Ψ una función positiva verificando la propiedad de balance detallado, esto es $\Psi(X) = e^{-X} \Psi(-X)$. La existencia de una distribución estacionaria caracterizada por un hamiltoniano tipo Ising está garantizada cuando la función Ψ es la rate V , es decir $\Psi(X) = e^{-X/2}$ [38].

¹Hamiltonianos que son idénticos a un hamiltoniano de Ising con los acoplamientos entre neuronas dados por una determinada regla de aprendizaje.

²Esta terminología se debe a que la mayoría de los modelos reticulares han estado asociados al estudio del magnetismo.

Vamos ver como se aplica el resultado anterior al modelo de red de neuronas introducido en el capítulo 2. En primer lugar dicho modelo de red de neuronas está basado en una ecuación maestra idéntica a (B.1) con

$$\bar{\omega}(\mathbf{s}; \mathbf{x}) = \int \left[\prod_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} dJ_{xy} f(J_{xy}) \right] \Psi \{2\beta s_{\mathbf{x}} h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})\} \equiv \overline{\Psi \{2\beta s_{\mathbf{x}} h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})\}} \quad (\text{B.3})$$

donde $h_{\mathbf{x}}(\mathbf{s}, \mathbf{J})$ es el campo local definido en (2.9) con $\theta_{\mathbf{x}} = 0$. Hay que notar que en este caso particular se tiene que A es cualquier subconjunto de dos posiciones cualquiera en la red, es decir $A = \{\mathbf{x}, \mathbf{y}\}$. Para el caso particular de $\Psi(X) = e^{-X/2}$ se tiene

$$\bar{\omega}(\mathbf{s}; \mathbf{x}) = \int \left[\prod_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} dJ_{xy} f(J_{xy}) \right] \prod_{\mathbf{y} \neq \mathbf{x}} \exp\{-\beta J_{xy} s_{\mathbf{x}} s_{\mathbf{y}}\}. \quad (\text{B.4})$$

Dada la propiedad de factorización de la función Ψ , al ser una exponencial, y dado que las variables J_{xy} son independientes se tiene

$$\bar{\omega}(\mathbf{s}; \mathbf{x}) = \prod_{\mathbf{y} \neq \mathbf{x}} \overline{\exp\{-\beta J_{xy} s_{\mathbf{x}} s_{\mathbf{y}}\}}. \quad (\text{B.5})$$

Utilizando el hecho de que $s_{\mathbf{x}} s_{\mathbf{y}}$ sólo puede tomar los valores ± 1 , las rates $\bar{\omega}(\mathbf{s}; \mathbf{x})$ quedan de la forma

$$\bar{\omega}(\mathbf{s}; \mathbf{x}) = \prod_{\mathbf{y} \neq \mathbf{x}} \left[\overline{\cosh \beta J_{xy}} - s_{\mathbf{x}} s_{\mathbf{y}} \overline{\sinh \beta J_{xy}} \right], \quad (\text{B.6})$$

que se puede poner como

$$\bar{\omega}(\mathbf{s}; \mathbf{x}) = \prod_{\mathbf{y} \neq \mathbf{x}} \mathcal{N}_{xy} \exp(-\beta \mathcal{J}_{xy} s_{\mathbf{x}} s_{\mathbf{y}}), \quad (\text{B.7})$$

con $\mathcal{N}_{xy}^2 = \overline{\exp(\beta J_{xy})} \overline{\exp(-\beta J_{xy})}$ y

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln \frac{\overline{\exp(\beta J_{xy})}}{\overline{\exp(-\beta J_{xy})}}. \quad (\text{B.8})$$

La probabilidad de transición dada por la expresión (B.7) verifica la propiedad de balance detallado en la misma forma que una rate de equilibrio, por lo tanto la distribución estacionaria a (B.1) es tipo Gibbs es decir

$$P_{st}(\mathbf{s}) \propto \exp(-\beta \mathcal{H}_{eff}(\mathcal{J}, \mathbf{s})), \quad (\text{B.9})$$

con

$$\mathcal{H}_{eff}(\mathcal{J}, \mathbf{s}) = -\frac{1}{2} \sum_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} \mathcal{J}_{xy} s_{\mathbf{x}} s_{\mathbf{y}}, \quad (\text{B.10})$$

que es un hamiltoniano efectivo tipo Ising, o lo que es lo mismo tipo Hopfield en el contexto de redes de neuronas. El llegar al resultado representado por la ecuación (B.10) sólo es posible suponiendo que la distribución de sinapsis presenta la propiedad de factorización

$$f(\mathbf{J}) = \prod_{\substack{\mathbf{x}, \mathbf{y} \\ \mathbf{x} \neq \mathbf{y}}} f(J_{xy}), \quad (\text{B.11})$$

y que la función Ψ tiene la forma exponencial $\Psi(X) = e^{-X/2}$. Para cualquier otra elección no está garantizada la existencia de tal distribución estacionaria. Como un caso particular está la situación en que tenemos fluctuaciones alrededor de la regla de Hebb que corresponde a tomar

$$f(J_{xy}) = \sum_{\mu=1}^P a_{\mu} \delta \left(J_{xy} - \frac{1}{a_{\mu} N} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} \right), \quad (\text{B.12})$$

que, utilizando (B.8) y tras un sencillo cálculo en el que se utiliza el hecho de que las variables $\xi_{\mathbf{x}}^{\mu}$ sólo pueden tomar los valores ± 1 , implica

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln \left\{ \frac{1 + \rho_{xy} \tanh(\beta\alpha)}{1 - \rho_{xy} \tanh(\beta\alpha)} \right\}, \quad (\text{B.13})$$

con $\rho_{xy} \equiv \frac{1}{P} \sum_{\mu} \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu}$ y $\alpha = PN^{-1}$. Aquí hemos supuesto $a_{\mu} = P^{-1} \forall \mu = 1, \dots, P$. A partir de (B.13) se puede ver que $\mathcal{J}_{xy} = \mathcal{J}_{yx}$ de forma que existe el hamiltoniano efectivo dado por (B.10). Asimismo, para el caso en que consideramos fluctuaciones alrededor de reglas de aprendizaje asimétricas, es decir

$$f(J_{xy}) = \sum_{\mu=1}^P a_{\mu} \delta \left(J_{xy} - \eta_{xy}^{\mu} \right), \quad (\text{B.14})$$

con $a_{\mu} = P^{-1}$ y, por ejemplo $\eta_{xy}^{\mu} = \alpha \left(A_1 \xi_{\mathbf{x}}^{\mu} \xi_{\mathbf{y}}^{\mu} + A_2 \xi_{\mathbf{x}}^{\mu} + A_3 \xi_{\mathbf{y}}^{\mu} + A_4 \right)$, se obtienen, utilizando (B.8) y tras un poco de álgebra, los acoplamientos sinápticos efectivos \mathcal{J}_{xy} dados en (2.39). Sin embargo, en este caso es necesario tener garantizado que $\mathcal{J}_{xy} = \mathcal{J}_{yx}$ para poder tener un hamiltoniano efectivo similar al caso de simetría.

Apéndice C

Diagonalización de matrices \mathcal{Q}

En este apéndice se desarrollan las técnicas llevadas a cabo para la diagonalización de la matriz \mathcal{Q} que aparece en (3.30), particularizada para diferentes *rates* $\Psi(X)$, cuando los estados solución a (3.20) son de la forma $\mathbf{m} = m_n(1, \dots, 1, -1, \dots, -1, 0, \dots, 0)$ ($m_n > 0$), es decir estados que tienen solapamientos con valor $+m_n$ con k patrones, solapamientos con valor $-m_n$ con $(n - k)$ patrones, y el resto de solapamientos son nulos. El estudio ha sido realizado para cada una de las situaciones particulares representadas por las *rates* K , M , y V . En el caso de la *rate* K el procedimiento es trivial pues la matriz resultante es diagonal; cf §3.4. El procedimiento utilizado para la diagonalización de \mathcal{Q} en las otras dos situaciones es el mismo en los dos casos pues dicha matriz tiene una forma similar que es

$$\mathcal{Q}_{\mu\nu} = \mathcal{A}_{\mu\nu} + \eta_\mu \delta_{\mu\nu}, \quad (\text{C.1})$$

donde los elementos de matriz $\mathcal{A}_{\mu\nu}$ y el parámetro η_μ dependen de la *rate* particular. Diagonalizar la matriz cuyos elementos vienen dados por (C.1) consiste en calcular las raíces λ_i , $i = 1, \dots, P$ solución a la ecuación característica $\det(\mathcal{Q} - \lambda\mathbf{I}) = 0$. Por cajas, la matriz $\mathcal{Q} - \lambda\mathbf{I}$ tiene la forma

$$\mathcal{Q} - \lambda\mathbf{I} = \left[\begin{array}{c|c} \mathcal{A}_{n \times n} - (\lambda - \eta_n)\mathbf{I}_{n \times n} & \mathbf{0}_{n \times (P-n)} \\ \hline \mathcal{B}_{(P-n) \times n} & -(\lambda - \eta_{P-n})\mathbf{I}_{(P-n) \times (P-n)} \end{array} \right], \quad (\text{C.2})$$

donde $\mathbf{0}$ es la matriz que tiene todos los elementos nulos y \mathcal{A} y \mathcal{B} son dos matrices cuyos elementos de matriz dependen de la *rate* particular. Hay que hacer notar que en el caso de *rate* V se tiene que \mathcal{B} tiene todos sus elementos de matriz iguales a cero.

En virtud de propiedades, por otra parte bien conocidas del álgebra matricial para matrices que tienen una estructura por cajas igual a la dada en (C.2), la ecuación característica queda en el forma

$$\det(\mathcal{Q} - \lambda \mathbf{I}) = \det \{ \mathcal{A}_{n \times n} - (\lambda - \eta_n) \mathbf{I}_{n \times n} \} \times \det \{ -(\lambda - \eta_{P-n}) \mathbf{I}_{(P-n) \times (P-n)} \} = 0. \quad (\text{C.3})$$

Las raíces de la ecuación característica original, es decir los autovalores, las obtenemos igualando a cero cada uno de los dos factores que aparecen en el segundo miembro de (C.3). El segundo factor da como solución a la ecuación característica $\lambda = \eta_{P-n}$, con multiplicidad $(P - n)$. Por otra parte, el primer factor no es sino el polinomio característico correspondiente a la diagonalización de la matriz \mathcal{A} con autovalores $\bar{\lambda} = \lambda - \eta_n$. Tras un poco de álgebra se obtiene que

$$\det \{ \mathcal{A}_{n \times n} - (\lambda - \eta_n) \mathbf{I}_{n \times n} \} = \Delta^n \det \{ \mathcal{J} - \hat{\lambda} \mathbf{I}_{n \times n} \}, \quad (\text{C.4})$$

donde Δ es una constante que depende de la rate particular, $\hat{\lambda} = (\lambda - \eta_n) \Delta^{-1}$ y \mathcal{J} es una matriz $n \times n$ cuya forma explícita es

$$\mathcal{J} = \begin{bmatrix} 1 & \cdots & 1 & \epsilon & \cdots & \epsilon \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & \epsilon & \cdots & \epsilon \\ \epsilon & \cdots & \epsilon & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \epsilon & \cdots & \epsilon & 1 & \cdots & 1 \end{bmatrix}, \quad (\text{C.5})$$

donde $\epsilon = 1$ para la rate M , y $\epsilon = -1$ para la rate V . De esta forma los k primeros vectores columna de la matriz \mathcal{J} tienen las k primeras componentes iguales a $+1$ y las $(n - k)$ restantes iguales a ϵ , y los $(n - k)$ vectores columna restantes tienen sus k primeras componentes iguales a ϵ y el resto iguales a $+1$. En cualquiera de las dos situaciones, es decir rate M y V , la matriz \mathcal{J} se puede diagonalizar de forma sencilla. Así tras un poco de álgebra sencilla, el determinante que aparece en el segundo miembro de (C.4) (que es el polinomio característico asociado a la matriz \mathcal{J}) se puede evaluar obteniendo

$$\det \{ \mathcal{J} - \hat{\lambda} \mathbf{I}_{n \times n} \} = (-1)^n \hat{\lambda}^{n-1} (\hat{\lambda} - n), \quad (\text{C.6})$$

expresión que tras igualarla a cero nos da otros dos autovalores, a saber $\lambda = \eta_n$ con multiplicidad $(n-1)$, y $\lambda = n\Delta + \eta_n$ con multiplicidad uno. Estos dos valores completan la diagonalización de \mathcal{Q} .

Bibliografía

- [1] J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982)
- [2] J.J. Torres, *Estudio de sistemas magnéticos con frustración dinámica: modelos dinámicos de campo medio de vidrios de espines*, Tesis de Licenciatura 1995.
- [3] D. O. Hebb, *The organization of Behavior: A Neurophysiological Theory*, Ed. Wiley, New York 1949.
- [4] D. Ferster, *Science* **273**, 1812 (1996)
- [5] B. Müller and J. Reinhardt eds., *Neural Networks, An Introduction*, Ed. Springer-Verlag, Berlin 1991.
- [6] C. R. Noback and R. J. Demarest, *Sistema nervioso humano, fundamentos de neurobiología*, Ed. McGraw-Hill México 1980.
- [7] P. Peretto, *An introduction to the modeling of neural networks*, Cambridge University Press 1992.
- [8] R. M. Berne y M. N. Levy, *Fisiología*, Times Mirror de España, Madrid 1993.
- [9] B. Katz, *Nerve, Muscle, and Synapse*, Ed. McGrawHill, New York 1966.
- [10] D.K. Smetters and A. Zador, *Current Biology* **6**, 1217 (1996)
- [11] C. Allen and C.F. Stevens, *Proc. Natl. Acad. Sci. USA*, **91**, 10380 (1994)
- [12] K.E. Sorra and K.M. Harris, *J. Neuroscience* **13(9)**, 3736 (1993)
- [13] A.I. Gulyas, R. Miles, A. Sik, K. Toth, N. Tamamaki, T.F. Freund, *Nature* **366**, 683 (1993)

- [14] N. Otmakhov, A.M. Shirke and R. Malinow, *Neuron* **10**, 1101 (1993)
- [15] Y. Goda and C.F. Stevens, *Proc. Natl. Acad. Sci. USA* **91**, 12942 (1994)
- [16] A. Mallart and A.R. Martin, *J. Physiol.* **196**, 593 (1968)
- [17] K.J. Stratford, K. Tarczy-Hornoch, K.A.C. Martin, N.J. Bannister and J.J.B. Jack, *Nature* **382**, 258 (1996)
- [18] C.F. Stevens and Y. Wang, *Nature* **371**, 704 (1994)
- [19] R. Penrose, *The Emperor's New Mind*, Oxford University Press, New York 1989.
- [20] D.J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev. A* **32**, 1007 (1985)
- [21] B. Derrida, E. Gardner and A. Zippelius, *Europhys. Lett.* **4**, 167 (1987)
- [22] H. Gutfreund and M. Mezard, *Phys. Rev. Lett.* **61**, 235 (1988)
- [23] F.A. Tamarit and E.M.F. Curado, *J. Stat. Phys.* **62**, 473 (1991)
- [24] D. Horn, *Physica A* **200**, 594 (1993)
- [25] C.R. Da Silva, F.A. Tamarit and E.M.F. Curado, *Int. J. Mod. Phys. C* **7**, 43 (1996)
- [26] F. Moss, *Phys. World* **10**, 15 (1997)
- [27] J.J. Torres, P.L. Garrido and J. Marro, to be published.
- [28] J.R.L. De Almeida and D.J. Thouless, *J. Phys. A: Math. and Gen.* **11**, 983 (1978)
- [29] J.J. Torres, J. Marro and P.L. Garrido, to be published.
- [30] F. Edwards and P.W. Anderson, *J. Phys. F* **5**, 965 (1975)
- [31] J.L. Van Hemmen and R. Kühn, *Phys. Rev. Lett.* **57**, 913 (1986)
- [32] J.L. Van Hemmen, *Phys. Rev. A* **36**, 1959 (1987)

- [33] J.L. Van Hemmen, D. Gensing, A. Huber and R. Kühn, *J. Stat. Phys.* **50**, 231 (1988)
- [34] A. Crisanti and H. Sompolinsky, *Phys. Rev. A* **37**, 4865 (1988)
- [35] P.L. Garrido and J. Marro, *J. Stat. Phys.* **74**, 663 (1994)
- [36] P.L. Garrido and J. Marro, *Phys. Rev. Lett.* **62**, 1929 (1989)
- [37] A.I. López-Lacomba, P.L. Garrido and J. Marro, *J. Phys. A: Math. and Gen.* **23**, 3809 (1990)
- [38] P.L. Garrido and M.A. Muñoz, *Phys. Rev. E* **48**, R4153 (1993)
- [39] A.I. López-Lacomba, J. Marro, *Europhys. Lett.* **25**, 169 (1994)
- [40] S. Kirkpatrick and D. Sherrington, *Phys. Rev. Lett.* **35**, 1972 (1975)
- [41] D.J. Amit, H. Gutfreund and H. Sompolinsky, *Ann. Phys.* **173**, 30 (1987)
- [42] L. Pastur, M. Shcherbina and B. Tirozzi, *J. Stat. Phys.* **74**, 1161 (1994)
- [43] D.C. Mattis, *Phys. Lett. A* **56**, 421 (1976)
- [44] P.L. Garrido and J. Marro, in “Artificial Neural Networks”, *Lecture Notes in Computer Science*, vol. 540, A. Prieto editor, Ed. Springer-Verlag, Berlin 1991, pp.25-33.
- [45] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*, Cambridge University Press, Cambridge, in press.
- [46] L. Elsgoltz, *Ecuaciones Diferenciales y Cálculo Variacional*, Ed. URSS, Moscú (1994).
- [47] T.M. Apostol, *Mathematical Analysis*, Ed. Addison-Wesley Publishing, Reading Massachusetts (1957).
- [48] J. von Neumann, in *Cerebral Mechanics of Behavior*, L.A. Jeffres editor, Ed. Wiley, New York 1951.

- [49] W. Feller, *An Introduction to Probability Theory and Its Applications*, Ed. John Wiley & Sons, New York 1968.
- [50] K. Binder and A.P. Young, *Rev. Mod. Phys.* **58**, 801 (1986)
- [51] M. Mezard, G. Parisi, and M.A. Virasoro, *Spin Glass Theory and Beyond*, Ed. World Scientific, Singapore 1987.
- [52] M.B. Weissman, *Rev. Mod. Phys.* **65** (1993)
- [53] G. Parisi, *Philos. Mag. B* **71**, 471 (1995)
- [54] P.L. Garrido and J. Marro, *Europhys. Lett.* **15**, 375 (1991)
- [55] M.F. Thorpe and D. Beeman, *Phys. Rev. B* **14**, 188 (1976)
- [56] H. Maletta, and P. Convert, *Phys. Rev. Lett.* **42**, 108 (1979)
- [57] S.M. Shapiro, C.R. Fincher, A.C. Palumbo, and R.D. Parks, *Phys. Rev. B.* **24**, 6661 (1981)
- [58] M.A. Manheimer, S.M. Bhagat, and H.S. Chen, *J. Magn. Magn. Mater* **38**, 147 (1983)
- [59] B. Gnedenko, *The Theory of Probability*, Ed. Mir Publishers, Moscow 1975.
- [60] J.M. Gozález-Miranda, A. Labarta, M. Puma, J.F. Fernández, P.L. Garrido, and J. Marro, *Phys. Rev. E* **49**, 2041 (1994)
- [61] J.J. Alonso, and J. Marro, *Phys. Rev. B* **45**, 10408 (1992)
- [62] J.J. Alonso and J. Marro, *J. Phys.: Condens. Matter* **4**, 9309 (1992)
- [63] H. Haken, *Synergetics*, Ed. Springer-Verlag, Berlin 1983.