



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

SIMULAZIONI NUMERICHE DI MODELLI A GRANA GROSSA
PER DNA A DOPPIA E TRIPLA ELICA

Candidato:
Matteo Carlassara
Matricola: **1026744**

Relatore:
Prof. Antonio Trovato

Indice

1	Introduzione	1
1.1	DNA e sue proprietà biofisiche	1
2	Modellizzazione teorica	3
2.1	Osservazioni sul DNA	3
3	Simulazioni	5
3.1	Metodo di Metropolis	5
3.2	Metodo di Wang-Landau	6
3.3	Reticoli diretti studiati	6
3.3.1	1+1	6
3.3.2	2+1	7
3.3.3	3+1	7
3.4	Implementazione dei metodi di simulazione	9
4	Risultati Termodinamici	15
	Conclusioni	21
	Bibliografia	23

Capitolo 1

Introduzione

1.1 DNA e sue proprietà biofisiche

L'acido desossiribonucleico, meglio noto con l'acronimo DNA, è una macromolecola organica che accomuna la maggioranza degli esseri viventi. È situato nel nucleo di tutte le cellule e ha il compito fondamentale di dirigere la cellula in tutte le fasi della vita, essendo depositario di tutta l'informazione genetica (questo ruolo in alcuni virus viene svolto dall'acido nucleico RNA). Riuscire a comprenderne la composizione chimica e da essa ricreare un modello molecolare adeguato (1953 per opera di Watson-Crick) richiese quasi un secolo di studi, che proseguono tutt'oggi nel tentativo di decodificarne l'intera informazione genetica.

Il DNA teorizzato nel modello Watson-Crick è un polimero formato da due filamenti complementari, ognuno creato dalla sovrapposizione di nucleotidi accoppiati; questi sono molecole composte a loro volta da una base azotata, un gruppo fosfato e uno zucchero pentoso.

Ogni parte del nucleotide svolge un ruolo ben preciso, la base azotata favorisce lo *stacking* tra nucleotidi vicini grazie alla formazione di legami fosfoesterici tra il carbonio 3' del desossiribosio e il 5' dello zucchero nel nucleotide successivo, generando un'asimmetria nel legame la cui conseguenza è che il DNA ha un verso di lettura. Le basi azotate sono quattro, suddivise in pirimidine, citosina (C) e timina (T), e purine, adenina (A) e guanina (G). In questo modello sono possibili solamente gli accoppiamenti del tipo A-T e C-G, il primo attraverso tre legami a idrogeno, il secondo con due. In presenza di due filamenti vicini, con basi azotate complementari per ogni coppia di nucleotidi presenti, la geometria dei legami fa assumere al polimero la struttura nota come doppia elica destrorsa, che presenta una bassa flessibilità e di cui si ha riscontro sperimentale. Il numero delle coppie di nucleotidi che compongono la catena dipendono dalla specie cui un organismo appartiene. Per alcuni batteri si hanno dalle $10^7 \div 10^{10}$ unità, con il raggiungimento della lunghezza macroscopica del centimetro se disteso linearmente, mentre nell'uomo si raggiungono le dimensioni del metro.

Da queste osservazioni si intuisce che devono esistere dei meccanismi che permettano al DNA di cambiare la sua natura rigida e lineare, per occupare un volume di qualche μm^3 nel nucleo, in cui lo si vede avvolto attorno a particolari proteine chiamate istoni, formando dei rocchetti, che si compattano in una struttura nota come cromatina. Attraverso gli esperimenti si è notato che la conformazione spaziale sopra descritta (nota come B-DNA) non è l'unica possibile, se il liquido nucleare ha un $pH < 5$ si osserva una catena sinistrorsa (A-DNA), oppure grazie a particolari modificazioni chimiche, prende la conformazione destrorsa a Z-DNA. In laboratorio inoltre si sono studiate altre sei conformazioni diverse.

La soluzione fu proposta negli anni sessanta da Karst Hoogsteen grazie alle osservazioni sulla cristallizzazione del DNA per $pH < 5$, nel quale la citosina protonata (C^+) forma un legame ad idrogeno in meno con G, mentre A-T legano in punti diversi rispetto al modello precedente. Entrambi gli accoppiamenti sono energeticamente sfavoriti rispetto a Watson-Crick nelle condizioni cellulari.

Il suo contributo è ancor più importante per spiegare l'esistenza in vivo del DNA a tripla elica, osservata per la prima volta nel 1957. In una conformazione di B-DNA, l'avvitamento dei filamenti crea due solchi, uno maggiore di $2.2nm$ in cui sono esposte le basi azotate e uno minore di $1.2nm$. Nel solco maggiore può unirsi in alcuni tratti un terzo filamento, nel caso in cui esso sia composto per la maggioranza di basi puriniche¹, formando legami Hoogsteen $T - (AT)_{WC}$, $C - (GC)_{WC}$ e $G - (GC)_{WC}$, $A - (AT)_{WC}$. La tripla elica è ancora argomento di studio e si ipotizza abbia un ruolo centrale nei processi di duplicazione e lettura, fornendo un punto di aggancio per molte proteine [6,7], tra cui l'RNA-polimerasi .

Il DNA si presenta come due filamenti legati per valori di $2 < pH < 12$, assumendo a seconda delle situazioni conformazioni diverse. Studiando il DNA a livello termodinamico si riscontra un altro fattore fondamentale per la stabilità del legame: la temperatura. Le catene sono unite se si trovano al di sotto di un valore critico chiamato temperatura di *melting*² T_m . Il suo valore sperimentale dipende dalla lunghezza della catena e dalla concentrazione delle basi azotate presenti: la rottura di due o tre legami a idrogeno richiede un salto energetico diverso, dunque la diversa concentrazione di Adenina e Guanina porta (nel limite termodinamico) ad una temperatura diversa. Nei pressi di T_m nel campione avviene una transizione netta di tipo sigmoidale, per temperature leggermente maggiori le catene sono slegate. Chimicamente in queste condizioni si assiste alla scissione dei legami tra le basi azotate e allo svolgimento dei filamenti; ogni filamento ha una grande flessibilità (aumenta l'entropia $\Delta S > 0$). L'ambiente circostante sfavorisce energeticamente la situazione: i filamenti sono costretti a esporre le regioni idrofobe alla soluzione acquosa e a formare legami con essa; inoltre lo *stacking* tra le basi è più debole (globalmente l'energia aumenta per la singola catena $\Delta E > 0$).

Queste semplici osservazioni permettono di valutare l'energia libera del sistema $\Delta F = \Delta E - T\Delta S$, la quale diventa negativa solamente quando la temperatura supera un valore critico.

In questa tesi si vogliono confrontare le proprietà di *melting* tra DNA a due e tre catene.

¹Le purine hanno due superfici di legame.

²Valore per cui il 50 % del campione è denaturato.

Capitolo 2

Modellizzazione teorica

2.1 Osservazioni sul DNA

Il DNA descritto dal modello Watson-Crick risulta a livello molecolare rigido e con una struttura aperiodica delle basi azotate. Se invece lo si osserva con una scala di ingrandimento minore, tale da non riuscire a distinguere le basi azotate, si presenta un certo grado di flessibilità e una struttura omogenea di filamenti attorcigliati che si possono pensare come formati da monomeri efficaci, ognuno con lo stesso numero di basi. Se si vuole procedere con una simulazione tenendo in considerazione i particolari sulla scala del nm, si deve procedere componendo una catena avente per unità (monomeri) i nucleotidi, che nel caso più semplice sono 10^7 elementi, e tutte le loro reciproche interazioni. Il voler ricercare le proprietà termodinamiche con questa tipologia di sistemi è ancora troppo dispendioso a livello computazionale.

La soluzione da adottare per queste ricerche è di focalizzare l'attenzione sul fattore che influenza maggiormente i valori medi termodinamici, nel nostro caso la flessibilità della catena, andando in seguito a stimare i pesi dei fattori microscopici trascurati da utilizzare per correggere la stima iniziale.¹ Si procede considerando come monomeri tratti interi della catena, la cui lunghezza (nota come lunghezza di Kuhn L_K) è tale da permettere piena libertà di movimento angolare sulle giunture tra due unità successive².

Ogni monomero viene considerato come un segmento, un filamento composto da N monomeri si dice che ha lunghezza L . In assenza di vincoli angolari, è possibile paragonare la posizione delle unità a un *random walk* di N passi. Tenendo in considerazione che ogni singolo filamento si trova in soluzione e la poca probabilità che a livello microscopico le basi azotate del primo monomero siano complementari a quelle di un secondo in una stessa catena, si esclude la formazione di un legame durante un ripiegamento generato dal movimento casuale. Nel modello si esclude questo genere di casistiche, imponendo che ogni monomero non si può posizionare sullo spazio occupato da un qualsiasi altro della stessa catena (principio di volume escluso). Si sta attuando una restrizione dei siti visitabili dai monomeri (*self-avoiding walk*).

Sempre considerando il caso di singolo filamento immerso in soluzione, esso non può avere forti interazioni a grandi distanze. Si decide di limitare le interazioni possibili a quelle di solo contatto tra catene diverse, andando ad attribuire per ogni tratto in comune un contributo in energia negativo $\epsilon < 0$.

L'insieme di tutte queste condizioni definisce un modello di *self and mutually avoiding walk*.

¹In questo lavoro di tesi ci si focalizzerà solo sulla prima parte.

²Nel caso di DNA a doppia elica si ha $L_K \cong 100\mu m$, nel caso di un singolo filamento $L_K \cong 6nm$.

Per questo lavoro di tesi si è deciso di attuare un'ulteriore restrizione ai possibili movimenti angolari delle giunzioni: si considerano i monomeri di lunghezza unitaria, i quali possono visitare i siti di un reticolo diretto a base quadrata (si veda Sezione 3.3). Per esempio, in un reticolo di due dimensioni un filamento di lunghezza $L-1$ può raggiungere la lunghezza L aggiungendo un monomero in direzione $(1,0)$ o $(0,1)$ rispetto alla posizione dell'ultimo elemento della catena.

In generale si è posto che il punto di partenza per tutte le catene fosse l'origine del reticolo e che, per ogni *step*, sia possibile scegliere di avanzare, aggiungendo un nuovo monomero con direzione arbitraria e verso positivo al termine di ognuno dei filamenti esistenti, oppure di annullare la mossa precedente, rimuovendo l'ultimo elemento dai filamenti. In aggiunta a tali condizioni, il contatto tra le catene può avvenire solo se:

1. due tratti di filamenti diversi occupano lo stesso sito reticolare ed essi sono entrambi all' n -esima componente, ovvero ognuna delle catene ha richiesto n unità per raggiungere quella posizione; in questo caso si pone che il contatto abbia un contributo in energia di -1 .
2. tre filamenti si legano con le stesse richieste precedenti; si è però distinto il contributo energetico nel caso il contatto sia triplo, con un contributo energetico -2 , oppure due filamenti legati e il terzo no, con un contributo pari a -1 . Indicando con 1, 2 e 3 le catene, si è posto di limitare i legami possibili a 1-2-3, 1-2 e 1-3, trascurando la possibilità 2-3, per meglio rappresentare la situazione reale.³

Le posizioni reticolari possono essere descritte in funzione della lunghezza L delle catene: considerando il fascio di rette parallele $y = -x + L$, i siti che appartengono all'intersezione tra il reticolo e una retta del fascio possono essere raggiunti dalle catene solo se queste sono composte da L monomeri. Essendo L una variabile del sistema, si possono descrivere i punti in funzione di x e L attraverso una semplice rotazione delle coordinate.

In generale, per questa tipologia di problemi, la dimensione D del sistema viene espressa in notazione $d + 1$ (es. $D = 2 \Rightarrow d = 1$ e la dimensione del reticolo si indica $1 + 1$) poiché sostituendo ad L il tempo immaginario, le catene rappresentano delle traiettorie di particelle quantistiche in dimensione d , per cui tramite l'equazione della diffusione, ci si potrebbe ricondurre - in linea di principio⁴- all'equazione di Schrödinger per una buca di potenziale, la cui profondità dipende dalla temperatura del sistema. In entrambe le visioni del problema, le particelle quantistiche/catene si trovano sempre in uno stato legato se $D = 2$, mentre per $D > 3$ ciò avviene solo se la buca è abbastanza profonda (quindi solo in questa situazione esiste una temperatura finita di *melting*).

Analizzando il modello di polimeri appena costruito, si riesce a trovare una soluzione matematica a meno di quadrature per due catene, indipendentemente dalla dimensione reticolare [3,4]; solamente nel caso $D = 2$ è possibile la risoluzione esatta del problema [5].

³Si sono effettuate anche simulazioni con tutti e quattro i legami, i risultati termodinamici ottenuti attraverso il metodo di Wang-Landau presentavano curve con evidenti fluttuazioni. La causa si presuppone essere la condizione di piattezza poco accurata.

⁴Il passaggio dal reticolo discreto al sistema continuo non è ovvio.

Capitolo 3

Simulazioni

Il modello descritto può essere simulato all'equilibrio termodinamico grazie ad un processo di Markov, in cui il sistema dato all' n -esima iterazione dipende solamente da quello dello stato precedente. Indicando con Γ_i lo stato iniziale in cui si trova il sistema e con Γ_j quello finale che si vuole raggiungere, si indica la probabilità di transizione tra questi due stati come P_{ij} , che ha una struttura matriciale data la generalità degli indici i e j . Questa matrice è tale che tutti gli elementi sulla diagonale siano definiti positivi, mentre quelli fuori diagonale possono assumere valori positivi o nulli nel caso in cui lo stato j non sia raggiungibile con un solo passo dallo stato i . Per garantire che i risultati della simulazione convergano alla distribuzione di Boltzmann $P_B(\Gamma_i) = \frac{e^{-\beta E(\Gamma_i)}}{\sum_j e^{-\beta E(\Gamma_j)}}$ dove $E(\Gamma_i)$ indica l'energia dello stato i , è necessario che vengano rispettate due proprietà:

1. *Ergodicità*: per ogni coppia i, j di stati iniziale e finale, $\exists n \in \mathbb{N}$ t.c. $(P_{ij})^n > 0$. Ovvero partendo da Γ_i la condizione è soddisfatta se in un numero n finito di iterazioni posso raggiungere Γ_j .
2. *Bilancio dettagliato*: la probabilità di avanzare dallo stato Γ_i allo stato Γ_j deve essere la stessa del caso inverso: $P_B(\Gamma_i)P_{ij} = P_B(\Gamma_j)P_{ji}$. Se soddisfatta questa condizione, allora si può dimostrare che il sistema è anche stazionario (ovvero che il valore della probabilità di essere in una configurazione tende asintoticamente al valore di $P_B(\Gamma_i)$ al procedere della simulazione).

3.1 Metodo di Metropolis

Si separa la probabilità di transizione in due componenti, quella di selezionare lo stato di prova partendo da quello iniziale - indicata con P_{ij}^{sel} - e la probabilità di accettare lo stato di prova dato lo stato di partenza - P_{ij}^{acc} -, tali per cui $P_{ij} = P_{ij}^{sel} P_{ij}^{acc}$. Nella prima è necessario accertarsi che sia soddisfatta la condizione di $P_{ij}^{sel} = P_{ji}^{sel}$, ma il metodo di Metropolis [1] non fornisce informazioni su come ricavarla. La seconda viene definita nel modo seguente:¹

$$P_{ij}^{acc} = \min \left\{ 1, \frac{P_B(\Gamma_j)}{P_B(\Gamma_i)} \right\}$$

Il metodo di Metropolis si utilizza a temperatura fissata.

¹Si dimostra che soddisfa la condizione di *bilancio dettagliato* e di conseguenza anche la stazionarietà.

3.2 Metodo di Wang-Landau

Per mezzo di una generalizzazione del metodo di Metropolis, il seguente algoritmo ricava la densità degli stati nello spazio dell'energia, indicata con $g(E)$.

Si vuole ottenere come stato stazionario $\frac{1}{g(E)}$ e si procede con il seguente approccio iterativo [2]:

1. essendo $g(E)$ una quantità incognita del problema, la si pone inizialmente 1 per ogni valore di energia;
2. si utilizza il metodo di Metropolis generalizzato per passare dalla configurazione iniziale E_i a quella finale E_j con la probabilità di transizione data da

$$p_{ij}^{acc} = \min \left\{ 1, e^{\frac{g(E_i)}{g(E_j)}} \right\}.$$

Una simulazione così realizzata dovrebbe riprodurre un *random walk* nello spazio delle energie;

3. alla fine di ogni iterazione, che venga accettata o rifiutata la configurazione di prova, si aggiorna la densità degli stati inerente all'energia in cui si trova il sistema, moltiplicandola per un fattore $f > 1$ definito inizialmente come $f_0 = e$; in un istogramma cumulativo $H(E)$ si salva la posizione nello spazio dell'energia in cui si trova il sistema, aggiungendo 1. In questo modo si permette alla simulazione di visitare tutto lo spazio delle energie;
4. si continua ad iterare sui punti precedenti fino a quando l'istogramma $H(E)$ non presenta alcun elemento nullo e soddisfa la condizione di piatezza: ogni suo valore deve essere superiore all'80% del valor medio $\langle H \rangle$ calcolato nell'ultima iterazione;
5. se le condizioni del punto precedente sono soddisfatte si procede azzerando $H(E)$ e si aggiorna il fattore di moltiplicazione $f_{i+1} = \sqrt{f_i}$;
6. si esce dal ciclo una volta che f raggiunge un valore molto prossimo all'unità; nel sistema studiato è stato posto come limite $f = \exp(10^{-8})$.

Il metodo non rispetta il *bilancio dettagliato* per la stima di $g(E)$, in quanto si procede aggiornando passo per passo la probabilità di selezione, ma converge molto rapidamente al valore corretto mano a mano che $f \rightarrow 1$.

Nota $g(E)$, si possono direttamente calcolare l'energia libera di Gibbs e l'entropia, da esse poi ricavare tutte le quantità termodinamiche d'interesse (si veda Capitolo 4).

3.3 Reticoli diretti studiati

3.3.1 1+1

È un reticolo quadrato di coordinate $(x, y) \in \mathbb{N}^2$. Per poter mettere in luce le proprietà derivanti dalla lunghezza della catena è stata effettuata una rotazione per porre l'asse x parallelo alla retta $y = -x$ attraverso il sistema:

$$\begin{cases} x_1 = \sqrt{2}(x - \frac{L}{2}) & 0 \leq x \leq L \\ y_1 = \frac{\sqrt{2}}{2}L \end{cases}$$

Nel caso si consideri il corrispettivo sistema quantistico, y_1 rappresenta direttamente il tempo immaginario. In Figura 3.1, 3.2 nelle immagini da 3. a 8. si sono utilizzate

le coordinate (x_1, y_1) con la proiezione degli istogrammi in funzione della lunghezza sull'asse x_1 .² Il numero di punti del reticolo che hanno la stessa lunghezza è $L + 1$.

3.3.2 2+1

È un reticolo cubico di coordinate $(x, y, z) \in \mathbb{N}^3$. Secondo gli stessi criteri posti per 1+1, il sistema della rotazione adottato è il seguente:

$$\begin{cases} x_1 = \frac{1}{\sqrt{2}} \left[L - \left(\frac{1}{\sqrt{3}} + 1 \right) (x + y) \right] \\ y_1 = \frac{1}{\sqrt{2}} (x - y) \\ z_1 = \frac{L}{\sqrt{3}} \end{cases}$$

Lo *slicing* del reticolo in funzione della lunghezza avviene attraverso i piani

$$z = L - x - y \quad \text{con } 0 \leq x \leq L, 0 \leq y \leq L.$$

Il numero di punti che stanno su ogni piano è $\frac{(L+1)(L+2)}{2}$

3.3.3 3+1

È un reticolo quadridimensionale di coordinate $(x, y, z, t) \in \mathbb{N}^4$. In questa situazione si è optato di tenere le tre coordinate spaziali (x, y, z) con le stesse considerazioni adottate per sistema ruotato in 2+1, mentre t si interpreta come un tempo³. Lo *slicing* avviene per spazi tridimensionali del tipo: $z = (L - t) - x - y$; ovvero fissato L e $0 \leq t \leq L$, la catena ha una lunghezza spaziale di $L' = L - t$; L'ultimo componente della catena appartiene al piano $z = L' - x - y$. I siti spaziali visitabili per ogni L sono tutti quelli compresi tra l'origine e il piano definito da L stessa, con un numero di punti dato da $\frac{(L+1)(L+2)(L+3)}{6}$.

²Nel caso di tre catene tutte tra loro interagenti, per la lunghezza unitaria il sistema si trova necessariamente in uno stato legato; per poter applicare Wang-Landau è stato necessario restringere lo spazio dell'energia, togliendo il caso con zero contatti per ogni lunghezza.

³Si ricordi che tutte le grandezze, compresi i punti del reticolo, sono trattate senza unità di misura.

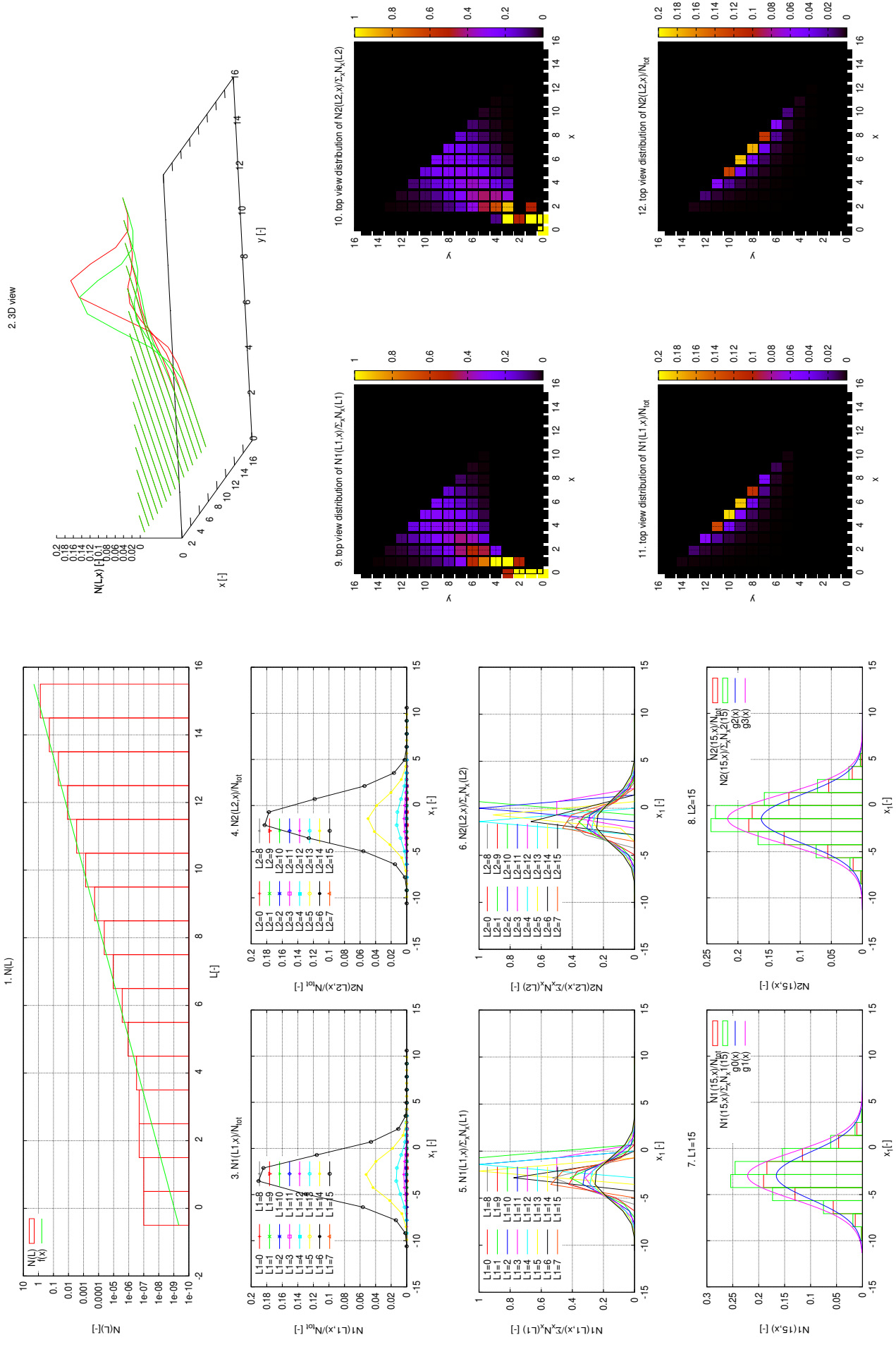


Figura 3.1: Simulazione con $N_C = 2, D = 2, T = \infty, p = \frac{4}{5}, L_{max} = 15, N_{tot} = 10^7$

3.4 Implementazione dei metodi di simulazione

In questo lavoro si utilizzano delle mosse Monte Carlo gran-canoniche, per cui la lunghezza L delle catene diviene un parametro fondamentale al pari dell'energia E . Come accennato nel Capitolo 2 per ogni *random walk* a tutte le catene si può:

1. aggiungere una nuova unità, la cui direzione viene determinata in modo casuale, sempre con verso positivo;
2. eliminare l'ultima unità, portando tutte le catene da lunghezza L a $L-1$;
3. se si trovano a lunghezza nulla (nell'origine), si impone che non possano decrescere ulteriormente; se si è raggiunta la lunghezza massima imposta per la simulazione, si vieta di crescere ancora⁴.

Nel primo punto, la probabilità di crescita deve essere la stessa per tutte le possibili combinazioni tra le direzioni assumibili dai monomeri aggiunti alle catene, che sono D^{N_C} ove $D = d + 1$ (modi di crescita nel reticolo diretto) e N_C il numero delle catene. Si indica con p la probabilità totale che il sistema scelga di crescere, data dalla sommatoria delle probabilità di raggiungere una determinata configurazione direzionale p_i , definita come $p_i = \frac{p}{D^{N_C}}$; infine con $1 - p$ si indica la probabilità di decrescere. Per soddisfare la condizione su P^{sel} è necessario che la probabilità di tornare indietro sia la stessa di avanzare, quindi che $1 - p = \frac{p}{D^{N_C}}$ da cui si ricava $p = \frac{D^{N_C}}{1 + D^{N_C}}$ ⁵.

In Figura 3.1 si può osservare un esempio di simulazione con $p = \frac{4}{5}$ per due catene non interagenti (a temperatura $T = \infty$). La notazione usata è la seguente: $N(L)$ è l'istogramma dei conteggi cumulativi per ogni lunghezza assunta dalle catene, nel grafico 1. viene normalizzato per il numero totale delle iterazioni N_{tot} e riportato in scala logaritmica nelle ordinate; $N_1(L_1, x)$ e $N_2(L_2, x)$ sono rispettivamente per la catena 1 e 2 le frequenze di tutte le posizioni reticolari (grafici da 2. a 12.), normalizzati con N_{tot} oppure per il valore cumulativo dato da $N(L)$.

Nel grafico 1. dopo un primo transiente iniziale l'istogramma segue un andamento esponenziale dato da $D^{N_C(L-L_{max})}$ come atteso a $T = \infty$ poiché il numero di stati per L_{max} fissato è $D^{N_C L_{max}}$. I grafici 2. 3. 4. 11. 12. fanno notare come il sistema, una volta raggiunta la lunghezza massima, rimanga in media fermo nel sito raggiunto, che risulta diverso per le due catene e spostato dalla posizione centrale.

I grafici 5. 6. 9. 10. mettono in luce che per $L \leq 5$ i siti visitati per ogni *slicing* sono al massimo due e al più si riscontra un solo passaggio (motivo per il quale in 1. le prime posizioni non seguono l'andamento esponenziale). A seconda di dove si trovano le catene per lunghezza 5, quasi tutti i siti raggiungibili con la scelta di crescita hanno almeno un passaggio (si vedano 9. 10.). La spiegazione sta nel fatto che la crescita è globalmente più probabile della decrescita, per cui il maggior numero di siti visitati sta a lunghezza massima secondo una crescita esponenziale, la cui posizione finale dipende non più solo dal passaggio precedente, ma anche dal sito a $L = 5$ da cui il fattore $1-p$ inizia a contribuire.

Il *bilancio dettagliato* ostacola la visita di tutti i siti reticolari e l'equiprobabilità che la catena possa trovarsi in una lunghezza compresa tra il valore minimo e quello massimo. Il problema si risolve ponendo $D^{N_C} = 1$ da cui risulta: $p = 0.5$ ⁶. Questa nuova scelta di probabilità è posta alla base di tutte le simulazioni successive (si osserva in Figura 3.2 che i profili sono perfettamente gaussiani come atteso per un cammino casuale e $N(L)$ risulta piatto).

⁴Si noti che in entrambi questi casi la catena può rimanere ferma.

⁵es.: $D = 2, N_C = 2 \rightarrow p = \frac{4}{5}$

⁶nel caso in cui si volessero confrontare i risultati con altre strategie simulative è necessario reintrodurre il fattore D^{N_C} dato che si è modificato il *bilancio dettagliato*.

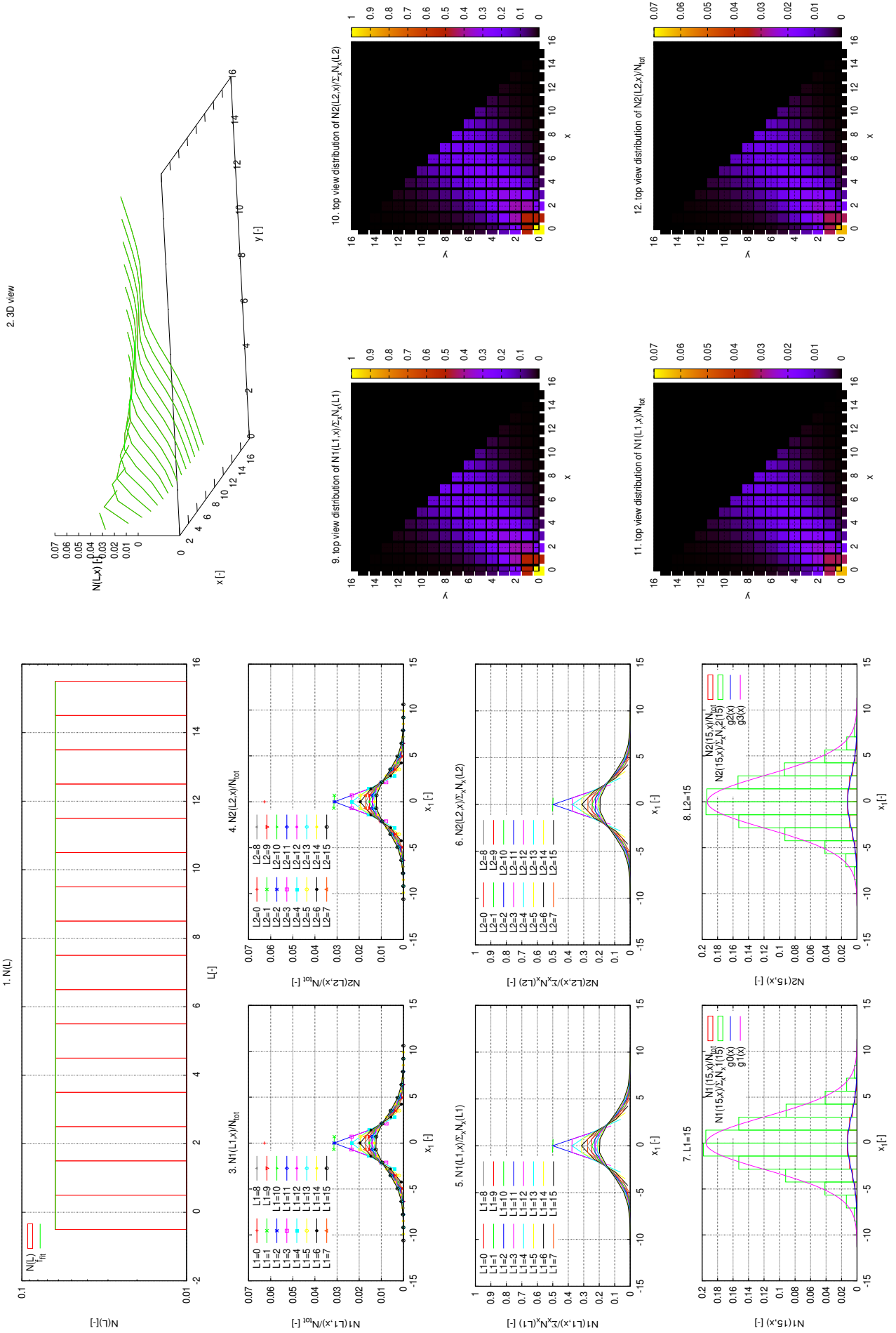


Figura 3.2: Simulazione con $N_C = 2, D = 2, T = \infty, p = \frac{1}{2}, L_{max} = 15, N_{tot} = 10^7$

Il *random walk* implementato non tiene conto dell'interazione tra le due catene, si è anzi nella condizione di accettare ogni configurazione di prova proposta essendo un Metropolis a $T = \infty$. Il risultato di Figura 3.2 permette di verificare che il processo di Markov è implementato correttamente.

Portando la temperatura ad un valore finito (per esempio $T = 1$) per la scelta di tornare indietro, se si sta rimuovendo un monomero di contatto si procede con una probabilità di selezione e^{-1} . Si osserva in Figura 3.3(a) una restrizione della larghezza a metà altezza rispetto al sistema libero e una crescita esponenziale di $N(L)$ del tipo: e^{aL+b} , i cui parametri sono ottenuti tramite un fit:

$$a_{fit} = 0.50992 \pm 0.00010 \text{ e } b_{fit} = -8.566 \pm 0.001481.$$

Ad a_{fit} corrisponde il rapporto tra l'energia libera di Gibbs per la coppia di basi e la temperatura del sistema. Nella fase legata tra due catene (in $D = 2$ la temperatura di *melting* è ∞) l'energia e quindi l'energia libera sono proporzionali al numero dei contatti e quindi al numero di coppie di basi L . Con la scelta fatta per p (nota 6) l'energia libera della fase *single-stranded* non dovrebbe invece crescere linearmente con L . La bontà del fit lineare è di nuovo un indice della corretta implementazione del processo di Markov.

Il metodo di Wang-Landau lavora con lo spazio delle possibili configurazioni di energia che per questo sistema si è assunta essere linearmente dipendente dal numero dei contatti n_c e segue le definizioni fornite alla Sezione 2.1 nei punti 1. e 2. Il valore massimo di n_c dipende dalla lunghezza raggiunta dai filamenti e può variare da zero ad un massimo proporzionale ad L con le seguenti possibilità:

1. in presenza di due catene il contributo energetico fornito da ogni nuovo contatto è -1 , n_c ha valore massimo di L ;
2. in presenza di tre catene il contributo vale -2 nel caso di contatto triplo, -1 se singolo, n_c ha valore massimo di $2L$.

La struttura creata è di tipo triangolare nel piano (n_c, L) come mostrato in Figura 3.4. Per ogni punto di questo spazio viene definito l'istogramma $H_L(n_c)$ e il logaritmo della densità degli stati $\ln(g_L(n_c))$ per contenere le divergenze dovute agli elevati valori assunti da $g(n_c)$ ⁷. La Figura 3.3(b) mostra una crescita lenta di $N(L)$ che forma, anche per piccoli valori di lunghezza, un picco ben definito. Per $L_{max} = 15$ i risultati termodinamici non sono molto accurati, essi migliorano se si aumenta la lunghezza massima raggiungibile per le catene; nelle simulazioni effettuate tale valore è stato posto $L_{max} = 32$.

Per tutte le dimensioni presentate nella Sezione 3.3 sono state simulate 1⁸, 2, 3 catene e implementato il metodo Wang-Landau per le ultime due.

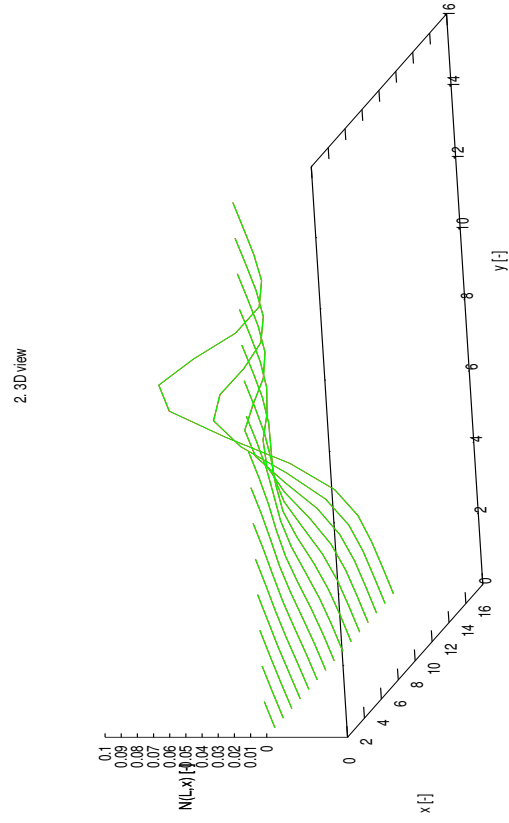
La Figura 3.5 mostra il confronto tra l'energia libera $F_L(T)$ ottenuta dal metodo Metropolis tramite il fit lineare per $L = 15$ per $T = 1$ e la rispettiva curva data dal metodo Wang-Landau attraverso la relazione: $F_L(T) = -T \ln(\sum_{n_c} g_L(n_c) e^{-\beta n_c})$. Per poter mettere in relazione i due risultati è stato necessario reintrodurre il fattore D^{LN_C} tramite le seguenti relazioni calcolate per due catene in due dimensioni:

$$a = -(a_{fit} + N_C \cdot \ln(2)) \cdot L$$

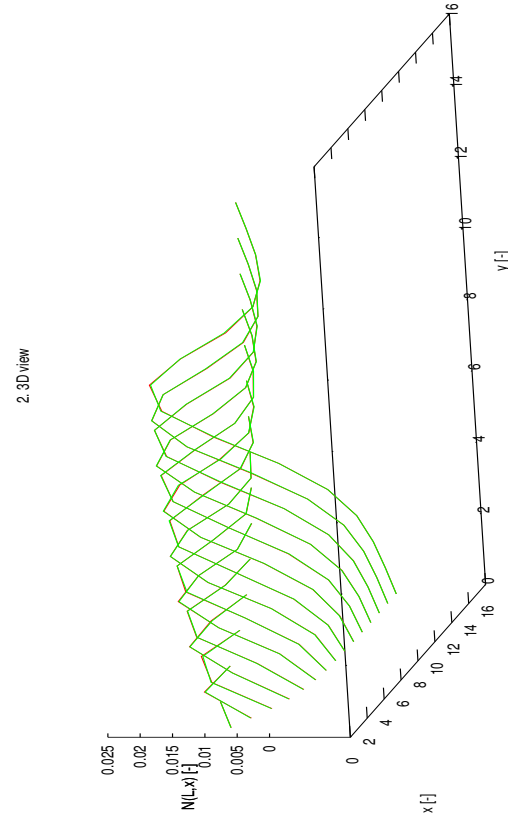
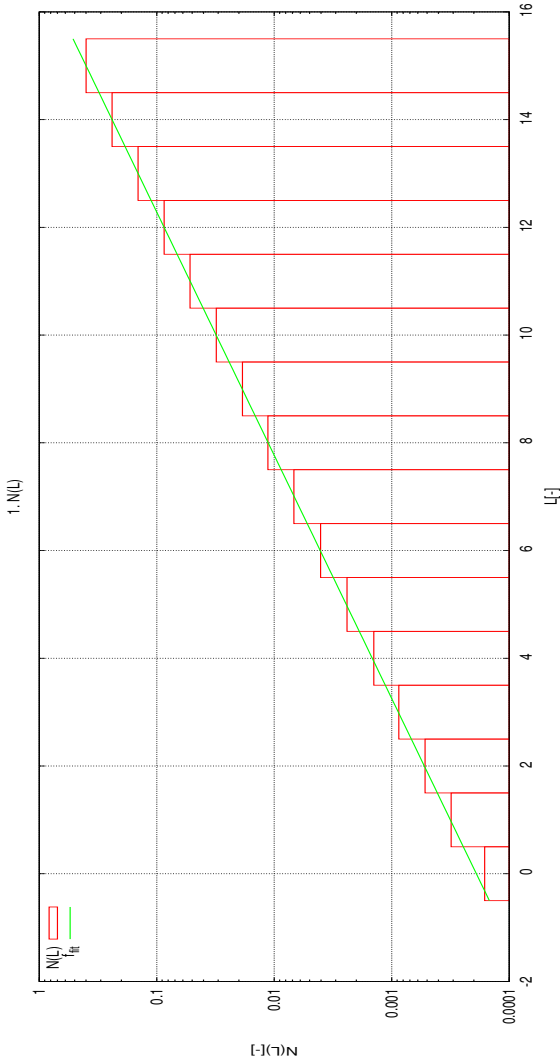
$$F_L^{new}(T) = F_L(T) + T \cdot (\ln(g_L(L)) - L \cdot \ln(D))$$

⁷Il fattore moltiplicativo viene sommato al logaritmo. Esso parte da 1 e viene dimezzato quando sono soddisfatte le condizioni di piattezza (si veda Sezione 3.2). La catena al termine di ogni ciclo Wang-Landau viene fatta ripartire dall'origine e si azzerava $N(L)$. N_{tot} diviene ora il numero di iterazioni necessarie per uscire dall'ultimo ciclo ed è diverso da quello posto nei casi precedenti.

⁸Il singolo filamento è stato utilizzato come prova della corretta implementazione dei metodi, considerando il passaggio per i siti vicini alla diagonale con un contributo in energia.



(a) Simulazione con metodo Metropolis a $T = 1, N_C = 2, D = 2, p = \frac{1}{2}, L_{max} = 15, N_{tot} = 10^7$



(b) Simulazione con il metodo Wang-Landau per $N_C = 2, D = 2, p = \frac{1}{2}, L_{max} = 15$

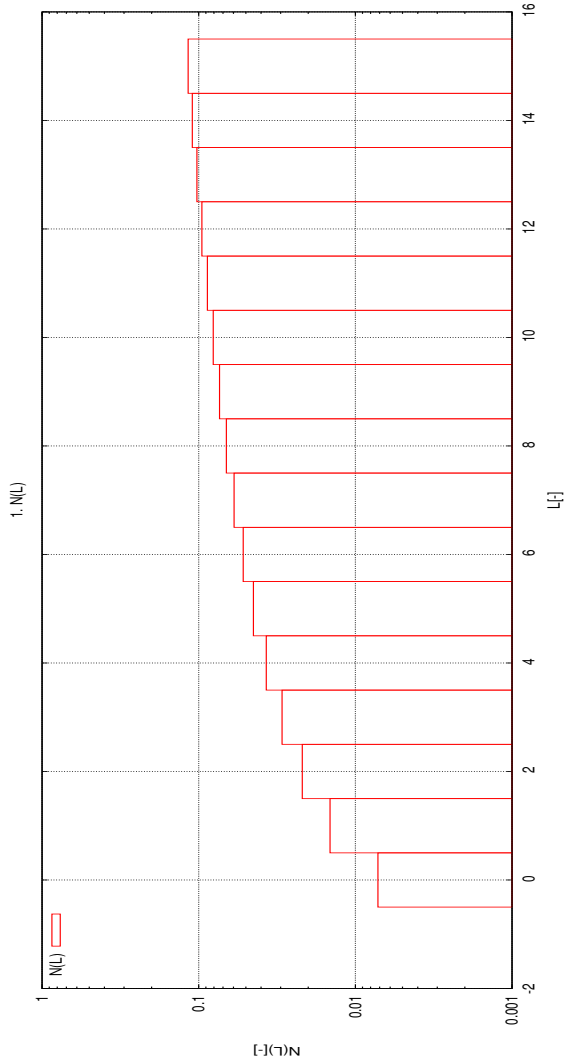


Figura 3.3: Confronto tra i due metodi di simulazione implementati.

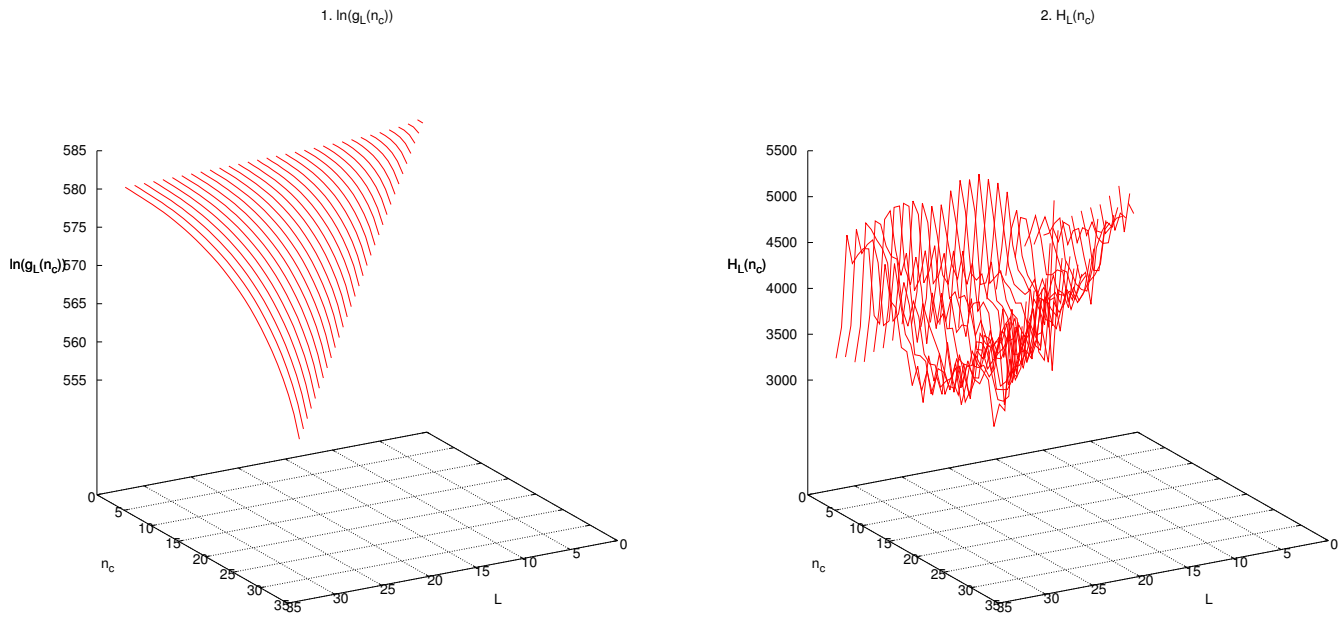


Figura 3.4: Simulazione con il metodo Wang-Landau per $N_C = 2, D = 2, p = \frac{1}{2}, L_{max} = 32$

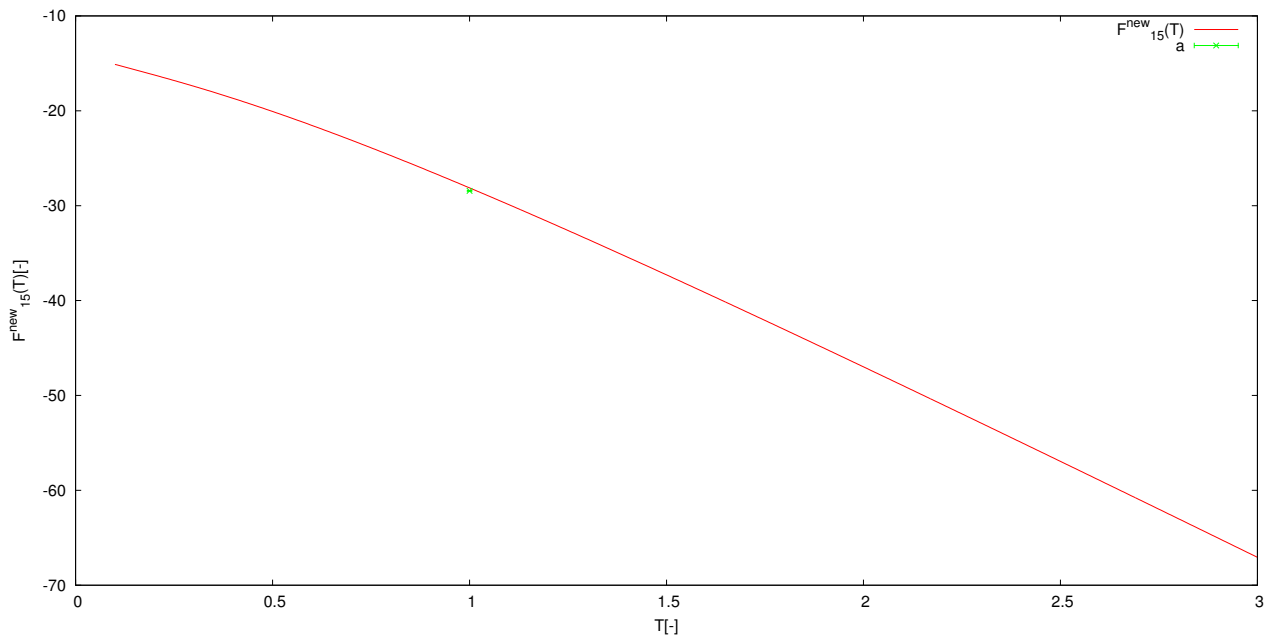


Figura 3.5: Confronto tra l'energia libera calcolata attraverso Wang-Landau e Metropolis per $L=15$

Capitolo 4

Risultati Termodinamici

Grazie al metodo di Wang-Landau si è ricavata la densità degli stati $g_L(n_c)$, da cui si ottengono i valori medi gran-canonici dell'energia e del calore specifico per ogni L:

$$\langle E_L \rangle_T = \frac{\sum_{n_c} n_c g_L(n_c) e^{-\beta n_c}}{\sum_{n_c} g_L(n_c) e^{-\beta n_c}} \quad , \quad C_L(T) = \frac{\langle E_L^2 \rangle_T - \langle E_L \rangle_T^2}{T^2}$$

Per il caso di due catene, il modello teorico sviluppato [3,4,5] dice che per un sistema in $D = 2$ o $D = 3$ non avviene una transizione di *melting* a temperatura finita, in corrispondenza il calore specifico non presenta picchi che divengono nel limite termodinamico e l'energia rimane sempre negativa. Nel caso di $D = 4$ invece esiste una transizione di *melting* per un valore finito della temperatura, nel limite termodinamico si attende che il picco del calore specifico diverga e che l'energia si annulli per $T > T_m$. Il sistema sviluppato in questa tesi, presenta una lunghezza massima delle catene finita, imposta per necessità computazionale, che non permette lo studio del limite termodinamico con $L_{max} \rightarrow \infty$. Si possono comunque ricavare delle informazioni sul comportamento del sistema nel limite, andando a studiare le variazioni in funzione della lunghezza L dei valori dell'energia media e del calore specifico.

I grafici nelle Figure 4.1 e 4.2 presentano dei valori riscaldati per la dimensione del reticolo e la lunghezza, nel caso del calore specifico, mentre solo per la lunghezza nel caso del valore medio dell'energia. In questo modo è resa più visibile la convergenza delle curve al crescere di L a quella curva ottenuta nel limite termodinamico. Nei grafici la lunghezza della catena è indicata dalla paletta a fianco di ogni figura.

All'avvicinarsi della temperatura allo zero, l'energia tende al valore assegnato come contributo energetico per i contatti (-1 singolo e -2 triplo), ovvero le due catene sono legate per ciascuna coppia di basi.

Per $T < 10$ il valor medio dell'energia assume un comportamento sigmoidale che all'aumentare della temperatura fa avvicinare le curve allo zero di energia, senza raggiungerlo. In questa regione nei grafici del calore specifico si osserva un picco che risulta sempre più definito all'aumentare della dimensione del reticolo. Più precisamente per $D = 2$ e $D = 3$ ¹ la larghezza del picco è più grande dell'altezza e all'aumentare di L l'altezza si riduce e la larghezza aumenta. Per $D = 4$ il picco è molto più definito, all'aumentare di L l'altezza aumenta e la larghezza diminuisce.

Per $T \rightarrow \infty$ la curva tende asintoticamente ad un valore di energia minore di zero che diminuisce in modulo all'aumentare di L . Si distinguono anche per questa regione delle

¹Per questa dimensione nel caso di tre catene si osservano delle leggere fluttuazioni sulla curva del calore specifico: si è verificato, controllando i dati riportati dalle simulazioni, che tali variazioni sono totalmente indipendenti dall'approssimazione fatta sul fattore moltiplicativo f ; la causa si ipotizza essere la bassa condizione di piatezza imposta per $H_L(n_c)$.

piccole differenze: nel caso di $D = 2$ e $D = 3$ - indipendentemente dal numero delle catene - si raggiungono valori di $\langle E_L \rangle < -0.1$, in $D = 4$ invece ci si avvicina maggiormente allo zero in energia senza però raggiungerlo.

In Figura 4.3 i grafici 1. 2. riportano i confronti tra i valori medi in energia e tra i calori specifici - che sono stati riscalati ulteriormente per il contributo energetico dei contatti - relativi a tutte le simulazioni effettuate. Si osserva che l'andamento dipende esclusivamente dalla dimensione D del reticolo e non dal numero delle catene N_C .

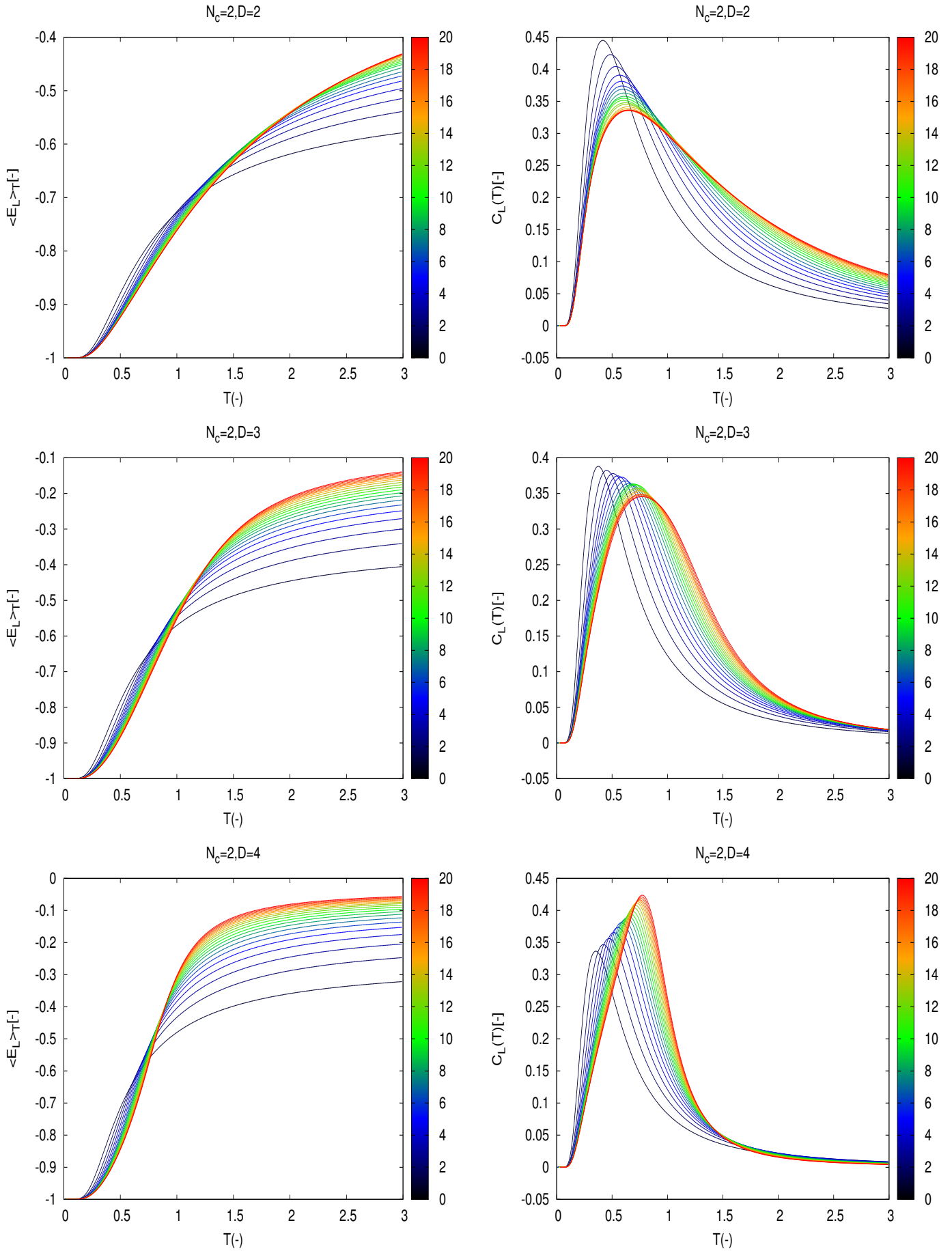


Figura 4.1: Curve del valor medio di energia e del calore specifico per $N_C = 2$ al variare di L e D .

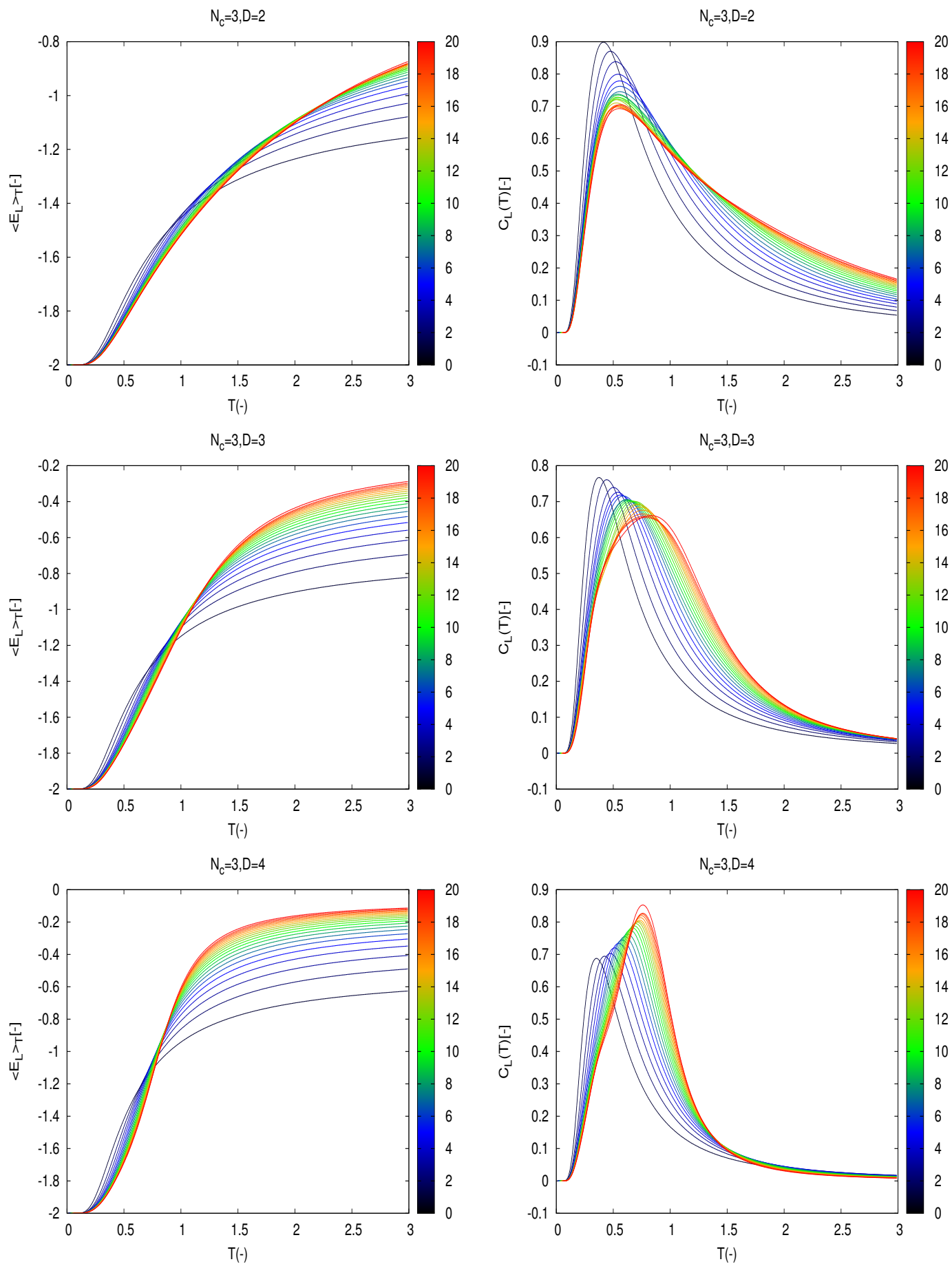


Figura 4.2: Curve del valor medio di energia e del calore specifico per $N_C = 3$ al variare di L e D .

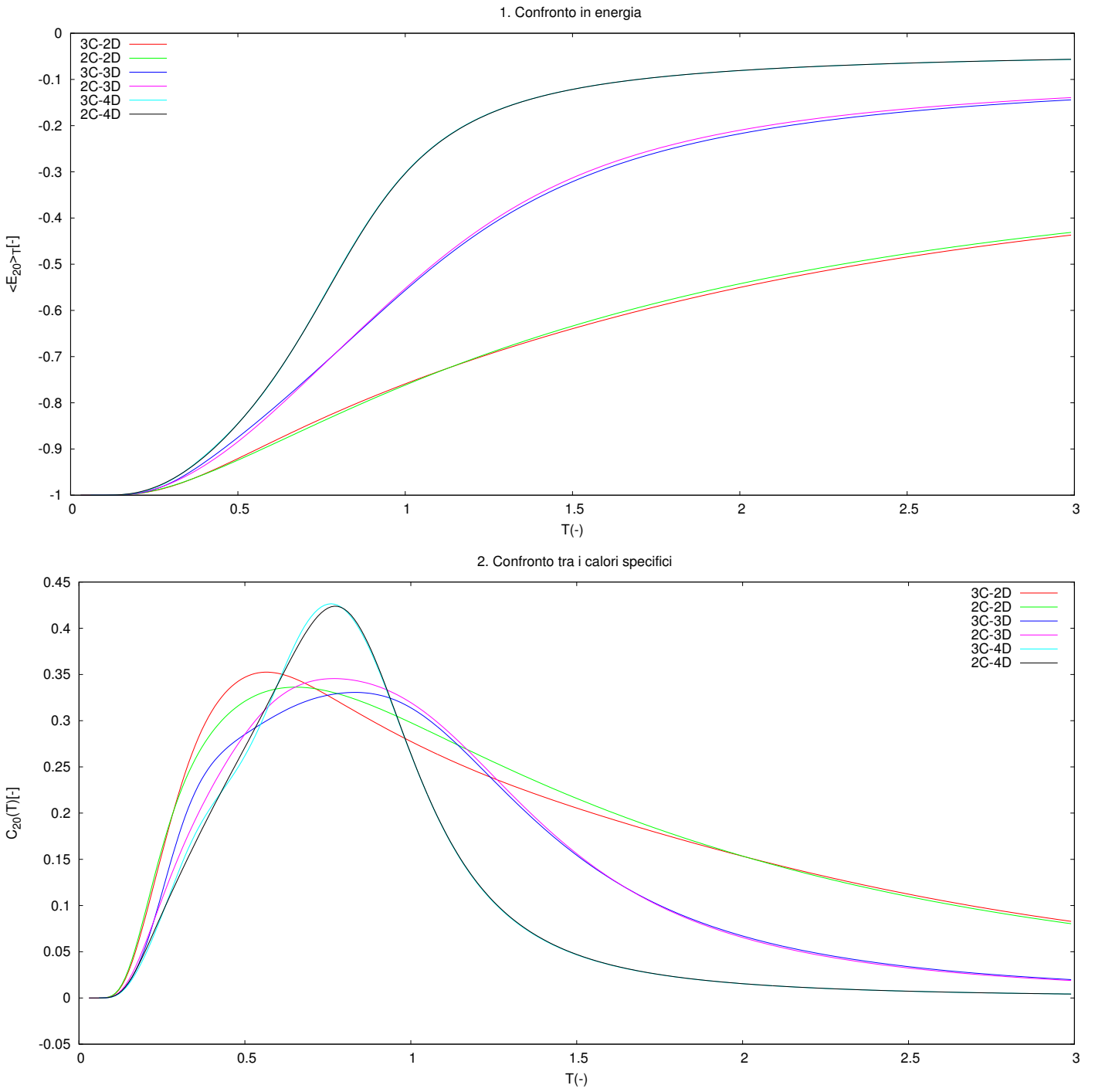


Figura 4.3

Conclusioni

I risultati termodinamici mostrano che il valore massimo delle curve del calore specifico si trova tra $0.5 < T < 0.9$. Nel caso di due catene in $D = 4$, il valore della temperatura trovato da altre strategie [3] è di $T_m = 0.79[-]$; In questa tesi si è riscontrato un valore di $T_m = 0.84[-]$.

I grafici dei valori medi di energia ci permettono di concludere che è avvenuto un salto energetico, con valori finali di energia negativi; non forniscono però informazioni su una possibile transizione del sistema a causa di valori di L troppo piccoli.

I grafici del calore specifico sono in accordo con quanto atteso dalla teoria [3,4,5] presentando un picco più marcato all'aumentare di L per $D = 4$ da cui si conclude che è avvenuta una transizione netta del sistema; per $D = 2$ e $D = 3$ il picco è sempre meno definito al crescere di L , quindi non è avvenuta una transizione netta del sistema.

La presenza di due o tre catene influenza il sistema attraverso il contributo energetico dei contatti - diverso nei due casi - che pesa sulle quantità termodinamiche come un fattore moltiplicativo e non cambia qualitativamente il *trend* delle curve, almeno per le lunghezze simulate in questo lavoro.

In sintesi, con questa tesi si è sviluppato un nuovo metodo di simulazione gran-canonico che permette di ricavare tutte le grandezze termodinamiche e verificato che riesce a riprodurre i risultati attesi da un metodo di Metropolis standard a temperatura infinita, Figura 3.1 e finita, Figura 3.2. Insieme al metodo di Wang-Landau è stato possibile verificare che la temperatura di *melting* del problema studiato ha un valore finito solo per $D \geq 4$.

Alcuni dei problemi osservati ma non approfonditi in questa tesi, possono essere punti di partenza per lavori futuri, per esempio migliorare la stima ottenuta per la temperatura di *melting* per due catene in $D = 4$ e cercare di confermare l'esattezza delle ipotesi effettuate in [4,5] e verificare che se soddisfatte, la temperatura di *melting* è più bassa per tre catene rispetto a due catene; si suggerisce di aumentare la lunghezza massima delle catene simulate e di porre una condizione di piattezza più restrittiva nel metodo di Wang-Landau.

Bibliografia

- [1] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. *Equations of State Calculations by Fast Computing Machines*. Journal of Chemical Physics. 21 (6): 1087–1092, 1953
- [2] Wang, F.; Landau, D. P. Phys. Rev. Lett. **86** 2050, 2001
- [3] Marenduzzo, D.; Trovato, A.; Maritan, A. Phys. Rev. E. **64** 031901, 2001
- [4] Maji, J.; Bhattacharjee, S.M.; Seno, F.; Trovato, A. *When a DNA triple helix melts: An analogue of the Efimov state* N. J. Phys. **12** 083057, 2010
- [5] Maji, J., Bhattacharjee, S.M., Seno, F., Trovato, A. *Melting behavior and different bound states in three-stranded DNA models* Phys. Rev. E **89** 012121, 2014
- [6] Frank-Kamenetskii, M.D.; Mirkin, S.M. An. Rev. Biochem. 1995. 64:65-95
- [7] Sinden, R.R. *DNA Structure and Function*. Academic Press, 1994