

AUTO-ENSAMBLAJE Y BALDOSAS DE ADN

7 de septiembre de 2004

Índice general

1	Introducción	2
1.1	Computación y embaledosados	3
1.2	ADN y reacciones básicas	3
1.3	Cristalización: un proceso de auto-ensamblaje	5
2	Definiciones y resultados básicos	12
2.1	El modelo de auto-ensamblaje	12
2.2	Resultados básicos	15
2.3	Complejidad de producir únicamente un \mathcal{T} -ensamblaje	22
3	Determinismo	23
3.1	Lattice distributivo	23
3.2	Condiciones suficientes para producir únicamente un \mathcal{T} -ensamblaje	25
4	Complejidad en la programación de rectángulos	30
4.1	Temperatura constante y rectángulos “delgados”	30
4.2	Temperatura constante y rectángulos “gruesos”.	35
4.3	Aumentando la Temperatura a $\mathcal{T} = 4$	38
5	Conclusiones	41

Capítulo 1

Introducción

Una bacteria *E. Coli* mide aproximadamente 1 nanómetro cúbico. Gruesamente, el mismo tamaño de un transistor en chips de computadores modernos. En vez de almacenar un único bit de información, la *E. Coli* almacena más de un megabyte de información genética en su ADN. Ésta es sólo una muestra de cómo la tecnología biológica puede superar vastamente la tecnología industrial actual.

Leonard Adleman descubrió que la densidad del almacenamiento de información en el ADN, combinada con el paralelismo de las reacciones químicas, tiene potencial para realizar cálculos matemáticos enormes. En 1994, Adleman [1] resolvió una pequeña instancia del problema del camino Hamiltoniano usando para realizar los cálculos, moléculas de ADN.

Este nuevo método posee numerosas ventajas prácticas. Por ejemplo, 1 milímetro cúbico de solución puede almacenar 10^{21} bits de información codificada como ADN, y reacciones químicas, tales como actividad enzimática o simple hibridación, pueden procesar esta información en paralelo.

El método que Adleman usó se basa en una “búsqueda combinatorial” a través de la creación de una librería de hebras de ADN en cuyas secuencias se encuentran codificadas las soluciones potenciales del problema. Por medio de técnicas de biología molecular tales como PCR (polimerasa chain reaction), separación magnética y gel electrophoresis, la librería puede ser “filtrada” para encontrar la solución. En esencia, esta técnica permite transformar un problema que requiere tiempo exponencial en uno que requiere espacio exponencial. Preguntas tales como cuáles son las mejores técnicas de laboratorio para la computación molecular o cómo usarlas de manera óptima, están aún abiertas.

En estos términos es particularmente interesante considerar la variedad de mecanismos que ocurren en el control de los procesos intracelulares, en aras de entender la estructura de los algoritmos que con este tipo de métodos se puede implementar. Una teoría general de computación molecular debería sugerir el cómo diseñar sistemas bioquímicos capaces de efectuar cualquier algoritmo.

1.1 Computación y embaldosados

El Problema del embaldosamiento (The tiling Problem) fue introducido por Hao Wang en 1961 [10] para estudiar una pregunta de lógica matemática. La pregunta es simple: dado un conjunto finito de baldosas geométricas (por ejemplo polígonos), determinar cuándo éstas pueden ser dispuestas (usando cada baldosa tantas veces como sea necesario) para cubrir el plano completo sin superponerse. La respuesta a esta simple pregunta es que no existe ningún algoritmo que entregue, para todos los casos, la respuesta correcta: el tiling Problem es indecidible [5]. La demostración de este resultado se basa en la construcción de patrones de embaldosamiento capaces de simular una máquina de Turing. Es precisamente en esta relación entre embaldosados y computación donde se centra nuestra atención y nuestro interés.

Para ilustrar mejor esta relación consideremos un ejemplo: en la figura 1.1 se muestra un conjunto de 7 baldosas, que embaldosan el plano al mismo tiempo que implementan un contador binario. El objetivo es simplemente contar 1, 2, 3, 4, 5, ..., pero en base 2. Es decir, 1, 10, 11, 100, 101, La forma de las baldosas representa la información usada en el cómputo: los lados de arriba y de abajo codifican el valor del bit en el contador, mientras que los lados de la derecha y de la izquierda se usan para indicar si el bit debe o no cambiar su valor. La baldosa semilla S desata el cómputo y las 2 baldosas de *acotamiento* proveen las condiciones iniciales. Las baldosas de *acotamiento* por abajo codifican los bits iniciales ...000. Las baldosas de *acotamiento* por la derecha producen una serie de “comandos” para incrementar el último bit en cada columna. Por último, las cuatro baldosas *de regla*, que representan la regla de cálculo, implementan la siguiente lógica: “si el bit de la derecha no es de cambio, el bit mantiene su valor, si el bit de la derecha es de cambio, 0 cambia a 1, y 1 cambia a 0.

Es importante notar que, para que en el ejemplo anterior el patrón resultante sea el correcto, es necesario que las baldosas *de acotamiento* se ensamblen a la baldosa semilla antes de usar las baldosas *de regla* y que las baldosas *de regla* sean agregadas solamente cuando ya están presentes las baldosas de la derecha y la de abajo. Estas condiciones previenen la formación de embaldosados en ausencia de las baldosas *de acotamiento* (como un embaldosado periódico que contenga únicamente baldosas con sus cuatro lados de tipo 0) como también previenen la formación de embaldosados incompletos. Eric Winfree [11] mostró cómo, usando la temperatura, estas condiciones pueden ser implementadas en un cierto tipo de “baldosas moleculares” o “baldosas de ADN” construidas en base a secuencias específicas de hibridación y diseñadas para “auto-ensamblarse” espontáneamente y formar embaldosados que realicen cálculos o adopten una forma o figura predefinida.

1.2 ADN y reacciones básicas

La química fundamental del ADN está basada en una doble hélice formada por dos hebras unidas bajo el principio de complementariedad de Watson-Crick. Cada hebra es una secuencia

1.2. ADN Y REACCIONES BÁSICAS

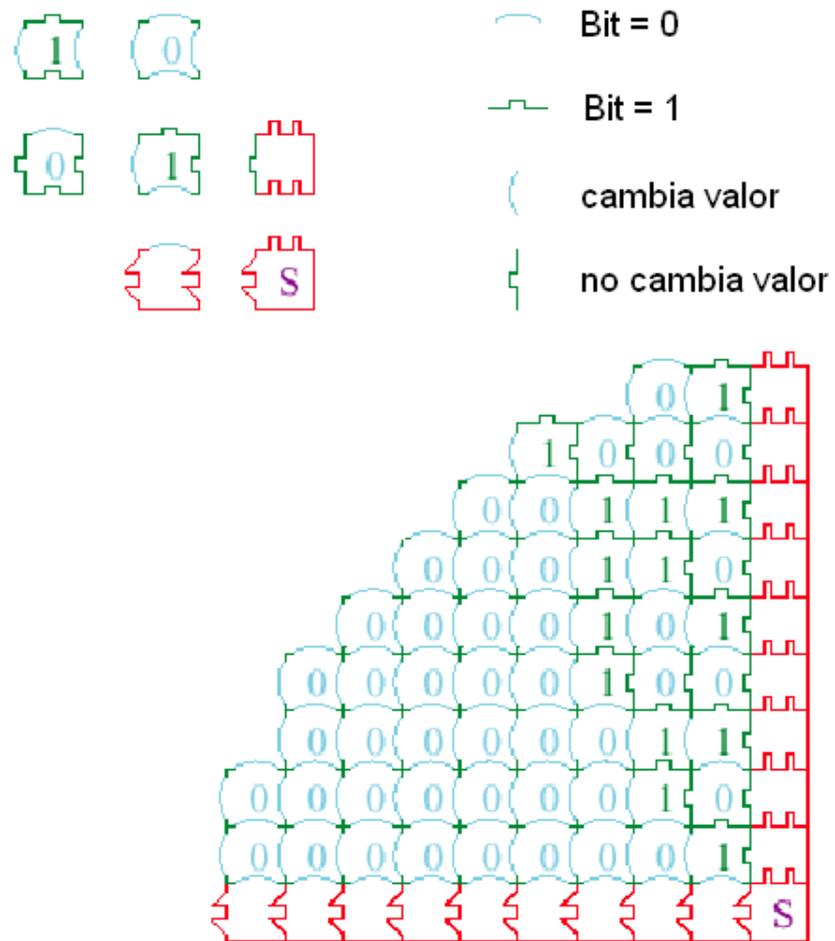


Figura 1.1: Un sistema de 7 baldosas que forman un embaledado aperiódico, las filas en el embaledado representan enteros, escritos en binario.

1.3. CRISTALIZACIÓN: UN PROCESO DE AUTO-ENSAMBLAJE

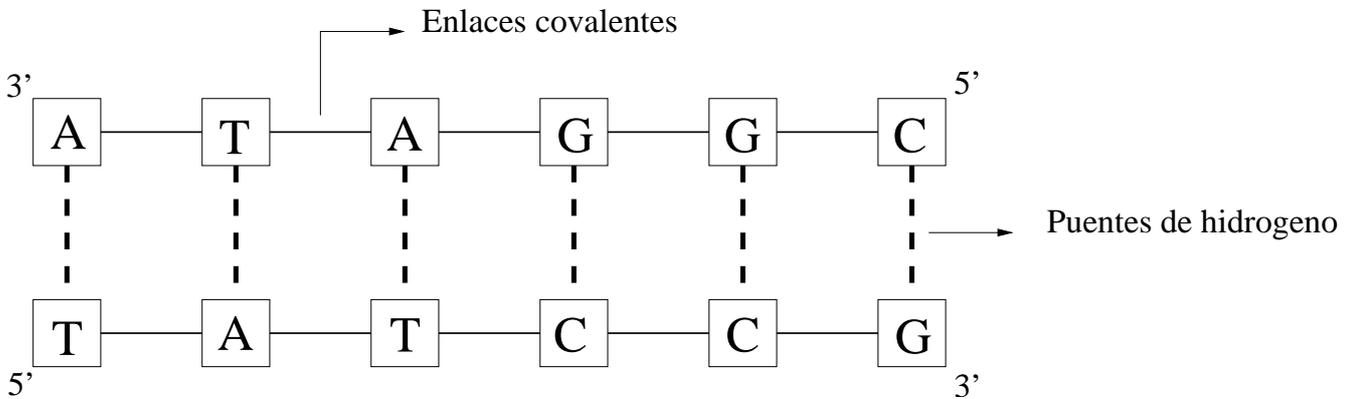


Figura 1.2: Los puentes de hidrógeno sólo pueden formarse entre los nucleótidos A y T o entre C y G. Para que dos hebras de ADN se unan (hibridación) deben ser anti-paralelas y complementarias.

de “unidades” enlazadas covalentemente. Cada unidad consiste de dos partes: la primera parte es común a todas las unidades y corresponde al azúcar-fosfato. La segunda parte es la que identifica a la unidad etiquetándola con una de las cuatro bases: adenina, timina, citosina y guanina (A, T, C, G). Las hebras son orientadas; tienen extremos 3' y 5'. Cuando el ADN da lugar a una hélice de doble hebra, las hebras deben ser antiparalelas, y las bases se deben alinear de manera complementaria (A con T, C con G). Tales hebras reciben el nombre de secuencias Watson-Crick complementarias (Figura 1.2).

Las hebras complementarias reaccionan formando puentes de hidrógeno, constituyendo una nueva molécula de ADN.

En las figuras 1.3 a 1.7 aparecen, junto con algunas reacciones que debemos considerar, las moléculas que estas reacciones forman.

1.3 Cristalización: un proceso de auto-ensamblaje

Para motivar el modelo de auto-ensamblaje, consideremos el proceso físico de cristalización. En términos generales, la cristalización es el nombre que recibe el proceso mediante el cual la materia cambia su estado desde gaseoso o líquido a sólido. Las partículas que componen la materia en estado gaseoso o líquido “flotan” libremente en el medio en el que se encuentran. Al disminuir la temperatura, las partículas (que reciben el nombre de monómeros) forman enlaces entre ellas organizándose según un patrón determinado que depende del tipo de monómeros que componen el sistema, dando lugar a un cristal. A medida que los enlaces se forman y el cristal va creciendo, los monómeros son agregados uno a uno a la superficie del cristal, enlazándose en sitios específicos. En el caso en que exista más de un tipo de monómero, pueden haber distintos tipos de sitios de enlace, cada uno con afinidad para distintos tipos de

1.3. CRISTALIZACIÓN: UN PROCESO DE AUTO-ENSAMBLAJE

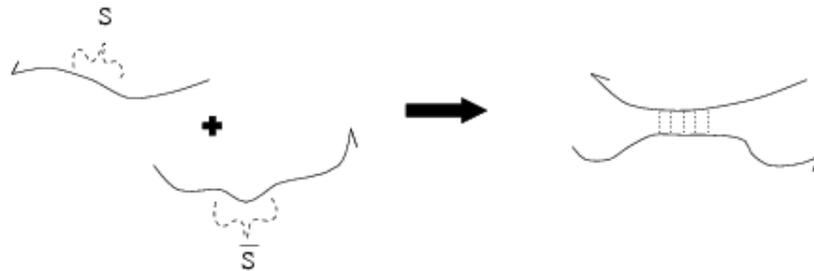


Figura 1.3: Esta es una reacción canónica para el ADN. Dos hebras con subsecuencias complementarias formarán puentes de hidrógeno (hibridación) entre los pares de bases complementarios, formando una nueva molécula. La estabilidad de esta reacción (la cual es reversible) depende principalmente de la temperatura y de la concentración de sales. La temperatura “umbral” (bajo la cual la molécula pierde su estabilidad, ya que se rompen los puentes de hidrógeno y las hebras se separan, denaturándose) depende a su vez del número de pares de bases enlazados.

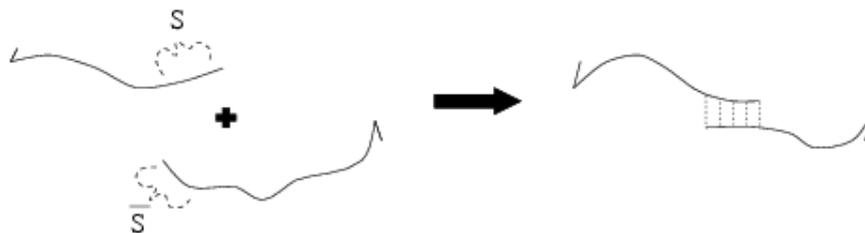


Figura 1.4: Esta reacción es un caso especial de la anterior donde las subsecuencias complementarias ocurren en los extremos de cada hebra. Note que las regiones no enlazadas (“extremos pegajosos”) están disponibles para reacciones posteriores con más ADN.

1.3. CRISTALIZACIÓN: UN PROCESO DE AUTO-ENSAMBLAJE

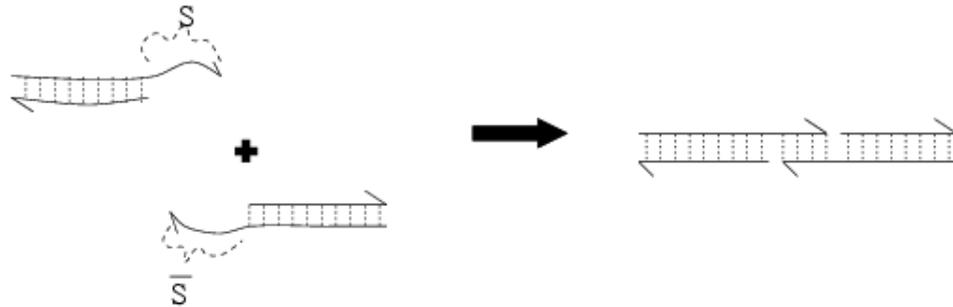


Figura 1.5: La reacción anterior puede ser usada para unir dos moléculas de doble hebra con extremos pegajosos complementarios. Si la enzima ligasa está presente en la solución, las “discontinuidades” que quedan en cada hebra pueden ser reparadas a través de la formación de enlaces covalentes obteniendo dos hebras continuas que forman una sola molécula de doble hebra.

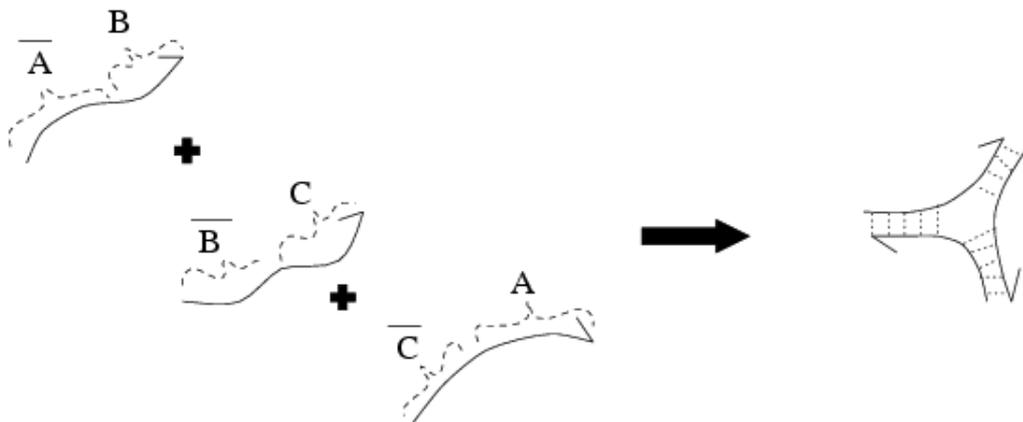


Figura 1.6: Ésta es la mas simple de las reacciones con más de dos hebras. El proceso de ensamblado de esta estructura es secuencial. Aquí solo se muestra la molécula final.

1.3. CRISTALIZACIÓN: UN PROCESO DE AUTO-ENSAMBLAJE

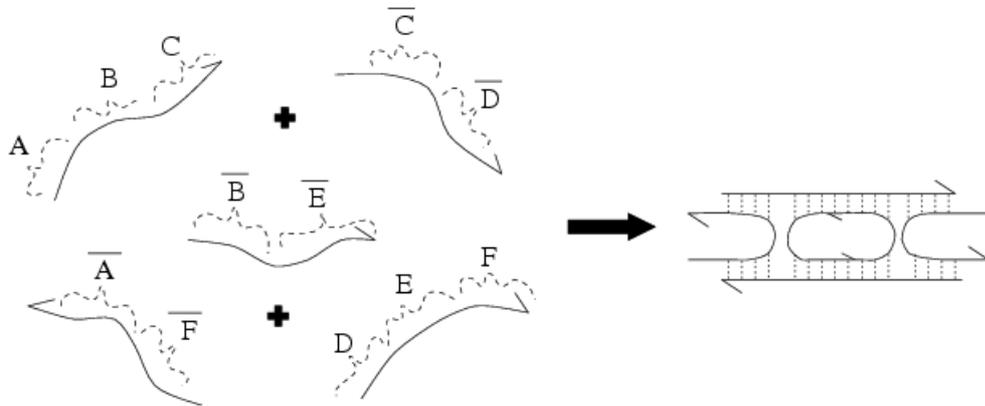


Figura 1.7: Ésta es la estructura mas complicada que consideraremos. Esta molécula, llamada “DAE” (hebras en hélice antiparalela con doble cruzamiento), posee buenas propiedades de rigidez y una disposición siempre plana. Note los extremos pegajosos.

monómeros. Típicamente, las unidades (monómeros) resultan en un ordenamiento periódico. En este caso el sólido final constituye un material homogéneo.

La pregunta de cuándo este ordenamiento resulta necesariamente periódico ha sido estudiada en matemáticas en el contexto de embañosados (tilings) en dos dimensiones. Un resultado sorprendente en la teoría de embañosados es que existen conjuntos de baldosas que admiten **sólo** embañosados aperiódicos del plano [5].

¿Será posible traducir estos resultados puramente matemáticos de vuelta al mundo físico, para producir cristales aperiódicos, o incluso cristales que realicen cómputos? Como ya mencionamos, Eric Winfree [11] mostró como cierto tipo de moléculas de ADN (DAE) puede ser usado para construir “baldosas moleculares” susceptibles de ser programadas para auto-ensamblarse (cristalizarse) en un cristal capaz de simular cualquier autómatas celular.

El modelo que consideraremos en este trabajo fue propuesto por Eric Winfree [9] con el nombre de Tile Assembly Model.

Las unidades fundamentales en este modelo (los monómeros) son baldosas cuadradas (tiles) con pegamento en cada una de sus aristas. Las baldosas se pegarán entre sí (se formarán enlaces entre los monómeros) sólo cuando las aristas adyacentes tengan pegamentos compatibles.

Eric Winfree [11] demostró experimentalmente que un modelo semejante puede ser implementado en un sistema molecular: cada baldosa es representada por una molécula de ADN con doble cruzamiento (DX) con cuatro extremos pegajosos cuyas secuencias representan los pegamentos de las aristas (Figura 1.8).

1.3. CRISTALIZACIÓN: UN PROCESO DE AUTO-ENSAMBLAJE

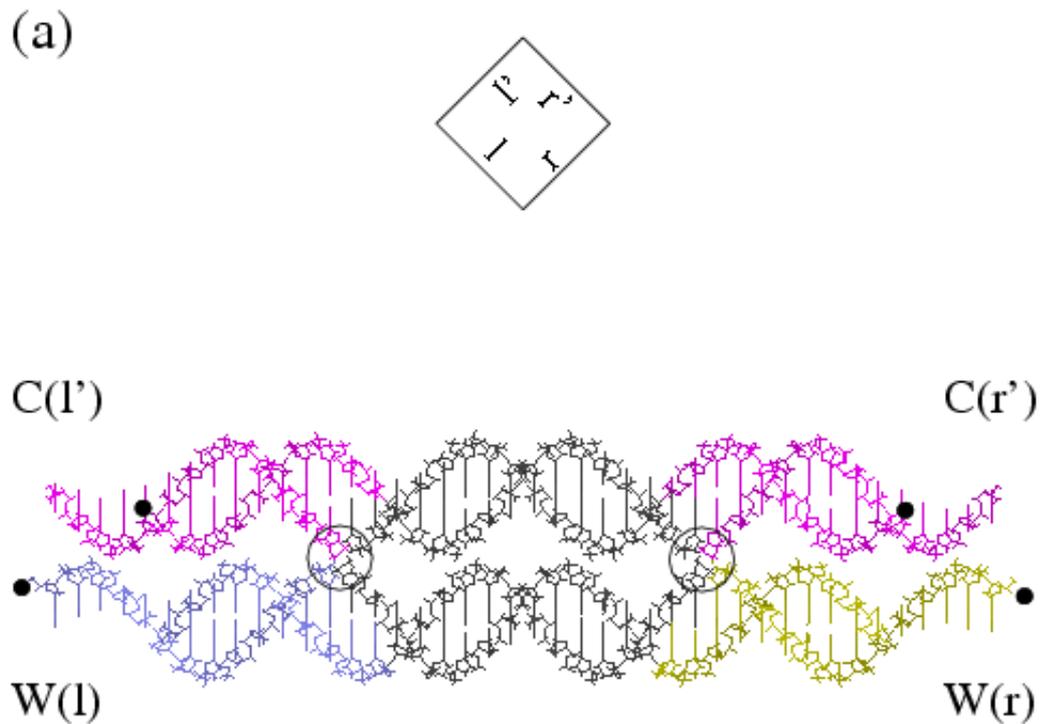


Figura 1.8: Representación de las baldosas con ADN. (a) Una baldosa molecular (molécula de doble cruzamiento). La molécula consiste de una región estructural interior (en negro) y cuatro brazos de doble hebra, cada uno terminado por un extremo pegajoso. Los pegamentos son implementados usando extremos pegajosos con secuencias únicas. Los cruzamientos están encerrados en círculos, y los puntos negros denotan los extremos 5' de cada hebra. Los colores se usan para indicar la etiqueta (el pegamento) de cada arista representada, por lo tanto cada hebra es multicolor.

1.3. CRISTALIZACIÓN: UN PROCESO DE AUTO-ENSAMBLAJE

Como vimos antes, la estabilidad de los enlaces entre moléculas de ADN depende fuertemente de un parámetro externo: la temperatura presente en el medio. En el modelo de auto-ensamblaje cada pegamento tiene asociado un poder que determina, en relación a la temperatura, la estabilidad de las uniones entre las baldosas. Las baldosas no pueden ser rotadas ni reflejadas.

De cada *tipo* de baldosa, el cual está completamente definido por los pegamentos de sus aristas, hay una cantidad infinita de ejemplares “flotando” en una superficie bidimensional.

Las baldosas se van agregando una por una a una estructura ya ensamblada. Cuando una baldosa colinda con el borde de la estructura, se agrega a ella si se satisface simultáneamente que los lados colindantes tengan pegamentos compatibles y que el poder de dichos pegamentos sea “suficiente” en relación a la temperatura. Se asume que los poderes de los pegamentos actúan de manera cooperativa. Es decir, si los lados colindantes que tienen pegamentos compatibles son más de uno, entonces la suma de dichos poderes es la cantidad que debe ser suficiente en relación a la temperatura para mantener estable la unión.

Las hipótesis biológicas para que dichas moléculas se comporten acorde al modelo de auto-ensamblaje son:

1. Las moléculas de doble cruzamiento pueden ser diseñadas en el laboratorio para autoensamblarse en cristales bidimensionales (figura 1.9). Esto ha sido demostrado en un sistema experimental ([11]).
2. El poder de los pegamentos puede ser implementado diseñando extremos pegajosos con energía específica de hibridación. El poder de hibridación del ADN depende principalmente del número de bases de pares. Así, por ejemplo, extremos pegajosos más largos pueden ser usados para representar pegamentos más poderosos (figura 1.9).
3. La hibridación de las moléculas DX, cuando dos extremos pegajosos son complementarios, es *cooperativa*. Luego, los poderes se “suman”. Evidencia empírica de esto se presenta en [11].

Para poder formular con claridad las preguntas de nuestro interés acerca del modelo debemos precisar, por ejemplo, qué significa que un poder sea “suficiente” para mantener estable un enlace. Para ello es necesario introducir el modelo matemático.

1.3. CRISTALIZACIÓN: UN PROCESO DE AUTO-ENSAMBLAJE

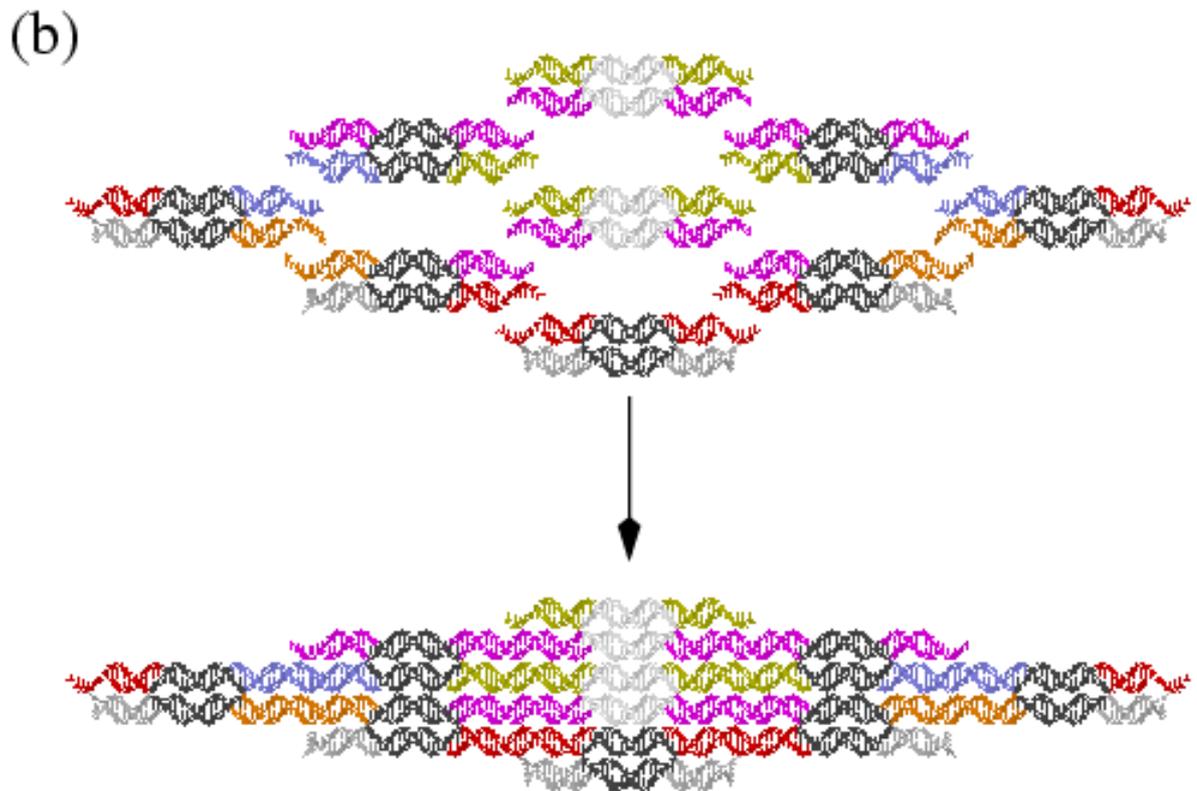


Figura 1.9: Representación de las baldosas con ADN. (b) El auto-ensamblaje de 9 baldosas moleculares, de 5 tipos distintos. Los extremos naranjos y rojos siendo más largos representan pegamentos mas poderosos.

Capítulo 2

Definiciones y resultados básicos

2.1 El modelo de auto-ensamblaje

Formalmente, una **baldoa** $t = (\sigma_N, \sigma_E, \sigma_S, \sigma_O)$, $\sigma_i \in \Sigma$, es una unidad cuadrada y orientada, con sus lados norte, este, sur y oeste etiquetados con algún pegamento de un alfabeto Σ . Asumimos que $null \in \Sigma$.

Diremos que una función $g : \Sigma \times \Sigma \rightarrow \mathbb{N} \cup \{0\}$ es **de poder** si $g(x, y) = g(y, x) \forall x, y \in \Sigma$ y $g(null, x) = 0 \forall x \in \Sigma$.

Para cada baldosa t denotaremos la etiqueta de su lado $i \in \Delta = \{N, E, S, O\}$ como $\sigma_i(t) \in \Sigma$. Representaremos la dirección opuesta a i como i^{-1} . Además usaremos N, S, E, O para representar funciones de $\mathbb{Z}^2 \rightarrow \mathbb{Z}^2$: $N(x, y) = (x, y + 1)$, $S(x, y) = (x, y - 1)$, $E(x, y) = (x + 1, y)$, $O(x, y) = (x - 1, y)$.

Sea T un conjunto finito de baldosas. Asumiremos que T contiene una baldosa especial **vacía** $= (null, null, null, null)$. Una **configuración** de T es una función $C : \mathbb{Z} \times \mathbb{Z} \rightarrow T$. Escribiremos $(x, y) \in C$ ssi $C(x, y) \neq \text{vacía}$. Una configuración C se dirá no vacía si existe $(x, y) \in C$. Dada una baldosa $t \in T$, y un par $(x, y) \in \mathbb{Z} \times \mathbb{Z}$, denotaremos por $C_t^{(x,y)}$ a la configuración que verifica:

$$C_t^{(x,y)}(i, j) = \begin{cases} t & \text{si } (i, j) = (x, y) \\ \text{vacía} & \text{en otro caso} \end{cases}$$

Definimos la suma C de dos configuraciones A y B como $C = A + B$ donde:

$$C(x, y) = \begin{cases} A(x, y) & \text{si } B(x, y) = \text{empty} \\ B(x, y) & \text{si } A(x, y) = \text{empty} \\ \infty & \text{en otro caso} \end{cases}$$

Note que $C = A + B$ será una configuración si y solo si A y B son configuraciones disjuntas. Decimos que $\{A, B\}$ es una partición de C .

2.1. EL MODELO DE AUTO-ENSAMBLAJE

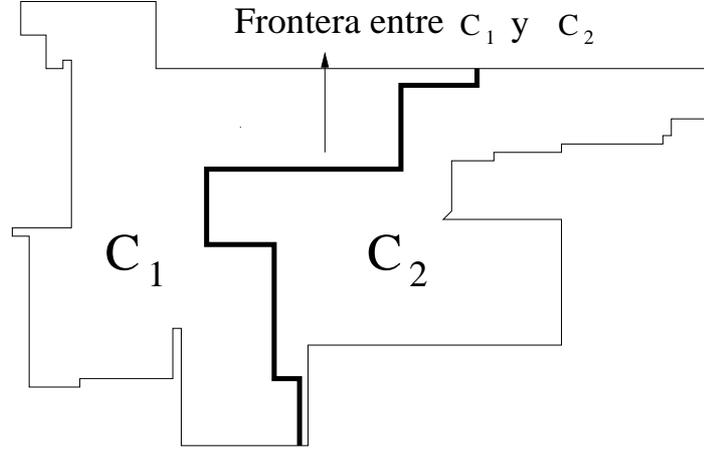


Figura 2.1: Frontera entre C_1 y C_2 .

Definimos la unión J de dos configuraciones A y B como $J = A \cup B$ donde:

$$J(x, y) = \begin{cases} A(x, y) & \text{si } A(x, y) = B(x, y) \text{ o } B(x, y) = \text{vacía} \\ B(x, y) & \text{si } A(x, y) = B(x, y) \text{ o } A(x, y) = \text{vacía} \\ \infty & \text{en otro caso} \end{cases}$$

Note que $A \cup B$ será una configuración si y sólo si coinciden en los puntos del plano en que se superponen.

Definimos la **energía** de una configuración C como la suma de todos los poderes de interacción entre baldosas:

$$G(C) = \frac{1}{2} \sum_{x, y \in Z} \sum_{i \in \Delta} g(\sigma_i(C(x, y)), \sigma_{i-1}(C(i(x, y))))$$

Decimos que una configuración C es un **\mathcal{T} -ensamblaje** (o bien que C es **\mathcal{T} -estable**) si **no** puede ser particionada en dos subconfiguraciones no vacías C_1 y C_2 tales que el poder de interacción en la frontera entre ellas sea menor que la temperatura \mathcal{T} , es decir, tales que $g(C_1, C_2) < \mathcal{T}$ (donde $g(C_1, C_2) = G(C) - [G(C_1) + G(C_2)]$ ver figura 2.1.)

Sea $C_{\mathcal{T}}$ el conjunto de todos los \mathcal{T} -ensamblajes de T .

Un **sistema de baldosas** \mathbf{T} es una 4-tupla $\langle T, s, g, \mathcal{T} \rangle$ donde T es el conjunto de baldosas, s es una baldosa especial que llamaremos *semilla*, g es la función *de poder* y $\mathcal{T} \geq 0$ es la temperatura. También denotaremos por s al \mathcal{T} -ensamblaje $S = \{C_s^{(0,0)}\}$.

Nuestra **dinámica** estará definida por una relación entre \mathcal{T} -ensamblajes de \mathbf{T} . Escribiremos $C \rightarrow_{\mathbf{T}} C'$ si existe una baldosa $t \in T$ y una posición (x, y) tal que $C' = C + C_t^{(x,y)}$.

2.1. EL MODELO DE AUTO-ENSAMBLAJE

En otras palabras, una baldosa t puede ser agregada a un \mathcal{T} -ensamblaje C si la suma de los poderes de interacción entre t y sus vecinos en C , es al menos la temperatura \mathcal{T} .

Sea $\rightarrow_{\mathbf{T}}^*$ la cerradura reflexiva y transitiva de la relación $\rightarrow_{\mathbf{T}}$.

Sea $\mathbf{T} = \langle T, S, g, \mathcal{T} \rangle$ un sistema de baldosas. Se define el conjunto de \mathcal{T} -ensamblajes **producidos** por \mathbf{T} como:

$$Prod(\mathbf{T}) = \{A \in C_{\mathcal{T}} \text{ t.q. } s \rightarrow_{\mathbf{T}}^* A\}$$

y el conjunto de \mathcal{T} -ensamblajes **terminales** de \mathbf{T} como:

$$Term(\mathbf{T}) = \{A \in Prod(\mathbf{T}) \text{ t.q. } \nexists B \neq A \text{ t.q. } A \rightarrow_{\mathbf{T}}^* B\}$$

Note que si C está en el $Prod(\mathbf{T})$, entonces C debe contener a la semilla.

Note que, en general, $Term(\mathbf{T}) \subsetneq Prod(\mathbf{T}) \subsetneq C_{\mathcal{T}}$.

En adelante llamaremos **nodos** a los elementos del $Prod(\mathbf{T})$, y **hojas** a los elementos del $Term(\mathbf{T})$.

Nosotros entenderemos \mathbf{T} con su dinámica como un grafo dirigido $\mathcal{G}(\mathbf{T}) = (V, E)$ donde:

- $V = \{ N / N \text{ es nodo de } \mathbf{T} \}$
- $E = \{ (N, N') \in V^2 / N \rightarrow_{\mathbf{T}} N' \}$

Si \mathbf{T} es tal que $|V| < \infty$ y existe una única hoja H diremos que \mathbf{T} **produce únicamente** H .

Para ilustrar este modelo, consideremos las 7 baldosas de la figura 2.2. Hay cuatro pegamentos diferentes, de poderes 0, 1, 1, 2. A temperatura $\mathcal{T} = 0$, cualquier monómero puede agregarse de manera estable, por lo que configuraciones aleatorias son producidas. A temperatura $\mathcal{T} = 1$, al menos un lado debe ser compatible para que el monómero pueda agregarse de manera estable, las configuraciones producidas dependen de la secuencia particular en que los monómeros fueron agregados. A temperatura $\mathcal{T} = 2$, para cada posición existe una única baldosa que puede agregarse en esa posición, independiente de la secuencia de eventos. Bajo estas condiciones, este conjunto de baldosas produce el Triángulo de Sierpinski, calculando el Triángulo de Pascal mod 2. A temperatura $\mathcal{T} = 3$, ninguna configuración es producida, pues ningún monómero puede agregarse de manera estable.

Estamos en condiciones de adentrarnos en las preguntas que son el objetivo puntual de esta memoria. Nos referimos a preguntas que consideran el auto-ensamblaje como una herramienta que podría permitir la construcción masiva de objetos a una escala molecular y con precisión nanométrica. En este contexto, un sistema de baldosas puede ser visto como un “programa” para construir una figura deseada.

2.2. RESULTADOS BÁSICOS

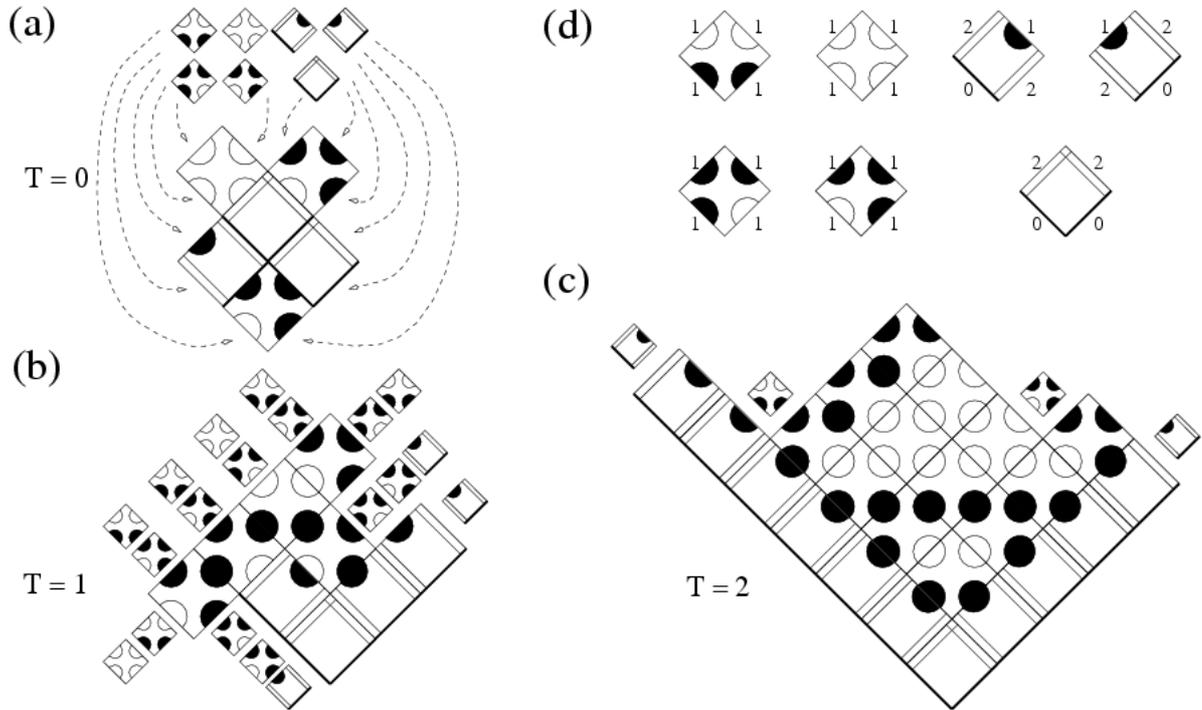


Figura 2.2: El Conjunto de baldosas se muestra en (d). Los poderes están indicados numéricamente, y los pegamentos están denotados gráficamente. En (a), (b) y (c) las baldosas pequeñas indican las posibilidades para agregar una baldosa de manera estable. (a) Cuando $\mathcal{T} = 0$, cualquier baldosa se puede agregar de manera estable, por lo tanto se producen 0-ensamblajes aleatorios. (b) Cuando $\mathcal{T} = 1$, típicamente también existen varias posibilidades para cada posición, por lo tanto se producen 1-ensamblajes aleatorios. (c) Cuando $\mathcal{T} = 2$, existe una única posibilidad para cada posición, resultando en un único patrón de formación.

2.2 Resultados básicos

Que un sistema de baldosas produzca una figura dada significa que, independientemente de la secuencia de eventos, siempre la estructura final debe ser dicha figura. Esto introduce el concepto de “auto-delimitación” (crecer sólo hasta un tamaño fijo), condición que debiera estar implícita en el diseño de las baldosas que constituyen el sistema de baldosas.

Una de las preguntas interesantes que surge de manera natural a propósito de lo anterior es cuándo un conjunto dado de baldosas permite un crecimiento incontrolable. Esto es, cuándo ensamblajes de tamaño arbitrariamente grande pueden ser producidos por un sistema de baldosas.

Sea $S = \{\mathbf{T} : \mathbf{T} \text{ es un sistema de baldosas que produce un } \mathcal{T}\text{-ensamblaje de tamaño}$

2.2. RESULTADOS BÁSICOS

arbitrariamente grande }.

Leonad Adleman [3] demostró que el conjunto S es indecidible para todo \mathcal{T} . La demostración consiste en probar, vía reducción desde el Tiling Problem, la indecidibilidad del conjunto $C = \{ \mathbf{T} : \mathbf{T} \text{ es sistema de baldosas que produce una "culebra" infinita} \}$, y luego mostrar que el conjunto S es idéntico a C .

El resultado anterior considera sistemas de baldosas arbitrarios que, en general, se comportan de manera no determinista (puesto que para una estructura en formación, si elegimos una posición libre en el borde, pueden existir más de una baldosa susceptibles de agregarse de manera estable a la estructura en esa posición). Si removemos esta condición, y consideramos Sistemas de Baldosas *deterministas* (en un sentido que precisaremos) es posible demostrar que ahora cuando $\mathcal{T} = 1$, el conjunto S es decidable. Más aún, existe un algoritmo polinomial que lo decide. Sin embargo, cuando $\mathcal{T} \geq 2$, el conjunto S sigue siendo indecidible incluso para el caso determinista.

Definición 2.2.1. *Un sistema de baldosas \mathbf{T} se dirá **determinista** si para todo par $(i, j) \in \mathbb{Z} \times \mathbb{Z}$ y para todo par de nodos N y N' de \mathbf{T} tales que (i, j) pertenece a N y a N' , se cumple que $N(i, j) = N'(i, j)$.*

En otras palabras, \mathbf{T} es determinista si para toda posición existe una única baldosa que puede, eventualmente, agregarse en esa posición.

Definiremos el **grafo de adyacencia** $G(C)$ de un \mathcal{T} -ensamblaje C como sigue. Los vertices estarán dados por $V = \{(i, j) \in \mathbb{Z} \times \mathbb{Z} / C(i, j) \neq \emptyset\}$. Habrá un arco entre dos posiciones (i_1, j_1) y (i_2, j_2) si y solo si $|i_1 - i_2| + |j_1 - j_2| = 1$.

Teorema 2.2.1. *Para sistemas de baldosas deterministas, si $\mathcal{T} = 1$ entonces el conjunto $S_d^1 = \{ \mathbf{T} : \mathbf{T} \text{ es un sistema de baldosas que produce un 1-ensamblaje de tamaño arbitrariamente grande} \}$ es decidable. Más aún, existe un algoritmo que lo decide en tiempo polinomial.*

Demostración:

Considere el siguiente algoritmo:

DECIDOR($\mathbf{T} = \langle T, s, g, 1 \rangle$)

- $A = s, n = 0$
- Mientras $n \leq (|T| + 1)^2$ hacer:
 - si (existe $t \in T$ tal que $A + t \in \text{Prod}(\mathbf{T})$)
 - $A \leftarrow A + t;$
 - $n++;$
 - si no **aceptar**
- **rechazar**

2.2. RESULTADOS BÁSICOS

Veamos que DECIDOR acepta una entrada w si y solo si $w \notin S_d^1$. En efecto, si DECIDOR acepta w entonces claramente A es un hoja de \mathbf{T} . El determinismo implica que \mathbf{T} produce únicamente A y por lo tanto $w \notin S_d^1$. Si DECIDOR rechaza w , entonces debe existir una posición (x, y) tal que $A(x, y) \neq \emptyset$ con $\max\{|x|, |y|\} > |T| + 1$, digamos $A(x, y) = t$. Sea P un camino entre la semilla s y la baldosa t en el grafo de adyacencia de A . Claramente $|P| > |T|$ luego deben existir al menos dos posiciones p_1 y p_2 tales que $A(p_1) = A(p_2) = i$. Ahora el determinismo de \mathbf{T} implica que el camino P' entre las posiciones p_1 y p_2 se repetirá indefinidamente, pues es 1-estable (figura 2.3). Luego $w \in S_d^1$. \square

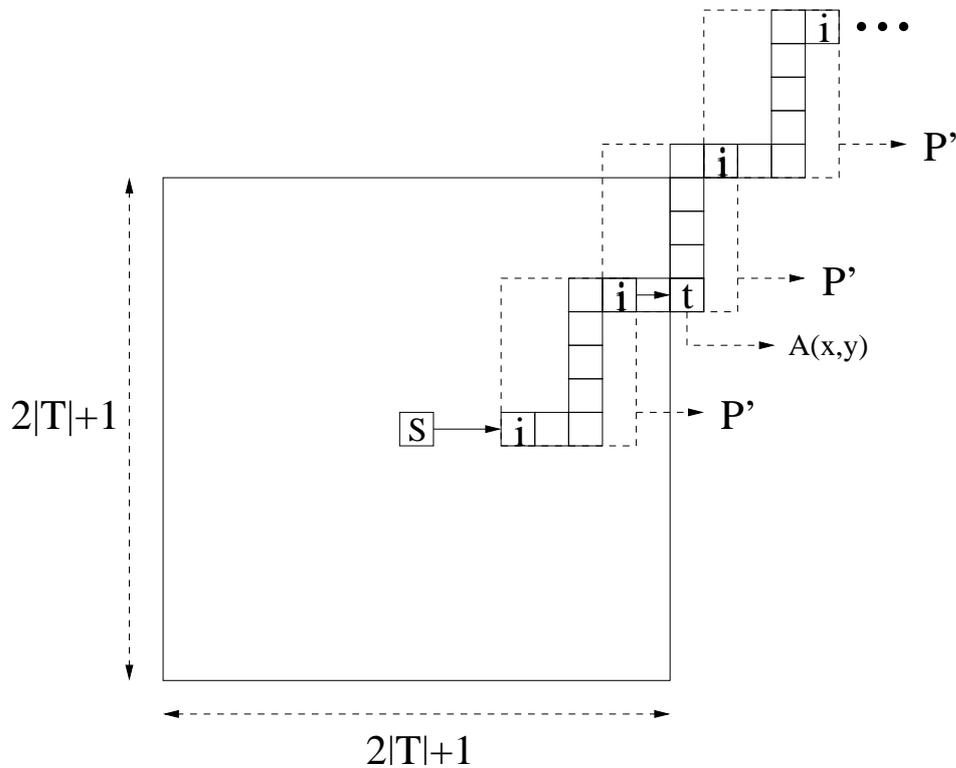


Figura 2.3: Cuando $\mathcal{T} = 1$, el determinismo de \mathbf{T} implica que el camino P' se repetirá indefinidamente.

Teorema 2.2.2. *Para sistemas de baldosas deterministas, si $\mathcal{T} \geq 2$ entonces el conjunto $S_d = \{\mathbf{T} : \mathbf{T} \text{ es un sistema de baldosas que produce un } \mathcal{T}\text{-ensamblaje de tamaño arbitrariamente grande}\}$ es indecidible.*

Demostración: La demostración se basa en la construcción de un sistema de baldosas \mathbf{T} que simula una máquina de Turing determinista arbitraria M . Asumiremos que la entrada para M es de la forma $\epsilon\sigma_1\sigma_2\dots\sigma_n$. El sistema de baldosas es tal que: $\mathbf{T} \notin S_d$ si y sólo si $M(\epsilon\sigma_1\sigma_2\dots\sigma_n)$ termina en tiempo finito. En las figuras 2.4 y 2.5 se exhiben las baldosas de \mathbf{T} . La temperatura es \mathcal{T} .

2.2. RESULTADOS BÁSICOS

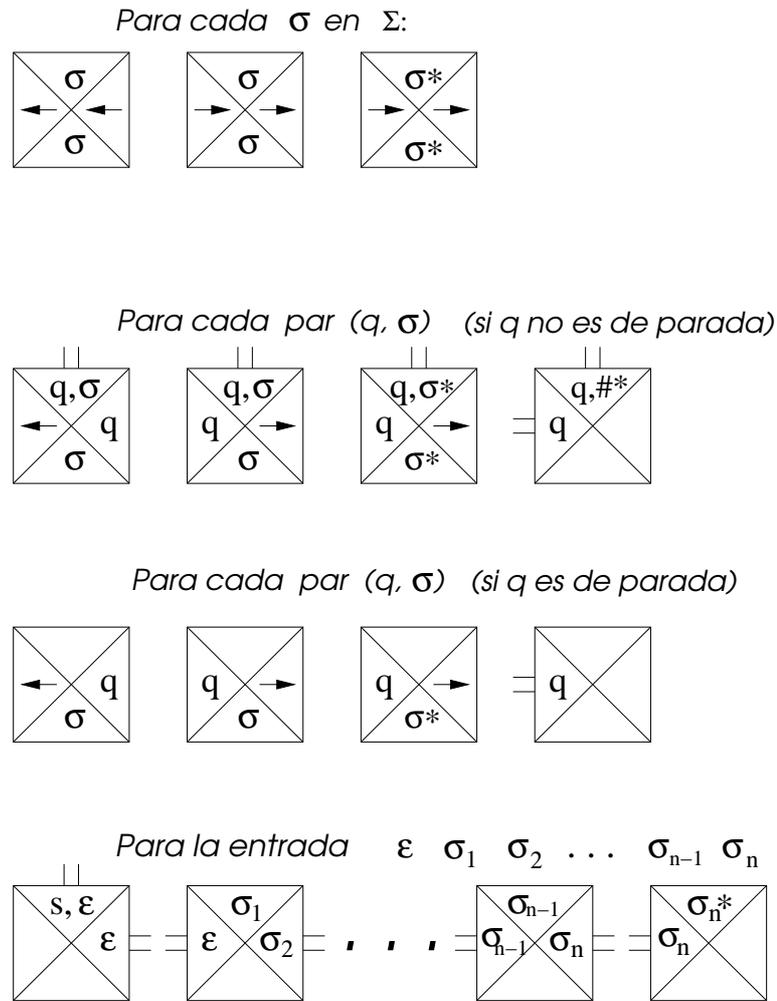
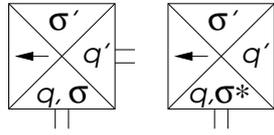


Figura 2.4: Si Σ y Q son el alfabeto y el conjunto de estados de una máquina de Turing M , entonces este conjunto de baldosas compone la base del sistema de baldosas T que simula M . Los poderes se asignan como sigue: Los lados norte y sur tienen poder 1, a menos que estén marcados con dos segmentos (en cuyo caso tienen poder T). Los lados este y oeste tienen poder $T-1$, a menos que estén marcados con dos segmentos (en cuyo caso tienen poder T).

2.2. RESULTADOS BÁSICOS

Para cada regla $(q, \sigma) \rightarrow (q', \sigma', \leftarrow)$:



Para cada regla $(q, \sigma) \rightarrow (q', \sigma', \rightarrow)$:

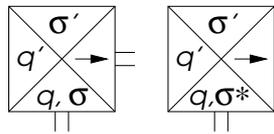


Figura 2.5: Si Σ y Q son el alfabeto y el conjunto de estados de una máquina de Turing M , entonces este conjunto de baldosas implementa las reglas de M . Los poderes se asignan como sigue: Los lados norte y sur tienen poder 1, a menos que estén marcados con dos segmentos (en cuyo caso tienen poder \mathcal{T}). Los lados este y oeste tienen poder $\mathcal{T} - 1$, a menos que estén marcados con dos segmentos (en cuyo caso tienen poder \mathcal{T}).

Cualquiera de las baldosas que codifican la entrada puede elegirse como baldosa semilla. Observe que a temperatura \mathcal{T} todas las baldosas que codifican la entrada se ensamblarán para formar una fila, que llamaremos fila semilla. Las baldosas se van agregando a la fila semilla formando nuevas filas. Cada fila representa un paso de M . Al terminar, la salida de M estará codificada en la última fila, y toda la historia de M se puede leer en el ensamblaje mirando por filas. Las baldosas con pegamentos marcados con $*$ sólo pueden agregarse en la última columna. \square

Dado que la demostración anterior consiste en simular directamente una Máquina de Turing con un sistema de baldosas determinista, el siguiente colorario es directo.

Corolario 2.2.1. *El modelo de auto-ensamblaje determinista con $\mathcal{T} \geq 2$ es Turing-Universal.*

En las figuras 2.6 y 2.7 se presenta un ejemplo. La máquina M , cuya función es insertar un $\#$ al principio de la entrada (si recibe como entrada $\epsilon \sigma_1 \dots \sigma_n$, devuelve $\epsilon \# \sigma_1 \dots \sigma_n$) es simulada por el sistema de baldosas T_M . En las figuras se muestran la máquina, las baldosas, y el ensamblaje final.

2.2. RESULTADOS BÁSICOS

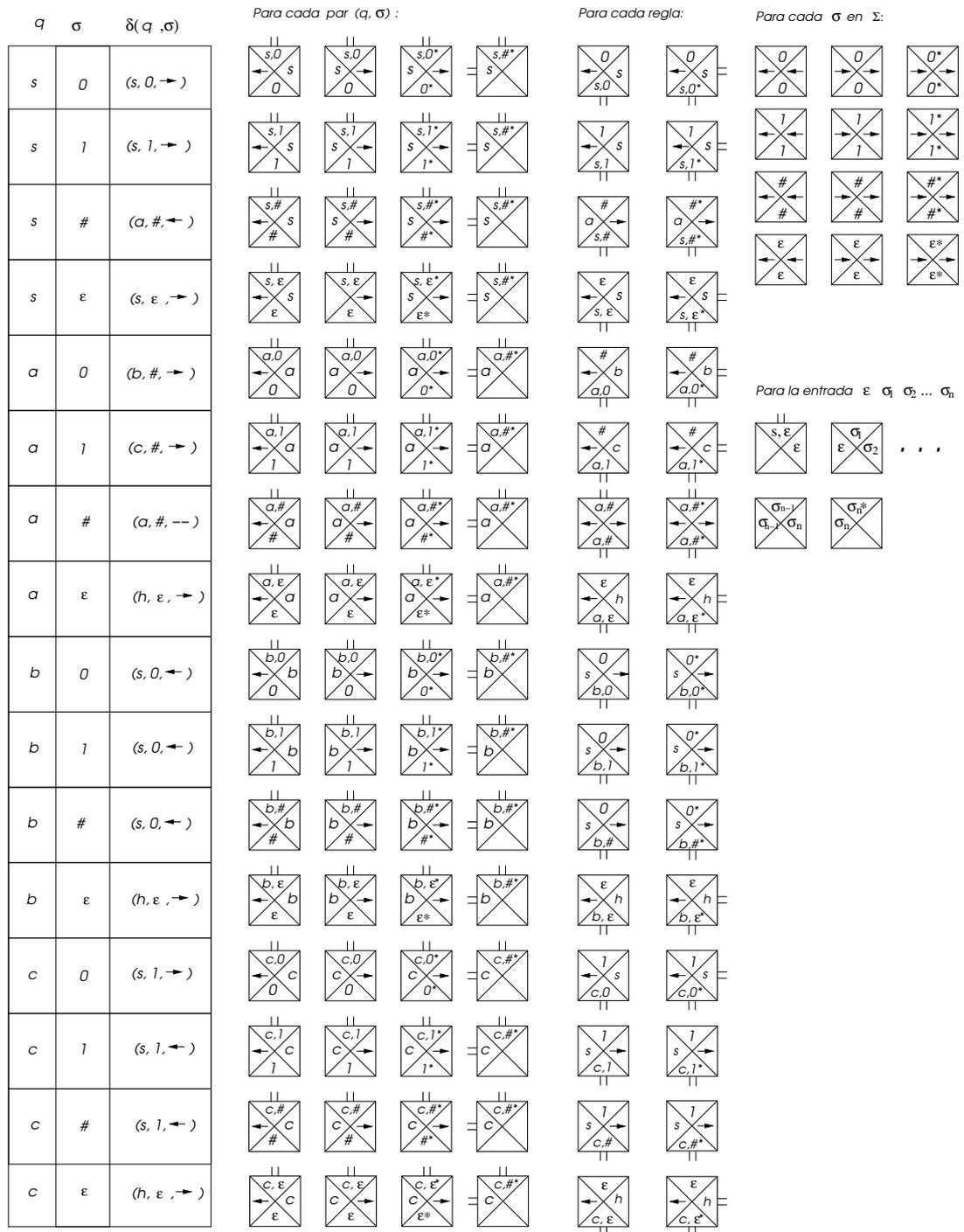


Figura 2.6: En el cuadro de la izquierda, el conjunto de reglas de una máquina de Turing M que inserta un $\#$ al comienzo de la palabra que recibe como entrada. A la derecha de cada regla, las baldosas que le corresponden.

2.2. RESULTADOS BÁSICOS

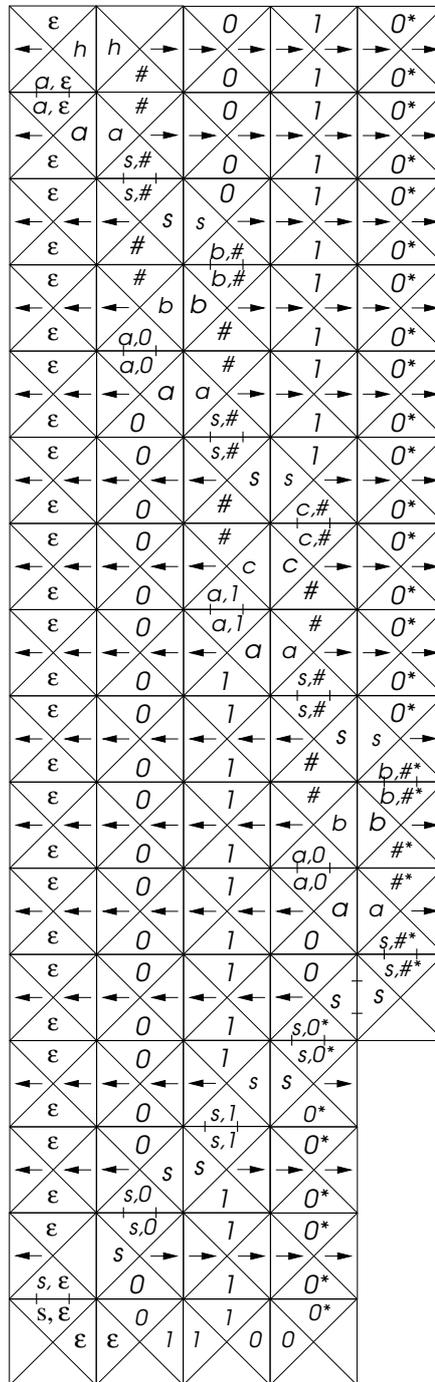


Figura 2.7: Ensamblaje final únicamente producido por el sistema de baldosas T_M en la entrada $\epsilon 010$.

2.3. COMPLEJIDAD DE PRODUCIR ÚNICAMENTE UN \mathcal{T} -ENSAMBLAJE

2.3 Complejidad de producir únicamente un \mathcal{T} -ensamblaje

Al pensar en la “programación” de un sistema de baldosas que produzca únicamente un ensamblaje específico, un asunto básico es el número de baldosas distintas que se requieren. Se define la *complejidad de programación* de un ensamblaje específico A como el número mínimo de tipos de baldosas distintos que se requieren para producir únicamente A .

Rothmund y Winfree [9] mostraron que la complejidad en la programación de un cuadrado de $n \times n$ cuando $\mathcal{T} = 1$ es n^2 . En general, si definimos el tamaño de un ensamblaje como el número de nodos de su grafo de adyacencia, no es difícil ver que si $\mathcal{T} = 1$, entonces la complejidad de programación de cualquier ensamblaje es del orden de su tamaño.

Para $\mathcal{T} \geq 2$ Rothmund y Winfree [9] probaron que la complejidad en la programación del cuadrado de $n \times n$ es $\Theta(\log(n)/\log(\log(n)))$. En su construcción usan $\mathcal{T} = 2$.

Adleman *et al.* [3] demostró que el problema de determinar un sistema de baldosas minimal que produzca únicamente un ensamblaje específico es NP-completo en general, pero que para ciertas familias de ensamblajes (cuadrados y árboles), puede ser resuelto en tiempo polinomial.

Nosotros analizaremos la complejidad en la programación de rectángulos y mostraremos cómo podemos disminuir dicha complejidad al permitir variaciones en la temperatura.

Capítulo 3

Determinismo

Si lo que queremos es encontrar un sistema de baldosas que se autoensamble en una figura dada, lo que pedimos es que tenga una cantidad finita de \mathcal{T} -ensamblajes producidos y que cualquiera que sea el camino partiendo desde la semilla, éste siempre llegue al mismo \mathcal{T} -ensamblaje final. En otras palabras si \mathbf{T} es un sistema de baldosas y $\mathcal{G}(\mathbf{T})$ es el grafo asociado, entonces queremos que éste tenga una única hoja. Si consideramos el orden $(V(\mathcal{G}(\mathbf{T})), \leq)$ -donde $A \leq B \Leftrightarrow A \rightarrow_{\mathbf{T}} B$ -, entonces lo anterior es equivalente a que (V, \leq) tenga un máximo global, lo cual se satisface para lattices finitos.

3.1 Lattice distributivo

Un corte de un \mathcal{T} -ensamblaje es un corte del grafo de adyacencia del \mathcal{T} -ensamblaje. Además para cada arco e_i en un corte, se define el **poder** s_i del arco e_i como el poder del pegamento que etiqueta las aristas de las baldosas (adyacentes) correspondientes a e_i . Definimos también el **poder de un corte** c como $\sum s_i$ para cada arco e_i en el corte.

Lema 3.1.1. *En todo sistema de baldosas determinista \mathbf{T} , la unión de \mathcal{T} -ensamblajes es \mathcal{T} -ensamblaje.*

Demostración:

Sean C_1 y C_2 dos \mathcal{T} -ensamblajes del mismo sistema de baldosas determinista \mathbf{T} tal que $C_1 \neq C_2$. El determinismo de \mathbf{T} nos dice que si $(x, y) \in C_1$ y $(x, y) \in C_2$ entonces $C_1(x, y) = C_2(x, y)$, luego $A = C_1 \cup C_2$ es una configuración de \mathbf{T} (queda bien definida). Veamos que A es \mathcal{T} -estable. Para ello debemos probar que todo corte c_A de A tiene poder mayor o igual a \mathcal{T} . Pero si c_A es un corte de A entonces $c_A \cap C_i$ es un corte de C_i para $i = 1, 2$. Puesto que el poder de c_A es mayor que el de $c_A \cap C_i$, la \mathcal{T} -estabilidad de C_i nos entrega el resultado. \square

3.1. LATTICE DISTRIBUTIVO

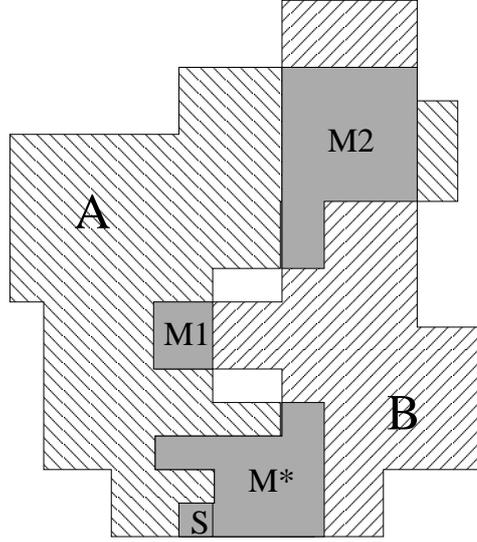


Figura 3.1: La intersección $M = M^* + M1 + M2$ de dos \mathcal{T} -ensamblajes A y B puede ser disconexa. M^* es la “componente conexa” que contiene a la semilla.

Lema 3.1.2. \mathbf{T} es determinista si y solo si (V, \leq) es un lattice.

Demostración:

Sea \mathbf{T} un sistema de baldosas determinista. Sean A y B nodos.

Sea $M = A \cap B$ definido como:

$$M(x, y) = \begin{cases} A(x, y) & \text{si } A(x, y) = B(x, y) \\ \text{vacía} & \text{si } A(x, y) = \text{vacía} \text{ o } B(x, y) = \text{vacía} \\ \infty & \text{en otro caso} \end{cases}$$

Notemos que M es una configuración de \mathbf{T} (queda bien definida) puesto que si A y B se intersectan en la posición (i, j) , entonces el determinismo nos dice que coinciden en esa posición.

Notemos que M podría no ser \mathcal{T} -estable, puesto que el grafo de adyacencia de M podría ser disconexo, lo que implica la existencia de un corte c de poder 0 (figura 3.1).

Sea M^* la configuración que corresponde a la componente conexa que contiene a la semilla, es claro que M^* es el *mínimo* (*meet*) del lattice para A, B . El lemma 3.1.1 nos dice que $A \cup B$ es un nodo de \mathbf{T} , luego es claro que $A \cup B$ es el *máximo* (*join*) del lattice para A, B . La distributividad es directa del hecho que cada paso en la dinámica requiere la adición de exactamente una baldosa.

Sea (i, j) una posición. Sean N y N' nodos no vacíos en (i, j) . El *join* de N y N' es un sucesor común, por lo que N y N' deben ser iguales en (i, j) . \square

3.2. CONDICIONES SUFICIENTES PARA PRODUCIR ÚNICAMENTE UN \mathcal{T} -ENSAMBLAJE

Si $(V(\mathcal{G}(\mathbf{T})), \leq)$ es un lattice, diremos simplemente que \mathbf{T} o que $\mathcal{G}(\mathbf{T})$ es un lattice.

Note que \mathbf{T} produce únicamente H si y sólo si \mathbf{T} es un lattice finito cuyo máximo global es H .

Aquí estamos interesados en dinámicas deterministas. Más aún, centraremos nuestra atención a lattices finitos cuyo máximo global corresponde al ensamblaje final. Recordemos que la forma de este ensamblaje final es el objeto en el cual estamos interesados.

Dado que no es posible saber a priori si un sistema de baldosas \mathbf{T} es un lattice finito, si lo que queremos es encontrar y exhibir Sistemas de Baldosas que produzcan únicamente un \mathcal{T} -ensamblaje A , necesitamos un método que nos permita verificar, dado A , que A es únicamente producido por \mathbf{T} . Es decir, necesitamos condiciones suficientes sobre A y sobre \mathbf{T} para que \mathbf{T} produzca únicamente A . Para esto debemos introducir nuevas nociones.

3.2 Condiciones suficientes para producir únicamente un \mathcal{T} -ensamblaje

Diremos que una baldosa en la posición (x, y) se encuentra **al norte** de otra baldosa en (x', y') si $y > y'$. Similarmente se define para las otras direcciones.

Diremos que una baldosa en la posición (x, y) se encuentra **directamente al norte** de otra baldosa en (x', y') , si $y > y'$ y $x = x'$. Similarmente se define para las otras direcciones.

Sea $t = (\sigma_N, \sigma_E, \sigma_S, \sigma_O) = A(x, y)$ con $A \in \text{Prod}(\mathbf{T})$ y $A(0, 0) = s$ donde $S = \{s\}$. Diremos que t es una **N -baldosa** si $g(\sigma_S(t), \sigma_S(t)) \geq 2$ y $y > 0$. Es decir si se encuentra al norte de la semilla y su lado sur es de poder al menos 2. De manera análoga se define **E -baldosa**, **S -baldosa** y **O -baldosa**. Note que ser una i -baldosa, $i \in \Delta$, es una propiedad que depende de la baldosa y de su posición en el ensamblado.

Diremos que un nodo A de \mathbf{T} cumple la propiedad **RC** si: cada fila al norte de la semilla tiene exactamente una N -baldosa, cada fila al sur de la semilla tiene exactamente una S -baldosa, cada columna al este de la semilla tiene exactamente una E -baldosa, y cada fila al oeste de la semilla tiene exactamente una O -baldosa.

Note que la propiedad RC es fácilmente verificable (polinomial).

Definición 3.2.1. Sea A un nodo de \mathbf{T} y $t \in T$. Definimos el conjunto de configuraciones T -estables de T que llegan a A y que no contienen a t como:

$$\text{Prod}_t^A(\mathbf{T}) = \{B \in \text{Prod}(\mathbf{T}) \text{ t.q. } t \notin B, B \rightarrow_T^* A \text{ y } B + t \in \text{Prod}(\mathbf{T})\}.$$

Definición 3.2.2. Una baldosa $t = A(x, y)$ se dirá **primera en y** si $\forall B \in \text{Prod}_t^A(\mathbf{T})$ se tiene $B(\bullet, y) = \text{vacía}$, y se dirá **primera en x** si $\forall B \in \text{Prod}_t^A(\mathbf{T})$ se tiene $B(x, \bullet) = \text{vacía}$.

3.2. CONDICIONES SUFICIENTES PARA PRODUCIR ÚNICAMENTE UN \mathcal{T} -ENSAMBLAJE

Informalmente, que una baldosa t en A sea primera en y , significa que en cualquier proceso de crecimiento desde la semilla hasta A , t será siempre la primera baldosa que se agrega en la fila y , y lo hará en la posición (x, y) .

Lema 3.2.1. *Sea A un nodo de \mathbf{T} tal que $s = A(0, 0)$ y A cumple la propiedad RC. Sea $t = A(x, y)$.*

- (a) t es N -baldosa ($y > 0$) o t es S -baldosa ($y < 0$) si y sólo si t es primera en y .
- (b) t es E -baldosa ($x > 0$) o t es O -baldosa ($x < 0$) si y sólo si t es primera en x .

Demostración: Para demostrar la parte (a) basta ver $y > 0$. Supongamos que t no es primera en y , entonces $\exists B \in \text{Prod}_t^A(\mathbf{T})$ t.q. $B(x', y) \neq \text{vacía}$ para algún $x' \in \mathbb{Z}$. Sin pérdida de generalidad podemos suponer que tal x' es el único, digamos $r = B(x', y)$. Dado que $\mathcal{T} = 2$, es claro que $g(\sigma_S(r), \sigma_S(r)) \geq 2$, por lo tanto se tiene que r es N -baldosa. Este hecho, dado que t es N -baldosa, viola la propiedad RC de A . La recíproca es evidente. Análogamente se tiene que: t es E -baldosa o t es O -baldosa si y sólo si t es primera en x . \square

Note que en A , toda fila y toda columna tiene una y sólo una baldosa primera.

Diremos que $A \in \text{Prod}(\mathbf{T})$ es **convexo** si y sólo si toda fila y toda columna de A está formada por un bloque completo. Formalmente, A debe cumplir que:

$\forall y$ si $(x_1, y) \in A$ y $(x_2, y) \in A$ entonces $\forall x$ tal que $x_1 < x < x_2$ se tiene que $(x, y) \in A$.

$\forall x$ si $(x, y_1) \in A$ y $(x, y_2) \in A$ entonces $\forall y$ tal que $y_1 < y < y_2$ se tiene que $(x, y) \in A$.

Lema 3.2.2. *Si A tiene la propiedad RC entonces A es convexo.*

Demostración: Por inducción en $|A|$. Claramente si $|A| = 1$ se tiene. Asumamos que se tiene para A_n donde n indica que $|A| = n$. Sea t la baldosa que se agrega a A_n . Si t es primera en x o primera en y entonces es claro que A_{n+1} es convexo. Supongamos entonces que t no es primera en x y que no es primera en y . Entonces por el lema 3.2.1, t no es ni N -baldosa ni S -baldosa, por lo que t debe pegarse en dos lados. La convexidad de A_n implica que los lados en que se pega t deben ser contiguos, es decir N-E, E-S, S-O, o, O-N. Se concluye la convexidad de A_{n+1} . \square

Definición 3.2.3. *Sea $t = A(x, y)$ tal que t no es i -baldosa, con $i \in \Delta$. Diremos que t **mira al N-E**, si $\forall B \in \text{Prod}_t^A(\mathbf{T})$, $\forall x' > x$ y $\forall y' > y$ se tiene que $B(x', y) = B(x, y') = \text{vacía}$. Similarmente se define mirar al S-E, S-O y N-O.*

Informalmente, que una baldosa t en A mire al N-E, significa que en cualquier proceso de crecimiento desde la semilla hasta A , t se agrega siempre pegandose por sus lados sur y oeste, es decir, en el instante en que t se agrega tiene "despejada la vista" solo hacia el norte y el este (figura 3.2).

3.2. CONDICIONES SUFICIENTES PARA PRODUCIR ÚNICAMENTE UN \mathcal{T} -ENSAMBLAJE

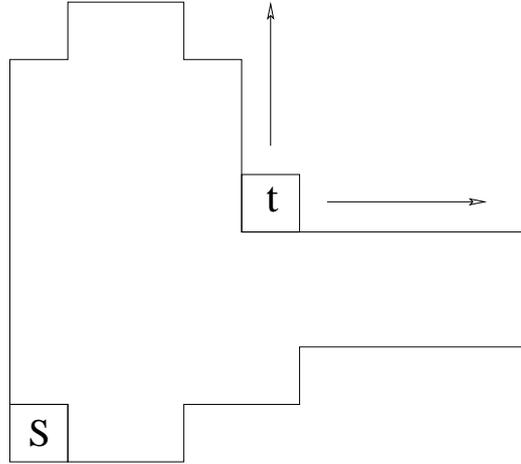


Figura 3.2: La baldosa t mira al N-E

Lema 3.2.3. Sea $t = A(x, y)$, tal que t no es ni N-baldosa, ni E-baldosa, ni S-baldosa ni O-baldosa, entonces t mira al N-E si y sólo si se tiene que:

- o bien t está al este de una N-baldosa y t no está al sur de una E-baldosa
- o bien t está al norte de una E-baldosa y t no está al oeste de una N-baldosa.

Similarmente se tiene la equivalencia para N-O, S-O y S-E.

Demostración:

(\Rightarrow)

Sea t mirando al N-E. Sea $B \in \text{Prod}_t^A(\mathbf{T})$, entonces se tiene que $B(x', y') = \text{vacía} \forall x' > x$ y $\forall y' > y$.

CASO $y > 0$, $x > 0$. Por el lema 3.2.1 (a), tenemos que t no es primera en y . Sea $t' = A(x', y)$ la primera en y ($x' < x$). Como $y > 0$, t' es N-baldosa, por lo que t está al este de una N-baldosa. Por el lema 3.2.1 (b), tenemos que t no es primera en x , sea $t' = A(x, y')$ la primera en x ($y' < y$), como $x > 0$, t' es E-baldosa, por lo que t está al norte de una E-baldosa, luego, t no está al sur de una E-baldosa.

CASO $y > 0$, $x < 0$. Como $y > 0$, ya tenemos que t está al este de una N-baldosa. Como $x < 0$, entonces no existe una E-baldosa en la columna x , luego, t no está al sur de una E-baldosa.

CASO $y < 0$, $x < 0$. Como $B(x', y') = \text{vacía} \forall x' > x \forall y' > y$, entonces $B + t$ es “disconexo”. Luego $B + t \notin \text{Prod}(\mathbf{T})$, que es una contradicción, por lo tanto, $\nexists t = A(x, y)$ mirando al N-E tal que $x < 0$ y $y < 0$.

3.2. CONDICIONES SUFICIENTES PARA PRODUCIR ÚNICAMENTE UN \mathcal{T} -ENSAMBLAJE

CASO $y < 0$, $x > 0$. Ya tenemos que $x > 0$ implica que t está al norte de una E -baldosa. $B(x', y) = \text{vacía} \forall x' > x \Rightarrow t$ no está al oeste de una N -baldosa.

(\Leftarrow)

Sea $t = A(x, y)$ tal que (t está al este de una N -baldosa y t no está al sur de una E -baldosa) o (t está al norte de una E -baldosa y t no está al oeste de una N -baldosa).

CASO en que t está al este de un N -baldosa, y que no está al sur de una E -baldosa. Estar al este de una N -baldosa implica que $y > 0$.

Sea $B \in \text{Prod}_t^A(\mathbf{T})$. Note que como A cumple la propiedad RC entonces B también la cumple. Luego por el lema 3.2.2, B es convexo, lo que permite concluir que $B(x', y) = \text{vacía} \forall x' > x$. Falta ver que $B(x, y') = \text{vacía} \forall y' > y$.

Supongamos $x > 0$ y sea t^* la primera en x . t^* es una E -baldosa. Como t no está al sur de una E -baldosa, entonces t debe estar al norte de t^* y, por la convexidad de B , concluimos.

Supongamos $x < 0$. Al igual que antes, estar al este de una N -baldosa, digamos $n = A(n_x, n_y)$, sumado a la convexidad de B , permite concluir que $B(x', y) = \text{vacía} \forall x' > x$. Además, como n es N -baldosa, se cumple que $\forall C \in \text{Prod}_n^A(\mathbf{T}) C(n_x, y') = \text{vacía} \forall y' > n_y$. Esto implica que $\exists y^*$ con $y^* < y$ tal que $B(x, y^*) \neq \text{vacía}$. La convexidad de B permite concluir.

CASO en que t está al norte de una E -baldosa y que no está al oeste de una N -baldosa. Los razonamientos son análogos. \square

El lema anterior, además de entregar una caracterización, nos dice que para determinar si una baldosa en un ensamblado con la propiedad RC *mira al N-E*, es suficiente conocer la posición relativa de la baldosa en el ensamblado. Además se puede verificar que toda baldosa que no sea primera en alguna fila o alguna columna, cumple alguna de las cuatro caracterizaciones. Así toda baldosa es o bien *i-baldosa* o bien mira al N-E o al S-E o al N-O o al S-O.

La siguiente es una condición necesaria para que un sistema de baldosas produzca un único ensamblado A .

Definición 3.2.4. Sea el sistema de baldosas $\mathbf{T} = \langle T, \{s\}, g, 2 \rangle$, y sea $A \in \text{Prod}(\mathbf{T})$ tal que A cumple la propiedad RC. Diremos que:

- Una N -baldosa $t = (\sigma_N, \sigma_E, \sigma_S, \sigma_O)$ está **\mathbf{T} -determinada** si:
 $[t' = (\sigma'_N, \sigma'_E, \sigma'_S, \sigma'_O) \in T] \Rightarrow [t' = t]$. Similarmente definimos para las otras clases.
- Una baldosa $t = (\sigma_N, \sigma_E, \sigma_S, \sigma_O)$ que mira al N-E está **\mathbf{T} -determinada** si:
 $[t' = (\sigma'_N, \sigma'_E, \sigma'_S, \sigma'_O) \in T] \Rightarrow [t' = t]$. Similarmente definimos para las otras clases.
- A es **\mathbf{T} -determinista**, si todas sus baldosas están \mathbf{T} determinadas.

3.2. CONDICIONES SUFICIENTES PARA PRODUCIR ÚNICAMENTE UN \mathbf{T} -ENSAMBLAJE

Note que el \mathbf{T} -determinismo es una propiedad fácilmente verificable (polinomial).

Teorema 3.2.1. *Sea $\mathbf{T} = \langle T, \{s\}, g, 2 \rangle$ un sistema de baldosas. Sea $A \in \text{Term}(\mathbf{T})$ un 2-ensamblaje \mathbf{T} -determinista con la propiedad RC. Entonces \mathbf{T} produce únicamente A .*

Demostración: Probaremos que si $\mathbf{T} = \langle T, \{s\}, g, 2 \rangle$ es un sistema de baldosas, y si $A \in \text{Term}(\mathbf{T})$ es un 2-ensamblaje \mathbf{T} -determinista con la propiedad RC, entonces si $C \in \text{Prod}(\mathbf{T})$, C es \mathbf{T} -determinista, tiene la propiedad RC y $C \rightarrow_{\mathbf{T}}^* A$.

Lo haremos por inducción en $|C|$. El caso base es trivial. Supongamos que el teorema es cierto para C con $|C| \leq n$, y $(s = C_1) \rightarrow_{\mathbf{T}} C_2 \dots \rightarrow_{\mathbf{T}} C_n \rightarrow_{\mathbf{T}} C_{n+1}$, donde $C_n + A_t^{(x,y)} = C_{n+1}$. Claramente, $|C_n| = n$, luego C_n es \mathbf{T} -determinista, tiene la propiedad RC y $C \rightarrow_{\mathbf{T}}^* A$. Mostraremos que lo mismo se tiene para C_{n+1} donde $t = C_{n+1}(x, y)$ es la nueva baldosa que se agraga a C_n .

Caso 1: t se pega con poder 2. Sin pérdida de generalidad supongamos que se pega por su lado sur, es decir: $g(\sigma_S(t), \sigma_S(t)) = 2$. $y \leq 0 \Rightarrow t' = C_n(x, y - 1)$ es una S -baldosa, pero en la columna $y - 1$ de C_n ya existe una S -baldosa (la primera en $y - 1$, que claramente no es t'), lo que viola la propiedad RC de C_n . Así $y > 0$, por lo que t debe ser una N -baldosa. El \mathbf{T} -determinismo de A implica que $C_{n+1} \rightarrow_{\mathbf{T}}^* A$ (t es la única baldosa que puede agregarse en esa posición), por lo que C_{n+1} tiene la propiedad RC, y es claramente \mathbf{T} -determinista.

Caso 2: t se pega con poder 1. Como la temperatura de \mathbf{T} es 2, t debe agregarse por dos de sus lados. Como C_n tiene la propiedad RC, C_n es convexo (lema 3.2.2), luego t se agrega en lados contiguos. S.p.g., asumamos que estos lados son el sur y el oeste. Claramente C_{n+1} es convexo y tiene la propiedad RC. Por la convexidad de C_n se tiene que $C_n(x', y') = \text{vacía} \forall x' > x$ y $\forall y' > y$. Luego t mira al N-E en cualquier 2-ensamblaje que extienda a C_n , en particular en C_{n+1} , y como A es \mathbf{T} -determinista se tiene que $t = A(x, y)$ ($C_{n+1} \rightarrow_{\mathbf{T}}^* A$). Luego C_{n+1} es \mathbf{T} -determinista. \square

Este teorema nos permite verificar, en las construcciones del capítulo 4, que las figuras objetivo son únicamente producidas.

Capítulo 4

Complejidad en la programación de rectángulos

4.1 Temperatura constante y rectángulos “delgados”

Una configuración R es un $k \times n$ -rectángulo si existe una posición (x_0, y_0) tal que $(x, y) \in R$ ssi $x_0 \leq x < x_0 + k$ y $y_0 \leq y < y_0 + n$.

Estamos interesados en cuales $k \times n$ rectángulos pueden ser únicamente producidos por sistemas de baldosas:

$R_{\mathcal{T}}(k, n) = \{N \in \mathbb{N} \text{ tal que existe un sistema de baldosas } \mathbf{T} = \langle T, S, g, T \rangle, |T| = N, \text{ y } \mathbf{T} \text{ produce únicamente un } k \times n \text{ rectángulo}\}$.

Definimos la complejidad $C_{\mathcal{T}}(k, n)$ de un $k \times n$ rectángulo como el mínimo número de baldosas distintas requeridas para producir únicamente un $k \times n$ -rectángulo (físicamente es el número de moléculas distintas que deben ser preparadas):

$$C_{\mathcal{T}}(k, n) = \min\{N \text{ tal que } N \in R_{\mathcal{T}}(k, n)\}$$

Escogemos, al igual que Rothmund y Winfree [9], temperatura $\mathcal{T} = 2$. Para mayor claridad comenzaremos con explicar en detalle el caso de un $3 \times n$ rectángulo.

Proposición 4.1.1. $C_2(3, n) \geq (n/2)^{1/3}$.

Demostración: Sea T el conjunto de baldosas de un sistema de baldosas \mathbf{T} que produce únicamente el $3 \times n$ -rectángulo. Consideremos, para el $3 \times n$ -rectángulo final, un sub-rectángulo de dimensiones $3 \times (n/2)$ que no contenga a la semilla. El número de filas (de largo 3 en este caso) distintas que podemos construir con $|T|$ baldosas es exactamente $|T|^3$. Luego, si $|T| < (n/2)^{1/3}$, en este sub-rectángulo debe existir al menos una fila que se repite, lo que gracias al determinismo de \mathbf{T} implica que el objeto crecerá indefinidamente. Esto es una contradicción (figura 4.1). \square

4.1. TEMPERATURA CONSTANTE Y RECTÁNGULOS “DELGADOS”

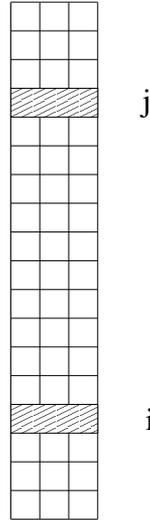


Figura 4.1: Si las filas i y j son iguales entonces el determinismo de T implica que las filas $i + 1$ y $j + 1$ también son iguales.

Proposición 4.1.2. $C_2(3, n) = O(n^{1/3})$.

Demostración: Exhibiremos un sistema de baldosas $T_{(3,n)}$ que produce únicamente el $3 \times n$ -rectángulo. La idea principal de la demostración es la de encontrar un conjunto de baldosas que funcione como “contador”. Ésta idea fué introducida por Rothmund y Winfree [9], es su caso, el conjunto de baldosas simulaba un contador binario en el cual la i^{ma} fila sobre el origen (fila semilla) representaba el entero i (escrito en binario). Este sistema tiene una limitación: el total de enteros distintos que pueden ser representados -que es lo que determina la altura máxima del rectángulo (2^3 en este caso)- depende del ancho. Para solucionar esto, usaremos una base b tal que la altura n escrita en b -ario pueda ser representada en términos del ancho (en este caso necesitamos $b = \lceil n^{1/3} \rceil$ y el sistema de baldosas cuenta en base b representando en cada fila un entero i entre 0 y $n - 1$).

Construiremos el $3 \times n$ -rectángulo usando $4b + 2$ baldosas ($4b$ baldosas constituirán el “contador” y 2 baldosas agradas a la semilla formarán la fila semilla). Cada baldosa tendrá, además de sus cuatro lados, una “etiqueta principal” que entregará cierta información semántica explicando su “función”.

El sistema de baldosas funciona como sigue:

La fila semilla codifica el número $b^3 - n$ y es formada por 2 baldosas especiales agregadas a la semilla. La fila $i + 1$ se ensambla a la fila i aumentando el contador en 1 . Las otras $4b$ baldosas aparecen en la figura 4.2. Están calificadas en 4 tipos: Type 1, Type 2, Type 3, Rightmost. Además, si miramos horizontalmente podremos observar otra clasificación: 0 -baldosas, 1 -baldosas, \dots , $(b-1)$ -baldosas.

4.1. TEMPERATURA CONSTANTE Y RECTÁNGULOS “DELGADOS”

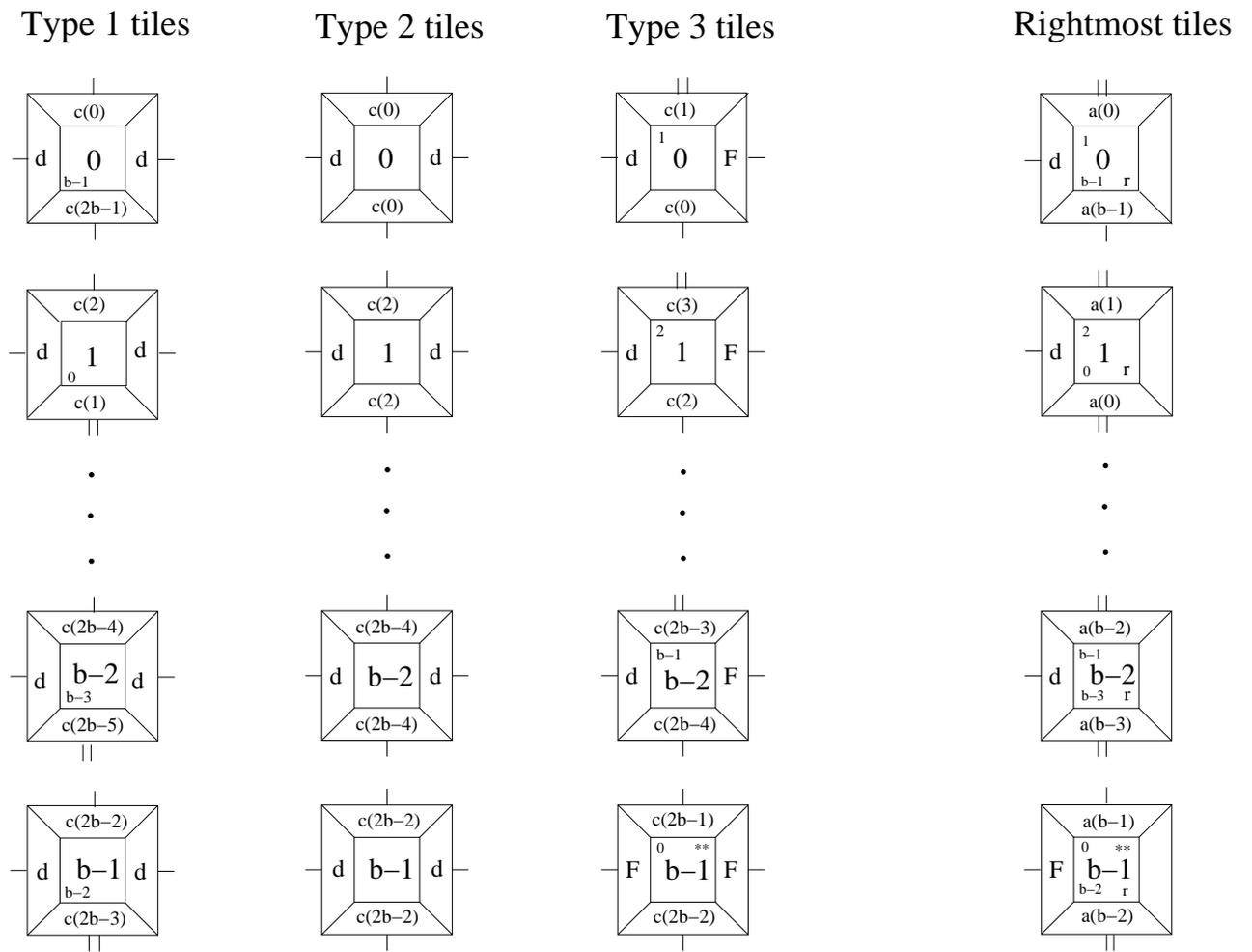


Figura 4.2: Conjunto de baldosas que simulan un contador en base b .

4.1. TEMPERATURA CONSTANTE Y RECTÁNGULOS “DELGADOS”

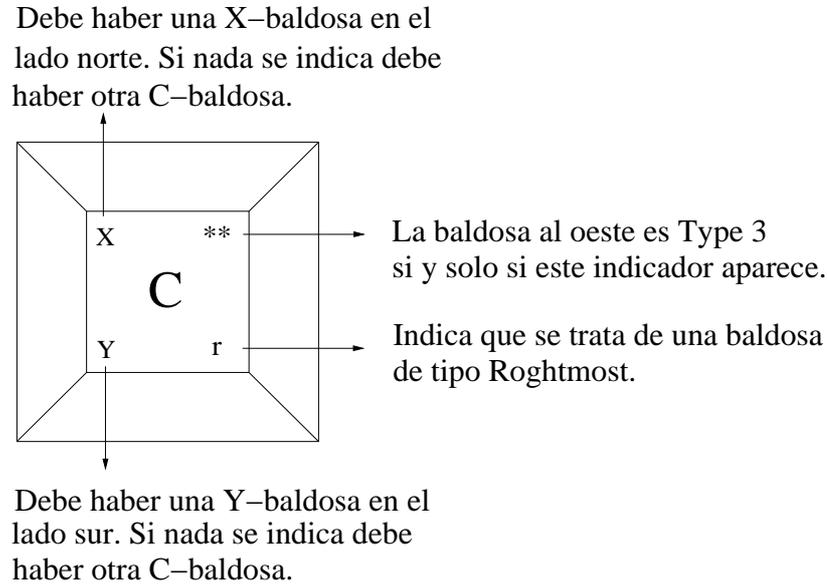


Figura 4.3: Un C-baldosa y su información semántica.

En la figura 4.3 se explica como debe ser entendida la información semántica de la “etiqueta principal”.

Como el poder del lado norte de las $(b - 1)$ -baldosas es 1, cuando el entero $n - 1$ es codificado, el contador para. Siguiendo las reglas semánticas no es difícil convencerse de que el rectángulo deseado es efectivamente ensamblado. \square

La figura 4.4 nos muestra el conjunto de baldosas para ensamblar un $3 \times n$ -rectángulo para el caso $n = 25$. La baldosa semilla puede ser cualquiera de las baldosas que forman la fila semilla.

Corolario 4.1.1. $C_2(3, n) = \Theta(n^{1/3})$

Para generalizar la construcción anterior a un $k \times n$ -rectángulo, basta con las $4b$ baldosas que forman el “contador” en base b (con $b = \lceil n^{1/k} \rceil$) más k baldosas especiales (que formen la fila semilla) que codifiquen el entero $b^k - n$. Así para tener complejidad $O(n^{1/k})$ es necesario que k sea a lo más $n^{1/k}$. Tenemos:

Teorema 4.1.1. Para $k \leq n^{1/k}$ se tiene: $C_2(k, n) = \Theta(n^{1/k})$.

4.1. TEMPERATURA CONSTANTE Y RECTÁNGULOS “DELGADOS”

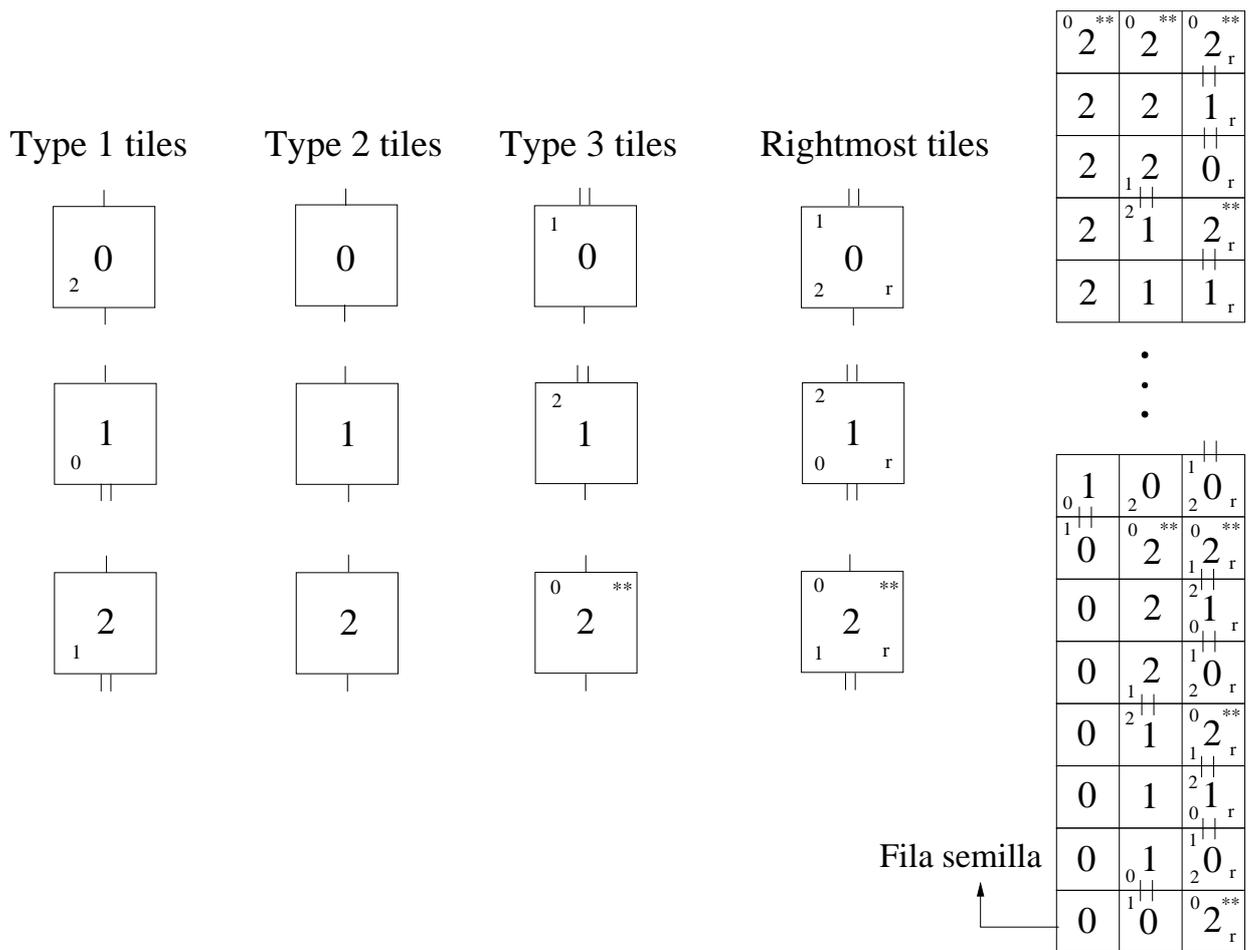


Figura 4.4: Conjunto de baldosas que produce únicamente un 3×25 -rectángulo. Los lados este y oeste tienen todos poder 1. En este caso $b = 3$. La fila semilla codifica el entero $3^3 - 25 = 2$ o $0\ 0\ 2$ si lo escribimos en 3-ario.

4.2. TEMPERATURA CONSTANTE Y RECTÁNGULOS “GRUESOS”.

4.2 Temperatura constante y rectángulos “gruesos”.

En la sección anterior vimos que para rectángulos “delgados” -esto es $k \times n$ -rectángulos con $k \leq n^{1/k}$ - el valor $C_2(k, n) = \Theta(n^{1/k})$. Mas precisamente, se necesitan $4\lceil n^{1/k} \rceil$ baldosas para el “contador” y k baldosas para la fila semilla, lo que dado que $k \leq n^{1/k}$ entrega una cota superior para la complejidad de $O(n^{1/k})$. Sin embargo, si no se trata de un rectángulo delgado, aún así nuestra construcción entrega un sistema de baldosas que lo produce únicamente, claro que la cota superior depende ahora de k . Así podemos establecer para todo k una cota superior del valor $C_2(k, n) = O(n^{1/k} + k)$. Note que si ponemos $k = n$ (cuadrado de $n \times n$) obtenemos $C_2(n, n) = O(n)$.

Si buscamos el ancho de un $k \times n$ -rectángulo que minimice el valor $C_2(k, n)$, encontraremos que corresponde a $k = \lceil \lambda_n \rceil$ donde λ_n es tal que $\lambda_n^{\lambda_n} = n$. En este caso el número de baldosas para contruir un $\lceil \lambda_n \rceil \times n$ -rectángulo es $O(n^{1/\lambda_n} + \lambda_n) = O(\lambda_n)$. En esta sección mostraremos que para rectángulos con ancho $k \geq \lceil \lambda_n \rceil$, se tiene que el valor $C_2(k, n) = O(\lambda_n)$ (esta cota superior no depende de k). Entonces tendremos que $C_2(n, n) = O(\lambda_n) = O(\log n / (\log(\log n)))$, valor que alcanza la cota inferior mostrada por Rothmund y Winfree [9].

Proposición 4.2.1. $C_2(\lambda_n, n) = \Theta(\lambda_n)$.

Demostración: La construcción de un $\lambda_n \times n$ -rectángulo es directa a partir de la construcción mostrada en la Sección 3. El conjunto de baldosas mostrado en la figura 4.2 también funciona en el caso general. Sólo necesitamos poner $b = \lceil n^{1/\lambda_n} \rceil$ y agregar $\lceil \lambda_n \rceil - 1 \leq b$ baldosas especiales para formar la fila semilla a partir de la baldosa semilla. Por lo tanto usamos $4b + \lambda_n - 1 \leq 5b$ baldosas en total. \square

Proposición 4.2.2. $C_2(\lambda_n + k, n) = O(\lambda_n)$.

Demostración: Lo primero que necesitamos es $2C_2(\lambda_n, n)$ baldosas para construir un $\lambda_n \times n$ -rectángulo particular. La particularidad está en el hecho que la columna formada por baldosas de tipo Rightmost está dividida en dos clases de baldosas: las k superiores y las $n - k$ inferiores con respecto a una baldosa especial t_* (es la única con poder 2 en su lado este). La idea es simple: La fila semilla codifica en sus lados norte el número $b^b - k$ y el número $b^b - (n - k)$ en sus lados sur (figura 4.5).

Para completar el rectángulo necesitamos otras 4 baldosas “de relleno”. Su trabajo es completar un $k \times n$ -rectángulo a la derecha de la columna Rightmost. La baldosa t_* es la clave para el proceso “de relleno” (figura 4.6). \square

4.2. TEMPERATURA CONSTANTE Y RECTÁNGULOS “GRUESOS”.

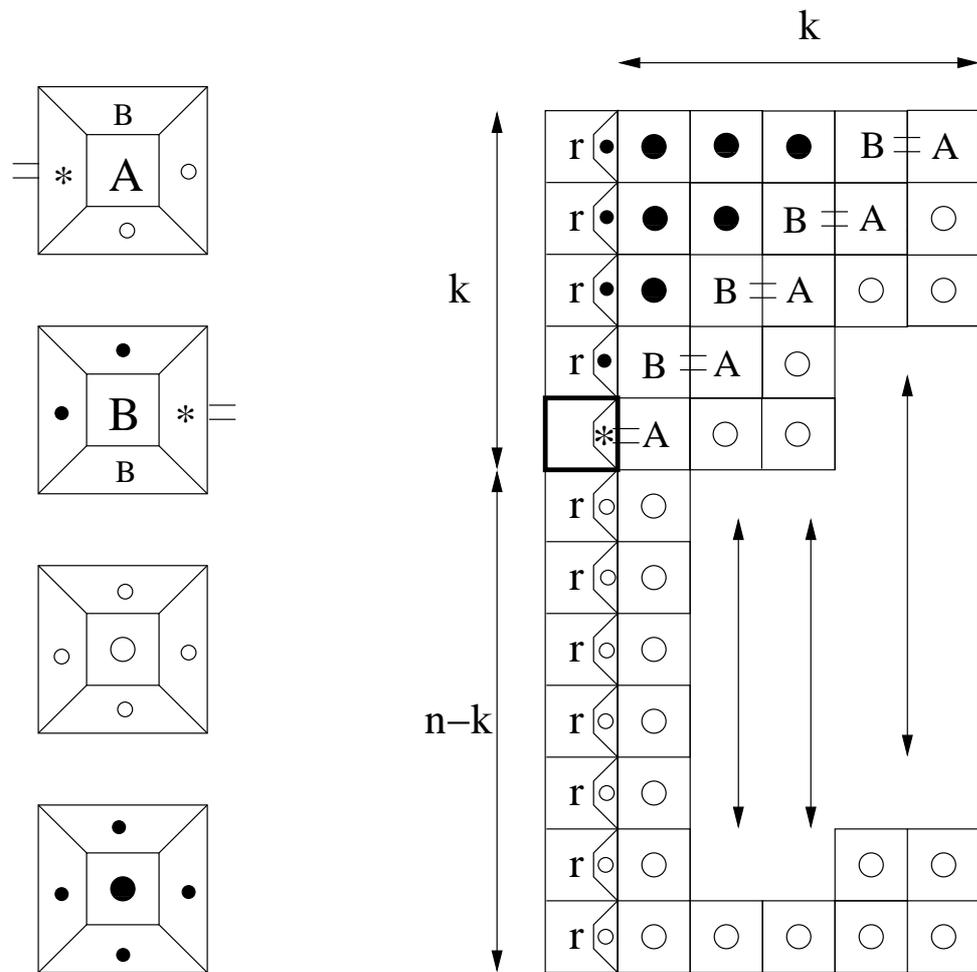


Figura 4.6: Las 4 baldosas “de relleno” y la manera en que actúan partiendo de la columna Rightmost.

4.3 Aumentando la Temperatura a $\mathcal{T} = 4$

Hasta ahora hemos considerado la programación de Sistemas de Baldosas deterministas que producen únicamente un \mathcal{T} -ensamblaje específico que representa la salida del programa. Hemos estudiado la complejidad de programar rectángulos, y hemos constatado que la dificultad mayor se presenta en la construcción de rectángulos delgados, pues su programación requiere de un alto número de baldosas. En esta sección presentaremos un mecanismo que permite disminuir este número, bajando dramáticamente la complejidad en la programación de rectángulos delgados.

Imagine que tenemos, bajo temperatura \mathcal{T} , una estructura \mathcal{T} -estable ya ensamblada. ¿Qué ocurre si aumentamos la temperatura a $\mathcal{T} + \Delta$? Dos cosas pueden ocurrir:

- La estructura tiene la energía suficiente para mantenerse estable bajo la nueva temperatura.
- La estructura no tiene la energía suficiente para mantenerse estable bajo la nueva temperatura. En este caso la estructura comenzará a desintegrarse hasta reducirse a una porción capaz de mantenerse unida.

El mecanismo que presentaremos para disminuir la complejidad en la programación de rectángulos delgados consistirá entonces de dos etapas:

- Construir un rectángulo grueso usando pocas baldosas.
- Calentar el sistema para provocar la desintegración del rectángulo grueso de modo que la estructura que “sobreviva” al proceso, sea nuestro rectángulo delgado objetivo.

La proposición 4.3.1 nos habla de la unicidad de la porción sobreviviente (que para nosotros será una porción $\mathcal{T} + \Delta$ -estable que contenga a la semilla) independiente de la secuencia de eventos que ocurrieron en el proceso de desintegración.

Proposición 4.3.1. *Sea D un \mathcal{T} -ensamblaje producido por un sistema de baldosas determinista $\mathbf{T} = (T, s, g, \mathcal{T})$. Entonces existe un único $M \subseteq D$, $\mathcal{T} + \Delta$ -subensamblaje maximal (c.r. a \subseteq), que es producido por \mathbf{T} .*

Demostración: La baldosa semilla s es un $\mathcal{T} + \Delta$ -ensamblaje. Sean C_1 y C_2 dos $\mathcal{T} + \Delta$ -ensamblajes producidos por el sistema de baldosas. El lemma 3.1.1 nos permite considerar $C_1 \cup C_2$ como un $\mathcal{T} + \Delta$ -ensamblaje producido por \mathbf{T} . \square

En este nuevo contexto podemos definir la complejidad de un $k \times n$ -rectángulo $C_{\mathcal{T}}^{\mathcal{T} + \Delta}(k, n)$, como el mínimo número de baldosas que producen únicamente una figura cuyo $\mathcal{T} + \Delta$ -subensamblaje maximal es el $k \times n$ -rectángulo.

Volvamos a nuestro problema original en el cual queremos construir un $3 \times n$ -rectángulo y apliquemos nuestro mecanismo. En el modelo con temperatura constante $\mathcal{T} = 2$ se necesitaban $\Theta(n^{1/3})$ baldosas distintas.

4.3. AUMENTANDO LA TEMPERATURA A $\mathcal{T} = 4$

Si construimos un $(\lambda_n + 3) \times n$ -rectángulo cuyo 4-subensamblaje maximal sea un $3 \times n$ -rectángulo y luego incrementamos la temperatura a $\mathcal{T} = 4$, al final del proceso, habremos construido un $3 \times n$ -rectángulo con sólo $O(\lambda_n)$ baldosas. En otras palabras, haciendo variable la temperatura, la complejidad de construir el rectángulo disminuye dramáticamente. Por ejemplo, si $n = 16,000,000$, la complejidad en el modelo a temperatura constante es 252 mientras que en el modelo a temperatura variable es 8.

Proposición 4.3.2. $C_2^4(3, n) = O(\lambda_n)$

Demostración: En la figura 4.7 se muestra como construir un $(\lambda_n + 3) \times n$ -rectángulo cuyo 4-subensamblaje maximal es un $3 \times n$ -rectángulo. Un $\lambda_n \times n$ -rectángulo es ensamblado usando la construcción previa. Baldosas especiales alargan la fila semilla para construir (con baldosas de mayor poder) un $3 \times n$ -rectángulo. \square

4.3. AUMENTANDO LA TEMPERATURA A $\mathcal{T} = 4$

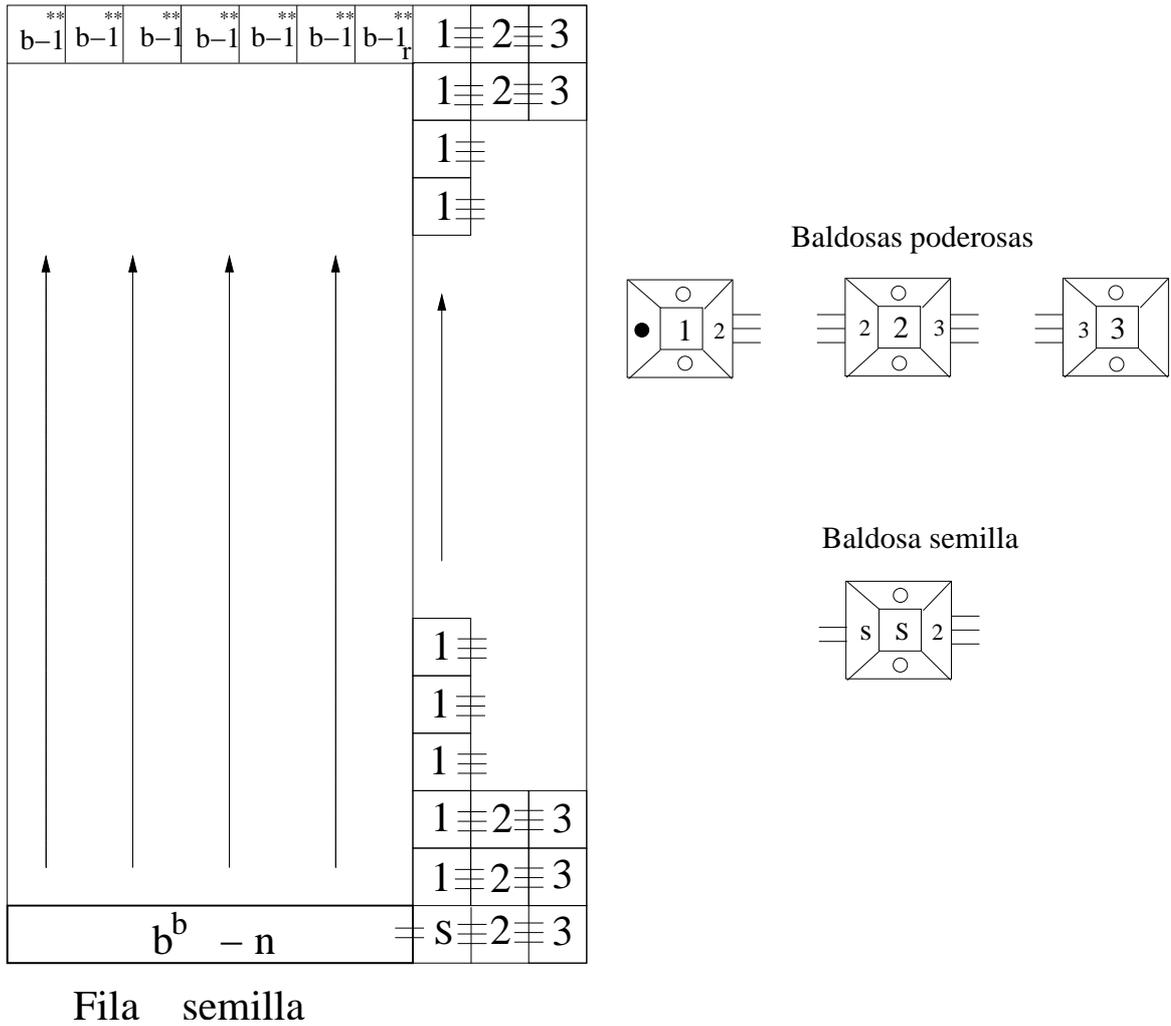


Figura 4.7: Las baldosas de mayor poder se ensamblan en un $3 \times n$ -rectángulo agregándose a la derecha del $\lambda_n \times n$ -rectángulo. Note que la semilla pertenece al $3 \times n$ -rectángulo.

Capítulo 5

Conclusiones

En nuestro estudio hemos mostrado que para Sistemas de Baldosas deterministas, cuando $\mathcal{T} = 1$, es posible distinguir Sistemas de Baldosas que permiten el auto-ensamblaje de estructuras arbitrariamente grandes de aquellos que no. Sin embargo, cuando $\mathcal{T} \geq 2$ ésto resulta imposible.

¿Será posible distinguir Sistemas de Baldosas deterministas de aquellos no deterministas?

También hemos estudiado la complejidad en la programación de $k \times n$ -rectángulos, y hemos constatado que aunque para rectángulo más angostos se requieren más baldosas en su construcción, un simple protocolo basado en un único cambio de la temperatura permite disminuir significativamente este número de baldosas.

¿Será de utilidad un protocolo que considere más que un único cambio de temperatura? En este caso, ¿es que la temperatura debe variar de manera monótona y creciente, o es que puede ser útil considerar aumentos y disminución en la temperatura?

¿Cuál es el tiempo que toma el sistema en ensamblar la estructura objetivo?. Adleman ha estudiado este problema [2] (the running time complexity). Sus resultados incluyen la obtención de cotas inferiores junto con construcciones que alcanzan esta cota para cuadrados de $n \times n$.

¿Cuál es la complejidad (en tiempo) de las construcciones exhibidas en este trabajo?
¿Como influyen los cambios en la temperatura en el tiempo que toma el ensamblaje?

Otra pregunta interesante, también abordada por Adleman es: ¿Cual es la concentración inicial óptima que garantiza que una figura dada se ensamble de la manera más rápida? Adleman [3] conjeturó que este problema (The Tile concentration Problem) es NP-completo, pero permanece aún abierto.

¿Que otros parámetros dependen de la concentración? ¿Puede la concentración ayudar a disminuir la complejidad en la programación de figuras auto-ensambladas?

Bibliografía

- [1] L. Adleman, Science 266, 1021-1024, 1994.
- [2] L. Adleman, Q. Cheng, A. Goel, M. Huang, “Runnun time and program size for self-assembled squares”, Proc. 33th ACM symposium on Theory of Computing (STOC), 2001, 740-748.
- [3] L. Adleman, Q. Cheng, A. Goel, M. Huang, D. Kempe, P. Moisset de Espanés and P. Rothmund, “Combinatorial optimization problems in self-assembly”, Proc. 34th ACM Symposium on Theory of Computing (STOC), 2002, 23-32.
- [4] L. Adleman, J. Kari, L. Kari and D. Reishus “On the decidability of self-assembly of infinite ribbons”, Proc. 43rd IEEE Symposium on Foundations of Computer Science (FOCS), 2002, 530-537.
- [5] Berger, R. “The Undecidability of the Domino Problem”, Memoirs of the American Mathematical Society, 66 (1966).
- [6] N. Jonoska, S. Karl and M. Saito, “Graph structures in DNA computing”. In *Computing with bio-molecules, theory and experiments*, (editor Gh. Pauen), Springer-Verlag (1998), 93-110.
- [7] M-Y Kao and V. Ramachandran, “DNA self-assembly for constructing 3D boxes”, Algorithms and Computation 12th International Symposium (ISAAC), 2001, Lecture Notes in Computer Science 2223, 429-440.
- [8] C. Mao, T. LaBean, J. Reif and N. Seeman, “Logical computation using algorithmic self-assembly of DNA triple-crossover molecules”, Nature 407, 493-496, (2000).
- [9] P. Rothmund and E. Winfree, “The program size complexity of self-assembled squares”, Proc. 33rd ACM Symposium on Theory of Computing (STOC), 2001, 459-468.
- [10] H. Wang, “Proving theorems by pattern recognition II”, Bell Systems Technical Journal, 40 (1961), 1-42.
- [11] E. Winfree, “Algorithmic self-assembly of DNA”, PhD thesis, California Institute of Technology, Pasadena, (1998).

BIBLIOGRAFÍA

- [12] E. Winfree, T. Eng and Grzegorz Rozenberg, “String tile models for DNA computing by self-assembly”, DNA Computing 2000, 63-88.
- [13] E. Winfree, X. Yang and N. Seeman, “Universal computation via self-assembly of DNA: some theory and experiments”, Proceedings of the Second Annual Meeting on DNA Based Computers, 1996, 172-190.