

# 行政院國家科學委員會專題研究計畫 成果報告

## DNA 生物晶片分析之光罩內部最少邊界點問題

計畫類別：個別型計畫

計畫編號：NSC93-2213-E-216-017-

執行期間：93年08月01日至94年07月31日

執行單位：中華大學資訊工程學系

計畫主持人：吳哲賢

計畫參與人員：顏勝茂 顏士勛

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 95 年 2 月 22 日

# 行政院國家科學委員會專題研究計畫成果報告

## 國科會專題研究計畫成果報告撰寫格式說明

### Preparation of NSC Project Reports

計畫編號：NSC93-2213-E-216-017-

執行期限：2004年8月1日至2005年7月31日

主持人：吳哲賢 中華大學資訊工程學系

E-mail: [jswu@chu.edu.tw](mailto:jswu@chu.edu.tw)

#### 一、中文摘要

生物晶片製造時，會因為光線的干擾造成晶片上面探針的錯誤。為了減少因為光線引起的錯誤，因此光罩邊界的總長度，便成為生物晶片光罩佈局好壞公認的評估標準。在本計畫中，我們針對單層生物晶片光罩最少邊界值問題，共推導出十二個定理，證明在固定光罩面積的條件下，最小邊界的值，並進一步提出最佳解之佈局演算法。

**關鍵詞：**生物晶片、光罩佈局、最佳演算法

#### Abstract

For biochip design process, probe errors will be happen by light influence. In order to reduce such error, the length of the mask border is the measurement standard of biochip mask layout for the minimum border problem of biochip mask. In this project, we totally derive 12 theorems to show the value of the minimum border length under the fixed mask area, and propose the optimal algorithm of biochip mask layout.

**Keywords:** biochip, mask layout, optimal algorithm.

#### 二、簡介

全球生物晶片在 1995 年展開研發熱潮，美國挾基因發展與半導體產業成熟獨步全球，生物晶片產業初期在美國即有 200 餘廠家投入此熱潮，至目前為止，約有 20 家大型規模的廠家。

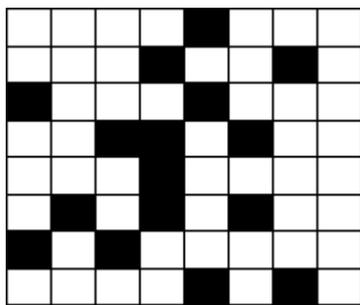
DNA 晶片對於生物科技產業的發展可媲美當年亨利福特所組織的汽車生產線對汽車工業的大革命，或是類似半導體對於電腦工業的革命性發展。總之，DNA 晶片是結合生物、電腦、機械等領域的研究人員所發展出來的最新生物科技，因此帶動了人類基因功能的自動化分析。DNA 晶片對於基因的檢測具有快速分析、準確率高、以及可以同時大量檢測基因的優點，因此改變了從前研究人員終其一生只鑽研一種基因的研究方式。

DNA 晶片的原理是將數千或數萬點 (spot)，以高密度的方點製作在拇指般大的晶片上，其材料可為玻璃片或是尼龍薄膜。每一小點為單股的 DNA，又稱為探針 (probe)。主要有兩種來源，核甘酸 oligonucleotide 和已經存在於基因庫中的互補核甘酸 cDNA。來源一的核甘酸晶片主要是由 Affymetrix(位於加州的 Santa Clara)這家生技公司所製造，他們主要是利用 DNA 的 A、T、C、G 四種鹼基，用類似築摩天高樓的方式，一個個不同的組合，築上約 20~25 個(層)核甘酸。來源二的互補核甘酸的晶片主要是利用從病人的檢體或是其他的生物體抽取出的已知的互補核甘酸，然後將這些互補核甘酸點在晶片上。

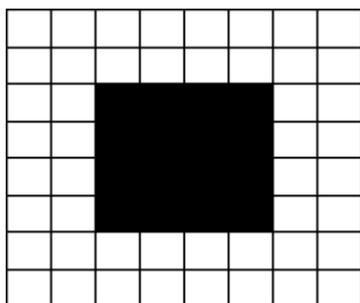
DNA 晶片的應用範圍廣泛，可應用於細胞生物化學的研究及疾病診斷上。例如癌症腫瘤學上的臨床診斷，或者在各種不同病毒感染疾病的診斷，像是 AIDS 和腸病毒的診斷。藉由 DNA 晶片的應用，可以快速又正確的診斷，以節省大量的人力及物力，並且可以搶在第一個時間點上救病人，以期做到早期發現，及早治療的功效。經由 Very Large-Scale Immobilized Polymer

Synthesis(VLSIPS)[20-22, 24-25] 這個技術來製造的生物晶片，會有因為光罩而產生的錯誤。這個錯誤主要原因是被光罩遮住的部份，會因為光線的散射及干擾等物理因素，讓本來不該被打掉的光敏保護基被打掉，使得晶片上面原本控制利用光罩遮住的位置，產生光反應，造成錯誤的結果。為了評估一片生物晶片製造的正確率，目前的研究，便利用光罩總邊界值來評估生物晶片製造的正確率[1-25]，總邊界越小，則錯誤率就越小。例如同樣利用光罩遮住 16 個位置，圖一光罩的邊界為 58；如果排成圖二，邊界就減少為 16，很明顯的減少很多。可以看出來，圖二的邊界減少後，可以減少光線干擾的範圍，所以錯誤的產生也會因此降低。

現在已知求光罩最小邊界的問題，是一個 NP-Hard 的問題[9, 14, 23]。因此目前大部分的研究成果都是啟發式(heuristic)的演算法，所求得解只是近似解。為了簡化题目的複雜度，本論文只考慮，在固定光罩遮住面積的條件下，一層探針的光罩佈局。本論文總共推導出十二個定理，來證明出最小邊界之值，並進一步提出得到這些最佳解之佈局演算法。



圖一：16 個光罩遮住位置，光罩邊界為 58



圖二：16 個光罩遮住位置，光罩邊界為 16

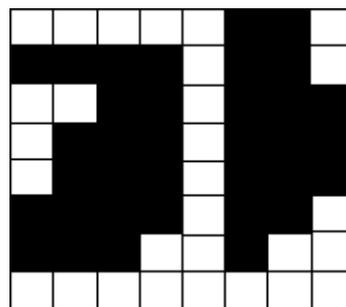
### 三、定理之推導與證明

光罩佈局的方式，類似在二維的方格紙上，把光罩遮住部分的方格塗黑。首先定義一塊區域光罩佈局的圖形，何謂凸多邊形，何謂凹多邊形。

[定義一]一塊區域光罩佈局的圖形，分為凹多邊形及凸多邊形兩種。

凹多邊形：存在兩個以上的邊可以互相正面彼此看到。

凸多邊形：不存在兩個以上的邊可以互相正面彼此看到。

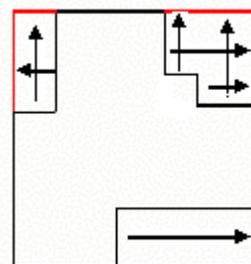


圖三：左圖為凹多邊形，右圖為凸多邊形

圖三為一範例。其中左圖為凹多邊形，右圖為凸多邊形凸多邊形。接著推導相關的定理。

[定理一]在一個區塊的凸多邊形下，邊界值一定是偶數。

證明：若此凸多邊形為矩形，邊長分別為  $a$  及  $b$ ，則邊界值為  $2(a+b)$ ，屬於偶數。若此凸多邊形不為矩形，則可以依據上下左右四個邊，圍出一個新的矩形(圖四)。原凸多邊形不在矩形上的邊，可分別一對一對映到矩形上，補成矩形。由圖可看出，原凸多邊形的邊界值和矩形的邊界值相等，所以亦屬於偶數。



圖四：凸多邊形依據上下左右四個邊，

### 圍出一個新的矩形

[定理二] 在一個區塊的凸多邊形下，邊界值為  $4L$  或  $4L + 2$ ,  $L \in \mathbb{Z}^+$ 。

證明：考慮光罩最小面積為 1 時，邊界值為 4。定理一已證明邊界值是偶數，所以邊界值為：4、6、8、10...。可以再細分為兩群： $\{4、8、\dots\}$  及  $\{6、10、\dots\}$ ，即邊界值為  $4L$  或  $4L + 2$ ,  $L \in \mathbb{Z}^+$ 。

[定理三] 在一個區塊的凸多邊形下，固定邊界值之最大面積的圖形，一定是矩形。

證明：假設在一個區塊的凸多邊形下，固定邊界值之最大面積的圖形不是矩形。由圖四我們可以看出，此凸多邊形可圍出一個新的矩形，邊界值相等，但是面積比原來凸多邊形還大。假設發生錯誤，所以圖形一定是矩形。

[定理四] 在一個區塊的凸多邊形下，邊界值為  $4L$  的最大面積為  $L^2$ ,  $L \in \mathbb{Z}^+$ ；且圖形為邊長  $L$  的正方形。

證明：由定理三得知，固定邊界值之最大面積圖形是矩形。假設矩形邊長分別為  $a$  及  $b$ ，則邊界值  $4L = 2(a + b)$ ，面積為  $ab$ 。利用幾何平均數小於等於算數平均數的定理，我們可以推導：

$$(ab)^{1/2} \leq (a + b) / 2. \quad (1)$$

$$ab \leq [(a + b) / 2]^2 = L^2. \quad (2)$$

得知面積的整數上限值為  $L^2$ 。

發現邊長為  $L$  的正方形，其邊界值為  $4L$ ，面積為  $L^2$ 。因此在一個區塊的凸多邊形下，邊界值為  $4L$  的最大面積為  $L^2$ ,  $L \in \mathbb{Z}^+$ ；且圖形為邊長  $L$  的正方形。

[定理五] 在一個區塊的凸多邊形下，邊界值為  $4L + 2$  的最大面積為  $L^2 + L$ ,  $L \in \mathbb{Z}^+$ ；且圖形為邊長分別為  $L$  及  $L + 1$  的矩形。

證明：由定理三得知，固定邊界值之最大面積圖形是矩形。假設矩形邊長分別為  $a$

及  $b$ ，則邊界值  $4L + 2 = 2(a + b)$ ，面積為  $ab$ 。利用(1)：幾何平均數小於等於算數平均數的定理，我們可以推導：

$$ab \leq [(a + b) / 2]^2 \\ = (L + 1/2)^2 = L^2 + L + 1/4. \quad (3)$$

得知面積的整數上限值為  $L^2 + L$ 。

發現邊長分別為  $L$  及  $L + 1$  的矩形，其邊界值為  $2[L + (L + 1)] = 4L + 2$ ，面積為  $L(L + 1) = L^2 + L$ 。因此在一個區塊的凸多邊形下，邊界值為  $4L + 2$  的最大面積為  $L^2 + L$ ,  $L \in \mathbb{Z}^+$ ；且圖形為邊長分別為  $L$  及  $L + 1$  的矩形。

[定理六] 在一個區塊的多邊形下，面積  $A + 1$  的最小邊界值，大於等於面積  $A$  的最小邊界值， $A \in \mathbb{Z}^+$ 。

證明：假設面積  $A + 1$  的最小邊界值，小於面積  $A$  的最小邊界值。令  $G$  為面積  $A + 1$  的最小邊界值之圖形。所以，

$$\text{邊界值}\{\text{圖形 } G : \text{面積 } A + 1\} \\ = \text{最小邊界值}\{\text{面積 } A + 1\} \\ < \text{最小邊界值}\{\text{面積 } A\}. \quad (4)$$

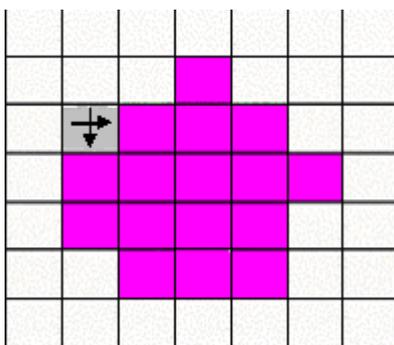
$$\text{最小邊界值}\{\text{面積 } A\} \\ \leq \text{邊界值}\{\text{所有圖形} : \text{面積 } A\}. \quad (5)$$

利用(4)及(5)，可推導出，

$$\text{邊界值}\{\text{圖形 } G : \text{面積 } A + 1\} \\ < \text{邊界值}\{\text{所有圖形} : \text{面積 } A\}. \quad (6)$$

面積  $A + 1$  之最小邊界值圖形  $G$ ，移除一凸點後，形成面積為  $A$  的新圖形，邊界值不變或減少(圖五)。

上述結果違背式子(6)，假設發生錯誤。因此在一個區塊的多邊形下，面積  $A + 1$  的最小邊界值，大於等於面積  $A$  的最小邊界值。



圖五：面積  $A + 1$  之最小邊界值的圖形  $G$ ，移除一凸點後，邊界值不變或減少，面積減為  $A$

[定理七] 在一個區塊的多邊形下，光單面積  $A$  大於光單面積  $B$ ，則面積  $A$  的最小邊界值大於等於面積  $B$  的最小邊界值； $A, B \in \mathbb{Z}^+$ 。

證明：因為光單面積： $A > B$ ，由定理六得知，最小邊界值{面積  $A$ }  $\geq$  最小邊界值{面積  $A - 1$ }  $\geq$  最小邊界值{面積  $A - 2$ }  $\geq \dots \geq$  最小邊界值{面積  $B + 1$ }  $\geq$  最小邊界值{面積  $B$ }。所以最小邊界值{面積  $A$ }  $\geq$  最小邊界值{面積  $B$ }。

[定理八] 固定光單面積下，最小邊界值的圖形，一定是一個區塊的凸多邊形。

證明：首先證明會形成一個區塊。假設固定光單面積下，最小邊界值的圖形不只一個區塊，針對任兩個區塊上下左右任一邊互相連接，形成一個新的圖形。則此一新圖形的光單面積不變，但是邊界值減少了互相連接的部份。假設發生錯誤，所以固定光單面積下，最小邊界值的圖形，一定是一個區塊。

接著證明是一個凸多邊形。假設固定光單面積  $A$  下，最小邊界值的圖形為凹多邊形。依照定義一，針對互相正面彼此看到的邊再加一個連接邊，形成一個新的圖形(圖六)。由圖可看出，新的圖形面積增加  $a$ ，邊界值減少互相對看的兩邊。所以，

$$\begin{aligned} & \text{最小邊界值}\{\text{面積 } A\} \\ &= \text{邊界值}\{\text{凹多邊形}\} \end{aligned}$$

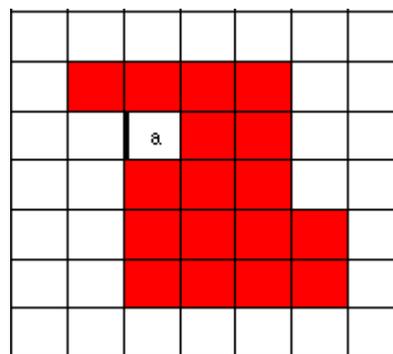
$$> \text{邊界值}\{\text{新的圖形}\}. \quad (7)$$

$$\begin{aligned} & \text{邊界值}\{\text{新的圖形}\} \\ &= \text{邊界值}\{\text{面積 } A + a\} \\ &\geq \text{最小邊界值}\{\text{面積 } A + a\}. \quad (8) \end{aligned}$$

利用(7)及(8)，可推導出，

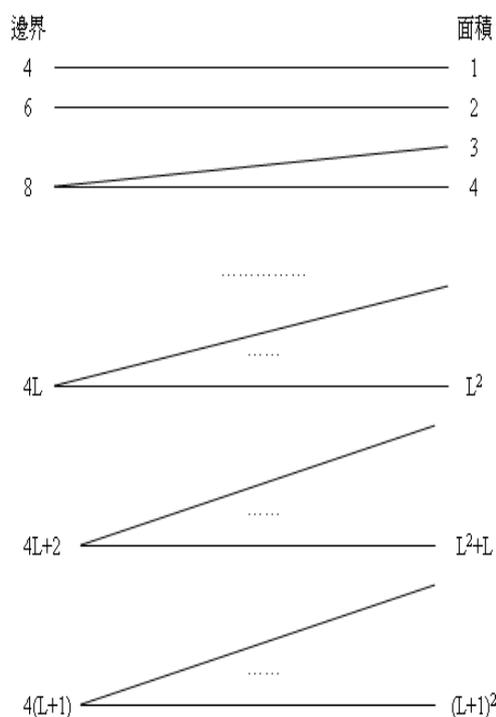
$$\begin{aligned} & \text{最小邊界值}\{\text{面積 } A\} \\ &> \text{最小邊界值}\{\text{面積 } A + a\}. \quad (9) \end{aligned}$$

上式結果違背定理七，假設發生錯誤。因此固定光單面積下，最小邊界值的圖形不是凹多邊形，而是一個區塊的凸多邊形。



圖六：凹多邊形，加一個連接邊，形成一個新的圖形。面積增加  $a$ ，邊界值減小

在一個區塊的凸多邊形下，由定理二：邊界值為  $4L$  或  $4L + 2$ ， $L \in \mathbb{Z}^+$ ；定理四：邊界值為  $4L$  的最大面積為  $L^2$ ， $L \in \mathbb{Z}^+$ ；以及定理五：邊界值為  $4L + 2$  的最大面積為  $L^2 + L$ ， $L \in \mathbb{Z}^+$ 。因此我們可以在圖七，畫出光單邊界和面積的對映關係。



圖七：一個區塊的凸多邊形下，光罩邊界和面積的對映關係

[定理九] 光罩面積為  $L^2$ ，則最小邊界值為  $4L$ ，圖形為邊長  $L$  的正方形， $L \in \mathbb{Z}^+$ 。

證明：由圖七我們知道，當光罩面積為  $L^2$ ，最小邊界值為  $4L$ 。而且由定理四可得知，此圖形為邊長  $L$  的正方形。

[定理十] 光罩面積為  $L^2 + L$ ，則最小邊界值為  $4L + 2$ ，圖形為邊長分別為  $L$  及  $L + 1$  的矩形， $L \in \mathbb{Z}^+$ 。

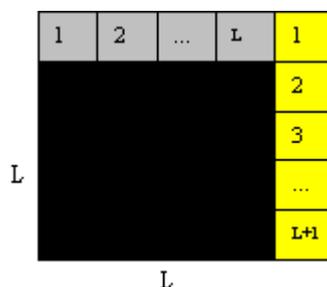
證明：由圖七我們知道，當光罩面積為  $L^2 + L$ ，最小邊界值為  $4L + 2$ 。而且由定理五可得知，此圖形為邊長分別為  $L$  及  $L + 1$  的矩形。

[定理十一] 光罩面積為  $L^2 + 1$  到  $L^2 + L$  時，則最小邊界值為  $4L + 2$ ， $L \in \mathbb{Z}^+$ 。

證明：由定理九得知，光罩面積為  $L^2$ ，最小邊界值為  $4L$ ，且圖形為邊長  $L$  的正方形。依序由此正方形的一邊，從頭到尾連續增加光罩面積從  $L^2 + 1$  到  $L^2 + L$  (圖八)。由圖可知，開始增加第一個點時，邊界值增加 2，變成  $4L + 2$ 。爾後增加點時，邊界

值都不變，維持  $4L + 2$ ，一直到面積增為  $L^2 + L$ ，形成邊長分別為  $L$  及  $L + 1$  的矩形。

由圖七我們知道，當光罩面積大於  $L^2$ ，最小邊界值至少為  $4L + 2$ 。因此光罩面積為  $L^2 + 1$  到  $L^2 + L$  時，最小邊界值為  $4L + 2$ 。



圖八：光罩面積從  $L^2 + 1$  依序增加，經由  $L^2 + L$  到  $(L + 1)^2$

[定理十二] 光罩面積為  $(L^2 + L) + 1$  到  $(L + 1)^2$  時，最小邊界值為  $4(L + 1)$ ， $L \in \mathbb{Z}^+$ 。

證明：由定理十得知，光罩面積為  $L^2 + L$ ，最小邊界值為  $4L + 2$ ，且圖形為邊長分別為  $L$  及  $L + 1$  的矩形。依序由邊長  $L + 1$  的一邊，從頭到尾連續增加光罩面積從  $(L^2 + L) + 1$  到  $(L + 1)^2$  (圖八)。由圖可知，開始增加第一個點時，邊界值增加 2，變成  $(4L + 2) + 2 = 4(L + 1)$ 。爾後增加點時，邊界值都不變，維持  $4(L + 1)$ ，一直到面積增為  $(L + 1)^2$ ，形成邊長為  $L + 1$  的正方形。

由圖七我們知道，當光罩面積大於  $L^2 + L$ ，最小邊界值至少為  $4(L + 1)$ 。因此光罩面積為  $(L^2 + L) + 1$  到  $(L + 1)^2$  時，最小邊界值為  $4(L + 1)$ 。

#### 四、最佳解演算法

考慮一層探針的光罩佈局，在固定光罩遮住面積的條件下，根據定理九至定理十二，我們可以推導出擁有最小邊界值之最佳解佈局演算法。

#### [最佳解演算法]

[步驟 1] 當光罩面積為  $L^2$ ， $L \in \mathbb{Z}^+$  時，則最小邊界值為  $4L$ ，且圖形為邊

長  $L$  的正方形。

- [步驟 2] 當光罩面積為  $L^2 + L$ ,  $L \in \mathbb{Z}^+$  時, 則最小邊界值為  $4L + 2$ , 且圖形為邊長分別為  $L$  及  $L + 1$  的矩形。
- [步驟 3] 當光罩面積為  $L^2 + k$ ,  $L \in \mathbb{Z}^+$ ,  $k \in \{1, 2, \dots, L - 1\}$  時, 則最小邊界值為  $4L + 2$ , 且圖形為在邊長為  $L$  的正方形一邊, 從頭到尾連續增加面積  $k$ 。
- [步驟 4] 當光罩面積為  $(L^2 + L) + k$ ,  $L \in \mathbb{Z}^+$ ,  $k \in \{1, 2, \dots, L\}$  時, 則最小邊界值為  $4(L + 1)$ , 且圖形為在邊長分別為  $L$  及  $L + 1$  的矩形上, 針對邊長為  $L + 1$  的一邊, 從頭到尾連續增加面積  $k$ 。

## 五、結論

考慮一層探針的光罩佈局, 在固定光罩遮住面積的條件下, 本計畫中, 我們證明十二個定理, 並根據定理推導出擁有最小邊界值之最佳解佈局演算法。考慮多層探針之光罩佈局最小邊界值問題, 將是我們未來值得研究的方向。

## 六、參考文獻

- [1] S. Akers, "On the use of the linear assignment algorithm in module placement", Proc. 18th Design Automation Conference, pp. 137-144, 1981.
- [2] C. Alpert and A. Kahng, "Geometric Embeddings for Faster (and Better) Multi-Way Netlist Partitioning" Proc. ACM/IEEE Design Automation Conf., pp. 743-748, 1993.
- [3] C. Alpert and A. Kahng, "Recent directions in netlist partitioning: A survey", Integration: The VLSI Jour. 19, pp. 1-81, 1995.
- [4] A. Caldwell, A. Kahng and I. Markov, "Can Recursive bisection Produce Routable Designs?", DAC, pp.477- 482, 2000.
- [5] C. Chang, J. Cong and M. Xie, "Optimality and Scalability Study of Existing Placement Algorithms", Proc. Asia South- Pacific Design Automation Conference, Jan. 2003.
- [6] J. Cong, M. Romesis and M. Xie, "Optimality, Scalability and Stability Study of Partitioning and Placement Algorithms", Proc. ISPD, pp. 88-94, 2003.
- [7] D. Cutting, D. Karger, J. Pederson and J. Tukey, "Scatter/Gather: A Cluster-Based Approach to Browsing Large Document Collections", (15th Intl. ACM/SIGIR Conference on Research and Development in Information Retrieval) SIGIR Forum, pp. 318-329, 1992.
- [8] K. Doll, F. Johannes, and K. Antreich, "Iterative Placement Improvement by Network Flow Methods", IEEE Transactions on Computer-Aided Design 13(10), pp. 1189-1200, 1994.
- [9] D. Franzblua and D. Kleitman, "An algorithm for covering polygons with rectangles", Information and Control, pp. 63-164, 1984.
- [10] W. Feldman and P. Pevzner, "Gray code masks for sequencing by hybridization", Genomics, 23 (1994), pp. 233-235, 1994.
- [11] C. Fiduccia and R. Mattheyses, "A Linear-Time Heuristic for Improving Network Partitions", Proc. Design Automation Conference, pp. 175-181, 1982.
- [12] S. Fodor, J. Read, M. Pirrung, L. Stryer, L. Tsai and D. Solas, "Light-Directed, Spatially Addressable Parallel Chemical Synthesis", Science 251, pp. 767-773, 1991.
- [13] D. Geschwind and J. Gregg (Eds.), "Microarrays for the neurosciences: an essential guide", MIT Press, Cambridge, MA, 2002.
- [14] M. Garey, D. Johnson and W. Freeman, "Computers and Intractability : A Guide to the Theory of NP-Completeness." San Francisco, CA, 1979.
- [15] L. Hagen, D. Huang and A. Kahng, "Quantified Suboptimality of VLSI Layout Heuristics", Proc. ACM/IEEE Design Automation Conf., pp. 216-221, 1995.
- [16] S. Hannenhalli, E. Hubbell, R. Lipshutz

- and P. Pevzner, “Combinatorial Algorithms for Design of DNA Arrays”, in *Chip Technology* (ed. J. Hoheisel), Springer-Verlag, 2002.
- [17] D. Huang and A. Kahng, “Partitioning-Based Standard-Cell Global Placement with an Exact Objective”, in *Proc. ACM/IEEE Intl. Symp. on Physical Design*, Napa, pp. 18-25, April 1997.
- [18] A. Kahng, I. Mçandoiu, P. Pevzner, S. Reda, and A. Zelikovsky, “Border Length Minimization in DNA Array Design”, *Proc. 2nd International Workshop on Algorithms in Bioinformatics (WABI 2002)*, R. Guigó and D. Gusfield (Eds.), Springer-Verlag Lecture Notes in Computer Science Series 2452, pp. 435–448, 2002.
- [19] A. Kahng, I. Mçandoiu, P. Pevzner, S. Reda, and A. Zelikovsky, “Engineering a Scalable Placement Heuristic for DNA Probe Arrays”, *Proc. 7th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2003)*, W. Miller, M. Vingron, S. Istrail, P. Pevzner and M. Waterman (Eds.), 2003, pp. 148–156, 2003.
- [20] T. Kozawa et al., “Automatic Placement Algorithms for High Packaging Density VLSI”, *Proc. 20th Design Automation Conference (DAC 1983)*, pp. 175–181, 1983.
- [21] R. Lipshutz, S. Fodor, T. Gingeras, D. Lockhart, “High density synthetic oligonucleotide arrays” *Nat. Genet.* 21 (1999), pp. 20–24, 1999.
- [22] K. Shahookar and P. Mazumder, “VLSI Cell Placement Techniques”, *Computing Surveys* 23(2) (1991), pp. 143-220, 1991.
- [23] A. Pevzner “*Computational Molecular Biology : An Algorithmic Approach*”, The MIT Press instead, 2000.
- [24] M. Sarrafzadeh and M. Wang, “NRG: GLobal and Detailed Placement”, *Proc. International Conference on Computer-Aided Design (ICCAD 1997)*, pp. 532–537, 1997.
- [25] M. Wang, X. Yang and M. Sarrafzadeh, “DRAGON2000: Standard-cell Placement Tool For Large Industry Circuits”, *Proc. International Conference on Computer Aided Design (ICCAD 2001)*, pp. 260–263, 2001.

