

行政院國家科學委員會專題研究計畫 成果報告

低維韻律陣列設計上封閉解之研究

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-009-090-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學資訊工程學系

計畫主持人：蔡中川

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 8 月 18 日

低維韻律陣列設計上封閉解之研究

On the Closed-Form Mapping Conditions in Designing Lower-Dimensional Systolic Arrays

計畫編號：NSC 92 - 2213 - E - 009 - 090 -

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：蔡中川 交通大學資訊工程系

計畫參與人員：黃為霖、林祖年 交通大學資訊工程系

一、中文摘要

當一 k 維韻律陣列用以執行 $k + 2$ 維以上之規律演算法時，我們稱其為低維韻律陣列，簡稱低維陣列。一低維陣列應避免發生通道相衝以確保其正確性。因此，通道相衝之檢查在低維陣列設計上一直扮演著相當重要的角色。過去已有許多通道相衝檢查方法被提出，但大部份方法之計算時間皆與規律演算法之問題大小相關，故當問題大小增大時，它們便顯得過於耗時。為解決此問題，已有幾位研究者針對映射 n 維規律演算法至 $n - 2$ 維陣列之常見情況，提出計算複雜度與問題大小無關之通道相衝檢查條件(稱為封閉解)。然而，這些封閉解限制了低維陣列內只能有鄰近資料通道，且只可用於平行多面體與梯形 I/O 空間，故它們具有應用範圍太窄之缺點。為處理此缺點，在本計劃中，我們對過去封閉解加以推廣，使其允許具有非鄰近資料通道之低維陣列。此外，對於映射 3 維規律演算法至線性陣列的情形，我們進一步提出了一新封閉解，其可適用於任意凸多邊形之 I/O 空間。

關鍵詞：封閉解、I/O 空間、通道相衝、低維陣列、韻律陣列

Abstract

A lower-dimensional systolic array ("lower-dimensional array" for short) is a k -dimensional systolic array that implement a regular algorithm of dimension greater than or equal to $k + 2$. Since a lower-dimensional array must avoid link conflicts to ensure its

validity, the checking of link conflicts is currently an important issue in designing lower-dimensional arrays. In the literature, several methods for checking link conflicts have been presented. However, most of them are time-consuming when the problem size is large, because their time complexities depend on the problem size of a regular algorithm. To overcome the problem, several closed-form mapping conditions have been proposed for checking link conflicts in the mapping of n -dimensional algorithms onto $(n - 2)$ -dimensional arrays. However, these conditions suffer from limited applicability since they are restricted to arrays with neighboring connections and to regular algorithms with parallelepiped or trapezoid I/O spaces. To conquer the limited applicability, in this project, we adapt these conditions to arrays with non-neighboring connections, and further propose a new closed-form mapping condition for the mapping of 3-dimensional algorithms with arbitrary convex I/O spaces onto linear arrays.

Keywords: closed-form mapping condition, I/O space, link conflict, lower-dimensional array, systolic array

二、緣由與目的

由於 VLSI 技術的進步，在一矽晶片上製作含有許多處理器之平行計算機已屬可能，而韻律陣列(systolic array)被認為是一種特別適合以 VLSI 技術加以製作的特殊平行計算機器[6,7]。

一韻律陣列是由許多規律排列之處理

器彼此局部相連所組成。由於其互連結構之規律性與局部性，韻律陣列常用於執行各種具有均勻(uniform)資料相依關係之規律演算法(regular algorithm)，如矩陣相乘與 LU 分解等演算法[8]。當一 k 維韻律陣列用以執行 $k+2$ 維以上之規律演算法時，我們稱此陣列為低維(lower-dimensional)韻律陣列，簡稱低維陣列。在韻律陣列的設計上，如何將一規律演算法映射至低維陣列執行，一直是個相當重要的問題[8,9,10]。

早期研究低維陣列的學者[9-11]指出，一低維陣列之設計必須避免發生通道相衝(link conflict)以確保其正確性，也就是，其必須避免有兩不同資料元素(data token)同時進入同一資料通道(data link)內。因此，在低維陣列的設計上，通道相衝之檢查扮演著相當重要的角色。過去已有許多通道相衝檢查方法被提出[1-4,9,10,13-16]，在不限制規律演算法之維度 n 與低維陣列之維度 k 的情況下，最好的方法[3,12,13,15]需列舉 $O((2N)^{n-k-1})$ 個整數點才可檢查出通道是否相衝；此處 N 為規律演算法之問題大小(problem size)。由於 N 通常相當大，故不限制 n 與 k 之通道相衝檢查方法在實際應用上皆有太耗時之缺點。

為了能更有效率地檢查通道相衝，幾位研究者針對低維陣列設計上的某些特殊場合，提出了對應的封閉解(closed-form mapping condition)；此處所謂封閉解，係指計算複雜度與 N 無關之通道相衝檢查條件。在[1,2]中，Ganapathy 和 Wah 分別針對矩陣相乘與遞連閉包問題(transitive closure)之規律演算法，提出了映射它們至線性陣列(linear array)情況下的封閉解。在[13,15]中，Xue 進一步針對所有 I/O 空間為平行多面體(parallelepiped)之 n 維規律演算法，提出了映射它們至 $n-2$ 維陣列情況下的封閉解。由於 Xue 所考慮的規律演算法包含矩陣相乘與遞連閉包問題之演算法，故 Ganapathy 和 Wah 的封閉解實際上為 Xue 的封閉解之一特例。此外，Xue 在[14,16]中也考慮了所有 I/O 空間為梯形(trapezoid)之 3 維規律演算法(三角形與平行多邊形為

梯形之特例)；他針對這些演算法映射至線性陣列的情形，提出了更為複雜的封閉解。

上述各封閉解皆有適用範圍太窄之缺點：不是僅適用於特定形狀之 I/O 空間，就是限制低維陣列內之資料通道必須為鄰近資料通道(neighboring link)。因而前述之封閉解無法滿足各種實際應用上的需求，為了克服此缺點，在本計劃中，我們從兩方面改進過去各封閉解。首先，基於[3]中對非鄰近互連陣列模式之研究成果，我們推廣了[13,15]中的封閉解，使其可應用在具有非鄰近資料通道的低維陣列上。接著，對於映射 3 維規律演算法至線性陣列的情形，我們提出了改進[14,16]之結果的新封閉解。我們的新封閉解不只適用於具有非鄰近資料通道之低維陣列，且也適用於所有 I/O 空間為任意凸多邊形之規律演算法。此外，為了測試新封閉解，我們已將其應用於各種實例上；我們的研究結果顯示，新封閉解滿足許多實際應用上的需求。

三、結果與討論

由於篇幅限制，本節僅簡要描述本計劃之研究成果。至於更詳細之內容，我們將另外撰寫論文發表。

3.1 基本假設與定理

令 V 為規律演算法內之一變數，且令 io_V, d_V, l_V 分別表示對應 V 之 I/O 空間、資料相依向量、與資料通道(以上各參數之詳細說明可參考[3,15,16])；因為我們允許非鄰近資料通道，故 $\gcd(l_V) = 1$ 非為必要條件，其中 $\gcd(l_V)$ 表 l_V 內各元素之最大公因數。在本報告中，我們假設 io_V 可表示為

$$io_V = \{M_V \overset{p}{x} + \overset{p}{q}_V \mid \overset{p}{x} \in R_V \cap \mathbf{Z}^{n-1}\}$$

之形式，其中， $M_V \in \mathbf{Z}^{n \times (n-1)}$ ， $\overset{p}{q}_V \in \mathbf{Z}^n$ 分別為一行滿秩(full column rank)矩陣與向量， R_V 則為一 $n-1$ 維凸形多面體(convex polytope)。上述假設使得我們能簡單地用一矩陣 $\Theta_V \in \mathbf{R}^{k \times (n-1)}$ 來描述變數 V 之所有輸

出入資料元素在低維陣列上的分佈； Θ_V 在 [13,15] 中被稱為分佈矩陣 (distribution matrix)，其第 i 行 ($i = 1, \dots, n-1$) 為

$$\theta_i^{\rho} = Sm_i^{\rho} - \frac{Hm_i^{\rho}}{Hd_V^{\rho}} Sd_V^{\rho},$$

其中， m_i^{ρ} 表 M_V 之第 i 行， $H \in \mathbf{Z}^{1 \times n}$ 與 $S \in \mathbf{Z}^{k \times n}$ 表時間排程向量 (time scheduling vector) 與空間映射矩陣 (space mapping matrix)。此外，我們也假設 d_V^{ρ} 對 i_{O_V} 滿足

$$d_V^{\rho} \notin \{M_V x^{\rho} \mid x^{\rho} \in \mathbf{R}^{n-1}\}$$

之關係，以避免當 N 增大時通道相衝總是會發生。(在 [15,16] 中，Xue 也有一與此類似之假設。)

我們接著描述本報告用以推導新封閉解之基本定理。令

$$\text{Diff}(X) = \{x_1^{\rho} - x_2^{\rho} \mid x_1^{\rho}, x_2^{\rho} \in X\}$$

表一集合 X 之差異集 (difference set)；基於前述兩假設與 [3, Theorem 4.4]，我們證明了在 $k = n - 2$ 之情況下，通道相衝檢查條件之形式如下：

定理 1 若 $k = n - 2$ 且 R_V 滿足 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1}) = \text{Diff}(R_V) \cap \mathbf{Z}^{n-1}$ ，則變數 V 之資料元素不發生通道相衝，若且唯若

$$z_{\min} \xi^{\rho} \notin \text{Diff}(R_V \cap \mathbf{Z}^{n-1}),$$

其中，

$$z_{\min} = \frac{\gcd(l_V^{\rho})}{\gcd(\gcd(SM_V \xi^{\rho}), \gcd(l_V^{\rho}))},$$

ξ^{ρ} 為 Θ_V 之零空間 (null space) 內一互質整數向量。

根據此定理，我們觀察到當 $k = n - 2$ 時，推導封閉解之問題可化簡為推導 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1})$ 的表示法之問題。在本計劃中，我們便是藉由分析不同情況下的 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1})$ 之表示法，而推導出對應於

不同情況的封閉解。

3.2 適用於平行多面體 I/O 空間之封閉解

本小節將推導映射 n 維規律演算法至 $n - 2$ 維陣列情況下的一個封閉解。此處所針對的 I/O 空間為平行多面體 I/O 空間，其根據 [13,15] 可定義如下：

定義 1 若定義 i_{O_V} 之凸形多面體 R_V 滿足

$$R_V = \{x^{\rho} \in \mathbf{R}^{n-1} \mid \mathbf{0} \leq x^{\rho} \leq b^{\rho}\},$$

其中 $b^{\rho} \geq \mathbf{0}$ 為一整數向量，則稱 i_{O_V} 為一平行多面體 I/O 空間。

由上述定義，不難直接導出 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1})$ 之表示法如下：

引理 1 若 i_{O_V} 為一平行多面體 I/O 空間，則 R_V 滿足 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1}) = \text{Diff}(R_V) \cap \mathbf{Z}^{n-1}$ 且 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1}) = \{x^{\rho} \in \mathbf{Z}^{n-1} \mid -b^{\rho} \leq x^{\rho} \leq b^{\rho}\}$ 。

基於此引理，我們可由定理 1 得到下列封閉解：

定理 2 若 $k = n - 2$ 且 i_{O_V} 為一平行多面體 I/O 空間，則變數 V 之資料元素不發生通道相衝，若且唯若

$$\max \left\{ \frac{|\xi_1|}{b_1 + 1}, \dots, \frac{|\xi_{n-1}|}{b_{n-1} + 1} \right\} \geq \frac{1}{z_{\min}},$$

其中， $\xi^{\rho} = [\xi_1, \dots, \xi_{n-1}]^T$ 與 z_{\min} 同定理 1 內之定義， $b^{\rho} = [b_1, \dots, b_{n-1}]^T$ 同定理 1 內之定義。

當 l_V^{ρ} 為滿足 $\gcd(l_V^{\rho}) = 1$ 之鄰近資料通道時，定理 2 之封閉解即成為 [13, Theorem 11] 及 [15, Theorem 14] 內所描述的封閉解。由此可知，定理 2 為 [13, Theorem 11] 及 [15, Theorem 14] 之推廣，其可用於設計含有非鄰近資料通道之低維陣列。

3.3 適用於凸多邊形 I/O 空間之封閉解

在本小節中，我們推導映射 3 維規律演算法至線性陣列情況下的一個封閉解。

此處考慮的 I/O 空間為任意凸多邊形 I/O 空間，其定義如下：

定義 2 若定義 io_V 之凸形多面體 R_V 為一頂點(vertex)皆為整數點之凸多邊形，則稱 io_V 為一凸多邊形 I/O 空間。

不失一般性，我們假設上述定義中之 R_V 被表示為以下形式：

$$R_V = \{[x_1, x_2]^T \in \mathbf{R}^2 \mid l_i \leq a_{i1}x_1 + a_{i2}x_2 \leq u_i, i=1, \dots, m\},$$

其中， $a_{i1}, a_{i2}, l_i, u_i \in \mathbf{Z}$ 且 l_i, u_i 滿足

$$l_i = \min\{a_{i1}x_1 + a_{i2}x_2 \mid [x_1, x_2]^T \in R_V\},$$

$$u_i = \max\{a_{i1}x_1 + a_{i2}x_2 \mid [x_1, x_2]^T \in R_V\}.$$

基於此假設，我們證明了 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1})$ 之表示法有一相當簡單之形式：

引理 2 若 io_V 為一凸多邊形 I/O 空間，則 R_V 滿足 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1}) = \text{Diff}(R_V) \cap \mathbf{Z}^{n-1}$ 且 $\text{Diff}(R_V \cap \mathbf{Z}^{n-1}) = \{[x_1, x_2]^T \in \mathbf{Z}^2 \mid l_i - u_i \leq a_{i1}x_1 + a_{i2}x_2 \leq u_i - l_i, i=1, \dots, m\}$ 。

令 $Hd_V \Theta_V = [\theta'_1, \theta'_2]$ ，則根據上述引理、定理 1、及 $\xi = \gcd(\theta'_1, \theta'_2)^{-1}[\theta'_2, -\theta'_1]^T$ 之事實，我們可以得到下列封閉解：

定理 3 在 $n=3, k=1$ 之情況下，若 io_V 為一凸多邊形 I/O 空間，則變數 V 之資料元素不發生通道相衝，若且唯若

$$\max\left\{\left|\frac{a_{i1}\theta'_2 - a_{i2}\theta'_1}{u_i - l_i + 1}\right| \mid i=1, \dots, m\right\} \leq \frac{\gcd(\theta'_1, \theta'_2)}{z_{\min}},$$

其中， z_{\min} 同定理 1 內之定義。

比較定理 2 與定理 3，不難發現定理 2 在 $n=3$ 之情況下為定理 3 之一特例。

3.4 複雜度分析

現在，我們描述定理 2 與定理 3 內各封閉解之複雜度。由於在一有限整數序列中取最大數只需線性時間之事實，定理 2

與定理 3 內各封閉解之複雜度分別為 $O(n+T(z_{\min}))$ 及 $O(m+T(z_{\min}))$ ，其中 $T(z_{\min})$ 為 z_{\min} 之計算時間。由於篇幅限制，我們將忽略 $T(z_{\min})$ 之分析。不過，藉由觀察 z_{\min} 之定義，可容易得知 $T(z_{\min})$ 為一與 N 無關之數，這說明了定理 2 與定理 3 內封閉解之複雜度也與 N 無關。

3.5 實例應用

在本小節中，我們舉出兩實例以說明新封閉解之應用。為方便，我們將分別以 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ 代表 \mathbf{R}^3 之標準基底 $[1,0,0]^T, [0,1,0]^T, [0,0,1]^T$ 。

例 1 考慮[15, Example 2]中計算 $C_{N \times N} = A_{N \times N} B_{N \times N}$ 之方陣相乘演算法。當 $N=4, \mathbf{l}_C = [2], H = [2,1,2], S = [1,1,-2]$ 時，Xue 曾說明通道 \mathbf{l}_C 上沒有發生通道相衝，但其提出之通道相衝檢查條件無法檢查出此事實[15, pp. 10–11]。現今，由於 io_C 為一如下之平行多面體 I/O 空間[15, p. 13]：

$$io_C = \{M_C \mathbf{x} + \mathbf{q}_C \mid 0 \leq \mathbf{x} < [N, N]^T, \mathbf{x} \in \mathbf{Z}^2\},$$

其中 $M_C = [\mathbf{e}_2, \mathbf{e}_1]$ ， $\mathbf{q}_C = [1,1,0]^T$ ，故 \mathbf{l}_C 無通道相衝之事實可應用定理 2 來檢查。

由於 $\mathbf{d}_C = \mathbf{e}_3$ ，我們可計算出 $\Theta_C = [2, 3]$ ，因此 $\xi = [3, -2]^T$ 。進而可算出 $z_{\min} = 2$ 。現在，由於下列不等式

$$\max\left\{\left|\frac{\xi_1}{N}\right|, \left|\frac{\xi_2}{N}\right|\right\} = \max\left\{\frac{3}{4}, \frac{2}{4}\right\} \geq \frac{1}{z_{\min}} = \frac{1}{2}$$

成立，故由定理 2 可知 \mathbf{l}_C 上沒有發生通道相衝。此例展示了定理 2 能成功地處理 Xue 的通道相衝檢查條件無法處理之情況。□

例 2 考慮[5, pp. 759–760]中計算 $C_{N_1 \times N_3} = A_{N_1 \times N_2} B_{N_2 \times N_3}$ 之帶狀矩陣(band matrix)相乘演算法，其中矩陣 A 與 B 之帶寬(bandwidth)分別以參數 p_1, p_2 與 q_1, q_2 表示。由於此演算法之 I/O 空間皆為六邊形，故過去文獻中的封閉解皆不適用於檢查此演算法在低維陣列上的通道相衝。以下，我們以[5,

Table 2]中之第 16 例展示定理 3 可適用於此帶狀矩陣相乘演算法。

[5, Table 2]中之第 16 例說明當 $N_1 = N_2 = N_3 = 4$, $p_1 = p_2 = 2$, $q_1 = 3$, $q_2 = 2$, $\overset{\leftarrow}{l}_A = \overset{\leftarrow}{l}_B = \overset{\leftarrow}{l}_C = [1]$, $H = [1, 1, 4]$, $S = [1, -1, -1]$ 時, 各資料通道上皆無通道相衝發生。為檢查此事實, 由[5, Table 1]可令 $\overset{\leftarrow}{d}_A = \overset{\leftarrow}{d}_2$ 及

$$io_A = \{M_A[x_1, x_2]^T \mid 1 \leq x_1 \leq N_1, 1 \leq x_2 \leq N_2, \\ -(p_1 - 1) \leq x_1 - x_2 \leq p_2 - 1, x_1, x_2 \in \mathbf{Z}\},$$

其中 $M_A = [\overset{\leftarrow}{e}_1, \overset{\leftarrow}{e}_3]$ 。我們由此可計算 $\theta'_1 = 2$, $\theta'_2 = 3$, $z_{\min} = 1$ 。現在, 檢查下列不等式

$$\max\left\{\frac{|\theta'_2|}{N_1}, \frac{|\theta'_1|}{N_2}, \frac{|\theta'_2 + \theta'_1|}{p_1 + p_2 - 1}\right\} = \\ \max\left\{\frac{3}{4}, \frac{2}{4}, \frac{5}{3}\right\} \geq \frac{\gcd(\theta'_1, \theta'_2)}{z_{\min}} = 1$$

成立, 故由定理 3 可知 $\overset{\leftarrow}{l}_A$ 上沒有發生通道相衝。相似地, 我們可檢查 $\overset{\leftarrow}{l}_B$ 及 $\overset{\leftarrow}{l}_C$ 上也沒有發生通道相衝。□

四、計劃成果自評

我們已達到本計劃之預期目標。我們已將過去文獻中的封閉解推廣至允許非鄰近資料通道的低維陣列上。我們也針對映射 3 維規律演算法至線性陣列的情形, 推導出一個適於任意凸多邊形 I/O 空間之新封閉解。我們已分析了新封閉解之複雜度並將其應用到各種實例上; 結果顯示出新封閉解滿足許多實際應用上的需求。

五、參考文獻

- [1] K. N. Ganapathy. *Mapping regular recursive algorithms to fine-grained processor arrays*. Ph.D. dissertation, University of Illinois, Urbana-Champaign, 1994.
- [2] K. N. Ganapathy and B. W. Wah. Optimal synthesis of algorithm-specific lower-dimensional processor arrays. *IEEE Trans. Parallel and Distri. Sys.*, 7(3):274–287, 1996.
- [3] W. L. Huang. *A virtual node approach to*

checking link conflicts in the mapping of dependence graphs into processor arrays. Master thesis, National Chiao Tung University, Taiwan, R.O.C., 2001.

- [4] J. Y. Ke and J. C. Tsay. An approach to checking link conflicts in the mapping of uniform dependence algorithms into lower dimensional processor arrays. *IEEE Trans. Comput.*, 48(7):732–737, July 1999.
- [5] J. Y. Ke and J. C. Tsay. Finding space-optimal linear array for uniform dependence algorithms with arbitrary convex index sets. *Journal of Information Science and Engineering*, 14:743–763, 1998.
- [6] H. T. Kung and C. E. Leiserson. Systolic arrays (for VLSI). In I. S. Duff and G. W. Stewart, editors, *Sparse Matrix Proceedings*, pages 256–282. SIAM, 1978.
- [7] H. T. Kung. Why systolic architectures? *Computer*, 15(1):37–46, January 1982.
- [8] S. Y. Kung. *VLSI Array Processor*. Prentice-Hall Int., Englewood Cliffs, NJ, 1988.
- [9] P. Z. Lee and Z. M. Kedem. Synthesizing linear array algorithms from nested for loop algorithms. *IEEE Trans. Comput.*, C-37(12):1578–1598, December 1988.
- [10] P. Z. Lee and Z. M. Kedem. Mapping nested loop algorithms into multidimensional systolic arrays. *IEEE Trans. Parallel and Distri. Sys.*, 1(1):64–76, January 1990.
- [11] I. V. Ramakrishnan, D. S. Fussell, and A. Silberschatz. Mapping homogeneous graphs on linear arrays. *IEEE Trans. Comput.*, C-35(3): 189–209, March 1986.
- [12] J. C. Tsay et al. On the data collision problem of the processor array with non-neighboring connections. Technical Report NSC89-2213-E009-139, Dep. Engineering & Applied Science, NSC, 2001.
- [13] J. Xue and P. Lenders. Avoiding data link and computational conflicts in mapping nested loop algorithms into lower-dimensional processor arrays. In

- Proc. Int. Conf. on Parallel and Distributed Systems*, pages 567–572, 1994.
- [14] J. Xue. A new formulation of mapping conditions for the synthesis of linear systolic arrays. In *Proc. Int. Conf. on Application Specific Array Processors*, pages 297–308, 1993.
- [15] J. Xue. A unified approach to checking data link and computational conflicts in the design of algorithm-specific processor arrays. Technical Report 94-100, Dep. Mathematics, Statistics and Computing Science, The University of New England, Australia, 1994.
- [16] J. Xue. Closed-form mapping conditions for the synthesis of linear processor arrays. *Journal of VLSI Signal Processing*, (10):181–199, 1995.