第18卷 第8期
1991年8月

中 国 激 光
CHINESE JOURNAL OF LASERS

Vol.18No. 8
August, 1991

光学并行矩阵乘法器的实验研究*

钱秋明 李庆熊 赵建明 王之江
(中国科学院上海光机所,201800)

提要:本文介绍全并行矩阵乘法的实验研究过程和结果,运算充分利用光的并行性,做一次三个矩阵乘法或二维变换只要一个系统时钟周期。文中讨论了误差的来源和误差的消除方法,给出了两个和三个(正实数)矩阵乘法的实验结果(模拟运算),两个矩阵相乘的精度为1.2%,三个矩阵相乘的精度为1.23%。

关键词:光计算技术,信息处理,图像变换

Experimental research of optical parallel multiple matrice multiplier

Qian Qiuming, Li Qingxiong, Zhao Jianming, Wang Zhijiang
(Shanghai Institute of Optics & Fine Mechanics, Academia Sinica, Shanghai)

Abstract: Expeirmental procedure of two or three matrice multiplication is introduced. The optical system make full use of optical parallelism. Matrix B is represented by a light source arrary, A and C are inputted by two SLM which can be replaced by two masks in this preliminary experiments. The mean error of $D=ABC$ in our experimental system is 1.23%. When B is a unity, the results of twoimatrice multiplication have been obtained with a mean error of 1.2%.

Key words: optical computation, information processing, image processing

目前计算机中,矩阵运算都是用串行方法完成的,因此矩阵运算中固有的并行性就完全丢失了[1,2]。光学计算具有极强的并行处理能力,因此人们一直寻求用光学方法实现全并行的矩阵运算。到目前为止,已有许多种光学矩阵乘法器件,这已在文献[3]中简要介绍过。文献[3]中提出用极少数成像系统完成全并行矩阵乘法,它基于光学全并行多矢量对外积运算器和矩阵乘法的相应分解实现了实数模拟矩阵乘法。本文将介绍这种全并行矩阵乘法的实验研究过程和结果,运算充分利用光的并行性,做一次三个矩阵乘法或二维变换只要一个系统时钟周期。

全并行矩阵乘法器原理概述

图1描述了文献[3]中基于矢量外积的全并行光学矩阵乘法系统。该系统采用$n\times n$个点光源放置于球透镜系统的焦面上作为照明光,同时它代表矩阵B,图中SL和CL分别为球透

收稿日期:1989年7月19日。
* 本工作得到国家自然科学基金和上海分院青年基金的资助。

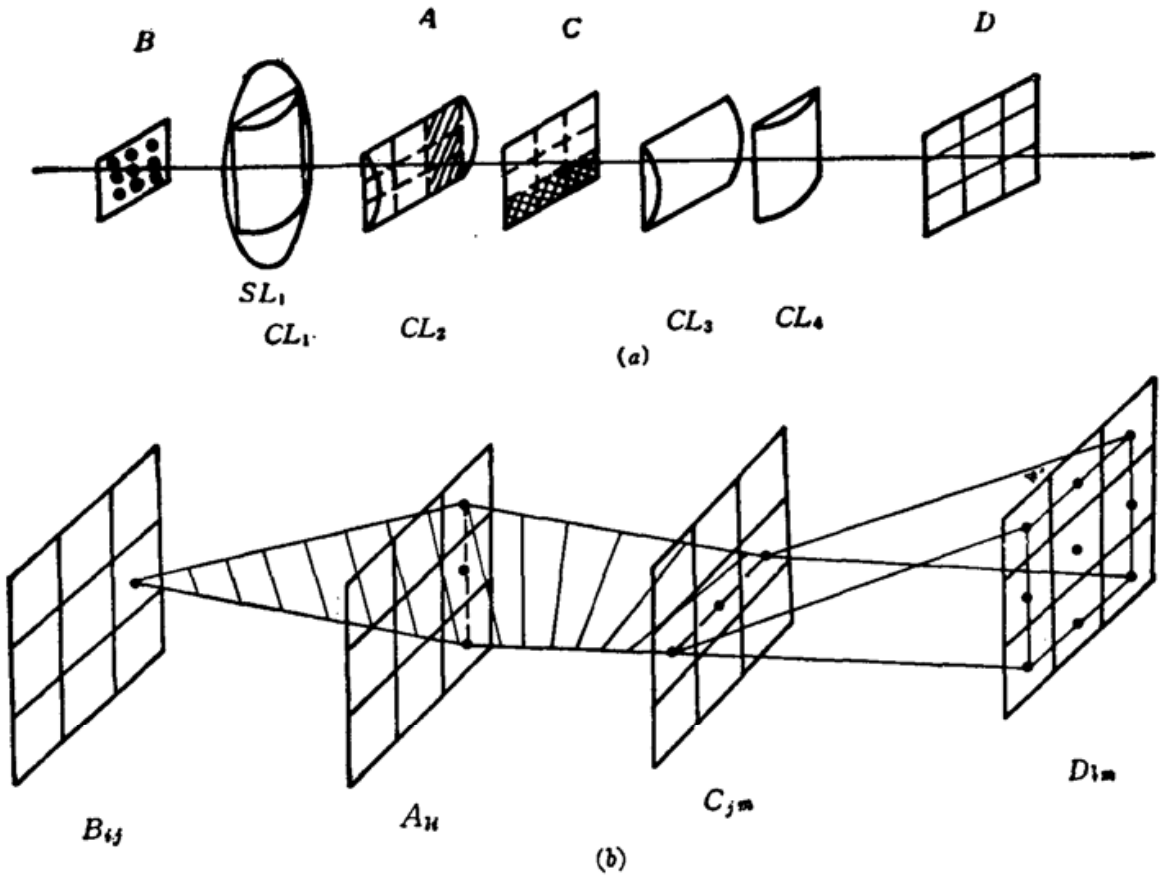

Fig. 1

(a) Optical Scheme of the three matrice multiplication system;
(b) Paths of the light rays

镜和柱透镜，它们的焦距有以下关系：

$$f_{SL1}=f_{CL1}=f_{CL2}=f_{CL4}=f, f_{CL3}=3f/4。$$

图中用于输入另外两个矩阵的空间光调制器用 A、C 表示，在初步实验中把它们用输入编码板代替（在二维变换中这两块编码板代表变换核），它们各单元的光强透过率与矩阵元素值成正比。因此这里必须首先假定所输入的矩阵每个元素都为正实数，对于非正实数的各种情形，可以采用各种算法，即使是复数也同样可以对输入矩阵做适当编码后变成正实数矩阵再输入[4]。在图 1 所示的系统中，光学乘法与加法的完成与内积做矩阵乘法时完全一样。图 1(a)给出此光学系统的原理图，图 1(b)给出了抽象的光线光路图，很显然，它可理解为正交成像的非球面光学系统。

假如光源为理想点源，而且所有透镜为理想成像透镜（不考虑任何像差，柱透镜只在一个方向成像），则矩阵 A 的编码板和 C 的编码板将按图 1(a)中的方式放置，如果点光源的强度分别为 B_{1m}，A、C 编码板的透过率为 (A_{ij})、(C_{ij})，则在接收器平面上得到以下强度分布：

$$D'_{ij}=G\Sigma A_{il}B_{1m}C_{mj},\ D_{ij}=D'_{ij}/G=\Sigma A_{il}B_{1m}C_{mj}\cdots\cdots \quad (1)$$

这就是矩阵 A、B、C 的乘法。当然透过率总是小于 1，实际上总是需要乘上一定的比例因子。

如果要进行两个矩阵相乘，则只要把上面的 B 矩阵改为单位矩阵即可，即把光源改为 n 个点源。

全并行光学矩阵乘法实验系统分析

前面所述的矩阵乘法器的主要误差来源为：(1)编码板各单元很小时产生衍射引起的误差；(2)各点光源所发的光角分布不均匀引起的误差；(3)背景光引起的误差；(4)编码板畸变引起的误差；(5)实际光学系统相对理想系统的偏差引起的误差。所有误差中编码板各单元尺寸较大时，影响较大的为(2)、(3)、(4)。第五种误差易减到所要求的精度。

1．编码板和光学元件尺寸及相互之间关系

假定各光源中心相距 s_x、s_y，A 和 B 编码板各单元中心相距为 d_{1x}、d_{1y}、d_{2x}、d_{2y}，接收器 D 的各单元为 $h_x \times h_y$ 的小方块。光学系统为理想系统，则根据波动光学理论，可得极限分辨率为

$$h_{x1} \approx 1.22\lambda / NA_x$$

$$h_{y1} \approx 1.22\lambda / NA_y$$

令

$$h_{x1} = h_{y1} = h_1$$

则可取

$$h_x = h_y = h > 2h_1$$

显然

$$S_x = S_y = d_{1x} = d_{2x} = d_{1y} = d_{2y} = h$$

系统的通光孔径应由编码板 A 和 B 的外框决定，因此各镜片的最小通光孔径都有一定的要求。

为了使接收器平面上结果图像各相邻单元之间无重迭区，则必需使编码板相邻单元之间有一定的不透光区，这个区域的宽度由系统各部分的分辨率决定。同样当各点光源并非理想的点时，则各光源的大小和相邻光源之间的间隙取决于系统的分辨率和编码板各单元的大小。由于用波动光学对这些数据只能作一近似估计(常常与实际情况偏离较大)，在此就不作仔细讨论，在实验中可以取两个相邻单元之间的空间间隙为每个单元的宽度的 R 倍(在接近系统分辨率使用极限时取 $R>1$，在远离系统分辨率极限时取 $R \ll 1$)。

2．各个点光源角分布不均匀引起的误差及其消除方法

光源角分布不均匀引起的误差是很显然的，下面考虑高斯分布的情况，它可以表示为

$$I(\theta) = I_0 e^{-2\theta^2}$$

当某个点光源发出的光束经过矩阵乘法系统到达 A 编码时，它可以表示为以下形式：

$$I_{AY} \sim e^{-\alpha y^2/f^2}$$

$$y = f \operatorname{tg}\theta \approx f\theta$$

透过 A 板后成为

$$I_{AY} \sim e^{-\alpha y^2/f^2} f_A(y_i)$$

到达 B 矩阵板时可以表示为

$$I_{BX} \sim I_{AY} e^{-\alpha x^2/f^2}$$

透过 B 板后成为：$I_{BX} \sim I_{AY} e^{-\alpha x^2/f^2} f_B(X_i)$，在接收器平面上有

$$I_{xiyi} = G e^{-\alpha(x_i^2+y_j^2)/f^2} f_A(y_j) f_B(x_i)$$

其中 G 与编码板的光阑有关。

为了使上述分布满足所要求的精度 δ，则

$$f/D = [18a/\ln(\delta+1)]^{1/2}$$

a 是确定高斯分布形状的参数；δ 是所要求的精度；D 为编码板的最大宽度，即 $D=\max(D_x, D_y)$，对实际情况，则可以根据测量结果选用不同的光源，当然这要在一定条件下才是可行的。

3．误差的消除方法

对于空间分布的相乘型误差，可以用很简单的方法加以消除，这只要在 A、C 两矩阵板上放上 $A_{ij}=C_{ij}=1$ 的编码板，取 $S_{ij}=1$，用接收器测出结果 D'_{ij}，求出 $\min(D'_{ij})/D'_{ij}=D_{ij}$，并用一透过率为 D_{ij} 的编码板放在接收器前，紧贴接收器，就可以消除相乘型误差。

类似地相加型误差可以通过接收器电子线路消除。用这样的方法可以使误差减到很小（1% 以下），因为光学系统中的主要误差都能用这两种方法消除。

系统调整、实验过程与实验结果

实验中用图 1 中的系统分别进行两个矩阵 $A_{10\times5}B_{5\times10}$ 相乘和三个矩阵 $U_{10\times5}$，$V_{5\times5}$，$W_{5\times10}$ 相乘。

1．系统调整

首先调整各个透镜和编码板的中心，使它们在一直线上，同时调整各透镜使它们与中心线垂直，然后调整两组柱透镜使它们相互正交。根据各光学元件、编码板的尺寸和精度要求可以得出以下参数：

$$\text{中心偏离}<(d-D)/4;$$

$$\text{垂直偏离}<h/D/2$$

其中 d 为光学元件的实际口径；$D=\max(D_x, D_y)$；h 为编码板单元的最小宽度。

2．实验过程

在实验中 $f=10\,\text{mm}$，编码板各单元的大小约为 $0.6\times0.6\,\text{mm}^2$，间隙为 0.4 mm 或 0.3 mm 宽，各柱透镜都固定在二维调整架上，编码板夹在三维调整架上，实验中首先让 B_{ij} 中的 5 个点光源照明，放入 $A_{ij}=1$，$C_{ij}=1$ 的编码板并调整，使接收器平面上接收到尽可能均匀的结果，然后放入实际的编码板进行测试，测量中显示器的读数误差为 1.0%。

3．实验结果

(1) 两个矩阵相乘：

$$A=\begin{bmatrix}1&1&0&0&1\\1&1&1&0&0\\0&1&1&1&0\\0&0&1&1&1\\1&0&0&1&1\\1&1&0&0&1\\1&1&1&0&0\\0&1&1&1&0\\0&0&1&1&1\\1&0&0&1&1\end{bmatrix}\qquad B=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$$

$$C=\begin{bmatrix}1&1&0&0&1&1&1&0&0&1\\1&1&1&0&0&1&1&1&0&0\\0&1&1&1&0&0&1&1&1&0\\0&0&1&1&1&0&0&1&1&1\\1&0&0&1&1&1&0&0&1&1\end{bmatrix}$$

$$ABC=\begin{bmatrix}3&2&1&1&2&3&2&1&1&2\\2&3&2&1&1&2&3&2&1&1\\1&2&3&2&1&1&2&3&2&1\\1&1&2&3&2&1&1&2&3&2\\2&1&1&2&3&2&1&1&2&3\\3&2&1&1&2&3&2&1&1&2\\2&3&2&1&1&2&3&2&1&1\\1&2&3&2&1&1&2&3&2&1\\1&1&2&3&2&1&1&2&3&2\\2&1&1&2&3&2&1&1&2&3\end{bmatrix}$$

(2) 三个矩阵相乘:

$$\begin{bmatrix}1&1&0&0&0\\1&1&1&0&0\\0&1&1&1&0\\0&0&1&1&1\\1&0&0&1&1\\1&1&0&0&1\\1&1&1&0&0\\0&1&1&1&0\\0&0&1&1&1\\1&0&0&1&1\end{bmatrix}\quad\begin{bmatrix}1&0&0&0&1\\1&1&0&0&0\\0&1&1&0&0\\0&0&1&1&0\\0&0&0&1&1\end{bmatrix}\quad\begin{bmatrix}1&0&0&0&1&1&0&0&0&1\\1&1&0&0&0&1&1&0&0&0\\0&1&1&0&0&0&1&1&0&0\\0&0&1&1&0&0&0&1&1&0\\0&0&0&1&1&0&0&0&1&1\end{bmatrix}$$

A B C

$$ABC=\begin{bmatrix}3&1&0&1&3&3&1&0&1&3\\4&3&1&1&3&4&3&1&1&3\\3&4&3&1&1&3&4&3&1&1\\1&3&4&3&1&1&3&4&3&1\\1&1&3&4&3&1&1&3&4&3\\3&1&1&3&4&3&1&1&3&4\\4&3&1&1&3&4&3&1&1&3\\3&4&3&1&1&3&4&3&1&1\\1&3&4&3&1&1&3&4&3&1\\1&1&3&4&3&1&1&3&4&3\end{bmatrix}$$

图 2 中给出了两个矩阵相乘的实验结果,给出的照片经过了放大(实验中加入了前面所述的误差校正板)。为了方便起见,实验中用各单元透过率为 1.0 或 0.0 的透射板,对图 2 的情况进行光电测量的结果分别如表 1 所示。

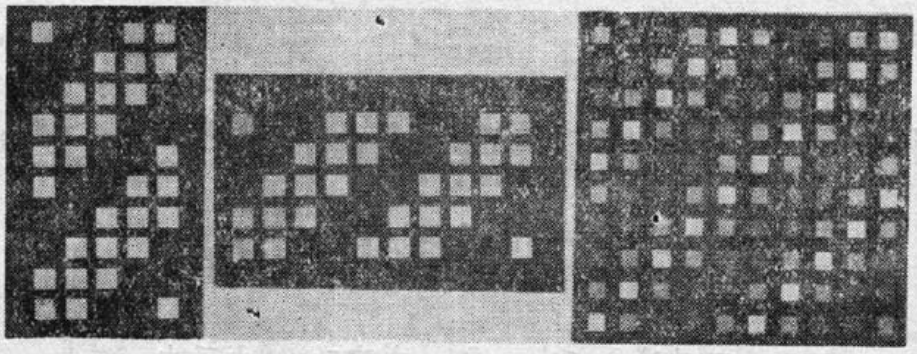

Fig. 2 Photographes of two matrice multiplication where A and C are masks and D is the results

Table. 1 Results of two matrice multiplication (detected by opto-electrical detector)

3.01	2.02	0.98	1.01	1.98	3.03	2.04	0.99	1.03	2.02
1.97	2.98	1.97	1.02	0.99	1.99	2.99	2.01	1.02	0.98
0.99	2.02	3.03	2.01	1.01	0.98	2.03	3.04	2.03	0.99
1.02	1.01	2.04	3.05	1.97	0.99	1.01	2.03	3.01	1.97
1.98	0.98	1.01	2.03	2.97	1.99	0.99	1.01	1.99	3.02
3.06	2.01	0.99	0.99	2.01	2.97	1.96	0.98	1.01	1.98
2.02	3.01	2.03	1.02	1.01	1.99	2.96	2.02	0.98	0.99
0.99	1.97	2.95	1.97	0.98	0.99	2.03	3.05	2.03	1.01
1.00	0.99	2.01	2.95	2.04	1.01	1.02	2.03	3.03	2.04
1.99	0.97	1.01	2.01	3.02	2.03	1.01	1.00	2.01	3.04

$\text{Error}_{max}=2\%$ $\text{Error}_{are}=1.2\%$

参 考 文 献

1 王之江, 中国科学院院刊, **2**(3), 198(1987)

2 *Proc. IEEE*, **72**(7), 1(1984)

3 钱秋明, 李庆熊, *et al.*, 中国激光, **18**. (7),540(1991)

4 H. J. Butterweck, Principles of Optical Data-Processing, Progress in Optics, Edited by E. Wolf, **24**, p213

(上接第 605 页)

有差别。这就是为什么在 r 很小时观测不到双稳态的原因。上述实验结果表明, I_1 在 I_p 中占的比例越大, 双稳态越明显, 具体表现就是滞后回线的宽度及上下支的差别越大。

如果先让 I_p 从大到小, 再从小到大变化, 则观测不到双稳态。这一点不难从上面的分析中加以理解。这也充分说明了双稳态起因于 $LiNbO_3$ 的记忆特性。$LiNbO_3$ 中的光栅可通过向其照射一束均匀的强光或给其加热来擦除。这种 PCR 的双稳输出特性, 使其可作为记忆元件来使用。当想回避这一特性时, 可通过适当选择 PCM 的前后向泵浦光之比来实现。

参 考 文 献

1 E. M. Wright *et al.*, *Opt. Commun.*, **51**(6), 428(1984)

2 G. C. Valley *et al.*, *Opt. Lett.*, **9**(11), 513(1984)

3 H. Rajbenbach *et al.*, *Opt. Lett.*, **10**(3), 137(1985)

4 H. Rajbanbach *et al.*, *Opt. Lett.*, **14**(1), 78(1989)