第34卷 第17期 计 算 机 工 程 2008年9月
Vol.34 No.17 Computer Engineering September 2008
·博士论文· 文章编号：1000—3428(2008)17—0004—03 文献标识码：A 中图分类号：TP391

区域分解对气象模式并行计算速度的影响

臧增亮[1,2]，饶宣锐[3]，潘晓滨[1]，张理论[4]，王春明[1]，何宏让[1]
(1. 解放军理工大学气象学院，南京 211101；2. 中国科学院大气物理研究所 LASG 实验室，北京 100029；
3. 二炮装备研究院，北京 100085；4. 国防科技大学计算机学院，长沙 410073)

摘 要：通过数值试验分析了区域分解策略对 ARPS 气象模式并行计算速度的影响，发现无论是否使用编译优化技术，均以分解后数据区域近似为正方形时具有最大的加速比和并行效率。在二级编译优化的情况下，并行速度还和分解方向有关，在 *y* 方向上的分解比在 *x* 方向上的分解更有利于提高并行效率，而在无优化情况下，并行速度和分解方向几乎无关。并从通信量和编译优化的角度对试验结果进行了讨论和分析。

关键词：区域分解；数值模式；并行计算

Influence of Domain Decomposition on Parallel Computation Speed of Meteorological Model

ZANG Zeng-liang[1,2], RAO Xuan-rui[3], PAN Xiao-bin[1], ZHANG Li-lun[4], WANG Chun-ming[1], HE Hong-rang[1]
(1. Meteorological College, PLA University of Science and Technology, Nanjing 211101; 2. LASG Lab, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029; 3. Equipment Research Institute of PLA's Second Artillery, Beijing 100085; 4. School of Computer Science, National University of Defense Technology, Changsha 410073)

【Abstract】The influence of domain decomposition on the parallel computation speed of ARPS model is analyzed on the base of some experiments. Results show that the subdomains similar to square possess the biggest acceleration rate and efficiency of parallel whether the optimization lever -o2 is applied to or not. Under the circumstances of optimization lever -o2, parallel speed also relates to decomposition direction. The decomposition along *y* direction is preferable to that along *x* direction for improving in parallel efficiency. However, under the circumstances without optimization, parallel speed is almost independent of decomposition direction. Furthermore, experimental results are discussed in the paper from the angles of communications traffic and optimization.

【Key words】 domain decomposition; numerical model; parallel computation

1 概述

长期以来，计算机的计算能力一直是制约数值天气预报发展的瓶颈。为了提高预报的准确率和尽可能地延长预报时效，越来越多的高性能计算机被应用到气象领域，数值模式的并行化也成为模式设计中不可少的内容之一[1-3]。对于格点大气模式而言，使用 MPI 方式对其并行化实际上就是将预报区域进行分解，同时对各子区域进行计算，通过消息传递交换边界，然后合并成一个整体。所以，分解策略是 MPI 并行必须面临的一个重要问题。

ARPS(The Advanced Regional Prediction System)是目前中尺度天气研究和预报领域中广泛应用的数值模式之一，该模式由美国俄克拉荷马大学风暴分析预报中心研制，以三维完全可压缩大气为研究对象，适用于研究尺度为几米到几百公里尺度的中尺度天气系统。模式的控制方程建立在三维非静力 Navier-Stokes 方程组基础上，包括：状态方程，动量方程，气压方程，位温方程，以及水汽、云水、雨水等水物质方程。空间微商在 Arakawa C 网格下，采用二阶或四阶有限差分格式。时间积分采用分离时间积分方案，小时间步长用来积分和声波相关的项，大时间步长用来积分其余项。本文以该模式为例，研究区域分解对气象数值模式并行计算速度的影响。

2 区域分解对并行计算速度影响的敏感性试验

2.1 试验方案

设模式水平维数为 N_x，N_y，网格距为 dx，dy，由于最外一层网格作为边界，因此内部的物理区域为 $(N_x-3)\times dx$，$(N_y-3)\times dy$。用 p 个处理机进行分解时，则要求 x 或 y 方向的处理机个数(记为 p_x，p_y)能被 (N_x-3) 和 (N_y-3) 整除。为方便分解，本文中用的 N_x，N_y 均为 195，则 (N_x-3) 和 (N_y-3) 可以被 2, 3, 4, 6 等整除。

试验在某国产高性能计算机上进行，并行环境为 MPICH(2.0 版)，采用的编译器为 Inter(8.1 版)。试验分为 2 组，第 1 组使用二级优化编绎选项(-O2)，分别在 1 个、4 个、8 个、16 个、32 个、48 个、64 个、96 个处理机上运行，并考虑不同的区域分解(见表 1)，如对于 4 个处理机(p=4)，p_x 和 p_y 可以有 4×1, 2×2, 1×4 等几种选择。第 2 组试验不

基金项目：国家自然科学基金资助项目(40505023, 40705020)；中国科学院大气物理研究所 LASG 实验室开放课题基金资助项目
作者简介：臧增亮(1977－)，男，博士，主研方向：气象数值模式；饶宣锐，高级工程师；潘晓滨，教授；张理论，副研究员；王春明、何宏让，副教授
收稿日期：2007-10-30 **E-mail：**zzlqxxy@163.com

使用任何优化编绎选项(0 级优化)，仅在 1 个处理机和 64 个处理机上试验，64 个处理机的区域分解包括 64×1, 32×2, 8×8, 2×32, 1×64 等(见表 2)。由于机时的限制，第 1 组试验模式积分为 1 h，第 2 组试验模式积分为 10 min。第 3 组对计算平流项的子程序进行优化试验，用 2 组数据分别在单一处理机上进行 0 级优化和 2 级优化，2 组数据的大小分别定义为 $A(20, 4)$和 $B(4, 20)$，由于此子程序比较简单，耗费机时很短，为方便比较，在这一组试验中对此子程序循环调用 10^6 次。

表 1　第 1 组试验参数和结果

p	p_x	p_y	T_p /s	$T_{p\min}$ /s	$T_{p\max}$ /s
1	1	1	49.61	49.61	49.61
4	4	1	19.34	14.47	19.34
	2	2	14.47		
	1	4	18.17		
8	8	1	10.64	8.50	10.64
	4	2	9.28		
	2	4	8.50		
	1	8	8.80		
16	16	1	9.34	4.18	9.34
	8	2	5.25		
	4	4	4.42		
	2	8	4.18		
	1	16	4.72		
32	32	1	7.02	2.37	7.02
	8	4	2.67		
	4	8	2.37		
	1	32	3.43		
48	48	1	6.68	1.88	6.68
	16	3	3.66		
	8	6	1.97		
	6	8	1.88		
	3	16	1.89		
	1	48	2.68		
64	64	1	6.32	1.56	6.32
	32	2	3.31		
	8	8	1.56		
	2	32	1.89		
	1	64	2.33		
96	96	1	5.26	1.16	5.26
	48	2	3.23		
	12	8	1.45		
	8	12	1.16		
	2	48	1.26		
	1	96	1.98		

表 2　第 2 组试验参数和结果

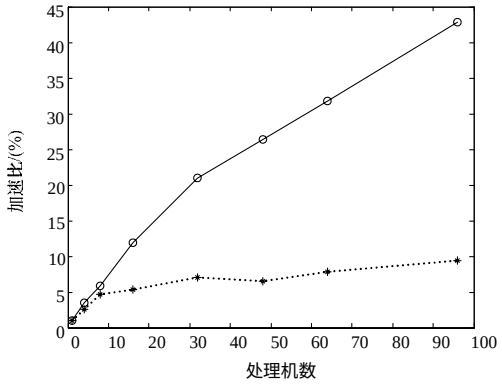
p	p_x	p_y	$T_p/(\times 10^3\cdot s^{-1})$
1	1	1	297.88
64	64	1	8.01
	32	2	6.63
	16	4	6.06
	8	8	5.97
	4	16	6.03
	2	32	6.61
	1	64	7.98

2.2　试验结果

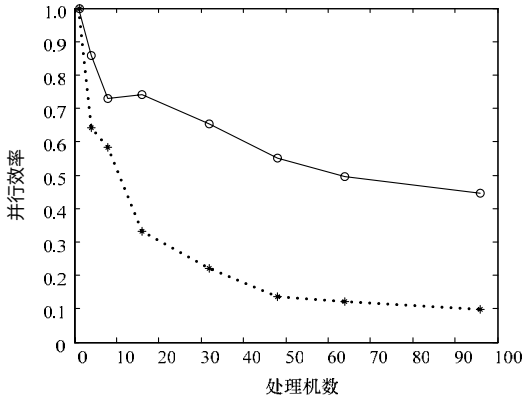
表 1 给出了第 1 组试验的墙钟时间(T_p)，由表可见，区域分解对计算速度有很大影响，例如对于 64 个处理机，使用 64×1，8×8 两种分解方式，墙钟时间分别为 6.32 s 和 1.56 s，两者相差 4.05 倍，表 1 中还给出了每一种分解方式的最快墙钟时间($T_{p\min}$)和最慢墙钟时间($T_{p\max}$)。仔细分析计算速度和区域分解的关系可以发现，速度最慢(墙钟时间最长)的分解总是对应于 p_y 为 1，即相当于仅在 x 方向进行一维分解；而速度最快(墙钟时间最短)的分解总是对应于 p_y 和 p_x 的近似相等，即正方形分解。

定义加速比和并行效率[2]分别为：$S_p = T_1/T_p$，$E_p=S_p/p$。记 $T_{p\min}$，$T_{p\max}$ 对应的加速比分别为 S_p，$S_p^{'}$，对应的并行效率分别为 E_p，$E_p^{'}$。图 1(a)给出了加速比 S_p 和 $S_p^{'}$ 随 p 的变化曲线，由图可见，在 1~8 个处理机之间，两者相差不大，当使用 8 个以上处理机时，两者的差距逐渐拉大，当 p 大于 8 时，S_p 中的曲线仍基本呈线性增加；而 $S_p^{'}$ 中曲线 p 大于 8 以后明显变缓。这说明在并行计算中使用的处理机越多，区域分解策略对并行加速比的影响越大，越有必要选择合理的区域分解方式。

图 1(b)中给出了并行效率 E_p 和 $E_p^{'}$ 随 p 的变化曲线，总体来看，随着 p 的增大，E_p 和 $E_p^{'}$ 都是减小的，但 $E_p^{'}$ 减小的要更快一些，在 p 为 16 时，就已经减小到 0.33 了。E_p 虽然在 p 小于 8 以前也是迅速减小，但在 p 为 16 时又略有提高，这表明选择合理的区域分解方式可以实现并行效率的相对超线性增长。



(a)第 1 组试验的 S_p (实线)和 $S_p^{'}$ (虚线)



(b)第 1 组试验的 E_p (实线)和 $E_p^{'}$ (虚线)

图 1　第 1 组实验中加速比和并行效率曲线图

图 2 为第 2 组试验中 S_p 和区域分解的关系，由图可见，8×8 的分解方式具有最大的加速比，这和第 1 组试验中的结果是一致的。值得注意的是，在最大加速比两侧的数值基本呈对称分布，这说明在不优化的情况下，并行计算的速度和分解的方向几乎无关，而在第 1 组使用编译优化时，在 y 方向上分解比在 x 方向上分解更有利于提高并行的计算速度。

平流项的计算模式中耗费机时较大的模块，试验 3 即针对平流项的计算试验，结果表明(表 3)对于具有相同规模的数组 $A(20, 4)$和数组 $B(4, 20)$，其 0 级优化时的计算时间接近相等，但使用 2 级优化时，两者所耗费的机时可以相差 3 倍左右。这进一步说明不同的分解方向可能造成并行效率的巨大

差异。

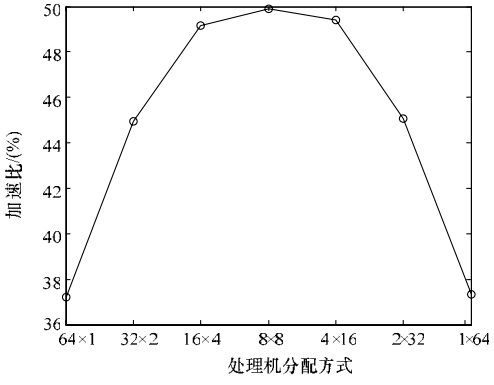


图 2　第 2 组试验 S_p 和区域分解的关系曲线

表 3　第 3 组试验参数和结果

数组类型	-O0/s	-O2/s
A (20,4)	2787.12	16.41
B (4,20)	2796.06	48.67

3　区域分解对并行计算速度影响的因素分析

3.1　区域分解对通信量的影响

很显然，通信量是制约并行计算速度的一个重要因素，而不同的区域分解方式又对通信量有重要影响。当处理机个数较少时，需要进行数据交换的边界少，故通信量也较少，一般不会成为制约并行计算速度的瓶颈，如在表 1 中，对于 4 个和 8 个处理机的情况，各种分解方式的并行计算速度相差都不大。但当处理机个数较多时，由于通信量的增加，通信量逐渐制约并行计算的速度，因此需要选择合理的区域分解方式，以实现最少的通信量。

若模式未分解时的维数固定为 $N_x \times N_y \times N_z$，总的处理机为定值，$p_x$ 和 p_y 可调。容量证明，在近似情况下，最少通信量的分解方式为

$$\frac{N_x}{p_x} = \frac{N_y}{p_y} \tag{1}$$

式(1)表明，从通信量的角度来讲，最优的分解方式是将每个计算子区域分割成正方形。由于本文中试验的 N_x 和 N_y 是相等的，因此最优的区域分解应该是 p_x 和 p_y 相等或近似相等，这和前 2 组试验结果都是吻合的。虽然在第 1 组试验中 16 个处理机时的最小墙钟时间出现在 $p_x = 2$，$p_y = 8$，但它和 $p_x = 4$，$p_y = 4$ 时的墙钟时间接近相同。

3.2　区域分解对编译优化性能的影响

在第 1 组试验中，不同的分解方向对模式运行的速度有很大影响，如 64 个处理机时，64×1 和 1×64 两种分解方式的加速比相差两倍多，而实际上这两者的通信量是相同的，这说明除了通信量以外，还有其他影响并行计算速度的因素。在第 2 组试验中，不同分解方式的加速比却几乎相等。由于第 1 组试验中采用了二级编译优化，而第 2 组试验中没有采用优化，因此可以认为，不同的区域分解方式可以通过编译优化影响并行计算速度。第 3 组试验则进一步证实不同类型数组对优化的敏感程度有明显差异。

编译优化有很多功能选项，比如二级优化中包括循环展开、指令调度、公用子串辨识等几种优化功能。由于机时的限制，没有逐个试验所有优化选项对不同分解计算速度的影响。下面仅对 2 种可能造成影响的优化功能进行讨论。

最典型的一种优化方法是循环展开，即将循环体内的程序展开成多个程序，并增加循环的步长。由于处理机可以同时进行多个浮点运算，将 1 行程序展开成多个程序后，暴露了更多的可同时执行的操作，从而提高计算速度。值得注意的是，循环展开通常都是在最内一层循环中进行，如果源程序最内一层循环的步数原来就比较少，则循环展开的效果将降低。在 ARPS 模式中，绝大部分的内层循环都是针对 *x* 方向的维数 n_x。当处理机总数 *p* 较少时，无论使用何种分解方式，每个处理机所承担子区域的 n_x 仍比较大。例如第一组试验中，在 *p*=4 时，几种分解结果中的 n_x 均大于 50，对循环展开的效果影响不大。但当 n_x 比较小时，如果再对其分解，循环展开的效率将降低，如 *p*=96 时，使用 96×1 分解时，n_x 仅为 5，故其计算速度较慢。第 3 组试验中，由于数组 *B*(4, 20) 在 *x* 方向的维数 n_x 较小，优化后的计算速度比相同规模数组 *A*(20, 4)慢 3 倍左右。

编译优化的另一个主要方面是充分利用 cache 提高计算速度。衡量 cache 效果的一个主要指标是 cache 的命中率，而提高 cache 命中率的主要做法之一就是提高循环体内部数据的局部性[4]。在 Fortran 语言中，数组是按列存放的，ARPS 模式中的内层循环都是 *x* 方向的，在访问过程中都可以做到按列访问，故当 *x* 方向的维数比较大时，访问的命中较高；当 *x* 方向的维数比较小时，虽然也是按列访问，但由于列比较短，在 catch 中要反复访问数组中不同的列，导致命中率降低。所以，从 cache 优化的角度来讲，分解后的子区域 *x* 方向维数较大有利于计算速度的提高。

4　结束语

本文研究了不同的区域分解对 ARPS 模式并行计算速度的影响，通过二级优化和不优化情况下的三组试验发现不同区域分解方式的并行计算速度有明显差别，以分解后的子区域近似于正方形具有最大的加速比和并行效率，并且在 *y* 方向上的分解比在 *x* 方向上的分解更有利于提高并行效率。本文还从通信量和编译优化的角度讨论了区域分解对计算速度影响的机制。

除了以上的机制分析之外，还可能存在其他影响并行计算的因素，如 MPI 消息传递的规模、节点间和节点内数据的通信速度等。实际上，对于 ARPS 这种复杂应用问题，定量完备地刻画并行计算的各类影响因素非常困难，本文以区域分解为着眼点进行研究，旨在为气象模式并行应用中的数据剖分策略提供技术参考。今后将通过更多的敏感性试验对更广泛的影响因素进行分析。

致谢：在模式的并行调试和试验分析过程中得到了国防科技大学宋君强研究员、赵军副研究员、朱小谦博士以及解放军理工大学李毅讲师的帮助，在此表示感谢。

参考文献

[1] 陈起英, 金之雁, 伍湘君, 等. 中期数值预报系统 T213L31 在 IBM/SP 高性能计算机上的建立[J]. 应用气象学报, 2004, 15(5): 523-533.

[2] 袁金南, 王在志, 薛纪善. 广州区域数值预报模式并行化计算[J]. 应用气象学报, 2004, 15(5): 556-561.

[3] 孙安香, 宋君强, 李晓梅. 数值气象预报中的并行计算研究[J]. 高技术通讯, 2001, 11(12): 33-36.

[4] 张理论, 宋君强, 李晓梅. LASG/IAP 全球海洋环流模式的并行计算及其优化[J]. 计算机工程, 2003, 29(14): 15-17.