

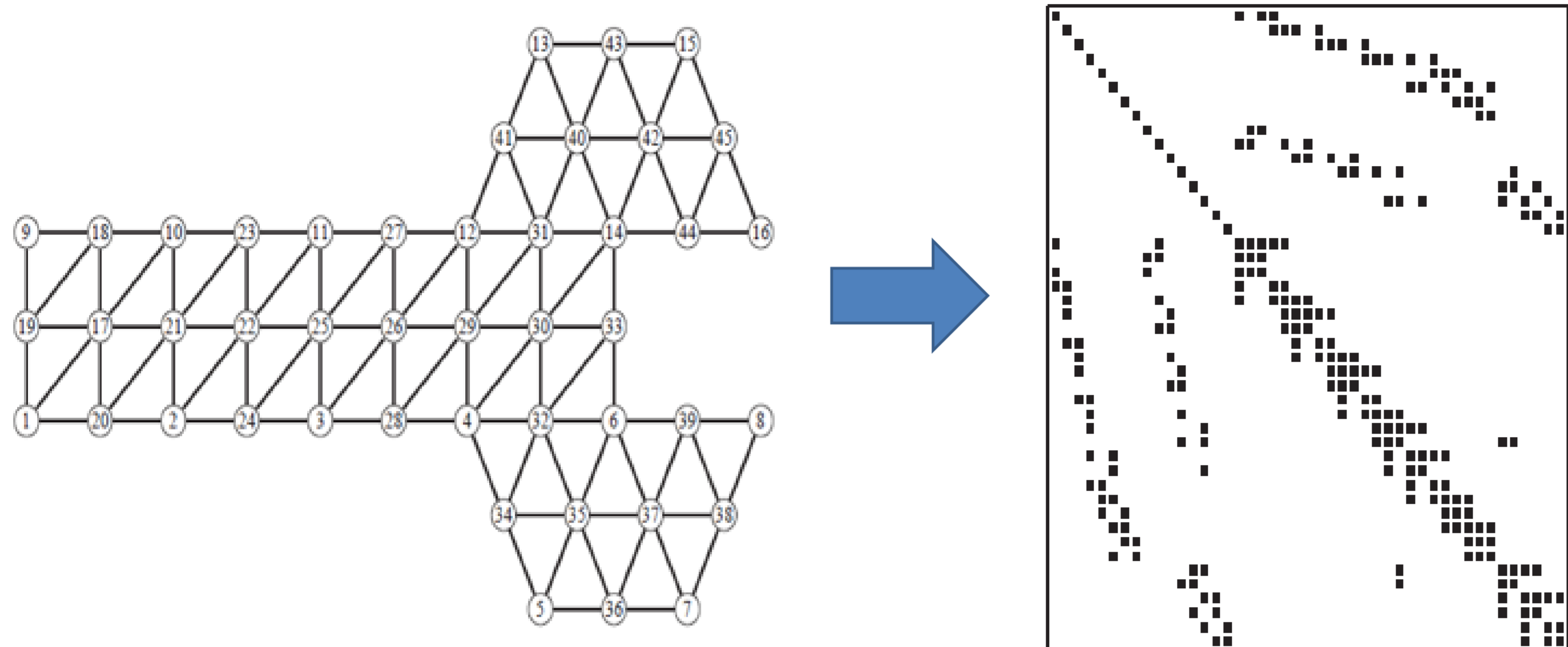
一种适合向量化的稀疏矩阵存储格式

谢佩珍, 刘杰, 徐涵, 龚春叶, 王庆林

(国防科学技术大学并行与分布式计算国家重点实验室, 长沙 410073)

Introduction

稀疏矩阵向量乘 (SpMV) 在很多科学和工程问题中起着十分重要的作用, 广泛地应用于科学领域和实际应用的工程问题中, 如粒子输运模拟、流体力学偏微分方程的求解、图的变换、天体物理等。在当前的计算机平台上, 主流的CSR存储格式对x的访问是不规则的, 不能直接采用向量化技术进行加速, 导致性能较低。目前在许多科学领域和工程中都出现了上千阶乃至几百万阶的大规模稀疏矩阵, SpMV的性能如果能够进一步提升, 整个相关领域问题的运算效率将会大大改善。因此对向量化稀疏矩阵存储格式的研究有着巨大的研究价值与意义。

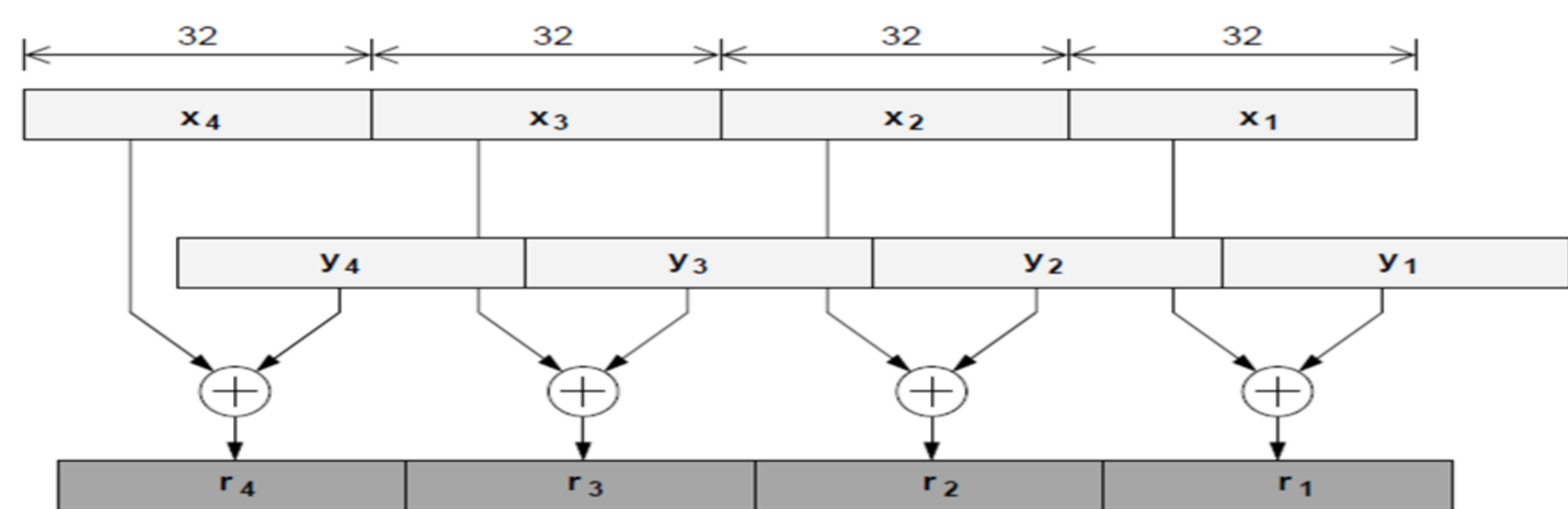


Target Stencils

1. 充分利用向量化加速提升SpMV性能

Intel Ivy bridge提供宽度为4的SIMD计算部件

Intel Phi (MIC) 提供宽度为8的SIMD计算部件

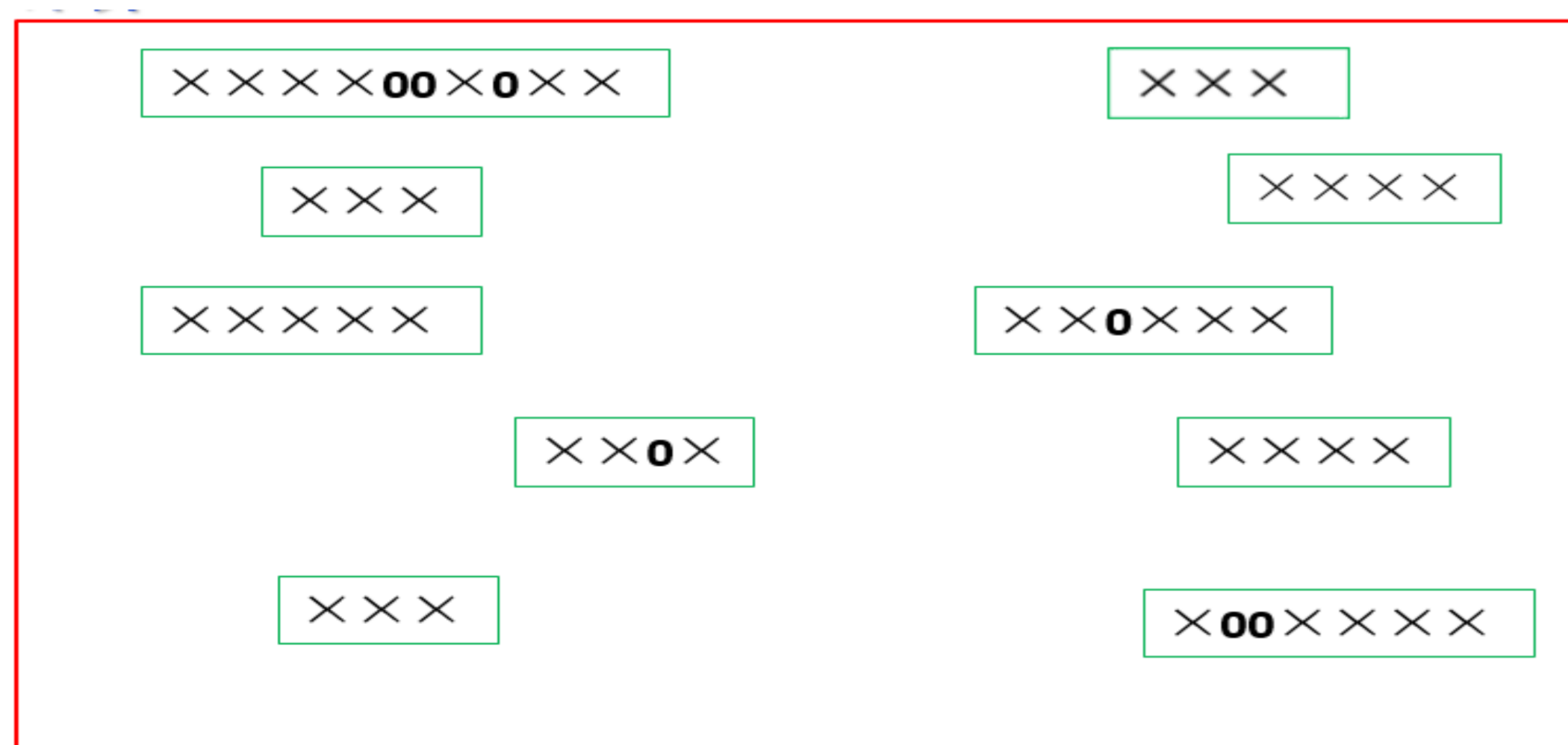


要求SPMV应用程序提供新型的存储格式, 提供向量化的机会

CSR(r)存储格式

1. CSR(r)格式特点

CSR(r)格式充分利用了数据的局部性, 将相邻或相近的非零元作为一个连续的数据片段进行存储。在该数据片段中, 非零元的访问是规则的、连续的, 可以直接使用向量化技术进行读取和计算, 从而大大提高了SPMV的性能。



2. CSR(r)格式算法介绍

(1) 从头到尾按行扫描稀疏矩阵的每一个数据, 从第一个非零元开始写入value数组, 其他非零元依次陆续写入。若两个非零元间隔零元的个数小于或者等于max, 则将两个非零元间的零元作为非零元处理, 和非零元一起写入value数组。value数组中存储的数据若在原稀疏矩阵中连续, 则所有这些连续的数据组成的片段称为r片段。

(2) 依次记录每个数据片段首个非零元的列下标index, 和数据片段包含数据的个数R, 存入col(R)数组中。

(3) 将稀疏矩阵A中每行第一个连续数据片段的索引信息 (每行第一个连续r片段首个非零元在整个Col(R)数组中的序号) 顺序存入数组ptr中。数组长度为m+1。

举例矩阵 A:

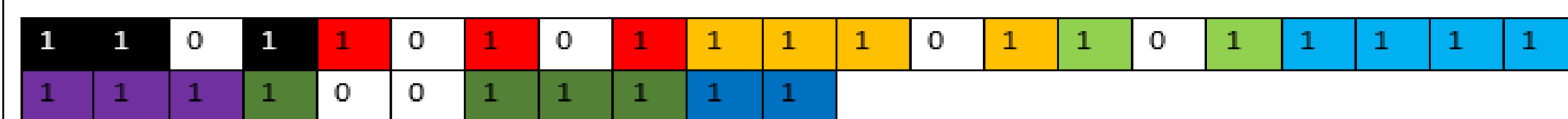
$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0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

A 矩阵图形表示:



CSR(r)数组图形化表示:

Value



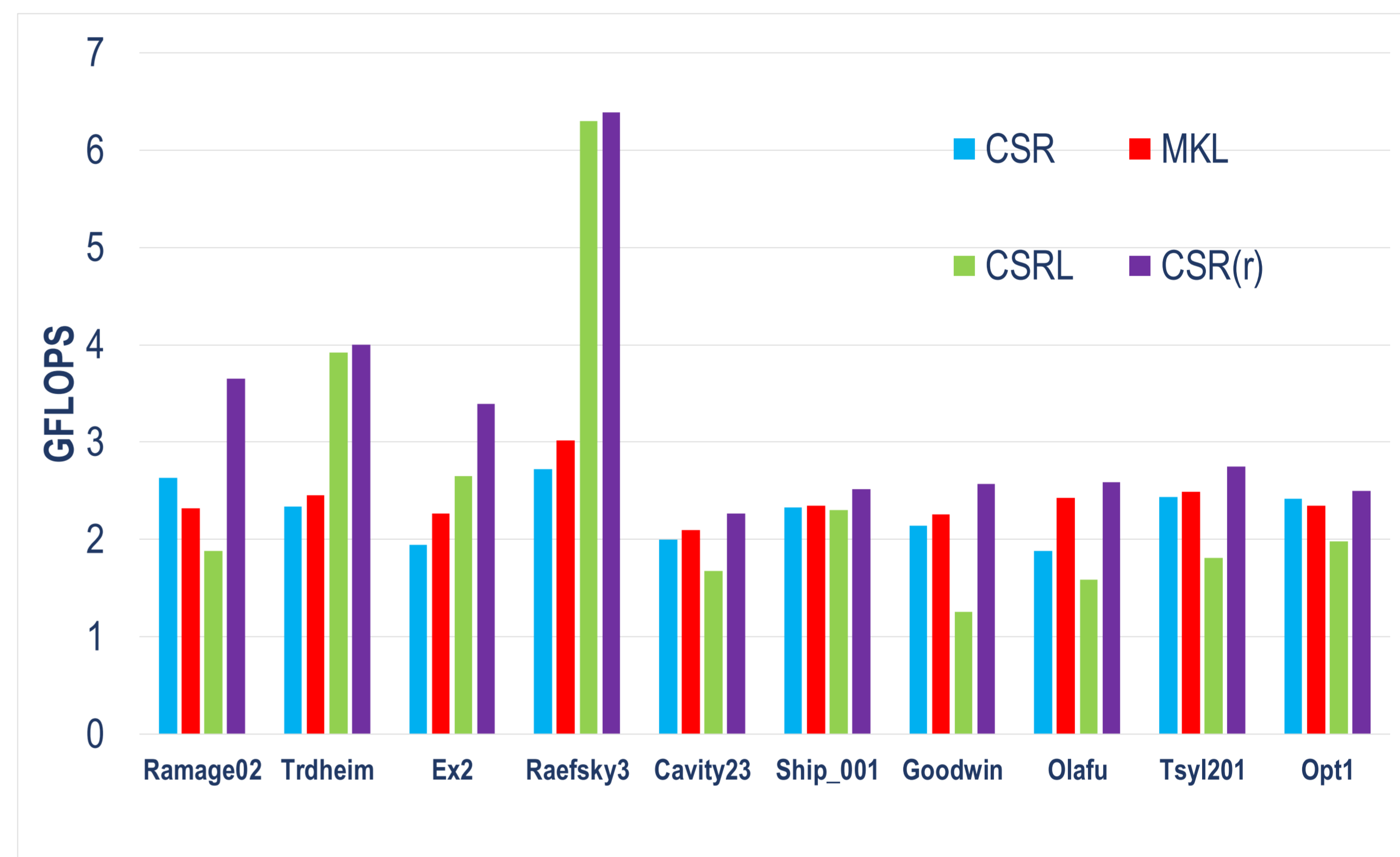
Col(R)



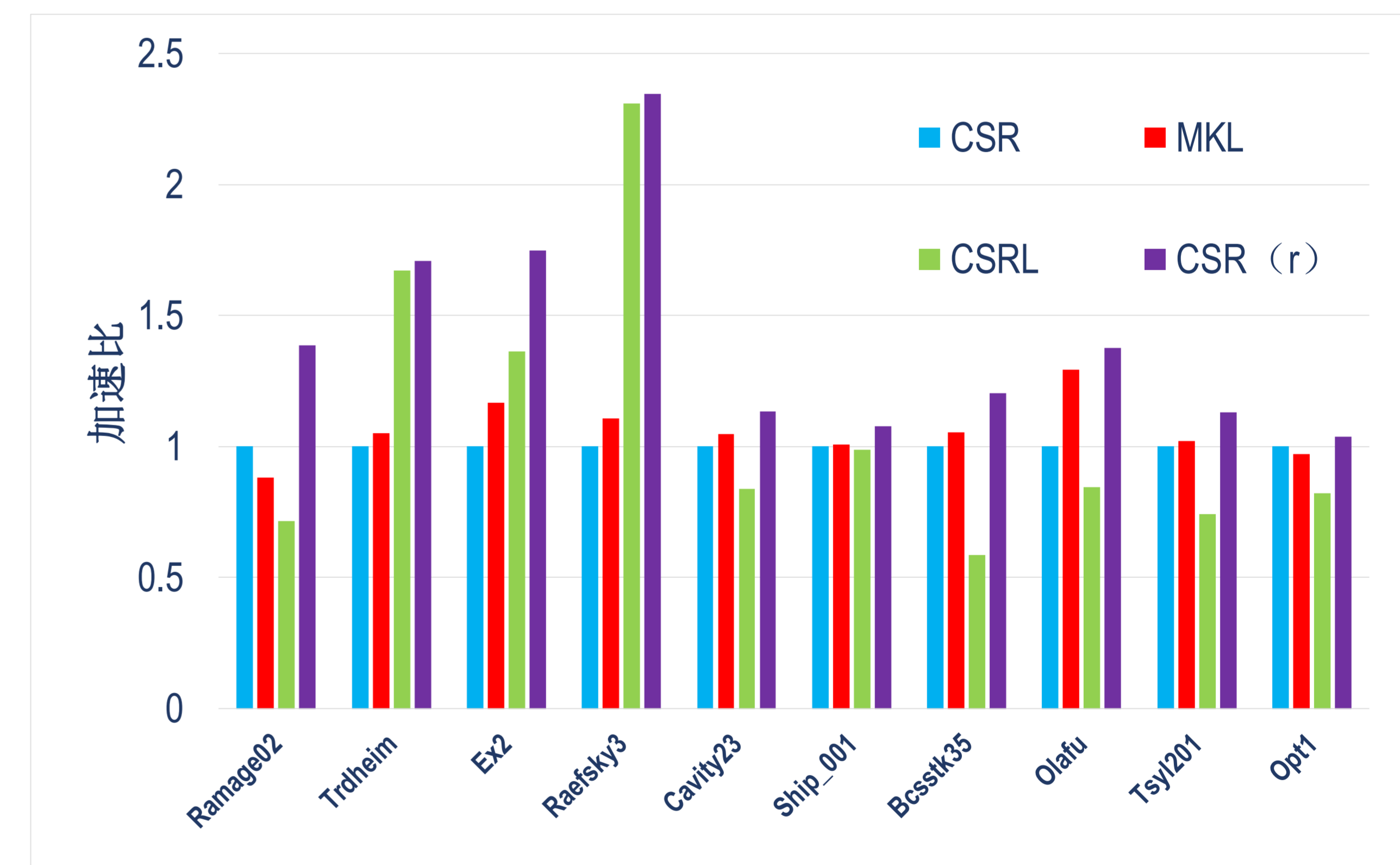
Ptr



Evaluations



不同稀疏矩阵CSR、MKL、CSRL和CSR(r)格式性能对比



不同稀疏矩阵CSR、MKL、CSRL和CSR(r)格式加速比

Conclusions

经统计分析, CSR(r)格式相比传统的CSR格式提高42%, 比国际著名的MKL库约有35%的提升。CSR(r)相对于向量化格式CSRL, 性能也有一定提升。另外CSR(r)格式比原CSRL结构优化效果不好的矩阵有显著地优化效果, 扩大向量化存储格式的优化范围。实验表明CSR(r)格式很好地发挥了向量化的加速效果, 大大提升了SPMV的性能并扩大了优化范围, 是一种适合向量化的稀疏矩阵存储格式。

References(selected)

- [1]. San X, Zhang Y, Wang T, et al. CRSD: application specific auto-tuning of SpMV for diagonal sparse matrices(M). Kura-Far 2011 Parallel Processing. Springer Berlin Heidelberg. 2011, 316-327.
- [2]. Liu X, Smelyanskiy M, Chow E, et al. Efficient sparse matrix-vector multiplication on x86-based many-core processors[C]. Proceedings of the 27th ACM International Conference on Supercomputing. ACM, 2013, 273-282.
- [3]. Vuduc Wilson, Demmel James, Yelick Katherine. (S)KI: A library of automatically tuned.