

曙光 TC4600 百万亿次超级计算系统新系统测试 1

广义本征值求解器 GenELPA

中国科学技术大学[超级计算中心](#) 沈瑜

一、 测试内容

测试使用 GenELPA 求解大规模矩阵的效率并与之前在原系统上做的测试结果进行对比。

广义本征值求解器 GenELPA 由中国科大超级计算中心自行开发[1], 该求解器基于面向 P 级应用的本征值求解库 ELPA[2], 用于解决第一性原理计算中 Kohn-Sham 方程的大规模并行计算工具。

GenELPA 分为两步, 第一步为右侧矩阵的分解, 此分解有三种方法, 分别包括使用 scalapack 提供的 pdotrf 进行 chelosky 分解 (m1), ELPA 自带的 chelosky 分解 (m2) 和对角化分解 (m3); 第二步包括调用 ELPA 求解本征值过程以及若干个并行矩阵乘积过程以进行广义本征值问题和一般本征值问题之间的转换。GenELPA1 和 GenELPA2 是分别使用 ELPA1 和 ELPA2 作为求解核心的求解器。

程序特点: 计算密集型, 通信密集型

二、 测试平台

- TC4600 原 E5v3 节点: 2×Intel Xeon E5-2680 v3 (2.5GHz), 单节点 24CPU 核心, 64GB DDR4 2133MHz, Mellanox FDR InfiniBand 56Gbps, CentOS 6.8, Intel 编译环境 2015 版 (编译器、数学库、MPI 并行库);
- TC4600 新 E5v4 节点: 2×Intel Xeon E5-2680 v4 (2.4GHz), 单节点 28CPU 核心, 128GB DDR 2400MHz, Intel OPA 100Gbps, CentOS 7.3, Intel 编译环境 2017 版 (编译器、数学库、MPI 并行库);
- TC4600 新 E3v5 节点: 1×Intel Xeon E3-1240 v5 (3.5GHz), 单节点 4CPU 核心, 32GB 2133MHz, Mellanox EDR InfiniBand 100Gbps, CentOS 7.3, Intel 编译环境 2017 版 (编译器、数学库、MPI 并行库)。

三、 GenELPA1 测试结果

程序设置:

矩阵大小: 20000x20000

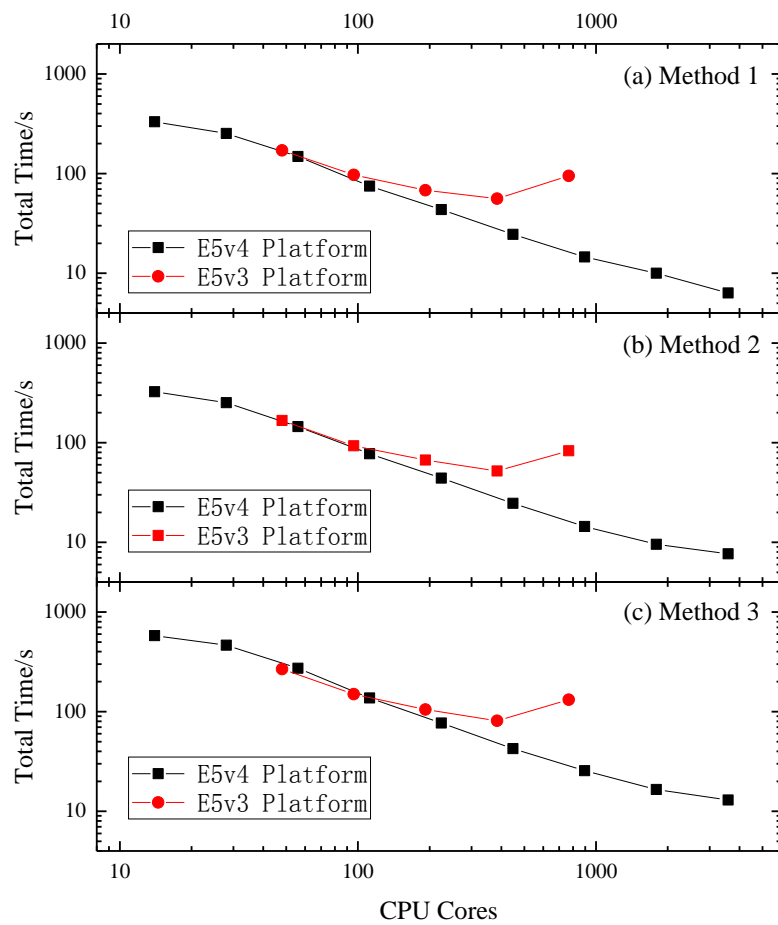
并行规模:

- 1) 使用 CPU 核心数为 2 的整数次方：16、32、64、128、256、512、1024、2048、4096；
- 2) 使用 CPU 核心数为单节点 28 核的倍数：14、28、56、112、224、448、896、1792、3584。

结果与讨论：

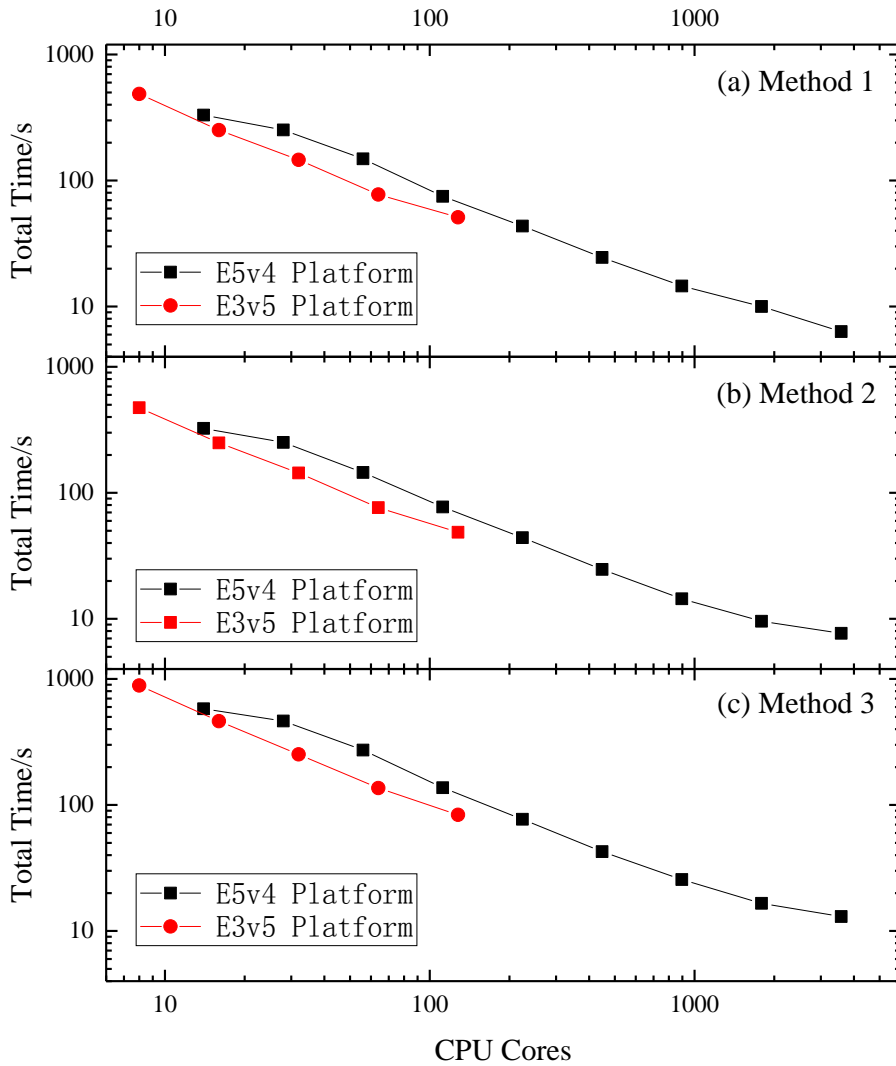
1. E5v4+OPA 计算网络系统与 E5v3 系统+InfiniBand 计算网络系统对比

- E5v3 测试条件：Intel 编译环境 2015 版，MPI 运行参数：-env I_MPI_FABRICS shm:ofa；
- E5v4 测试条件：Intel 编译环境 2017 版，MPI 运行参数：-env I_MPI_FABRICS shm:tmi。



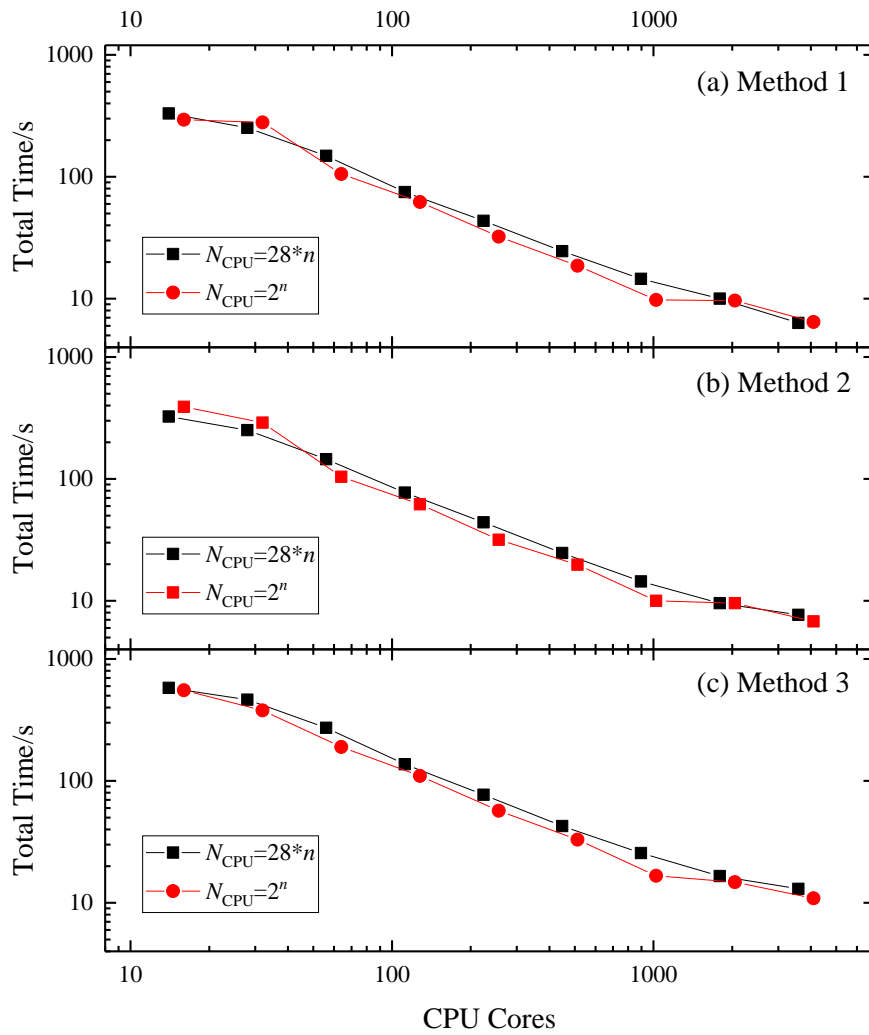
- 使用核心数较少时 (<100 核心)，新老系统速度接近；
- 超过 100 核心时，新系统优势明显；
- 新系统一方面网络性能更好，另一方面单节点具有更多的计算核心，因此整体并行扩展性可达到千核级；老系统测试时可能存在其他作业干扰（运行其他作业的节点与测试节点共用同一个 InfiniBand 交换机）等因素，因此并行扩展性不理想。

2. E5v4 与 E3v5 对比



- E5v4 系统能支持更大规模的并行；
- 在 E3v5 系统的规模之内 (160 个 CPU 核心), E3v5 速度更快, 并行效率相当；
- 与 E5v4 相比, E3v5 主频更高, 单 CPU 核心性能更强, 同时采用了节点内高效计算网络 Multi-Host (4 个节点共享 100Gbps, 每个节点 25Gbps), 能够在同等并行规模下具有更快的计算速度, 以及在系统规模下具有高效的并行扩展性。

3. 并行规模的方式对比（按整节点或者 2 的指数次方分配 CPU 资源的对比）



- 并行性能很好，在使用 4000 核心仍然有较好的并行效率；
- 大部分情况下，使用 CPU 核心数为 2 的整数次方（64、128、...）会具有更好的性能，与文献中结论一致[2]。

四、 GenELPA 2 结果

程序设置：

矩阵大小：20000x20000

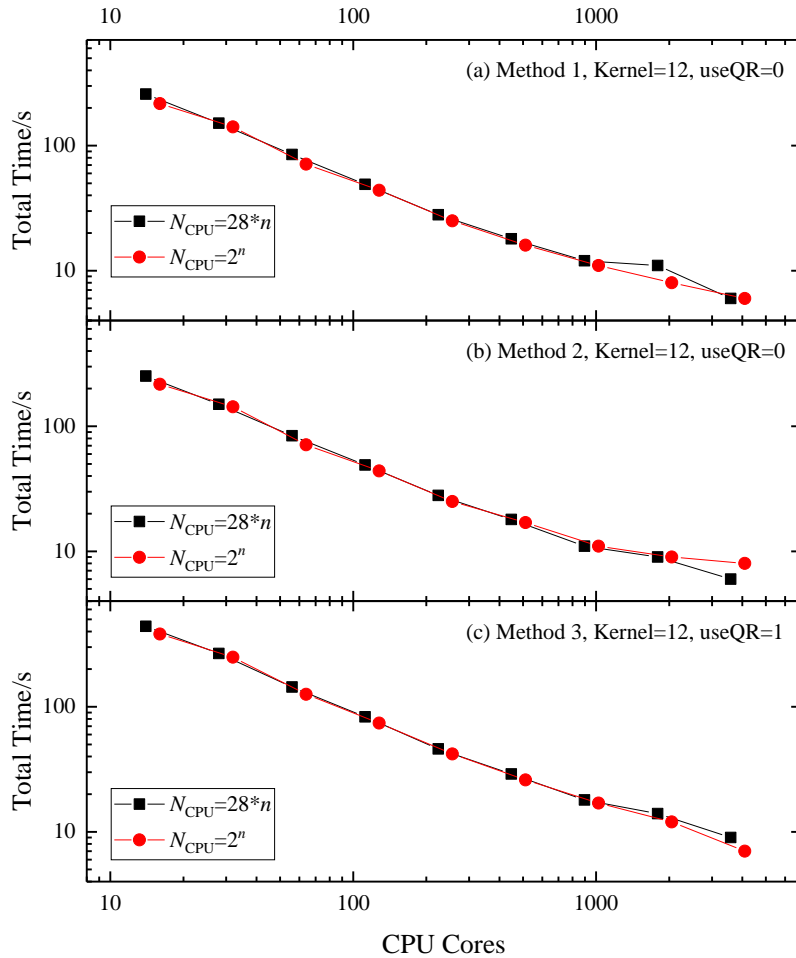
并行规模：

- 1) 使用 CPU 核心数为 2 的整数次方：16、32、64、128、256、512、1024、2048、4096；
- 2) 使用 CPU 核心数为单节点 28 核的：14、28、56、112、224、448、896、1792、3584。

使用计算速度最快的 kernel 和 useQR 参数

测试平台：E5v4+OPA 计算网络

测试内容：使用 CPU 核心数为 2 的整数次方与整节点（28 的倍数）从十核到千核并行度范围内的计算速度和扩展性。



结果讨论：

- 在 E5v4 系统上，采用 kernel=12，useQR=0（方法 1 和 2），useQR=1（方法 3）的参数搭配的速度会更快，主要是由于 kernel=12 使用了 AVX 扩展指令集，对速度影响较大，更多的数据发现 useQR=0 和 useQR=1 对速度影响较小，实际使用时可以都设置 useQR=0 以便简化参数；
- ELPA2 在 E5v4 系统上也具有很好的并行扩展性，这表明 E5v4 系统的计算网络非常高效；
- 大部分情况下，使用 CPU 核心数为 2 的整数次方（64、128、……）与整节点相比，性能基本一致或略好，与文献中结论一致[2]。在千核以上时，会稍有差异（2~3 秒）可认为测试误差。

五、 参考资料

- [1] 沈瑜, 孙广中, 并行广义本征值求解器 GenELPA[J/OL]. 计算机科学与探索 <http://www.cnki.net/kcms/detail/11.5602.TP.20170116.1702.012.html>; GenELPA Project: https://git.ustclug.org/yshen/GenELPA_C
- [2] Andreas Marek, Volker Blum, Rainer Johanni, Ville Havu, Bruno Lang, Thomas Auckenthaler, Alexander Heinecke, Hans-Joachim Bungartz, and Hermann Lederer: The ELPA Library - Scalable Parallel Eigenvalue Solutions for Electronic Structure Theory and Computational Science The Journal of Physics: Condensed Matter 26, 213201 (2014). <http://stacks.iop.org/0953-8984/26/213201> ELPA project: <https://elpa.mpcdf.mpg.de/>