

doi: 10.3969/j.issn.1000-8349.2024.03.06

基于 GPU 集群的空间 VLBI 射电源条纹搜索研究

段学铭^{1,2}, 童力^{1,3,5}, 郑为民^{1,3,4,5}, 张娟^{1,3,4,5}

(1. 中国科学院上海天文台, 上海 200030; 2. 中国科学院大学, 北京 100049; 3. 中国科学院射电天文重点实验室, 北京 100101; 4. 国家基础学科公共科学数据中心, 北京 100190; 5. 上海市空间导航与定位技术重点实验室, 上海 200030)

摘要: 探月工程四期将发射嫦娥七号“鹊桥二号”中继星, 搭载 4.2 m 口径的抛物面望远镜, 与地面射电望远镜构成首个月轨空间 VLBI。空间 VLBI 受中继星轨道扰动和星载设备时延等因素影响, 射电源的预报时延模型无法引导相关处理机正常工作, 需通过基于实测信号条纹搜索的方法找到符合要求的高精度时延模型。为此, 提出一种基于 GPU 集群的空间 VLBI 射电源条纹搜索算法, 研究在 GPU 集群上负载均衡的任务分配方法, 并行实现相关处理模块和残余值搜索模块。通过 RadioAstron 1 min 观测数据验证, 相较于目前 CPU 集群平台, GPU 集群处理速度提升了 27.0 倍。

关键词: 空间 VLBI; 条纹搜索; GPU 集群; 相关处理

中图分类号: P111.44 **文献标识码:** A

1 引言

甚长基线干涉测量 (very long baseline interferometry, VLBI) 技术是一种高精度、高角分辨率的射电干涉技术, 广泛应用于大地测量、天文观测和深空导航等领域^[1]。该技术关键指标之一是 VLBI 角分辨率, 表示为: $\theta \approx \frac{\lambda}{B}$, 其中, θ 为角分辨率, λ 为波长, B 为基线长度 (即两个射电望远镜之间的距离)。基线越长, 角分辨率越高。地基 VLBI 由地球表面的射电望远镜构成, 基线长度不超过地球直径。科研人员计划将射电望远镜发射到空间, 与地面射电望远镜构成空间 VLBI (space VLBI, SVLBI)。空间 VLBI 基线长度可超过地球直径, 可实现更高角分辨率。

收稿日期: 2024-01-11; 修回日期: 2024-03-04

资助项目: 国家自然科学基金 (11973011); 国家基础科学数据共享服务平台 (U1931135)

通讯作者: 童力, tongli@shao.ac.cn

空间 VLBI 是未来 VLBI 重要发展方向之一。1963 年 Matveenko 等人^[2]提及将射电望远镜发射到空间, 旨在实现地基 VLBI 无法达到的角分辨率。20 世纪 70 年代至 20 世纪 80 年代初, 包括欧空局主导的 QUASAT (Quasar Satellite) 项目^[3]、苏联的 RACSAS (Radio Astronomical Cosmic System for Aperture Synthesis) 项目^[4]以及 IVS (International VLBI Satellite) 项目^[5]在内的空间 VLBI 项目, 虽未最终实施, 但在早期研究阶段积累了宝贵经验, 为后续空间 VLBI 的发展奠定了基础。1986 年美国利用 TDRSS (Tracking and Data Relay Satellite System) 卫星的 4.9 m 口径望远镜, 配备 S 和 Ku 波段的接收机, 与地面望远镜构成首个空间 VLBI 试验系统^[6], 将基线长度延伸至 2.2 倍地球直径, 并成功获得射电源干涉条纹, 首次验证了空间 VLBI 的可行性。1997 年日本成功发射首个专用于空间 VLBI 的 VSOP (VLBI Space Observatory Programme)/HALCA (Highly Advanced Laboratory for Communications and Astronomy) 卫星^[7]。该卫星搭载 8.8 m 口径抛物面望远镜, 配备 L、C 和 K 波段的接收机, 与地面望远镜构成空间 VLBI 系统, 将基线长度延伸至 3 倍地球直径, 开展了对银河和外星系射电源致密结构的研究。2011 年俄罗斯成功发射专用于空间 VLBI 的 RadioAstron 项目 Spektr-R 望远镜^[8], 其 10 m 口径的抛物面望远镜, 配备 P、L、C 和 K 波段的接收机, 与地面望远镜构成空间 VLBI 系统, 将基线长度延伸至 28 倍地球直径, 在 K 波段实现了微角秒的角分辨率。中国也在积极推动空间 VLBI 的发展, 计划发射嫦娥七号“鹊桥二号”中继星。此星将搭载 4.2 m 口径 X 波段抛物面望远镜, 与地面望远镜构成月球轨道空间 VLBI 系统, 首次进行月地超长基线空间 VLBI 试验。

空间 VLBI 数据处理依赖高精度时延模型, 但空间 VLBI 受中继星轨道扰动和星载设备时延等因素影响^[9], 射电源预报的时延模型无法引导相关处理机正常工作, 需通过基于实测信号条纹搜索的方法找到符合要求的高精度时延模型。目前已有一些针对射电源条纹搜索的算法, 例如 CAF-W (cross ambiguity function-wavelet boosting) 条纹搜索算法^[10]、多重网格条纹搜索算法^[11]和二阶多项式时延模型条纹搜索算法^[12]。

CAF-W 条纹搜索算法通过定位 CAF 幅度峰值估计时延和时延率, 并通过小波增强技术提高 CAF 峰值的信噪比, 实现无预报时延模型的条纹搜索。但该算法搜索范围与基线长度成正比, 不适用于长基线的空间 VLBI。

多重网格条纹搜索算法通过粗细两级网格进行条纹搜索, 粗网格搜索确定范围, 细网格搜索确定最佳时延和时延率。但该算法搜索得到线性时延模型, 不适用于信号弱、相关处理长时间积分得到干涉条纹的空间 VLBI。

二阶多项式时延模型条纹搜索算法是通过搜索二维时延和时延率网格, 确定获得相关幅值最大的时延和时延率网格, 拟合高精度的二阶时延模型。该算法在线性时延模型基础上引入时延率变化项, 提高了时延模型精度, 可适用于空间 VLBI。基于 CPU 集群设计的该算法, 在处理 RadioAstron 1 min 单基线观测数据时, 用时 45 min。以此估计, 若处理月轨空间 VLBI 1 h 多基线观测数据, 用时超过 100 h, 数据处理速度较慢。

GPU (graphic processing unit) 已发展为一种高性能处理平台, 具有高度并行、多线程、大数据带宽以及数百个计算核心等特点^[13]。除在科学计算等领域应用外, 在射电天文学中, 特别是 VLBI 领域, GPU 也得到了广泛应用。中国科学院上海天文台研究了

GPU 集群并行软件相关处理架构^[14]，用于 CVN (Chinese VLBI Network) 数据相关处理。俄罗斯科学院 IAA (International Academy of Astronautics) 研发了基于 GPU 集群的软件相关处理机^[15]，用于 VGOS (VLBI Global Observing System) 数据相关处理。ngEHT (next-generation EHT) 和 SMA (Submillimeter Array) 正在开发基于 GPU Tensor 核的软件相关处理机^[16]，用于处理大规模宽频的数据。

目前空间 VLBI 条纹搜索数据处理速度较慢，为此，本文提出一种基于 GPU 集群的空间 VLBI 射电源条纹搜索算法，研究在 GPU 集群上负载均衡的任务分配，探索提高处理速度的优化方法，并行实现相关处理模块和残余值搜索模块。

2 空间 VLBI 射电源条纹搜索算法

空间 VLBI 射电源条纹搜索算法由相关处理模块、残余值搜索模块和时延模型拟合模块组成。相关处理模块是将两台站的信号进行相关处理，得到台站间的互相关功率谱。残余时延和残余时延率统称为残余值。残余值搜索模块是将互相关功率谱进行二维傅里叶变换，得到残余值及相关幅值。时延模型拟合模块是根据残余值对时延和时延率进行修正，并拟合高精度二阶时延模型。图 1 展示了空间 VLBI 射电源条纹搜索算法流程。首先，设置以预报时延模型 $(\tau_{g0}, \dot{\tau}_{g0})$ 为中心的等步长二维时延和时延率搜索网格，搜索时延和时延率步长分别为 d_τ 、 $d_{\dot{\tau}}$ ^[12]：

$$d_\tau = \frac{N_f}{10B}, \quad d_{\dot{\tau}} = \frac{1}{2f_0T_c}, \quad (1)$$

其中， N_f 为 FFT 点数， B 为数据带宽， f_0 为天空频率， T_c 为相关处理的积分时间。不同网格对应不同的时延和时延率，表示经不同时延补偿的数据处理。二维搜索网格设置如图 2 所示，红色网格对应预报时延模型。每个网格的时延和时延率是以预报时延模型为中心等步长计算得到。其次，对每个网格进行相关处理和残余值搜索，得到残余值及相关幅值。接着，计算每个网格的相关幅值均值，确定相关幅值均值最大的网格为最优网格。最后，根据最优网格的残余值，对该网格的时延和时延率进行修正，并利用修正后的时延和时延率进行拟合，得到高精度的二阶时延模型。

2.1 相关处理模块

选取某台站为参考站，一般选取地面口径最大的望远镜，将其称为地面站。在空间中的望远镜称为空间站。设地面站射频信号为 $d_1(t)$ ，空间站射频信号为 $d_2(t)$ ，同一波前信号到达两台站的时延为 τ_g ，则：

$$d_2(t) = d_1(t + \tau_g). \quad (2)$$

将射频信号下变频得到基频信号：

$$s_1(t) = d_1(t)e^{-j2\pi f_0 t}, \quad (3)$$

$$s_2(t) = d_2(t)e^{-j2\pi f_0 t} = d_1(t + \tau_g)e^{-j2\pi f_0 t}, \quad (4)$$

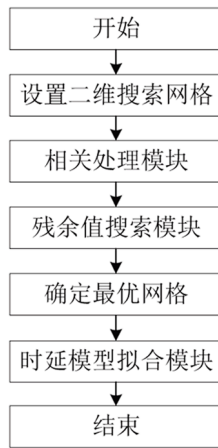


图 1 空间 VLBI 射电源条纹搜索流程图

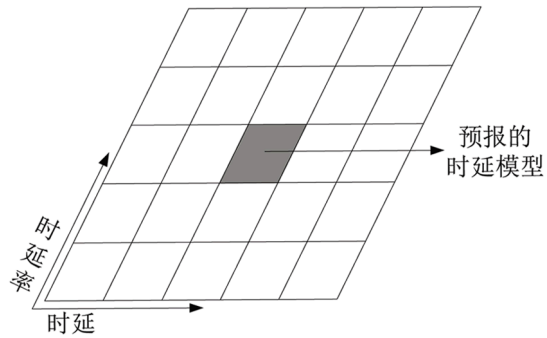


图 2 二维搜索网格示意图

其中, f_0 为天空频率。设预报时延为 τ_{g0} , 以地面站为参考站, 对空间站进行时延补偿:

$$s_2(t - \tau_{g0}) = d_2(t - \tau_{g0})e^{-j2\pi f_0(t - \tau_{g0})} = d_1(t + \tau_g - \tau_{g0})e^{-j2\pi f_0(t - \tau_{g0})} . \quad (5)$$

通过条纹旋转, 消除由于对射频信号做时延补偿引入的相位 $e^{j2\pi f_0 \tau_{g0}}$ 偏差:

$$s_{2f_s}(t) = s_2(t - \tau_{g0})e^{-j2\pi f_0 \tau_{g0}} = d_1(t + \tau_g - \tau_{g0})e^{-j2\pi f_0 t} . \quad (6)$$

设 $d_1(t)$ 通过傅里叶变换为 $D_1(f)$, 对 $s_1(t)$ 和 $s_{2f_s}(t)$ 分别进行傅里叶变换:

$$S_1(f) = D_1(f + f_0) , \quad (7)$$

$$S_{2f_s}(f) = D_1(f + f_0)e^{j2\pi(f_0 + f)(\tau_g - \tau_{g0})} . \quad (8)$$

将 $S_1(f)$ 与 $S_{2f_s}(f)$ 共轭相乘累加, 得到两台站间的互相关功率谱:

$$P_{1,2}(f) = S_1(f)S_{2f_s}^*(f) = |D_1(f + f_0)|^2 e^{-j2\pi(f_0 + f)(\tau_g - \tau_{g0})} . \quad (9)$$

当 $\tau_{g0} = \tau_g$ 时, 两台站间的互相关功率谱最大。

2.2 残余值搜索模块

相关处理得到的互相关功率谱 $P_{1,2}(f)$ 是时变的, 表示为 $P_{1,2}(f, t)$ 。对其进行二维傅里叶变换并取其绝对值, 得到一个关于残余时延和残余时延率的相关幅值函数, 如图 3 所示, 其公式为:

$$F(\Delta\tau, \Delta\dot{\tau}) = \left| \int_0^B \int_0^T P_{1,2}(f, t) e^{-j2\pi f_0 \Delta\dot{\tau} t} e^{-j2\pi \Delta\tau f} dt df \right| , \quad (10)$$

其中, $\Delta\tau$ 为残余时延变量, $\Delta\dot{\tau}$ 为残余时延率变量, $F(\Delta\tau, \Delta\dot{\tau})$ 为关于残余时延和残余时延率的相关幅值, B 为数据带宽, T 为每组二维傅里叶变换的数据时长, f_0 为天空频率。当 $F(\Delta\tau, \Delta\dot{\tau})$ 最大时, $\Delta\tau$ 为残余时延, $\Delta\dot{\tau}$ 为残余时延率。

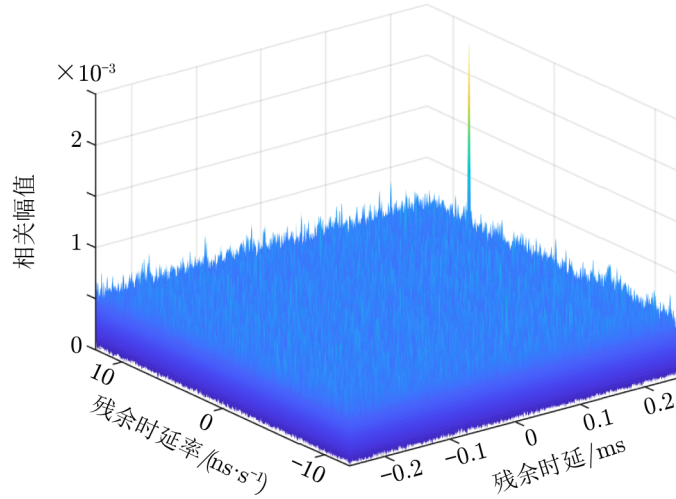


图3 残余值搜索相关幅值

2.3 时延模型拟合模块

对于空间 VLBI, 使用线性时延模型难以满足相关处理机的精度要求^[9]。考虑将观测数据分成较短数据块进行处理, 通过拟合二阶时延模型的方法来提高时延精度。将观测数据划分为 L 组, 对每组数据进行相关处理和残余值搜索, 得到残余时延 $\Delta\tau_l$ 和残余时延率 $\Delta\dot{\tau}_l$ 。设预报时延为 τ_g , 预报时延率为 $\dot{\tau}_g$, 每组数据时长为 T 。利用残余值对预报值进行修正, 得到时延 τ_l 和时延率 $\dot{\tau}_l$:

$$\begin{cases} \tau_l = \tau_g + \Delta\tau_l \\ \dot{\tau}_l = \dot{\tau}_g + \Delta\dot{\tau}_l \end{cases} \quad (l = 0, 1, 2, \dots, L-1) \quad . \quad (11)$$

对分组得到的时延值二阶时延模型拟合时, 需要根据连续性进行, 保证相邻两组时延相等, 即

$$\begin{cases} \tau'_0 = \tau_0 \\ \tau'_l = \tau'_{l-1} + \dot{\tau}_l T \end{cases} \quad (l = 1, 2, 3, \dots, L-1) \quad . \quad (12)$$

设待定二阶时延模型系数为 $[a_0, a_1, a_2]$, 时延的二阶多项式与时延率的一阶多项式建立超定方程组:

$$\begin{cases} \tau'_l = a_0 + a_1 t_l + a_2 t_l^2 \\ \dot{\tau}_l = a_1 + 2a_2 t_l \end{cases} \quad (l = 0, 1, 2, \dots, L-1) \quad . \quad (13)$$

其中, $t_l = lT$ 。求解模型系数 $[a_0, a_1, a_2]$ 。

3 基于 GPU 的条纹搜索算法

不同数据块不同网格间的相关处理与残余值搜索相互独立, 因此适合利用 GPU

加速运行。图 4 为条纹搜索算法各模块在 CPU 上的耗时百分比。我们考虑将大规模计算的相关处理模块和残余值搜索模块置于 GPU, 将时延模型拟合模块置于 CPU。相关处理模块包括地面站数据的解码、空间站数据的解码及条纹旋转、地面站和空间站数据的快速傅里叶变换 (fast Fourier transform, FFT)、空间站数据的小数比特补偿、地面站与空间站数据的共轭相乘累加。残余值搜索模块包括互相关功率谱的 2D-FFT 和残余值及相关幅值的确定。相关处理模块和残余值搜索模块在 GPU 上实现被称为内核计算。空间 VLBI 射电源条纹搜索在 GPU 上实现框架如图 5 所示。

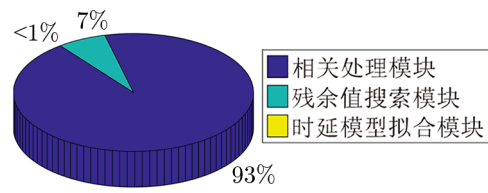


图 4 各模块搜索耗时百分比

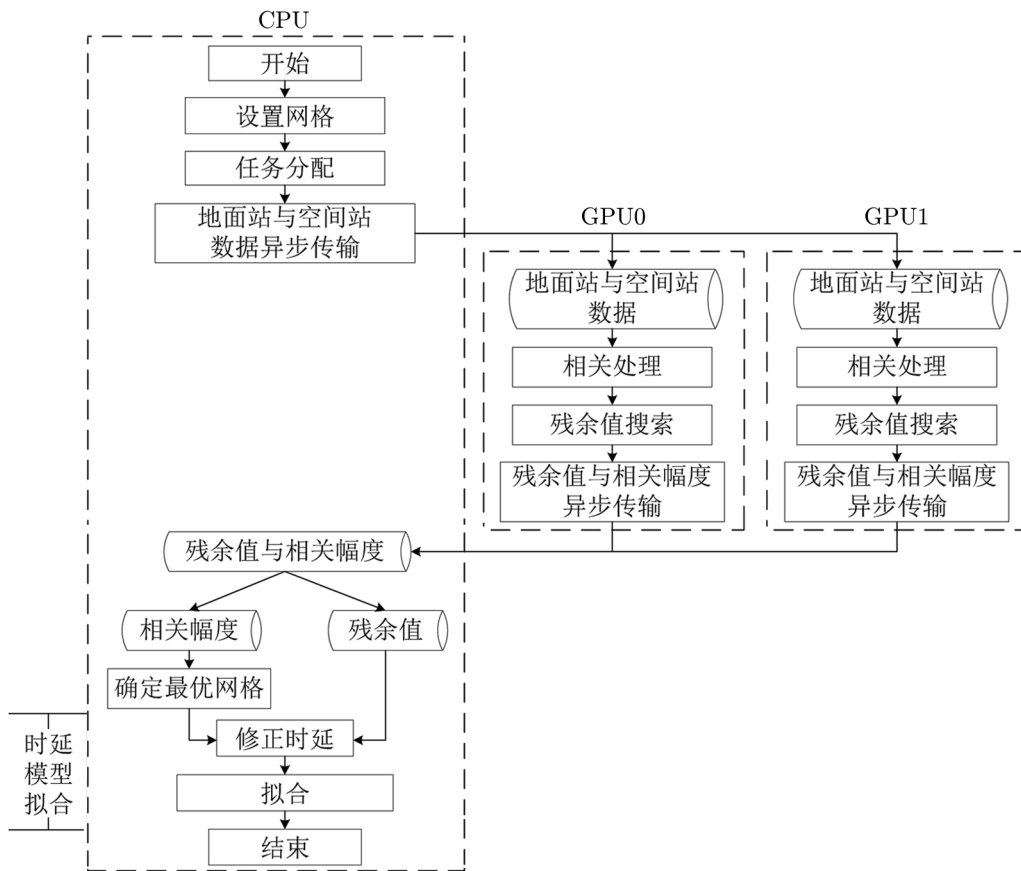


图 5 空间 VLBI 射电源条纹搜索在 GPU 上实现框架图

3.1 任务分配

在 Linux GPU 集群平台下, 空间 VLBI 射电源条纹搜索算法通过 MPI+POSIX 线程混合编程方式实现。MPI 控制计算节点, POSIX 线程控制计算节点内的多 GPU, 实现节点间的并行和节点内的多 GPU 并行。该算法的两级并行架构要求高效的任务分配策略。图 6

展示了条纹搜索任务分配方案，以 3×3 搜索网格为例，将观测数据进行分组，其中不同数据块对应图中网格内不同颜色块。射电源条纹搜索任务分配以数据块为最小分配粒度，将不同数据块不同网格的任务负载均衡地分配到每个计算资源中，确保每个计算资源的任务为相同数据块不同网格的任务。

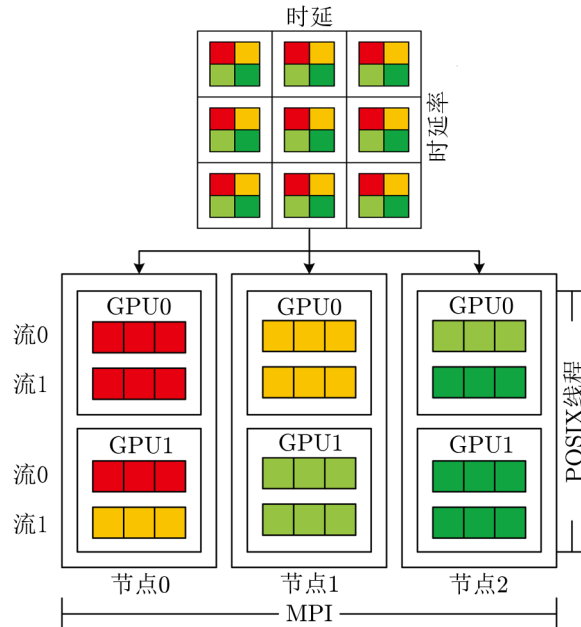


图 6 任务分配方案

以地面站为参考站，地面站数据无需时延补偿和条纹旋转，只需解码和FFT。地面站数据处理包括地面站数据传输和内核计算。与地面站不同，空间站的数据需要时延补偿，而时延补偿取决于二维搜索网格的时延和时延率。在不同搜索网格中，空间站数据需要进行不同的整数比特补偿、解码及条纹旋转、FFT 和小数比特补偿。在某个任务中，经相关处理和残余值搜索后，地面站数据处理的结果存储在 GPU 显存中。在后续相同数据块不同网格的任务中，无需再处理地面站数据，只需处理空间站数据。在地面站与空间站的数据共轭相乘累加时，地面站的数据可直接从 GPU 内存中提取，有效减少了地面站数据处理的时间。地面站数据处理优化原理如图 7 所示。



图 7 优化前后地面站数据处理时间对比

任务分配过程中，我们引入了并发流（其原理如图 8 所示），因此实现了数据的异步传

输和内核计算的异步执行, 在优化地面站数据处理的基础上, 进一步提高处理速度。GPU 可同时启动多条流, 使数据传输和内核计算在每条流中独立进行。异步执行的方式, 允许某条流在传输某个任务的地面站数据和空间站数据同时, 其他流执行其他任务的相关处理和残余值搜索。有助于最大程度地隐藏内核计算的开销, 提高整体处理速度。



图 8 引入并发流前后的数据处理时间对比

3.2 并行实现

本节研究内核计算在 GPU 上的实现, 包括相关处理模块实现和残余值搜索模块实现。图 9 为内核计算在 GPU 上的并行实现框架图。并发流的应用可以实现数据的异步传输和内核计算的异步执行, 其中数据异步传输通过 `cudaMemcpyAsync()` 函数^[13]实现。FFT 和 2D-FFT 通过利用 `cuFFT` 库^[17]中 `cufftExecC2C()` 函数实现, 通过 `cufftPlanMany()` 函数创建的 `plan` 区分 FFT 维数。残余值搜索模块的残余值通过 `cublasIsamax()` 函数^[18]确定, 相关幅度通过残余值位置确定。由于地面站和空间站的数据占内存较大, 只能将其存储在全局内存中。但空间站条纹旋转参数和小数比特补偿参数占内存较小, 可将其存储在共享内存中, 提高内存读写速度。

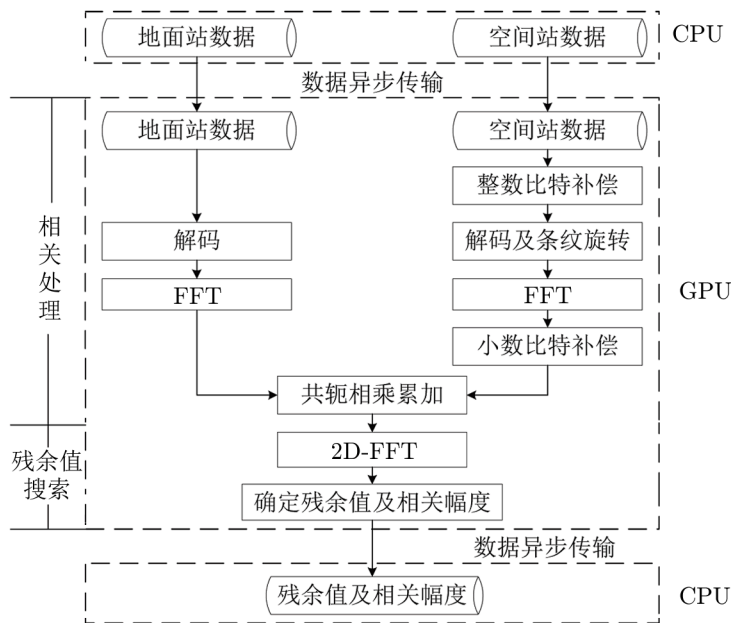


图 9 内核计算在 GPU 上的并行实现框架图

4 测试结果与性能分析

测试使用的 Linux GPU 集群计算平台, 包括 7 个计算机节点, 每个节点配置 4 块 Nvidia Tesla K80 GPU。测试数据为射电源 0823 + 033 的 RadioAstron 观测数据, 其中地面站为阿雷西博射电望远镜 (Arecibo Radio Telescope, Ar), 空间站为 RadioAstron 空间 Spektr-R 望远镜 (Ra), 由地面站 Ar 与空间站 Ra 构成空间 VLBI。观测时间为 2014 年 4 月 11 日 23 时 7 分 0 秒至 2014 年 4 月 11 日 23 时 8 分 0 秒, 共 1 min 观测数据, 将其划分为 61 组, 中心频率为 4836 MHz, 数据带宽为 16 MHz, 积分时间为 8.192×10^{-3} s, FFT 点数为 16384。预报时延模型为 $[7.36875 \times 10^{-3}$ s, 2.88779×10^{-6} s/s], 搜索时延和时延率步长分别为 d_τ 、 $d_{\dot{\tau}}$:

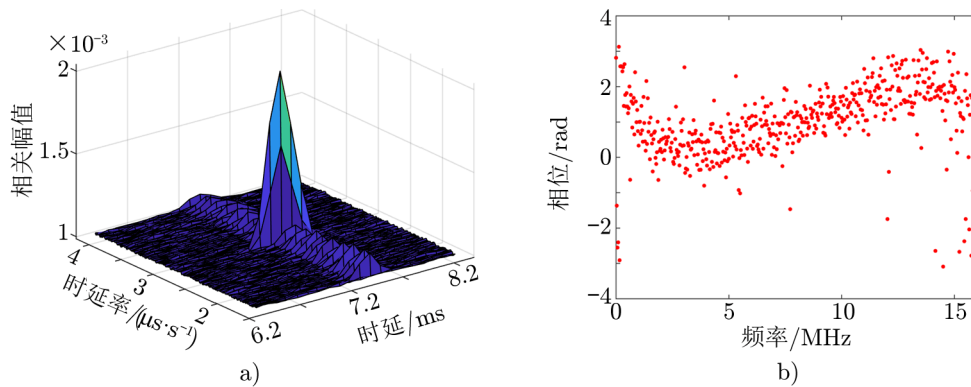
$$d_\tau = \frac{N_f}{10B} = 1.02400 \times 10^{-4} \text{ s} , \quad (14)$$

$$d_{\dot{\tau}} = \frac{1}{2f_0 T_c} = 1.26210 \times 10^{-8} \text{ s/s} . \quad (15)$$

测试中建立不同搜索范围、以预报时延模型为中心的等步长二维时延和时延率搜索网格。

4.1 测试结果

设置搜索网格为 20×200 , 时延搜索范围为 6.34475×10^{-3} s \sim 8.29035×10^{-3} s, 时延率搜索范围为 1.62569×10^{-6} s/s \sim 4.13727×10^{-6} s/s。通过图 10a) 可以确定最佳网格点位于 (11, 101), 时延为 7.36875×10^{-3} s, 时延率为 2.88779×10^{-6} s/s, 相关幅值为 2.14279×10^{-3} 。随后, 通过时延模型拟合模块得到高精度二阶时延模型 $[7.36871 \times 10^{-3}$ s, 2.88784×10^{-6} s/s, -4.31429×10^{-11} s/s²]。将空间 VLBI 射电源条纹搜索得到的二阶时延模型引入到相关处理机^[19]中, 可以得到如图 10b) 所示的互相关功率谱的相位谱。



注: a) 20×200 网格的相关幅值; b) Ar-Ra 互相关功率谱的相位谱。

图 10 条纹搜索结果

我们使用 7 个计算节点的 GPU 集群, 同时优化地面站数据处理、应用并发流。通过对

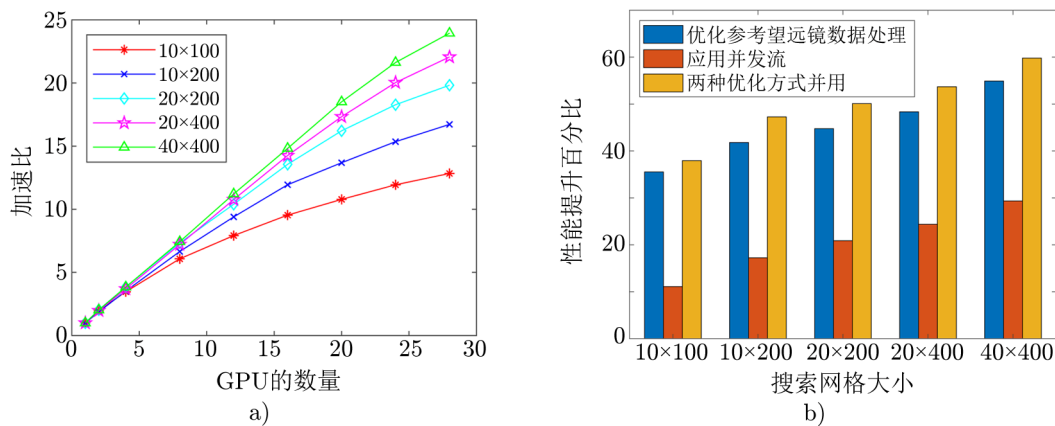
比表 1 中 20×200 搜索网格运行耗时可知, 当对 1 min 观测数据进行条纹搜索时, 相较于 CPU 集群, GPU 集群数据处理速度提升了 27.0 倍, 处理速度有了数量级的提升, 满足工程需求。

表 1 CPU 集群与 GPU 集群的运行耗时

搜索网格 (时延 \times 时延率)	加速方式	运行耗时	加速比
20×200	CPU 集群	45 min ^[12]	$\times 1$
20×200	GPU 集群	1 min 40 s	$\times 27.0$

4.2 性能分析

图 11a) 展示了使用不同数量 GPU 计算资源的加速比 (相较于单 GPU) 曲线。当计算资源一定时, 大搜索范围的加速比通常高于小搜索范围的加速比, 这是因为大搜索范围在 GPU 上的任务量较大, 可以充分展现 GPU 高性能计算的优势; 随着计算资源的增加, 大搜索范围和小搜索范围的加速比差异逐渐扩大, 这是因为 GPU 任务量越大, GPU 负载越大, 处理速度越快。总的来说, 基于 GPU 集群的空间 VLBI 射电源条纹搜索具有良好的可扩展性。



注: a) 不同 GPU 数量及不同搜索范围的加速比; b) 不同搜索范围及不同优化方式的性能提升百分比。

图 11 程序性能分析

图 11b) 展示了使用不同优化方式性能提升百分比 (相较于未优化)。随着搜索范围的扩大, 优化地面站数据处理的性能百分比不断提高, 因为 GPU 分配到的相同数据块不同网格的任务量越大, 地面站数据处理的次数减少越多, 处理速度提升越明显; 随着搜索范围的扩大, 应用并发流的性能提升百分比也不断提高, 因为搜索范围越大, GPU 的任务量越大, 负载越大, 资源利用率越高, 内核并发的可能性越大。虽然优化地面站数据处理和应用并发流都可以提高性能, 但这并不等于同时应用两种优化方式的效果, 因为减少地面站数据处理提高了 GPU 资源利用率, 同时也降低了内核并发的可能性。相较于只应用并发流, 在优化地面站数据处理的基础上应用并发流, 性能提升更显著。

5 结论

本文提出了一种基于 GPU 集群的空间 VLBI 射电源条纹搜索算法。根据算法实现的两级并行架构, 本文研究了负载均衡的任务分配策略。在任务分配基础上, 本文又提出了优化地面站数据处理和应用并发流两种优化方法, 在 GPU 上实现了相关处理模块和残余值搜索模块的并行化。最后, 利用 RadioAstron 观测数据进行验证, 相较于 CPU 集群, GPU 集群处理速度有了数量级的提升, 可用于工程事后数据处理。此外, 本文对此程序进行了性能分析, 结果显示, 该条纹搜索程序具有良好的可扩展性。该研究成果可用于后续的空间 VLBI 项目。

参考文献:

- [1] 钱志翰, 李金岭. 甚长基线干涉测量技术在深空探测中的应用. 北京: 中国科学技术出版社, 2012: 1
- [2] Matveenko L I, Kardashev N S, Sholomitskii G B. *Soviet Radiophysics*, 1965, 8(4): 461
- [3] Schilizzi R T, Burke B F, Booth R S, et al. *The QUASAT Project*. Cambridge: Cambridge University Press, 1984, 110: 407
- [4] Kardashev N S, Pogrebenko S V, Tsarevskii G S. *Soviet Astronomy*, 1980, 24: 366
- [5] Pilbratt G. *IVS-A Second Generation Space VLBI Mission*. Cambridge: Cambridge University Press, 1991, 131: 102
- [6] Levy G S, Linfield R P, Ulvestad J S, et al. *Science*, 1986, 234(4773): 187
- [7] Hirabayashi H, Hirose H, Kobayashi H, et al. *Science*, 1998, 281(5384): 1825
- [8] Kardashev N S, Khartov V V, Abramov V V, et al. *Astronomy Reports*, 2013, 57: 153
- [9] Likhachev S F, Kostenko V I, Girin I A, et al. *Journal of Astronomical Instrumentation*, 2017, 6(03): 1750004
- [10] Zhang T, Meng Q, Chen C, et al. *Publications of the Astronomical Society of the Pacific*, 2017, 129(977): 074501
- [11] 孙晓彤, 童力, 郑为民, 等. *天文研究与技术*, 2021, 18(01): 52
- [12] 朱晓娜, 童力, 郑为民, 等. *天文研究与技术*, 2022, 19(06): 568
- [13] *CUDA C Programming Guide*. https://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf, 2024
- [14] 陈中. 博士学位论文, 北京: 中国科学院大学, 2015: 46
- [15] Ken V, Surkis I, Kurdubova Y, et al. *Polarization*, 2015, 1: 2
- [16] Yu W, Romein J W, Dursi L J, et al. *Galaxies*, 2023, 11(1): 13
- [17] *CUFFFT Library*. https://docs.nvidia.com/cuda/pdf/CUFFFT_Library.pdf, 2024
- [18] *CUBLAS Library*. https://docs.nvidia.com/cuda/pdf/CUBLAS_Library.pdf, 2024
- [19] 张娟, 郑为民, 刘磊, 等. *中国科学: 物理学力学天文学*, 2021, 51(11): 37

Research on Space VLBI Radio Source Fringe Search Based on GPU Cluster

DUAN Xue-ming^{1,2}, TONG Li^{1,3,5}, ZHENG Wei-min^{1,3,4,5}, ZHANG Juan^{1,3,4,5}

(1. Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Beijing 100101, China; 4. National Public Science Data Center for Basic Disciplines, Beijing 100190, China; 5. Shanghai Key Laboratory of Space Navigation and Positioning Techniques, Shanghai 200030, China)

Abstract: The fourth phase of the Lunar Exploration Project will launch the Chang'e-7 relay satellite, carrying a 4.2-meter parabolic telescope. Together with ground-based radio telescope, it will form the first Lunar Orbit Space VLBI system. Space VLBI is influenced by the relay satellite's orbit perturbations and onboard equipment delays, and the predicted time delay model for radio source may not effectively guide the normal operation of the correlator. It is necessary to employ a fringe search based on actual measurement signals to find a high-precision time delay model that meets the requirements. Currently, fringe searches are accelerated using CPU cluster, but the search time is prolonged, lasting up to 2700s, which fails to meet the practical engineering requirements. A space VLBI radio source fringe search algorithm based on GPU cluster is proposed. The research involves studying task distribution for load balancing on GPU cluster and parallel implementation of correlation modules and residual value search modules. Validated with RadioAstron observation data, the processing speed of GPU cluster has improved about 27.0 times compared to CPU cluster. The research outcomes can be applied to subsequent space VLBI projects.

Key words: space VLBI; fringe search; GPU cluster; correlation