

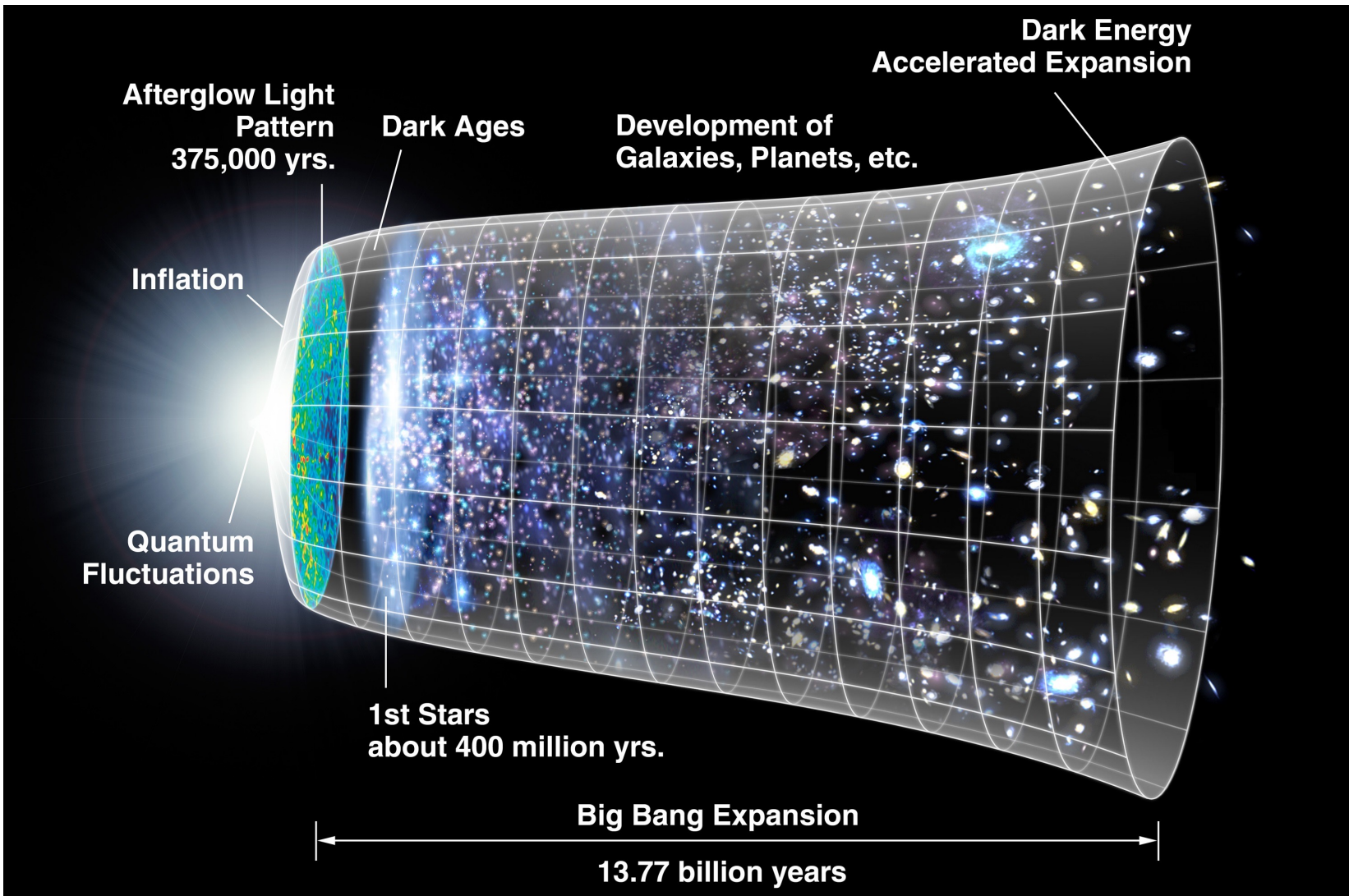
宇宙微波背景辐射 偏振模式简介

中国科学技术大学－粒子物理与原子核物理学学术报告

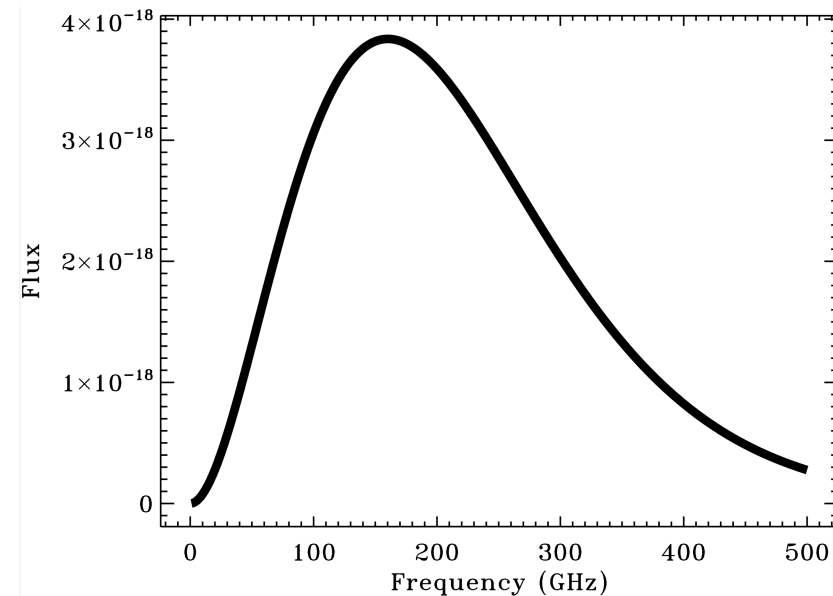
报告人：刘浩

2023-07-07

大爆炸与宇宙的历史



$$B(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(k_B T)} - 1}$$

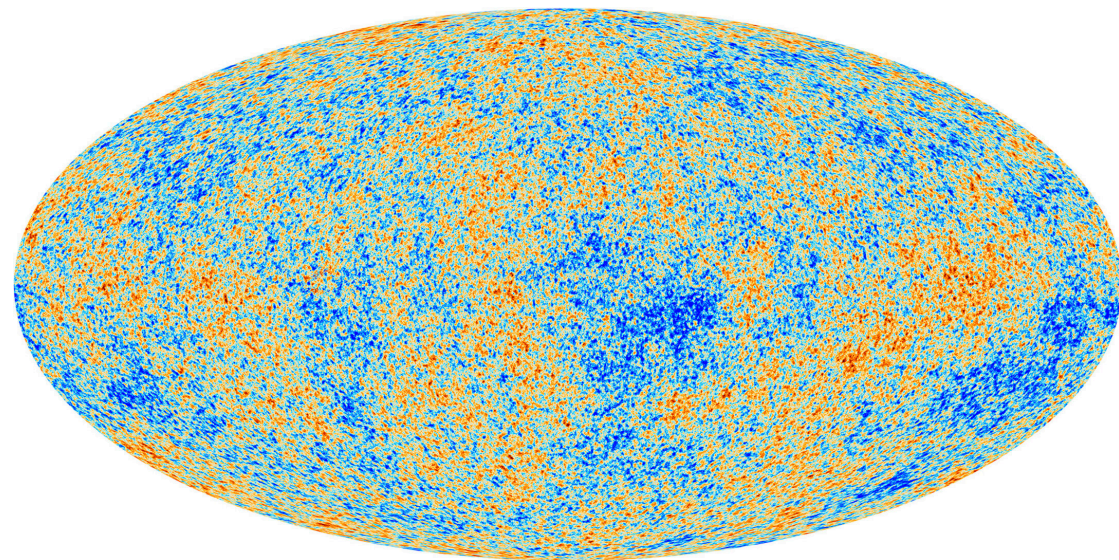
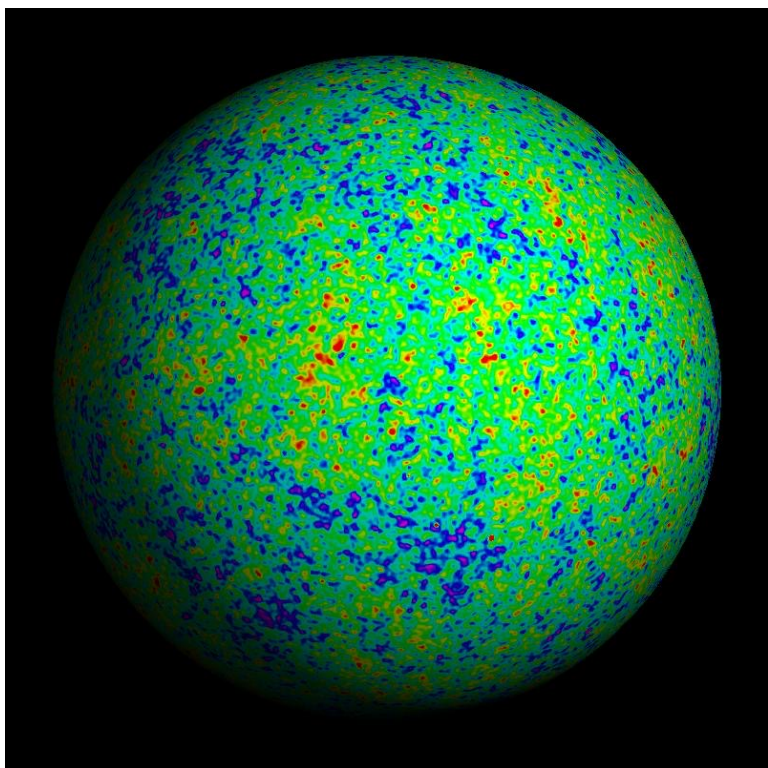


图源自：维基百科，Big Bang 词条

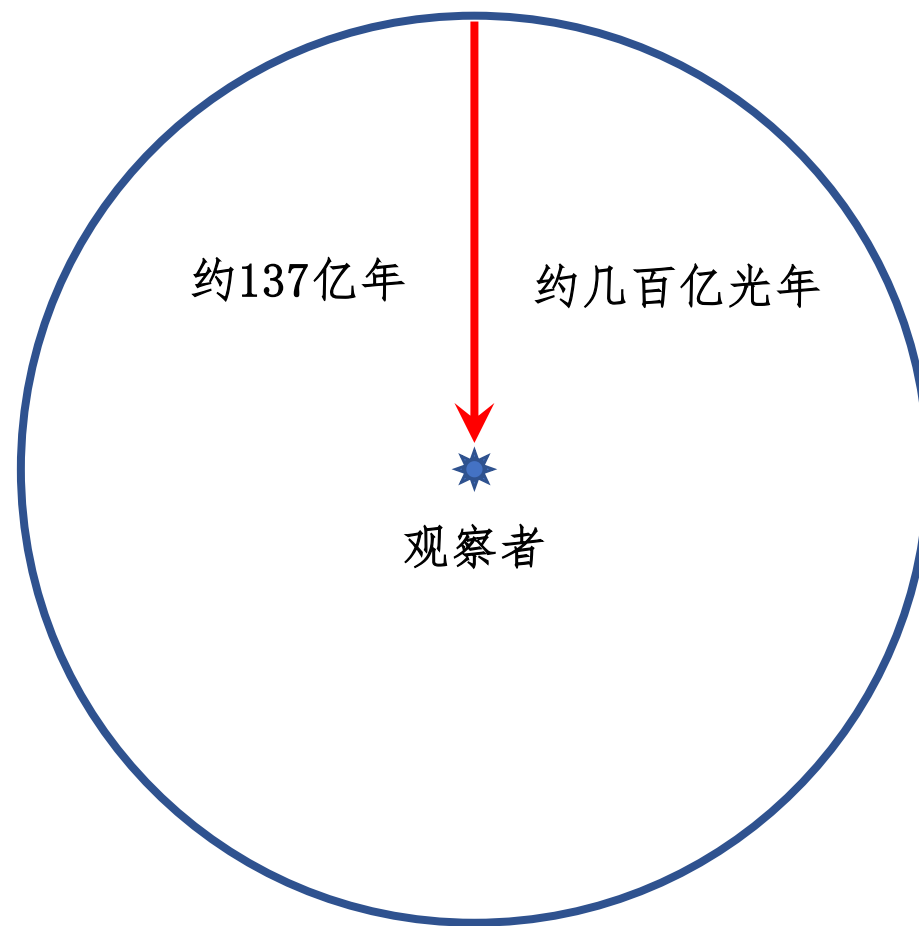
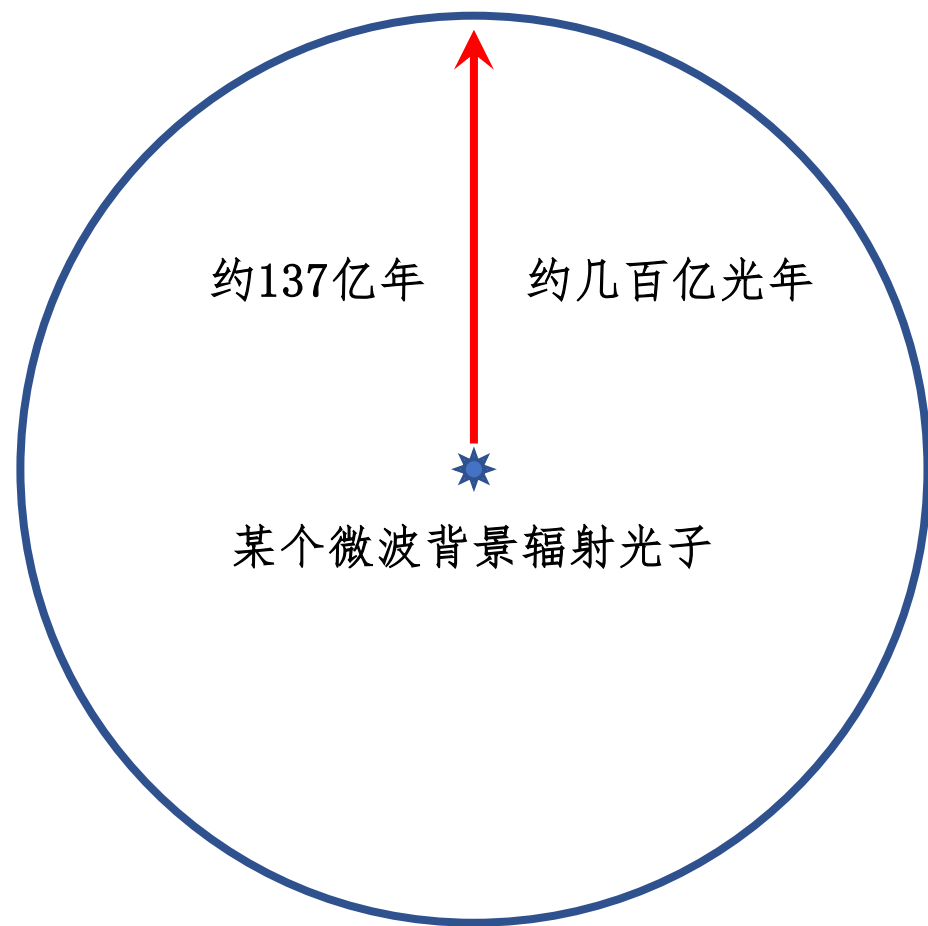
宇宙中最古老的信号：微波背景辐射

图源自：<https://space.mit.edu/>

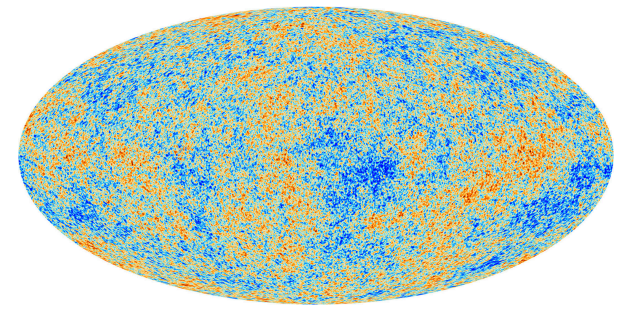
Credit: ESA Planck mission



微波背景辐射的传播






CMB 与球面



- 但是这里立刻就会产生一堆问题：
 1. 我们看到的CMB只在一个球面上，但它在刚诞生时是仅存在于球面上吗？
 2. 我们今天看到的 CMB 是怎样从三维空间被“压缩”到球面的？
 3. 假设在“当年”离我十米远处有一个 CMB 光子，今天它在哪？我能看到它吗？
 4. 如果我们在同一时刻从宇宙中的另一个地方看 CMB，结果是一样还是不同？
 5. 如果我们从同一个地方但是在不同时刻去看 CMB，结果是一样还是不同？
 6. 如果以上两个问题的答案是“不同”，你能够从上一页的图粗略估计一下，我们需要跑多远，或者等多久，才能看到比较明显的差异吗？
 - 无需给出具体数值，只要有合理的思路即可。
 7. 如果我们返回到 CMB 诞生的时刻，以上“不同地方”和“不同时刻”的差异在那个时候，在几何上分别对应于什么？
 8. 这些差异背后会不会有某种“相同”的东西？
 - 暂时只针对“不同地方”来讨论
 9. 暂时不考虑这个“相同”的东西究竟是什么，而是先思考：
 - 如果有，它需要具备一些什么特性？你能够设想一下吗？

宇宙学原理、协方差矩阵与旋转不变性

- 对 CMB 来说最重要的一条原理是宇宙学原理（哥白尼原理）：
 - 宇宙是均质和各向同性的。
 - 地球在宇宙中的位置并不特殊。
 - 无论从何处观测，也无论向哪个方向观测，得到的统计特征都一致。
- 然而有趣的是，虽然宇宙学原理极其重要，但它从未被证明过，也不可能被证明。
 - 理论证明 
 - 实验证明 
 - 通过观测逐步验证 

宇宙学原理、协方差矩阵与旋转不变性

- 基于宇宙学原理，在 CMB 研究中，我们常会考虑以下几个问题并给出默认的回答。而回答基本上决定了研究的方向：
 - 球面上任何两点的 CMB 温度/偏振的关联是否和两点的次序有关？
 - 以上关联是否和第一个点的空间位置有关？
 - 如果第一个点已确定，以上关联是否和两点之间的方位角有关？
- 在默认情况下，考虑到宇宙学原理，对以上三个问题的答案一般为：否、否、否。

宇宙学原理、协方差矩阵与旋转不变性

- 用稍微数学化一点的语言来讲，以上三个问题可以分别归纳为 CMB 涨落的两点相关函数的下列属性：
 - 交换对称性（Commutative property，即两点相关满足交换律）。
 - 均质性（Homogeneity，即两点相关与位置无关）。
 - 各向同性（Isotropy，即两点相关与方向无关）。
- 同时需要注意：原始的宇宙学原理（哥白尼原理）不包含宇称对称（镜像反演对称），但 CMB 的标准理论一般认可宇称对称。
- 这会导致 CMB 的不同种类角功率谱有明显差异（后话）。
 - 特别是对偏振有明显影响。

宇宙学原理、协方差矩阵与旋转不变性

- 从纯粹技术的角度来说，以上提到的，基于宇宙学原理的均质性和各向同性可以简单地归结为以下原则：
 - CMB 的协方差矩阵 $C(a, b)$ 仅取决于 a 、 b 两点的间距。
- 这一结论对 CMB 温度涨落的理论研究提出了严格的要求：旋转不变性。这一原则虽然简单，却意味着数学上的一系列后果。
 - 球面上的旋转对称性从属于 $SO(3)$ 旋转群。
 - 角动量算符构成 $SO(3)$ 旋转群的生成元。
 - 球谐函数是角动量算符的本征函数。
 - 角动量平方算符是卡西米尔算符，天生具备旋转不变性
 - 所以球谐函数是 CMB 温度涨落研究的主要数学工具之一。
 - CMB 温度涨落研究的主要目标是 CMB 角功率谱。

宇宙学原理、协方差矩阵与旋转不变性

- 但在 CMB 研究中真正具有根本地位的实际上是魏格纳 D-矩阵 (Wigner D-matrix) :
 - 魏格纳 D-矩阵在非偏振情况下依照宇宙学原理带来的对称性依次退化为球谐函数 (球面温度) 和勒让德多项式 (角功率谱)。
 - 在偏振情况下复原并变形为自旋球谐函数 (远比球谐函数复杂), 用于描述球面偏振。
 - 但任何情况下我们仍然要求旋转不变性, 这就对球面偏振的分析带来了一些要求。

魏格纳D-矩阵与旋转的数学变换

$$Y_{\ell m}(\theta', \varphi') = \sum_{m'=-\ell}^{\ell} Y_{\ell m'}(\theta, \varphi) D_{m'm}^{(\ell)}(\alpha, \beta, \gamma), \quad (56)$$

非偏振情况:

其中 (θ, φ) 代表旋转前的点, (θ', φ') 代表旋转后的点, (α, β, γ) 代表旋转的欧拉转角, $D_{mm'}^{(\ell)}(\alpha, \beta, \gamma)$ 称为对应于 (α, β, γ) 的魏格纳 **D**-矩阵。

偏振情况:

$${}_s Y_{\ell, m_1}(\theta', \varphi') = \sum_{m_2=-\ell}^{\ell} D_{m_1, -m_2}^{(\ell)}(\alpha, \beta, \gamma) [{}_s Y_{\ell, m_2}(\theta, \varphi)]. \quad (121)$$

共同特点 (关键点): ℓ 分量在任何旋转下都是保守的 \rightarrow 无论偏振还是非偏振, 基于宇宙学原理的 CMB 研究的核心部分最终都会归结到和 ℓ 分量相关的功率谱问题上。

偏振与非偏振：实数场与复数场

- 球面上非偏振与偏振信号的旋转变换非常相似，原因是它们的区别只不过是一个对应于球面实数场，一个对应于球面复数场。
 - 实数温度场： $T(\theta, \phi)$ 。
 - 复数偏振场： $P_{\pm}(\theta, \phi) = Q(\theta, \phi) \pm iU(\theta, \phi)$ 。
 - Q 、 U 指的是偏振的 Q 、 U 斯托克斯参数。
- 由于 Q 、 U 斯托克斯参数的旋转变换天生具有 2 倍变换关系，复数偏振场 $P_{\pm}(\theta, \phi)$ 自动对应于自旋为 2 的自旋球谐函数。
- 这使得 CMB 偏振场虽然是宏观场，但却和自旋为 2 的大量粒子的叠加系统有些类似。
 - CMB 偏振功率谱 \rightarrow 可类比于居于各个态的粒子的数目分布。
 - 旋转不变性 \rightarrow 可类比于 l 相同但 m 不同的态中的粒子独立（非相干）同分布。

复数偏振场的宇称态

- 一般的球面偏振场没有特定的宇称，也无需考虑这个问题。但微波背景辐射对我们提出了不一样的要求：
 - CMB 的偏振主要来自 CMB 光子在最后散射面上同自由电子的汤姆森散射。
 - 没有圆偏振，无需考虑Intensity。
 - CMB 涨落源于暴胀场的量子涨落，包含标量（密度涨落）和张量（原初引力波）两部分。
 - 标量涨落通过汤姆森散射产生的复数偏振场只能是偶数字称的。
 - 具有奇数字称的复数偏振场只有通过张量涨落才能产生。
 - 所以对 CMB 偏振场的宇称的研究是原初引力波探测的数学前提。

CMB偏振的字称态

基于以上需求，人们习惯上把 CMB 偏振分为具有偶数字称的 E 模式和具有奇数字称的 B 模式。

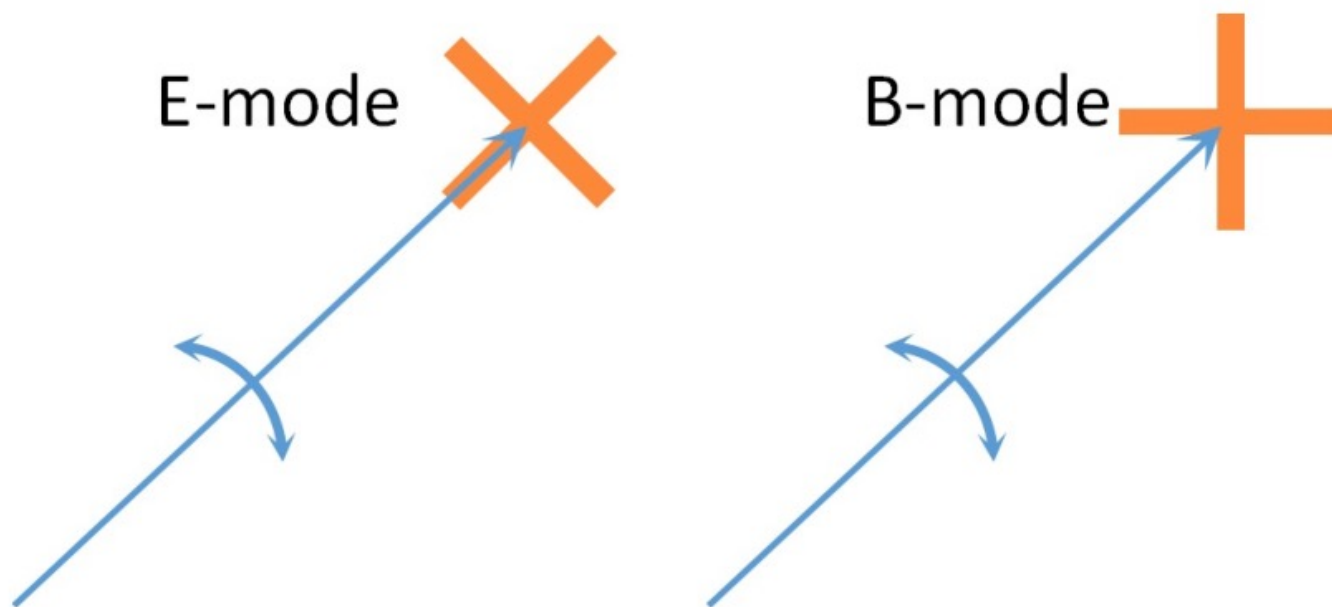


图 5: E 模式和 B 模式的偏振方向在实空间的展示。图源自 [10]。

各种宇称态是完全独立的吗？

$$a_{E,lm} = \frac{a_{2,lm} + a_{-2,lm}}{2}$$

$$ia_{B,lm} = \frac{a_{2,lm} - a_{-2,lm}}{2},$$

- 宇称态的数学构建基本上都具有以上形式。
- 镜像变换时自旋改变符号。
- 于是以上两种形式自然地具有镜像不变和镜像改变符号的属性。
- 小插曲：不变源于变
 - 自旋天生是镜像变号的。
 - 但是这种变号看似规则，实际会连基函数一起变。
 - 单纯自旋在实空间其实没有对称性。

CMB偏振宇称态的变换：微分变换形式

- CMB 偏振宇称态的数学变换比较复杂，一般有两条路线：
 - 微分变换形式

$$\bar{\partial}_+[{}_s f(\theta, \varphi)] = -\sin^{+s}(\theta) \left[\frac{\partial}{\partial \theta} + i \csc(\theta) \frac{\partial}{\partial \varphi} \right] \sin^{-s}(\theta) [{}_s f(\theta, \varphi)]$$

$$\bar{\partial}_-[{}_s f(\theta, \varphi)] = -\sin^{-s}(\theta) \left[\frac{\partial}{\partial \theta} - i \csc(\theta) \frac{\partial}{\partial \varphi} \right] \sin^{+s}(\theta) [{}_s f(\theta, \varphi)].$$

$$\bar{\partial}_-^2[\mathbf{P}_+] = \sum_{\ell m} \left[\frac{(\ell+2)!}{(\ell-2)!} \right]^{1/2} a_{+2, \ell m} Y_{\ell m}(\theta, \varphi)$$

$$\bar{\partial}_+^2[\mathbf{P}_-] = \sum_{\ell m} \left[\frac{(\ell+2)!}{(\ell-2)!} \right]^{1/2} a_{-2, \ell m} Y_{\ell m}(\theta, \varphi),$$



线性组合 → 用普通球谐函数反变换 → 标量场 $\mathbf{E}(\mathbf{n})$ 、 $\mathbf{B}(\mathbf{n})$ → 利用已有的标量场理论讨论偏振的旋转不变性的问题。

CMB偏振宇称态的变换：微分变换形式

- 小尺度发散问题

$$\begin{aligned}\delta_-^2[\mathbf{P}_+] &= \sum_{\ell m} \left[\frac{(\ell+2)!}{(\ell-2)!} \right]^{1/2} a_{+2,\ell m} Y_{\ell m}(\theta, \varphi) \\ \delta_+^2[\mathbf{P}_-] &= \sum_{\ell m} \left[\frac{(\ell+2)!}{(\ell-2)!} \right]^{1/2} a_{-2,\ell m} Y_{\ell m}(\theta, \varphi),\end{aligned}$$

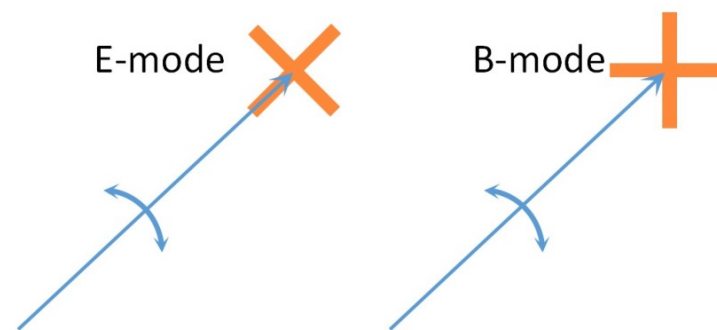


图 5: E 模式和 B 模式的偏振方向在实空间的展示。图源自 [10]。

CMB偏振宇称态的变换：积分变换形式

- 从上面的旋转变换公式可以看出，自旋球谐函数天生契合旋转不变的要求。这引出了 CMB 偏振宇称态变换的另一条路线：积分变换形式。

$$a_{\pm 2, \ell m} = \iint \mathbf{P}_{\pm}(\theta, \varphi) [{}_2Y_{\ell m}^*(\theta, \varphi)] \sin \theta d\theta d\varphi,$$

$$a_{E, \ell m} = \frac{a_{2, \ell m} + a_{-2, \ell m}}{2}$$

$$i a_{B, \ell m} = \frac{a_{2, \ell m} - a_{-2, \ell m}}{2},$$

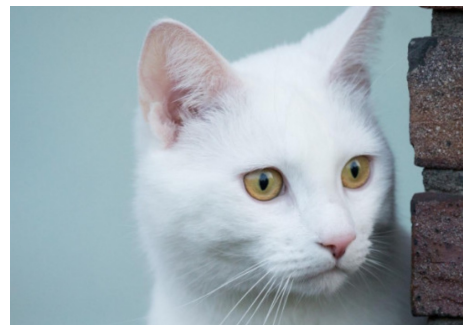
- 如果仅满足于在球谐函数空间得到 E 和 B 模式，那么到这里就差不多了。但更重要的问题是：E 和 B 模式在实空间究竟是什么样的？

黑猫、白猫、黑狗、白狗

黑

白

猫



狗



CMB偏振宇称态的变换：积分变换形式

- 这一工作在 2018 年才由我们首先完成 (Hao Liu et al., JCAP05(2018)059)

$$Q + iU = \sum_{\ell m} a_{+2,\ell m} [{}_2Y_{\ell m}]$$

$$Q - iU = \sum_{\ell m} a_{-2,\ell m} [{}_2Y_{\ell m}].$$

$$F_{+, \ell m} = \frac{1}{2} ({}_2Y_{\ell m} + {}_{-2}Y_{\ell m}) \quad a_{E, \ell m} = \frac{a_{2, \ell m} + a_{-2, \ell m}}{2}$$
$$F_{-, \ell m} = \frac{1}{2} ({}_2Y_{\ell m} - {}_{-2}Y_{\ell m}). \quad ia_{B, \ell m} = \frac{a_{2, \ell m} - a_{-2, \ell m}}{2},$$

$$Q = \frac{1}{2} \left(\sum_{\ell m} a_{2, \ell m} [{}_2Y_{\ell m}] + \sum_{\ell m} a_{-2, \ell m} [{}_{-2}Y_{\ell m}] \right), \quad \begin{cases} Q_E = \sum_{\ell m} a_{E, \ell m} F_{+, \ell m} \\ iU_E = \sum_{\ell m} a_{E, \ell m} F_{-, \ell m} \end{cases}; \quad \begin{cases} Q_B = \sum_{\ell m} (ia_{B, \ell m}) F_{-, \ell m} \\ iU_B = \sum_{\ell m} (ia_{B, \ell m}) F_{+, \ell m} \end{cases}$$
$$iU = \frac{1}{2} \left(\sum_{\ell m} a_{2, \ell m} [{}_2Y_{\ell m}] - \sum_{\ell m} a_{-2, \ell m} [{}_{-2}Y_{\ell m}] \right).$$

$$\begin{pmatrix} Q \\ U \end{pmatrix} = \begin{pmatrix} Q_E \\ U_E \end{pmatrix} + \begin{pmatrix} Q_B \\ U_B \end{pmatrix}.$$

CMB偏振宇称态的变换：积分变换形式

- 以上公式虽然不算太复杂，但需要借助预先算好的球谐函数空间的 E、B 模式。
- 所以我们进一步给出了纯实空间的变换公式：

$$\begin{aligned}
 a_{E,\ell m} &= \frac{a_{2,\ell m} + a_{-2,\ell m}}{2} \\
 &= \frac{1}{2} \iint \{(Q + iU)[{}_2Y_{\ell m}^*] + (Q - iU)[{}_{-2}Y_{\ell m}^*]\} d\sigma \\
 &= \iint (QF_{+,\ell m}^* + iUF_{-,\ell m}^*) d\sigma
 \end{aligned}$$

$$\begin{aligned}
 ia_{B,\ell m} &= \frac{a_{2,\ell m} - a_{-2,\ell m}}{2} \\
 &= \frac{1}{2} \iint \{(Q + iU)[{}_2Y_{\ell m}^*] - (Q - iU)[{}_{-2}Y_{\ell m}^*]\} d\sigma \\
 &= \iint (QF_{-,\ell m}^* + iUF_{+,\ell m}^*) d\sigma.
 \end{aligned}$$

$$\begin{aligned}
 Q_E(\sigma) &= \sum_{\ell m} a_{E,\ell m} F_{+,\ell m}(\sigma) = \iint d\sigma' \quad (142) \\
 &\left\{ Q(\sigma') \sum_{\ell m} [F_{+,\ell m}(\sigma) F_{+,\ell m}^*(\sigma')] + iU(\sigma') \sum_{\ell m} [F_{+,\ell m}(\sigma) F_{-,\ell m}^*(\sigma')] \right\},
 \end{aligned}$$

$$\begin{aligned}
 \sum_{\ell m} F_{+,\ell m} F_{+,\ell m}^* &= \frac{1}{4} \sum_{\ell m} \quad (143) \\
 &[{}_2Y_{\ell m}][{}_2Y_{\ell m}^*] + [{}_2Y_{\ell m}][{}_{-2}Y_{\ell m}^*] + [{}_{-2}Y_{\ell m}][{}_2Y_{\ell m}^*] + [{}_{-2}Y_{\ell m}][{}_{-2}Y_{\ell m}^*] \\
 \sum_{\ell m} F_{+,\ell m} F_{-,\ell m}^* &= \frac{1}{4} \sum_{\ell m} \\
 &[{}_2Y_{\ell m}][{}_2Y_{\ell m}^*] - [{}_2Y_{\ell m}][{}_{-2}Y_{\ell m}^*] + [{}_{-2}Y_{\ell m}][{}_2Y_{\ell m}^*] - [{}_{-2}Y_{\ell m}][{}_{-2}Y_{\ell m}^*],
 \end{aligned}$$

CMB偏振宇称态的变换：积分变换形式

$$\sum_{\ell m} F_{+, \ell m}(\sigma) F_{+, \ell m}^*(\sigma') = \quad (145)$$

$$\sum_{\ell=0}^{\ell_{\max}} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell, -2}(\beta, \alpha) + {}_2Y_{\ell, 2}(\beta, \alpha) + {}_{-2}Y_{\ell, -2}(\beta, \alpha) + {}_{-2}Y_{\ell, 2}(\beta, \alpha)}{4} =$$

$$\text{Re} \left(\sum_{\ell=0}^{\ell_{\max}} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell, -2}(\beta, \alpha) + {}_2Y_{\ell, 2}(\beta, \alpha)}{2} \right)$$

$$\sum_{\ell m} F_{+, \ell m}(\sigma) F_{-, \ell m}^*(\sigma') =$$

$$\sum_{\ell=0}^{\ell_{\max}} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell, -2}(\beta, \alpha) - {}_2Y_{\ell, 2}(\beta, \alpha) + {}_{-2}Y_{\ell, -2}(\beta, \alpha) - {}_{-2}Y_{\ell, 2}(\beta, \alpha)}{4} =$$

$$i \text{Im} \left(\sum_{\ell=0}^{\ell_{\max}} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell, -2}(\beta, \alpha) - {}_2Y_{\ell, 2}(\beta, \alpha)}{2} \right).$$

$$\mathcal{Y}_+(\sigma, \sigma') = \sum_{\ell=0}^{\ell_{\max}} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell, -2}(\beta, \alpha) + {}_2Y_{\ell, 2}(\beta, \alpha)}{2} \quad (146)$$

$$\mathcal{Y}_-(\sigma, \sigma') = \sum_{\ell=0}^{\ell_{\max}} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell, -2}(\beta, \alpha) - {}_2Y_{\ell, 2}(\beta, \alpha)}{2},$$

CMB偏振宇称态的变换：积分变换形式

- 最后就能得到一套相对简单而且纯实空间的矩阵变换（卷积）公式：

$$\begin{pmatrix} Q_E \\ U_E \end{pmatrix} = \iint \begin{pmatrix} \text{Re}(\mathcal{Y}_+) & -\text{Im}(\mathcal{Y}_-) \\ \text{Im}(\mathcal{Y}_+) & \text{Re}(\mathcal{Y}_-) \end{pmatrix} \begin{pmatrix} Q \\ U \end{pmatrix} d\sigma'.$$

$$\begin{pmatrix} Q_B \\ U_B \end{pmatrix} = \iint \begin{pmatrix} \text{Re}(\mathcal{Y}_-) & -\text{Im}(\mathcal{Y}_+) \\ \text{Im}(\mathcal{Y}_-) & \text{Re}(\mathcal{Y}_+) \end{pmatrix} \begin{pmatrix} Q \\ U \end{pmatrix} d\sigma'.$$

- 紧接着的问题就是：它还能变得更加简单吗？
 - 为什么 E、B 的计算要分开成两个式子？
 - E、B、Q、U到底是什么东西？

统一的代数形式

- E、B、Q、U正如黑白猫狗，但这是不够的。
 - 黑、白、猫、狗之间并没有什么特别的数学变换。
 - 但Q、U、E、B之间有非常紧密的数学关联。
- 能否将E、B、Q、U统一在同一个代数结构之下？
 - 形式和内容的统一。
 - 自然地回答它们之间的联系。
 - 变换公式的大幅简化。

偏振四元数的引入

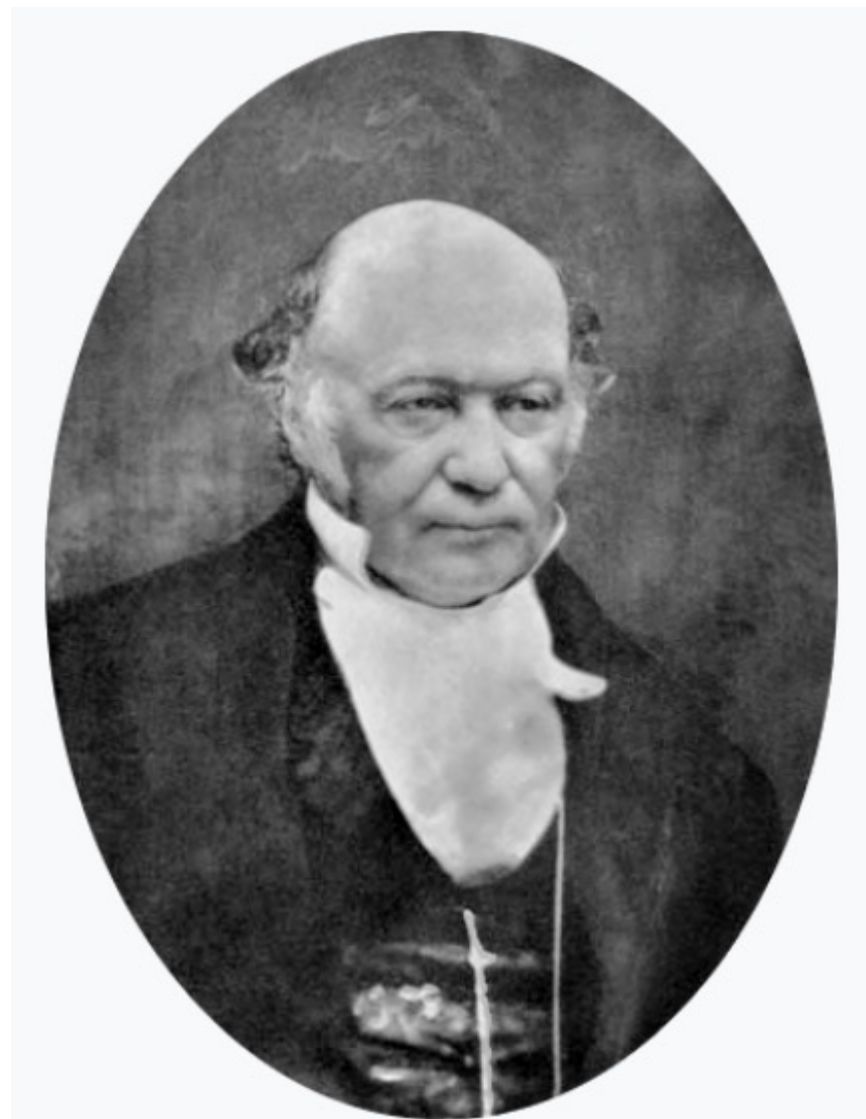
- 四元数理论的创建人：威廉·哈密顿。
- 实数 \rightarrow 复数 \rightarrow 四元数
- 最后一个结合除代数。

$$\mathbf{q} = (a, b, c, d) = a + bi + cj + dk = z_1 + z_2j,$$

$$1. \quad ii = jj = kk = -1.$$

$$2. \quad ij = k, \quad jk = i \text{ and } ki = j.$$

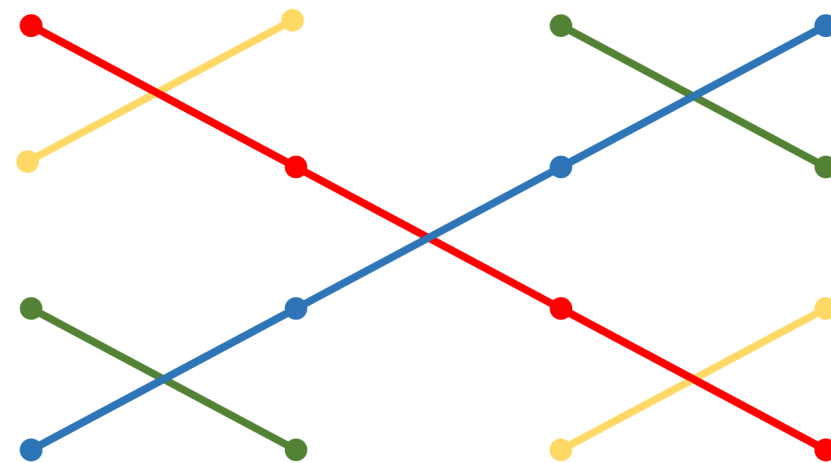
$$3. \quad ij = -ji, \quad jk = -kj \text{ and } ki = -ik.$$



偏振四元数的引入

$$\begin{aligned} \mathbf{q}_1 \mathbf{q}_2 = & (a_1 a_2 - b_1 b_2 - c_1 c_2 - d_1 d_2) + \\ & (a_1 b_2 + b_1 a_2 + c_1 d_2 - d_1 c_2) \mathbf{i} + \\ & (a_1 c_2 - b_1 d_2 + c_1 a_2 + d_1 b_2) \mathbf{j} + \\ & (a_1 d_2 + b_1 c_2 - c_1 b_2 + d_1 a_2) \mathbf{k}, \end{aligned}$$

- 红、黄、绿、蓝: a_2, b_2, c_2, d_2



偏振四元数的引入

- 接下来我们定义偏振变换中需要用到的五个四元数：

$$\mathcal{F}_+(\mathbf{n}, \mathbf{n}') = \sum_{\ell} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell,-2}(\beta, \alpha) + {}_2Y_{\ell,2}(\beta, \alpha)}{2}$$
$$\mathcal{F}_-(\mathbf{n}, \mathbf{n}') = \sum_{\ell} \sqrt{\frac{2\ell+1}{4\pi}} \frac{{}_2Y_{\ell,-2}(\beta, \alpha) - {}_2Y_{\ell,2}(\beta, \alpha)}{2}.$$

$$\mathcal{F}_{\ell m} = F_{+, \ell m} + F_{-, \ell m} \mathbf{j}$$

$$\mathcal{G} = \mathcal{F}_+ + \mathcal{F}_- \mathbf{j}$$

$$\mathcal{D} = P_E + P_B \mathbf{j}$$

$$\mathcal{A}_{\ell m} = a_{E, \ell m} + (a_{B, \ell m} \mathbf{i}) \mathbf{j} = a_{E, \ell m} + a_{B, \ell m} \mathbf{k}$$

$$\mathcal{P} = Q + (U \mathbf{i}) \mathbf{j} = Q + U \mathbf{k},$$

偏振四元数的引入

- 二元实空间到四元球谐函数空间的正反变换为：

$$\mathcal{A} = \mathcal{F}^{H_{101}} \mathcal{P}$$

$$\mathcal{P} = \mathcal{F}^{*010} \mathcal{A},$$

- 二元实空间到四元实空间的变换为： $\mathcal{D} = \mathcal{G}\mathcal{P}$,

- 四元实空间到四元球谐函数空间的变换为：

$$\mathcal{D} = \mathcal{Y}_2 \mathcal{A}$$

$$\mathcal{Y}_2 = {}_2Y_{\ell m}(\mathbf{n})$$

$$\mathcal{A} = \mathcal{Y}_2^H \mathcal{D},$$

偏振四元数的引入

- 以上变换公式具有如下特点：
 - 每个公式都只涉及一次四元数乘法。
 - 涉及 E、B、Q、U 的全部变换是通过四元数乘法同时完成的。
 - 所有变换函数都是固定的，不随球面偏振场而变。

可能的启迪

- 对球面偏振场的真正完备的描述不是 E 、 B 、 Q 、 U ，而是由它们组合成的偏振四元数 $\mathbf{D} = (Q_E, U_E, Q_B, U_B)$
- 类似的规律也存在于具有其它自旋的系统之中。即不仅限于自旋为 2。
- 我们可以借助四元数进一步从二元实空间直接变换到二元宇称空间（更好的形式对称性）：

$$\mathbf{A} = \mathcal{F}^{H_{101}} \mathcal{P} = \mathbf{a}_{E,lm} + \mathbf{a}_{B,lm} \mathbf{k},$$

$$\mathcal{Y}_0 \mathbf{A} = \mathbf{E} + \mathbf{Bk},$$

$$(\mathbf{E} + \mathbf{Bk}) = [\mathcal{Y}_0 \mathcal{F}^{H_{101}}] (\mathbf{Q} + \mathbf{Uk}).$$

可能的启迪

- 从二元实空间直接变换到二元宇称空间后，系统的本征值问题也变得简单明了：
 - 原始的自旋系统本征值问题涉及升降自旋算符，是比较麻烦的：

$$\bar{\partial}_+[{}_s Y_{\ell m}(\theta, \varphi)] = + [(\ell - s)(\ell + s + 1)]^{\frac{1}{2}} [{}_{s+1} Y_{\ell m}(\theta, \varphi)]$$

$$\bar{\partial}_-[{}_s Y_{\ell m}(\theta, \varphi)] = - [(\ell + s)(\ell - s + 1)]^{\frac{1}{2}} [{}_{s-1} Y_{\ell m}(\theta, \varphi)]$$

$$\bar{\partial}_- \bar{\partial}_+[{}_s Y_{\ell m}(\theta, \varphi)] = - [(\ell - s)(\ell + s + 1)] [{}_s Y_{\ell m}(\theta, \varphi)].$$

- 当转换到二元宇称空间后，本征值问题自然地变为大家熟知的球谐函数的本征值问题：

$$\hat{L}_z(\mathbf{E}_{\ell m}) = m \mathbf{E}_{\ell m}$$

$$\hat{L}^{(2)}(\mathbf{E}_{\ell m}) = \ell(\ell + 1) \mathbf{E}_{\ell m}$$

$$\hat{L}_z(\mathbf{kB}_{\ell m}) = m \mathbf{kB}_{\ell m}$$

$$\hat{L}^{(2)}(\mathbf{kB}_{\ell m}) = \ell(\ell + 1) \mathbf{kB}_{\ell m}$$

可能的启迪

- 我们还可以进一步讨论一个一般的线性自旋系统的本征值问题。
 - 根据以上结果，一般情况下的线性自旋系统对应的矩阵应该是一个四元数矩阵，从而其本征值问题对应于四元数矩阵的本征值分解。
 - 一个相对简单的特例是：如果这个线性自旋系统的对角化恰好是自旋球谐函数空间完成，那么它的本征值问题就是下式，其中每个对角元是系统的一个四元数本征值：

$$\mathcal{M} = \mathcal{Y}_2 \begin{pmatrix} \mathcal{A}_1 & 0 & 0 & \cdots \\ 0 & \mathcal{A}_2 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathcal{A}_n \end{pmatrix} \mathcal{Y}_2^H,$$

- 依此类推，一个更加复杂的自旋系统只要是线性的，那么就可以用四元数矩阵本征分解的方式来完善地解决其本征值问题。

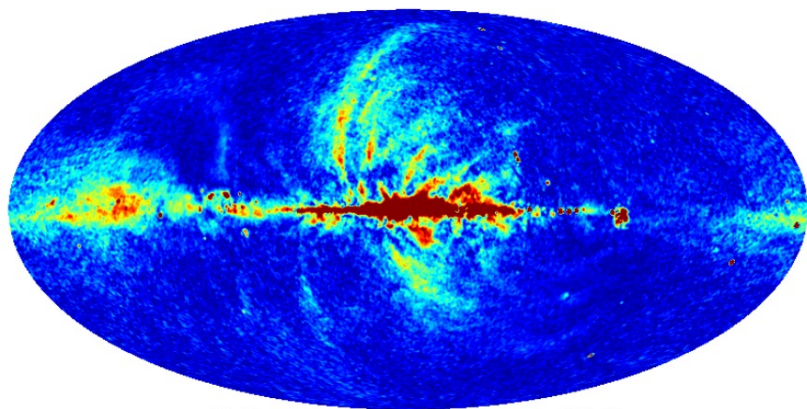
在原初引力波探测中的应用

- 偏振四元数的基本形式为 $\mathbf{D} = (Q_E, U_E, Q_B, U_B)$
- 于是在天图空间中实际存在的应该是 Q_E, U_E, Q_B, U_B ，而非简单的 Q 、 U 。
- 对噪声来说，这一分解没有太大意义。
- 如果存在特定的物理机制，这四个分量有可能呈现出完全不同的结构。这一点尤其在偏振前景的研究中会带来一些有趣的结果。

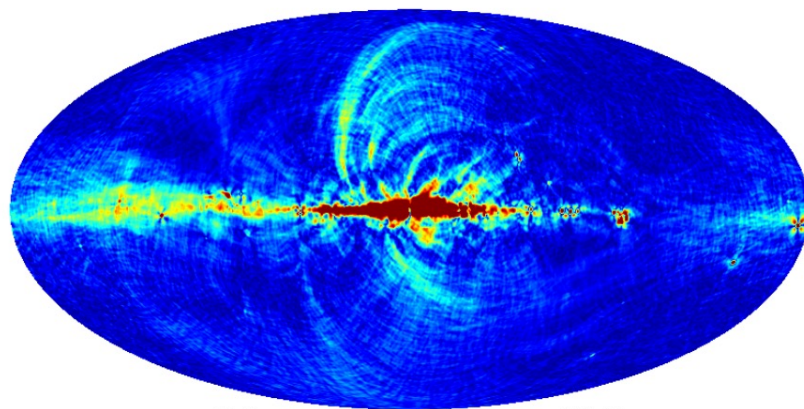
前景辐射中 E、B 模式分离的偏振信号

- 超新星爆发导致的环状辐射的实空间分解，可以看出信号主要由 E 模式主导，其物理机制已经得到了合理的解释（Hao Liu, 2018, A&A 617, A90）。

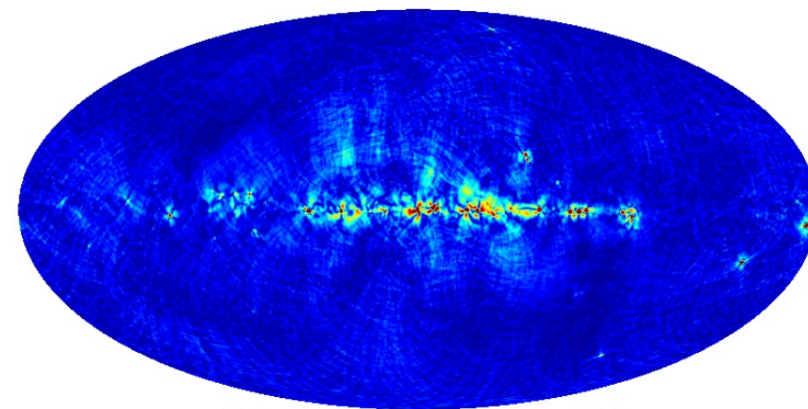
E+B



E-only



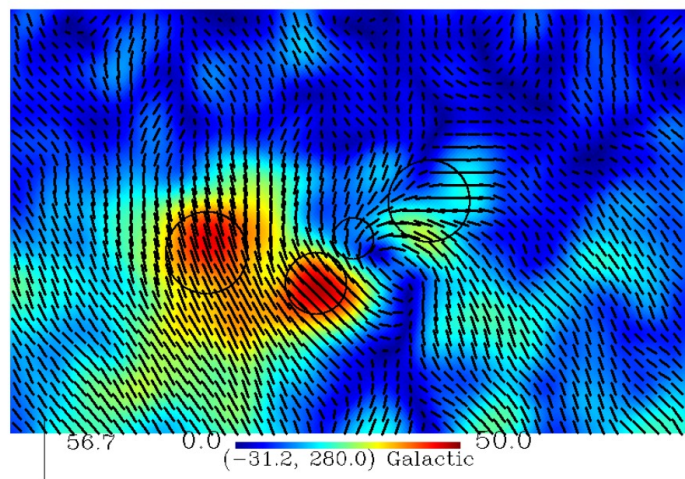
B-only



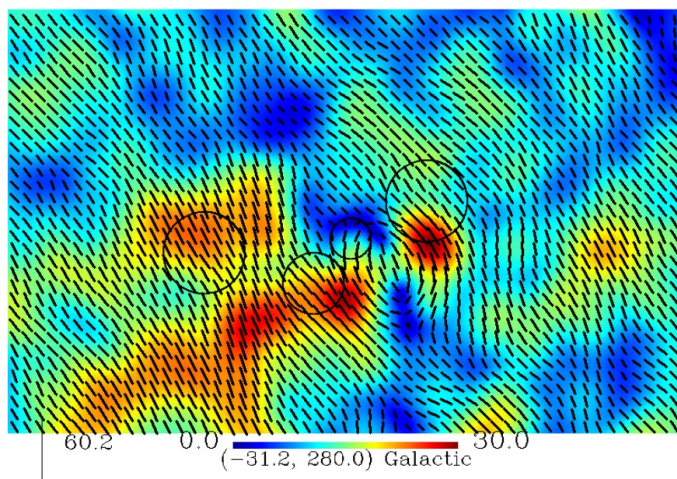
前景辐射中 E、B 模式分离的偏振信号

30 GHz

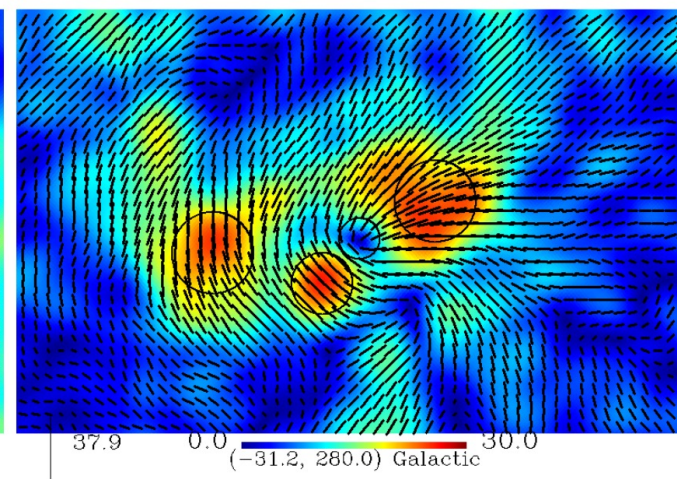
30, both families



30, only E family

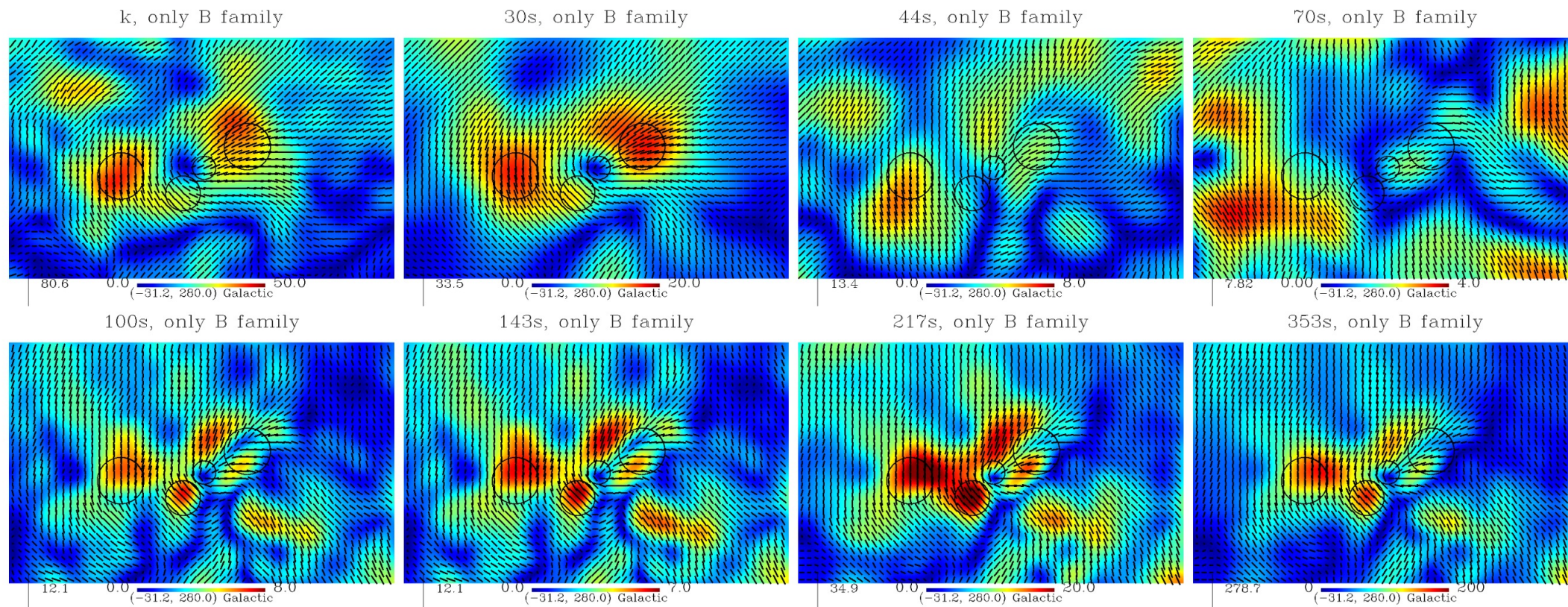


30, only B family



前景辐射中 E、B 模式分离的偏振信号

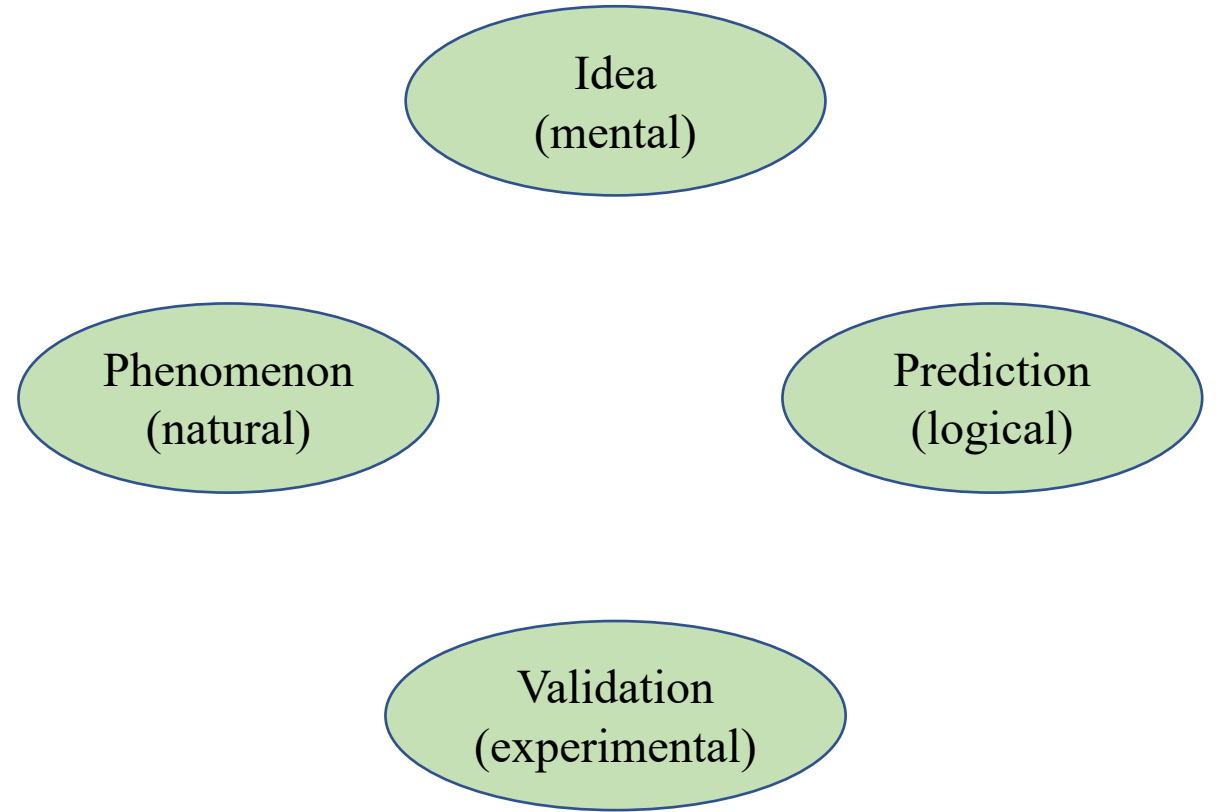
更多频段



由繁而简和由简而繁

- 以上我们在引入偏振四元数时主要考虑的是变换的形式要简单、优美、根本。
 - 删繁就简，泼墨山水
- 但是在现实中，我们往往需要考虑更加实际的效率问题：
- 这个时候我们往往会走到问题的另一个极端：
 - 通过精密复杂的设计来达成高效。
 - 精描细画，工笔仕女

A typical pattern of cognition



- The one minute principle

When we come to the CMB science...

- We are dealing with phenomenon on the sphere.
- We need to use the spherical harmonic transforms.
- CMB is only one realization of the allowed ensemble.
 - The cosmic variance.
 - The axis of evil.
 - The north-south asymmetry problem.
 - The odd-even parity problem.
 - All requires to study the ensemble, not one realization.
 - Most of them cannot be done in a pure theoretical way.
- Therefore, a **fast SHT** (spherical harmonic transform) for **a large number of sky maps** is crucial for the quality of most CMB-related work.
 - Maybe other fields of astrophysics

The main idea

The forward SHT decomposes a scalar field on a sphere, such as the CMB temperature map $T(\theta, \phi)$, into spherical harmonic coefficients $a_{\ell m}$, following the equation

$$a_{\ell m} = \sum_{\theta\phi} T(\theta, \phi) Y_{\ell m}(\theta, \phi) d\sigma, \quad (1)$$

$$a_{\ell m}^k = \sum_{\theta\phi} T^k(\theta, \phi) P_\ell^m[\cos(\theta)] e^{-im\phi} d\sigma$$

$$= \sum_{\theta\phi} T^k(\theta, \phi) P_\ell^m[\cos(\theta)] [e^{-i\frac{2\pi mn}{N}} e^{-im\phi_0(\theta)}] d\sigma$$

$$= \sum_{\theta} P_\ell^m[\cos(\theta)] \left[\sum_n T^k(\theta, n) e^{-i\frac{2\pi mn}{N}} \right] e^{-im\phi_0(\theta)} d\sigma$$

$$= \sum_{\theta} P_\ell^m[\cos(\theta)] \mathcal{T}_m^k(\theta),$$

$$(a_\ell^k)_m = (P_\ell^\theta \cdot \mathcal{T}_\theta^k)_m.$$

Definition of spherical harmonics

A ring-like sky division (pixelization)

Change the order of summation

Use FFT for acceleration (**point 1**)

Turns out to be a matrix multiplication (**point 2**)

The main idea

- Comparison with the traditional idea (**large number of sky maps**):

$$(a_\ell^k)_m = (P_\ell^\theta \cdot \mathcal{T}_\theta^k)_m.$$

- If $k = 1$ (one map at one time), this is a matrix-vector multiplication.
- It is well known that a matrix-matrix multiplication is much more efficient than a matrix-vector multiplication.

The main idea

- A similar idea can be found for the inverse-transform:

$$\begin{aligned} T^k(\theta, \phi) &= \sum_{\ell m} a_{\ell m}^k P_{\ell}^m[\cos(\theta)] e^{im\phi} \\ &= \sum_{\ell m} a_{\ell m}^k P_{\ell}^m[\cos(\theta)] e^{i\frac{2\pi mn}{N}} e^{im\phi_0(\theta)} \\ &= \sum_{m=0}^{l_{\max}} e^{i\frac{2\pi mn}{N}} \left[e^{im\phi_0(\theta)} \sum_{l=m}^{l_{\max}} a_{\ell m}^k P_{\ell}^m[\cos(\theta)] \right] \\ &= \sum_{m=0}^{l_{\max}} \mathcal{T}_m^k(\theta) e^{i\frac{2\pi mn}{N}}, \end{aligned}$$

Definition of spherical harmonics

A ring-like sky division (pixelization)

Change the order of summation (point 2)

Use inverse-FFT (point 1)

Inversed order of
MM and FFT

The main idea

- What is the most efficient way of matrix multiplication?
- GPU-computation (**point 3**).
- Therefore we have three key points:
 - FFT.
 - Matrix-vector multiplication → Matrix-matrix multiplication.
 - GPU accelerated matrix-matrix multiplication.
 - → fastSHT (<https://github.com/liuhao-cn/fastSHT>)
- These form the idea of fast SHT for a large number of sky maps.

Technical details supporting the main idea

- Real-real computation (Legendre is real). $(a_\ell^k)_m = (P_\ell^\theta \cdot \mathcal{T}_\theta^k)_m.$
- Batch FFT with full customization.
 - Full IO, size, plan, and stride control, cannot be done by e.g., python-FFT.
- Asynchronous memory copying during FFT.
- Specially designed memory scheme (compact, suitable for MM).
- OpenACC assisted FFT-mapping procedure.
- Skip FFT during iteration (FFT is lossless).
- Separated E and B computation.
- Ring optimization (Schaeffer 2013).
- Customized precision control of the Legendre polynomial.
- Q-to-EB and then U-to-EB rather than QU-to-E and then QU-to-B.
 - Saves 50% FFT-mapping operations.

Hardware and installation

- **Double-precision** GPU card (Tesla H100, A100 or V100)
 - Especially, H100 and A100 provides much better double precision performance with its tensor cores.
 - Also works for other Nvidia card but can be less efficient.
- Architecture:
 - Fortran core (cuBLAS) + python wrapper
- System requirement:
 - Ubuntu 20.04 with Nvidia GPU installed (optional).
- Quick three-line installation

```
git clone https://github.com/liuhao-cn/fastSHT.git
cd fastSHT
./configure.sh
```

Running the test code

```
Testing nside=32, lmax=31
Max relative map-T error for the alm2t test is: 1.323184388022305e-07
Max relative map-Q error for the alm2t test is: 1.885257816975172e-11
Max relative map-U error for the alm2t test is: 1.8595229765633962e-11
Max relative cl-TT error for the t2alm test is: 9.004056644864834e-07
Max relative cl-EE error for the t2alm test is: 8.166956469583631e-07
Max relative cl-BB error for the t2alm test is: 8.25748603167834e-07

Testing nside=32, lmax=32
Max relative map-T error for the alm2t test is: 1.3226391934300326e-07
Max relative map-Q error for the alm2t test is: 1.8880449584812636e-11
Max relative map-U error for the alm2t test is: 1.8516923603874962e-11
Max relative cl-TT error for the t2alm test is: 9.285166864869648e-07
Max relative cl-EE error for the t2alm test is: 8.420544533654261e-07
Max relative cl-BB error for the t2alm test is: 8.514068485278809e-07
```

- `cd ./scripts`
- `python test_comprehensive.py`
- Will test several combinations of resolution and ℓ_{max} and compare the results with Healpy

Running the benchmark code

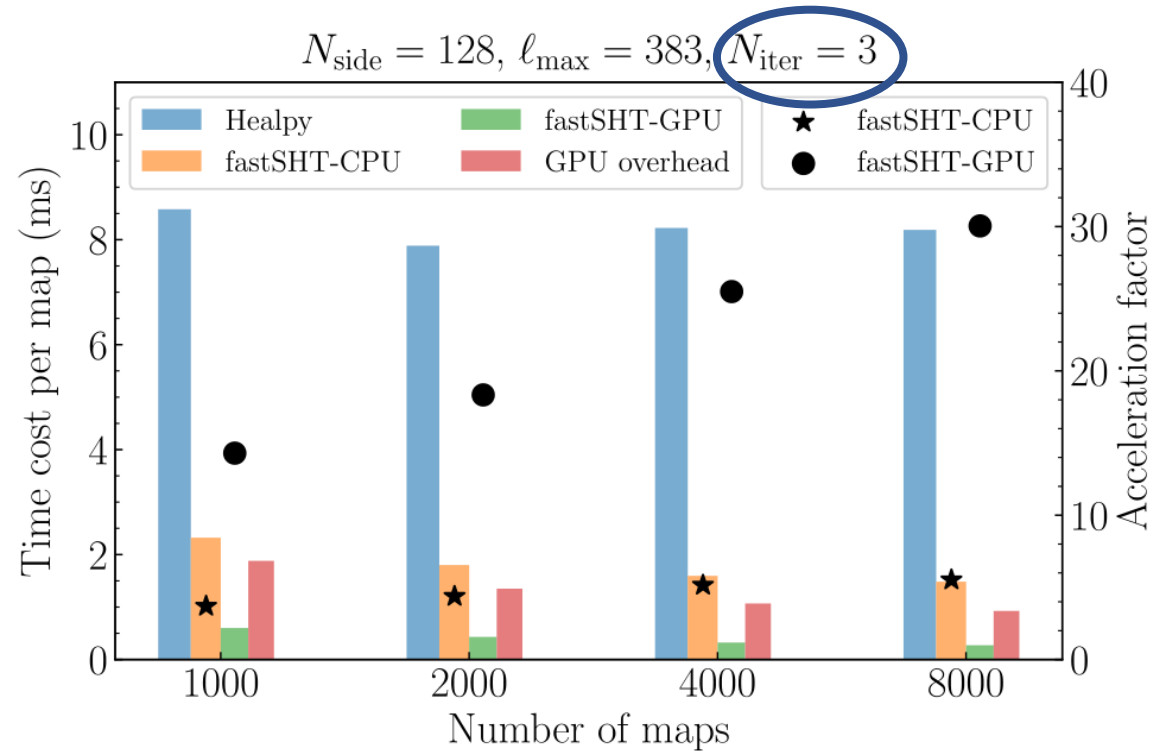
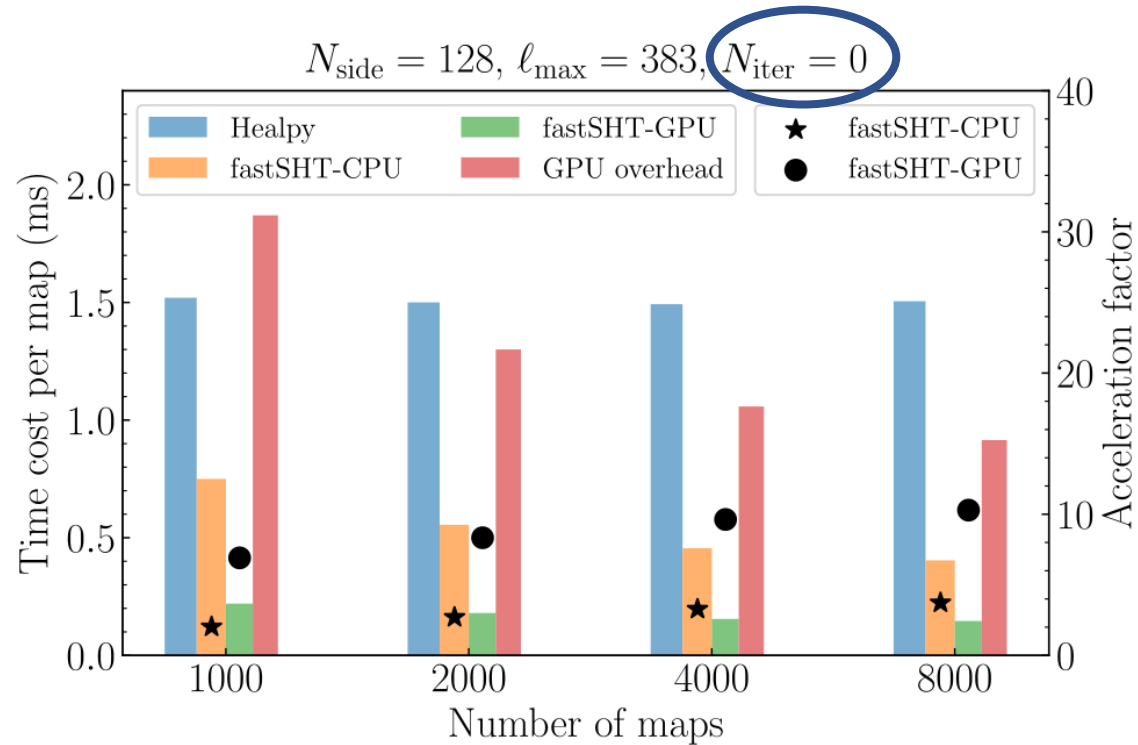
- `python benchmarks.py 128 1000 8 3 t2alm True`
- Which means to run the benchmark for:
 - `nside=128`
 - For a batch of 1000 maps
 - Using 8 CPU cores
 - Use 3 rounds of iterations
 - Test for the conversion from temperature map to the harmonic domain
 - Compare the time cost with Healpy
- Note: the results can be different on different machines

A CPU-only test

```
Working with the following parameters:
Nside = 128, Nsim = 8000, n_proc = 8, Niter = 3, comparison with HEALPix = True, test type=t2alm
Note: this test will only show the time cost. For accuracies use test_comprehensive.py

Testing t2alm...
Time cost for memory initialization is 0.01016545295715332
Calculation time cost for fastSHT is 24.552080631256104
Time cost for memory initialization is 0.02204608917236328
Calculation time cost for healpy is 71.70923089981079
```

A more comprehensive test

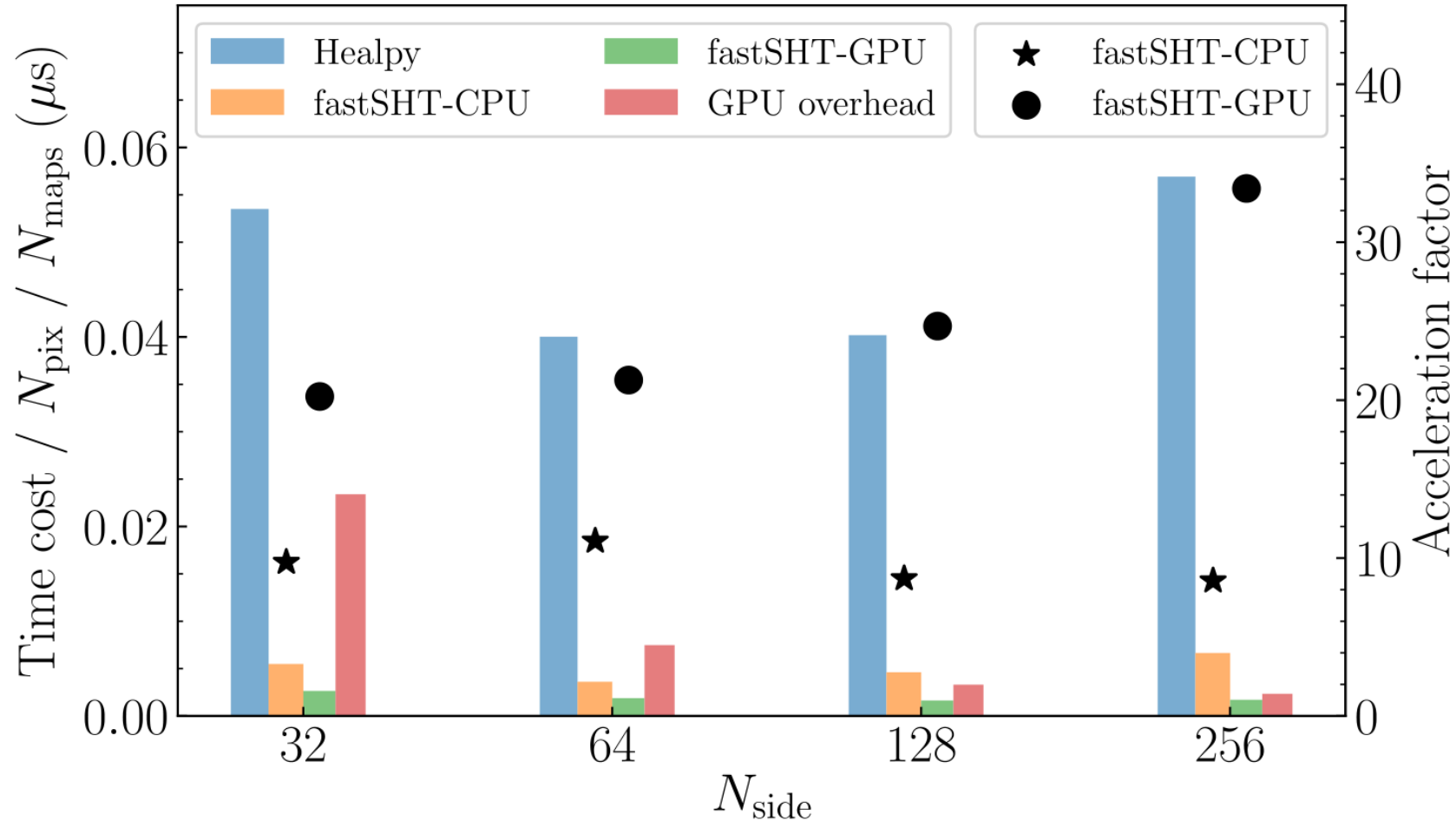


Fix the resolution, ℓ_{max} , number of iteration, only change the size of batch from 1000 to 8000 maps
The result show that:

- For CPU-only, the speed increases by 2-6 times.
- When GPU is employed, up to 30 times.
- More complicated computation brings higher acceleration factor.

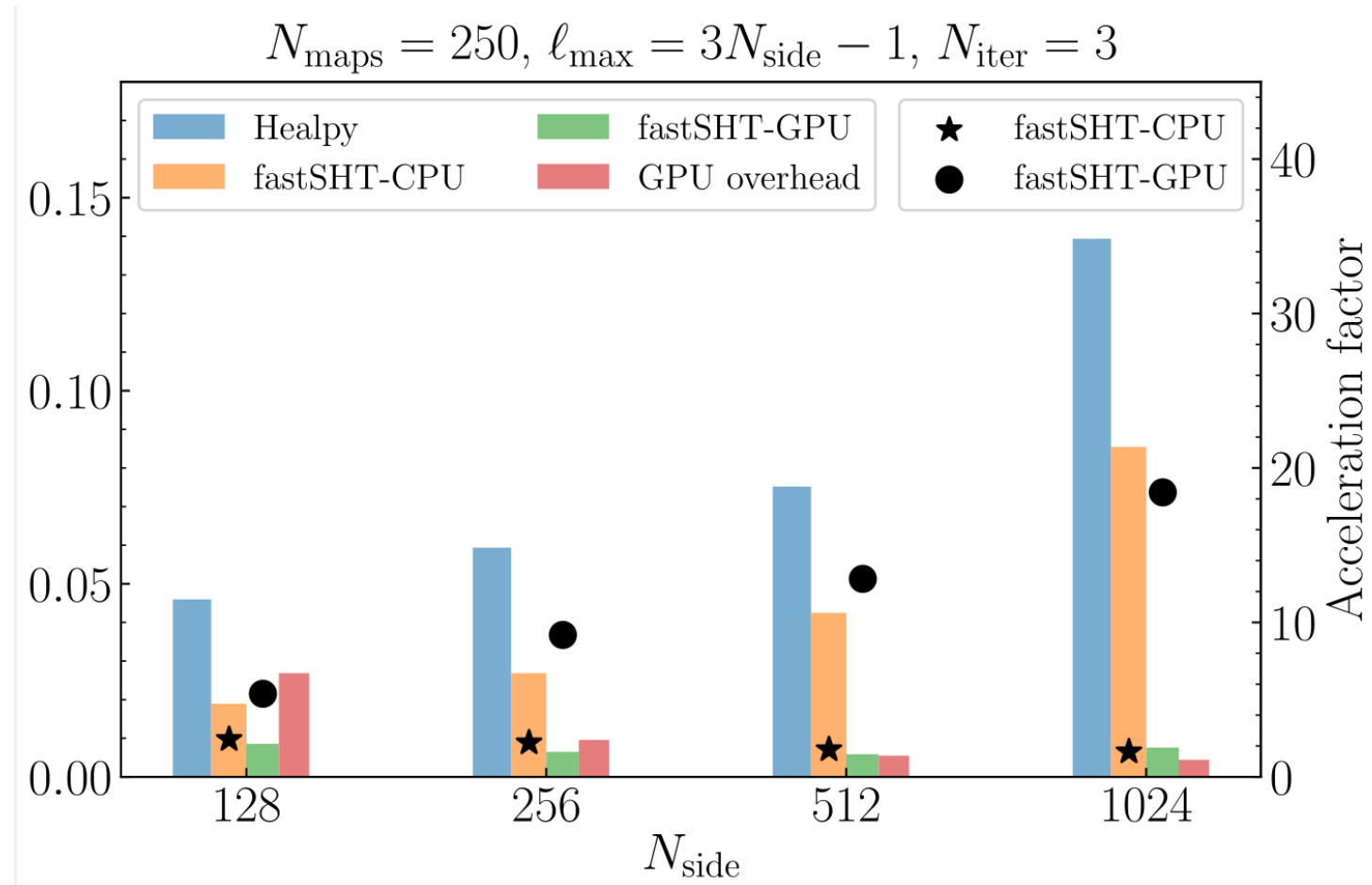
A more comprehensive test

$$N_{\text{maps}} = 4000, \ell_{\text{max}} = 3N_{\text{side}} - 1, N_{\text{iter}} = 3$$



- Change only the resolution

A more comprehensive test



- Same to above but allows higher resolutions.
- Here we start to have a problem with the GPU memory.

A more comprehensive test

Table 2

The Time Costs for *Healpy* (t_1), *fastSHT-CPU* (t_2), and *fastSHT-GPU* (t_3) in a Pipeline that Fixes the *EB* Leakage

N_{side}	N_{maps}	t_1	t_2	t_3	Overhead
256	2000	815 s	181 s (4.5 \times)	62 s (13 \times)	4.8 s
512	500	1325 s	416 s (3.2 \times)	95 s (14 \times)	4.9 s

Note. The CPU part is done with 24 cores. The GPU overhead is presented in the last column, and the acceleration factors relative to *Healpy* are given in brackets.

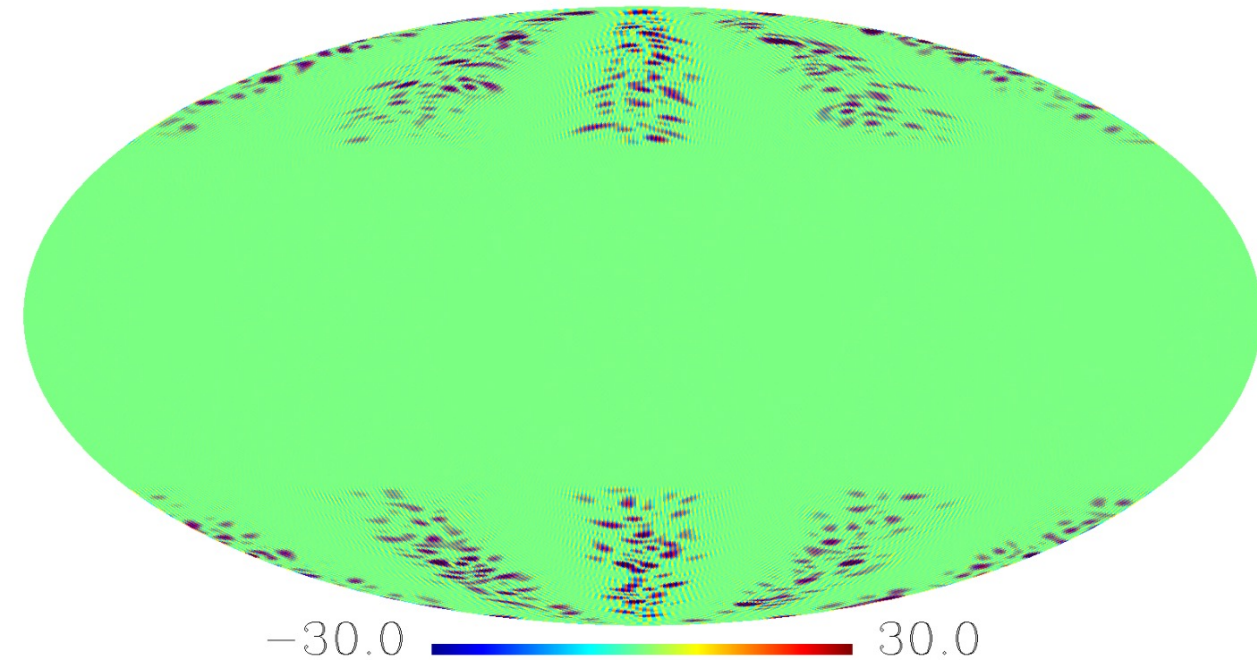
- Time cost of a relatively complicated job: fix the EB-leakage.
- With fastSHT and GPU, we are within the one-minute principle.
- And the number of maps is also typical for a CMB simulation task.

An interesting improvement of convergence

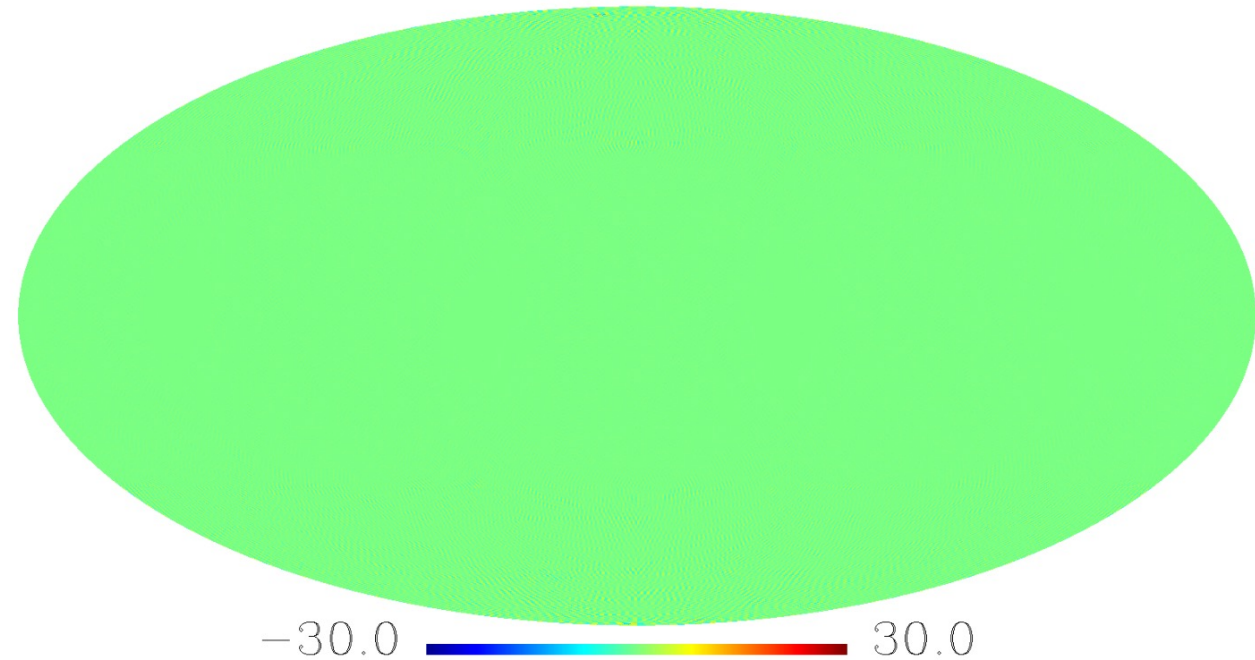
- The convergence problem in SHT due to pixelization
 - FFT: perfect orthogonality even for discrete FFT.
 - SHT: no available pixelization scheme can ensure orthogonality.
- Improvement by Jacobian iteration:
 - $T \rightarrow a_{\ell m}$
 - $a_{\ell m} \rightarrow T'$
 - Get the T-difference for iteration.
- Two possible schemes:
 - $T \rightarrow a_{\ell m}$ for all m , and then $a_{\ell m} \rightarrow T'$ for iteration.
 - $T \rightarrow a_{\ell m}$ for one m , and immediately $a_{\ell m} \rightarrow T'$ for iteration.
- A strict theoretical analysis of which one is better is not done yet

An interesting improvement of convergence

- However, we can still see the effect by tests:



The residual after 3 rounds for the traditional iteration

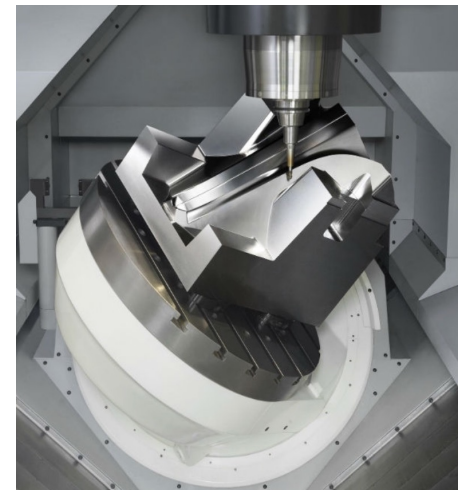


The residual after 3 rounds for the immediate iteration

The result is interesting and might be useful, but still needs more theoretical study and tests.

A few perspectives

- Different strategies for different stages of development
 - Early stage (The scale is everything)
 - Later stage (The devil is in the details)
 - Exponential increasing has to stop somewhere
- We need to use almost all functions provided by the complicated low-level interfaces
 - E.g., gemm (matrix multiplication): all 13 parameters are in use
 - Exhaustive optimization
- Like the difference between LEGO and machine tool
 - **Can we provide an interface to the "machine tool"?**





Accelerating Spherical Harmonic Transforms for a Large Number of Sky Maps

Chi Tian¹ , Siyu Li², and Hao Liu^{1,2} 

¹ School of Physics and Optoelectronics Engineering, Anhui University, 111 Jiulong Road, Hefei, Anhui, 230601, People's Republic of China; ustc_liuhao@163.com

² Key Laboratory of Particle and Astrophysics, Institute of High Energy Physics, CAS, 19B YuQuan Road, Beijing, 100049, People's Republic of China

Received 2022 August 23; revised 2022 September 16; accepted 2022 September 26; published 2022 November 16

Abstract

The spherical harmonic transform is a powerful tool in the analysis of spherical data sets, such as cosmic microwave background data. In this work, we present a new scheme for spherical harmonic transforms that supports both CPU and GPU computations, which is especially efficient on a large number of sky maps. By comparing our implementation with the standard *Libsharp-HEALPix* program, we demonstrate a 2–10 times speedup for the CPU implementation, and an up to 30 times speedup when a state-of-the-art GPU is employed. This new scheme's software package is available via an open-source GitHub repository.

Unified Astronomy Thesaurus concepts: [Cosmic microwave background radiation \(322\)](#); [Astronomy data analysis \(1858\)](#); [GPU computing \(1969\)](#)

<https://github.com/liuhao-cn/fastSHT>

Thank you very much for attention!