精益、可靠且可复现：Popcorn 空洞查找器的算法分析与重新实现

Schachner、Alvaro Roy [a,b,*]、Paz、Dante Javier [a,c]、Rico Gallego、Juan Antonio [d]、Correa、Carlos Mauricio [e]、Ruiz、Andrés Nicolás [a,c]、Gualpa、Sebastián Rogelio [a,f]、Cabral、Juan Bautista [g,f,b]

[a]理论与实验天文学研究所，*CONICET-UNC*，*Laprida 854*，*X5000BGR*科尔多瓦，阿根廷[b]数学、天文学、物理与计算学院，科尔多瓦国立大学，*Av. Medina Allende s/n*，*X5000HUA* 科尔多瓦，阿根廷[c]科尔多瓦天文台，科尔多瓦国立大学，*Laprida 854*，*X5000BGR* 科尔多瓦，阿根廷[d]*COMPUTAEX*，埃斯特雷马杜拉超级计算中心，*Ctra Nacional521*，*km 41.8*，*10071* 卡塞雷斯，西班牙[e]马克斯·普朗克地外物理研究所，*Postfach 1312*，*Giessenbachstr*，*D-85748* 加兴，德国[f]国家科学与技术研究委员会（*CONICET*），*Godoy Cruz 2290*，*C1425FQB* 布宜诺斯艾利斯自治市，阿根廷[g]技术转移管理处，国家空间活动委员会（*CONAE*），卡涅特山麓，科尔多瓦，阿根廷

arXiv:2611.28882v1 [astro-ph.IM] 21 Nov 2026

摘要

宇宙空洞是宇宙中广袤的低密度区域，蕴含着有关其大尺度结构和演化的宝贵信息。通过将球形空洞扩展为重叠球体结构，Popcorn 空洞查找器改进了对宇宙空洞族群的统计表征。然而，其源于研究原型的背景逐渐阻碍了其可维护性和可扩展性。此外，发表后的算法修改尚未得到系统评估，其统计和计算后果也因此未得到充分表征。在本工作中，我们对 Popcorn 空洞查找器进行了全面的重新实现，优先考虑算法透明度、可维护性和可复现性。我们记录了完整的识别流程，重新审视了原始算法选择，并评估了这些选择带来的统计和计算后果。不同实现之间的比较揭示了空洞大小函数行为的显著变化：该函数如今表现为从幂律区间过渡到类谢克特分布。这些变化不仅更好地表征了空洞尺寸，还使我们能够采用一种更新颖且更稳健的方法来确定散粒噪声主导区间，同时避免依赖先前工作中隐含的内部查找器参数。性能基准测试表明，球形空洞识别的速度提升了 1.62× 倍，重叠移除吞吐量也得到显著改善。我们确定 Popcorn算法受到负载不均衡的负面影响，并在一项测试配置中发现，采用新引入的提前终止增长准则后，速度提升了 63× 倍。这些结果为爆米花空洞识别奠定了更易维护且更具可复现性的基础，同时阐明了制约其性能的统计和计算权衡。

程序概要

程序标题： Popcorn 空洞查找器

*CPC*库程序文件链接：（由技术编辑添加）开发者代码仓库链接：

`https://gitlab.com/void-research-labs/popcorn-void-finder`许可条款：BSD三条款许

可证编程语言：C++外部例程/库：OpenMP、MPI、HDF5（可选）

问题性质： 宇宙空洞边界无法仅由可观测量唯一界定，因此其识别需要明确的模型。球形空洞寻找器提供了一种广泛使用的方法，可对空洞群体进行统计表征，但其几何假设也带来了重要限制。尤其是，当复杂的低密度区域受到周围高密度结构的约束时，可能会被分割成多个球形空洞。这种分割可能扭曲推断出的空洞尺度分布和空洞丰度，同时丢失有关底层低密度区域几何结构与连通性的信息。这些限制推动了空洞识别方法的发展：在保留球形方法计算和统计优势的同时，为复杂空洞形态提供更加灵活的表示。

求解方法：Popcorn 空洞查找器将球形空洞推广为由任意数量的重叠球体并集表示的结构，因此无需设定固定的全局几何结构，便能描述复杂且广延的低密度区域。识别过程从球形空洞目录开始，这些空洞作为爆米花结构的初始种子。随后，通过识别并添加满足空洞低密度判据、且能为该结构带来最大额外体积的球体，逐步扩展每个种子。

这一过程会逐步构建更复杂的配置，直到无法再纳入合适的球体。最终形成的结构保留了球形空洞识别的统计基础，同时能够灵活地表示非球形空洞形态。

关键词：宇宙空洞、宇宙大尺度结构、空洞识别算法、计算宇宙学、高性能计算、科学软件

1. 引言

宇宙空洞是空间中广袤的低密度区域，呈现为星系丝状结构之间巨大且近乎空旷的空间，构成了宇宙中可观测到的最大尺度结构 [例如文献 1、2 及其中的参考文献]。人们利用其统计和几何性质来约束宇宙学参数、研究结构增长、检验替代引力理论，并探究暗能量的本质[3, 4, 5, 6, 7, 8]。此外，空洞环境较少受到若干重子过程的影响 [9] ，而这些过程会使高密度区域的解读变得复杂，因此空洞环境对精密宇宙学分析尤其具有吸引力 [10, 11, 12]。

从宇宙空洞种群中推导出的若干统计观测量，已被证明是现代宇宙学中颇具价值的工具。其中最相关的例子包括空洞尺度函数（空洞尺度函数）和空洞-星系关联函数（*VGCF*）。空洞尺度函数 描述空洞丰度随其有效半径变化的函数，与用于暗物质晕的质量函数十分相似。其形状和演化对底层宇宙学模型、物质涨落幅度以及大尺度结构的增长历史都很敏感[1, 11, 13]。类似地，*VGCF* 刻画了空洞中心周围星系的统计分布，提供了关于低密度区域周围密度和速度剖面的信息。由于对红移空间畸变和大尺度速度流敏感，这一观测量在观测宇宙学中变得尤为重要 [14, 15, 16, 12]。通过 Alcock-Paczynski检验和红移空间畸变测量，空洞尺度函数 和 *VGCF* 已被用于约束宇宙的膨胀历史和宇宙结构的增长率，并检验替代引力模型与暗能量情景 [14, 3, 5, 6, 8]。

综上，这些应用凸显了构建稳健且可靠的空洞目录的重要性，从而能够准确刻画宇宙网中的低密度成分。然而，与晕或星系团等高密度结构不同，空洞并不存在唯一界定的边界，也没有得到普遍认可的物理定义。因此，不同的空洞搜寻方法往往会根据所采用的示踪体分布、密度阈值、拓扑判据或几何假设，识别出不同的结构 [17, 18, 19]。这种内在的模糊性促成了各种空洞识别算法的发展，而每种算法都针对不同的科学目标和观测条件进行了优化 [20, 21, 22, 10]。

一种广泛使用的方法基于对物质密度场 [20, 21, 22] 的分析。通过借助分水岭变换将密度场类比为地形景观，这些方法将空洞定义为随着水位逐步上升而被淹没的盆地。这类空洞识别算法的优点在于，无需预设形状，就能自然地还原宇宙空洞的复杂形态，同时还能识别较大低密度区域中的预定义的形状，同时还允许识别较大低密度区域中的层级子结构。

另一种方法是直接搜寻积分密度对比度低于预设阈值的区域 [23, 24, 25, 26, 27]。这类方法定义一种模板形状（通常为球形或近似球形的几何形状），并在物质或示踪体密度场上进行卷积，直至所包围的区域满足低密度判据。这种几何描述的简洁性能够更直接地控制所识别空洞的密度属性，同时也便于解释其有效大小和体积。由于积分密度对比度是在大尺度上定义的，并且与空洞丰度理论密切相关 [13, 1], ，因此这类空洞搜寻器无需任何过滤判据 [28]。然而，使用预定义的几何模板可能会限制这些方法准确恢复真实宇宙空洞高度不规则形态的能力。近期的方法通过允许将球形区域合并成更大的非球形结构，解决了这一局限。较早的例子包括 Popcorn 空洞查找器，它通过迭代扩展并合并低密度球体，构建能够共同满足低密度判据的自由形态空洞区域 [28], ；以及近期开发的 VERSUS，它采用合并判据来识别嵌套的类空洞结构 [29]。特别是，Popcorn 空洞查找器为球形空洞搜寻算法提供了几何扩展，能够在保持计算简便性和几何可解释性的同时，恢复高度非球形的空洞区域。

宇宙学模拟和观测数据集的规模与复杂性不断增加，为天文学带来了显著的计算挑战 [30, 31]。现代分析需要处理规模越来越大的示踪体目录和分辨率越来越高的密度场，因此计算性能、可扩展性和内存使用效率成为空洞识别方法实际应用中的关键因素。与此同时，科学软件日益复杂，这凸显了构建可维护、可扩展代码库的重要性，使其能够支持长期开发、可复现性以及新分析策略的整合 [32, 33]。然而，许多空洞查找算法的原始介绍往往只对这些方面略作讨论。

从这个意义上说，Popcorn 空洞查找器的初始实现更像是一个软件原型 [34], ，旨在展示将 Popcorn 爆米花空洞作为球形空洞几何扩展的可行性。尽管这种方法成功奠定了基础，但后续开发产生的代码库包含过时的实现、实验遗留内容，而且错误处理也很少。由于缺乏标准化的使用流程和全面的用户文档，很难确定配置和运行该软件的预期方式。本文介绍了 Popcorn 空洞查找器的完整重新实现，重点关注现代软件设计、计算性能、并行执行和算法可扩展性。我们特别详细介绍了其算法结构和实现策略，从而比以往更清晰、更可复现地阐述该方法。

*通讯作者。电子邮件地址：alvaro.schachner@unc.edu.ar

该框架旨在促进未来的开发、比较研究和大规模宇宙学分析，同时保持灵活且易于维护的软件架构。

本文结构如下：第 2 节正式介绍 Popcorn 空洞查找器，并描述完整的爆米花空洞识别过程。第 3 节重点讨论计算积分密度对比时面临的挑战。第 4 节介绍一系列启发式优化，这些优化采用宽松的约束，专为我们的特定宇宙学情景而设计。接下来，第 5 节和第 6 节分别介绍 Popcorn 空洞查找器的软件架构，着重说明其并行模型和可复现性，并通过与原始软件进行定量比较来评估本研究的贡献。第 7 节总结全文，并展望未来工作。

2. 大尺度结构中的 Popcorn 空洞

球形空洞方法假设，通常可以将空洞近似为低密度的球形区域。由于其属性简单明确，这类方法已成功用于研究空洞群体的统计属性 [14, 25, 35, 36, 4, 10] ，以及星系在这些极低密度环境中的演化 [26, 37, 38, 9, 39]。然而，其主要局限恰恰在于球对称性这一假设，因为它无法捕捉真实宇宙空洞中复杂且不规则的形态。因此，球形方法可能会将延展的低密度区域分割成多个较小的空洞，或无法完整描述其几何结构 [28]。

Popcorn 空洞查找器（以下简称*PVF* ）在球形方法的基础上，放宽了对固定几何形状的假设。该算法不再将空洞限定为单个球体，而是通过对联合体积施加相同的欠密度判据，将空洞定义为相互重叠的球形区域的连通并集。这种构造使空洞能够迭代生长并适应底层物质分布，从而有效形成任意形状的结构，同时保留明确定义的基于密度的选择判据。从这个意义上说，爆米花空洞可视为球形空洞的自然推广。

识别爆米花空洞本质上是一个多步骤过程，结合了物理假设、几何约束和算法设计选择。我们不将该方法描述为一个单一的整体流程，而是将其分解为主要组成部分，并说明各阶段在完整流程中所发挥的作用。整个过程由两个核心算法构成。第一个算法识别球形空洞，为后续的爆米花空洞增长阶段提供初始种子；第二个算法则尝试通过迭代添加球形成员，将这些种子扩展为扩展结构，同时保持积分密度对比度准则。两个阶段均由同一组参数控制：最小和最大搜索半径，$[R_{\min}$、$R_{\max}]$，以及空洞的积分密度对比阈值 Δ_v。

2.1. 球形空洞算法

球形空洞查找器通过将连续体积离散为规则立方网格来运作。除了能够使用最近邻搜索算法加速计算外，每个单元还将作为初始候选区域，潜在的球形空洞将由此形成。该流程概述于算法 1：对每个单元分别应用三步流程，每个候选区域至多生成一个空洞。

算法1 球形空洞查找过程：将连续体积分割为候选单元，并尝试在每个单元中拟合一个低密度球体。

要求： 搜索范围 $[R_{\min},R_{\max}]$，密度阈值 Δ_v
确保： 球形空洞
构建网格G
对于每个网格单元$g_c \in G$ **执行**
  估算候选半径 r_i
  当$r_i \geq R_{\min}$ **时执行**
    在C 中生成g_c内的局部中心
    缩小搜索范围 $[R_{\inf}, R_{\sup}] \leftarrow$ 界定(R_i)
    空洞 $\leftarrow \{$ 对$c_i \in C$ 中的每个点，执行 FindVoid$(c_i, R_{\inf}, R^{\sup}, \Delta_v)\}$
    如果 找到有效空洞**则**
      接受最大值 v。半径
      v∈空洞
      断开
    结束 if
    减小候选半径 R_i
  结束 while
结束循环

候选半径估计：在执行详细搜索之前，先找出当前单元可能容纳的最大空洞半径。具体做法是在单元中心处使用离散同心球，快速计算密度的下界近似值。搜索从$R_{\max}$开始，以$R_{\max}/N^{\text{step}}$为步长递减，直至低于$R_{\min}$ ；其中N^{step}为离散化步数。

$$N_{\text{step}} = \left\lceil \frac{R_{\max}}{R_{\min}} \right\rceil. \tag{1}$$

对于固定的离散半径r_i，可根据位于单元内所有可能球心范围内的特定示踪粒子子集，得到质量$M_{\inf}$ 的下界；无需直接进行距离计算即可快速确定该子集。然后，计算得到的密度下限为：

$$\rho_{\inf} = \frac{3M_{\inf}}{4\pi r_i^3}. \tag{2}$$

如果在整个过程中都未满足密度阈值，则舍弃该单元；否则，进入下一阶段，并记录r_i作为潜在球形空洞的初始尺寸估计。

局部优化：此时，我们可以开始对空洞进行详细搜索。利用上一步得到的离散半径

（及其后继值），我们可以推导出一个缩小的范围 $[R_{\text{inf}}、R_{\text{sup}}]$，并将精细搜索限制在该范围内。

该过程首先在单元的体积内生成固定数量的随机中心，然后对每个中心执行有界搜索过程：对于每个中心 $\vec{c}$，扫描位于距离 $\vec{t}$ 不超过 $[R_{\text{inf}}、R_{\text{sup}}]$的所有示踪体，计算以 $\vec{c}$ 为中心、半径为$\|\vec{t} - \vec{c}\|$的精确球形密度。记录满足密度阈值 Δ_v的最大距离。总体而言，这将得到一组以各个随机中心为根的潜在空洞。

渐进式半径回溯：如果上一步找到了有效空洞的子集，则接受其中半径最大的空洞，并停止该单元的过程；否则，减小初始离散半径r_i，然后重试局部优化。如果上限半径低于$R_{\min}$，则停止该过程且不接受任何空洞。

2.2. 重叠球形空洞

由于球形空洞查找器会独立应用于每个候选单元，相邻候选区域经常会收敛到几乎相同的球形空洞，而这些空洞对应于同一个欠密区域。因此，原始输出中包含大量高度重叠的球体，需要专门的后处理阶段来消除这些冗余。

需要注意的是，非重叠球体配置对冲突消解的顺序十分敏感。在我们的情形中，为了保留代表每个空洞的最大球体，必须严格按照半径递减的顺序进行冲突消解。每次迭代时，选取当前尚存的球体作为主要候选项，识别与其重叠的邻球，并有选择地移除相交的球体以解决冲突。由此，我们推导出以下不变量：任何存活到轮到自己处理的球体，必定是其重叠集合中最大的；否则，它早已在之前的迭代中被移除。因此，无须进行大小比较，邻域搜索时只需考虑球心距离小于 $2R_v$ 的球体（其中 R_v 是当前候选区域的半径）。被丢弃的球体不再参与后续的重叠消解。

总体而言，这一过程确保当多个相互重叠的候选区域代表同一片区域时，会优先保留最大的球形空洞。然而，这一保证并不适用于一般的空洞。实际上，物理空洞会被分解为彼此不重叠的球体；由于其复杂几何形状无法用球形来描述，这些球体无法充分生长到足够大。PVF 正是为了解决这些问题而设计的。

2.3. Popcorn 空洞扩展算法

生成球形空洞目录后，即可开始 Popcorn 构建过程。该生长过程总结于算法 2，并对集合中的每个空洞分别独立应用。算法不再搜索任意的三维边界，而是将多个球形成员的并集构建为一个爆米花空洞。在每次迭代中，它会识别局部在保留积分密度阈值的同时，可与现有结构连接的最佳球体。此过程如下：

算法 2 Popcorn 空洞识别：从一组球形空洞开始，对每个球体应用迭代扩展过程，将它们转换为一组爆米花空洞。

要求： 球形空洞 S
确保： 爆米花空洞

对每个根球$s \in S$ **执行**
  初始化爆米花空洞种群 Pop ← {s}
  重复
    用种子 ← FibonacciSeed(Pop) 填充种子表面
    当种子不为空**执行**
      候选集合 ← {inflate(种群,种子) for 种子 ∈ 种子}
      最佳候选项 ← $\max_{c\in\text{候选集合}} c$.体积
      如果 最佳候选项过小**则**
        中断 ▷ 停止爆米花形成
        结束条件判断
      种群.append(最佳候选项)
      种子.remove(最佳候选项.seed)
    结束循环
  直到 未添加任何球体
  接受种群
结束循环

爆米花空洞初始化：每个爆米花空洞都以一个球形成员作为初始成员，该成员对应于根球。在此阶段，爆米花结构等同于一个球形空洞，并且在构建时已满足密度阈值。

表面采样：为扩展空洞，算法首先在当前爆米花结构的边界上识别一组候选生长位置。这些位置称为种子，代表可供新球形成员生长的潜在中心。

恰当分布种子至关重要，既要确保探索所有可能的生长方向，又要避免空洞表面出现偏好性取向或采样不足的区域。为此，使用斐波那契格对每个球形成员的表面进行采样。斐波那契格是一种广泛存在于自然界的图案 [，例如向日葵种子、菠萝鳞片；见40, 41]。斐波那契格点能够在球体表面近乎均匀地分布，并实现最佳堆积效率。此外，该方法允许任意选择点的总数，因此采样密度可随球体半径自然缩放，同时保持近乎恒定的空间分辨率。

这些特性使斐波那契格点特别适用于迭代生长阶段，因为此阶段要求对爆米花边界进行各向同性且均匀的覆盖。

卫星的局部生长：对于每个表面种子点，尝试围绕种子位置扩大一个低密度球体，以此添加一个新的球形成员。此过程与球形空洞 Finder（不重新居中）类似，分为两个步骤：

1. 初步筛选阶段会估算以种子为中心的球体是否能够满足联合体积的积分密度对比阈值。2. 若满足条件，则通过有界细化步骤确定此类球体的最大允许半径。

在此过程中，密度评估会考虑爆米花结构中已经包含的示踪粒子，从而确保球体并集保持欠密。每次成功的 inflate 都会生成一个半径、体积和密度对比均定义明确的候选球形成员。

在所有膨胀候选体中，我们选择能够为爆米花结构贡献最大额外体积的球体。如果该球体的半径低于最小允许尺度，则停止生长过程。否则，将选定的球体加入爆米花空洞，并从候选集中移除相应的种子。其余种子予以保留，使算法能够在后续迭代中探索其他生长方向。

这种贪心策略优先选择最显著的几何扩展，同时保持计算效率。

迭代生长与终止：表面采样和生长过程会反复迭代，直到无法再生长出其他球形成员。当没有种子能够产生满足密度阈值的有效球体时，爆米花空洞即被视为完成。值得注意的是，如果无法再添加其他球体，则初始根球本身就构成一个有效的爆米花空洞。

2.4. 重叠的爆米花空洞

球形空洞目录中的碎片化问题会直接导致：由碎片化空洞生长而来的爆米花空洞会产生几乎相同的重叠结构。这种情况是预期之中的，否则就意味着扩张过程未能正确捕捉空洞的几何结构。因此，很容易检测并移除冗余空洞，因为它们通常具有相对较大的重叠率。

清理流程的第一步是识别所有重叠的爆米花空洞对。如果第一个结构中的至少一个球形成员与第二个结构中的一个球形成员发生重叠，则认为这两个爆米花空洞彼此重叠。对于每一对重叠的空洞，我们计算由两个爆米花空洞的所有球形成员构成的并集体积。随后，根据容斥关系得到交集体积

$$V(A \cap B) = V(A) + V(B) - V(A \cup B),$$

其中，A 和B表示两个爆米花空洞。之后，逐一处理每一对空洞。如果该空洞对中的空洞此前均未被移除，且重叠体积大于最小球体，即$V(A \cap B) > (4/3)\pi R_{min}^3$，则认为这两个空洞“等效”，并应用筛选标准。

筛选标准的直接选择是保留体积最大的空洞，这与球形空洞查找器所采用的贪心策略相一致。相反，*PVF* 保留根球半径最大的空洞。这一选择源于这样的观察：较大的根球通常位于低密度结构的中心区域，而较小的根球则更可能代表同一空洞的外围碎片。因此，从较大根球生长而来的爆米花空洞往往能够捕捉低密度区域的整体结构，而从较小根球生长而来的空洞则主要从外向内复现该结构。这种行为也促成了第4.1小节所讨论的提前终止生长准则。

到目前为止，我们已经介绍了如何从物质分布的示踪粒子集中构建爆米花空洞，其中支配空洞识别的核心标准是候选区域内的积分密度对比度。在接下来的章节中，我们将把重点转向这一标准，以及超大型目录和复杂几何结构所带来的挑战。

3. 积分密度对比度的计算

在研究宇宙大尺度结构时，物质分布通常通过密度场来刻画，其中我们将密度对比度定义为物质密度 $\delta(\mathbf{x})$ [1] 相对于平均宇宙密度 $\bar{\rho}$ 的局部偏差：

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}} - 1\,. \tag{3}$$

在这一框架下，超密区域对应于 $\delta(\mathbf{x}) > 0$ 的位置，而低密度区域满足 $\delta(\mathbf{x}) <0$。宇宙空洞表现为延展的连通区域，其物质含量显著低于宇宙平均密度，代表宇宙网中密度最低的环境。

由于大尺度结构源于引力不稳定性导致的初始微小密度扰动，给定区域内的积分密度决定了该区域相对于膨胀宇宙背景度规 [42] 是膨胀还是坍缩。空洞是预计会膨胀并随时间推移变得更加低密度的低密度区域，因此，空洞通常通过有限体积V上的积分（平均）密度对比度来刻画。从形式上看，当一个区域的包围密度对比度 Δ 低于预设阈值 Δ_v 时，可以将其视为空洞样区域：

$$\Delta = \frac{1}{V}\int_V \delta(\mathbf{x})\, d^3x = \frac{M/V}{\bar{\rho}} - 1 \le \Delta_v\,, \tag{4}$$

其中，M 是V内的总质量。

因此，计算候选区域的积分密度对比度可归结为计算质量和体积，这两项计算都需要特别考虑性能和几何因素。我们先从质量累积开始，这一过程高度依赖空间结构，以避免在粒子查询时遍历整个模拟。

[1]在我们的分析中，将隐含的时间依赖性视为常数。

3.1. 最近邻搜索

高效估算局部示踪粒子密度，需要将距离计算限制在有限的空间邻域内。在*PVF* 中，这些估算通过规则的空间网格加速，该网格能够高效执行最近邻搜索（*NNS*）查询。通过将搜索空间划分为立方单元，可以把距离计算限制在每个关注点周围的有限邻域内，从而避免对无法影响局部密度估计的远处示踪粒子进行不必要的计算。这一策略等同于粒子和流体动力学模拟中常用的链接单元法，可确保在有限半径内完整地进行邻居搜索 [43, 44]。

除了加速 *NNS* 查询外，同一网格还用于定义球形空洞查找器所探索的候选中心。这一双重作用引入了相互制约的设计要求：一方面，单元大小应足够大，以最大限度地提高邻居搜索的效率；另一方面，网格必须提供足够的空间分辨率，以采样大尺度结构，同时避免遗漏潜在的空洞候选体。为平衡这些约束，网格分辨率通过结合基于密度的判据和几何约束来确定：首先为单元数量选择一个初始猜测值，然后在单元过大时进行调整。因此，这一约束在大多数情况下会覆盖初始的基于密度的估计值，我们将在第 4.2 小节中重新讨论这一局限性。

为了尽量减少距离计算，我们采用了一种算法，可以快速判断相邻单元位于候选球体之外、之内，还是仅有部分包含于其中。在最佳情况下，该算法允许我们直接丢弃或累积单元中的所有对象；而对于与候选球体表面接触的单元，则退回到逐粒子计算距离的方法。其基本思想是：对于中心单元 $\vec{c}$，存在一个最小半径的球体，无论相邻单元位于 $\vec{c}$内部的何处，该球体都不可能覆盖该相邻单元。类似地，可以定义一个最大半径的球体，它总能覆盖相邻单元。图1通过一个二维示例说明了这些判据的几何含义。

在 popcorn 空洞扩展过程中，候选球体会被追加到一个包含不变量集合的现有结构中，该集合由
具有累积质量的粒子组成，而与这些候选区域的大小和位置无关。为避免重复计算这些不变量，增量 popcorn 结构会维护一个粒子向量。此外，还引入了一个掩码，以防止累积过程中同一对象被重复计算；当粒子位于不断扩展的 popcorn 结构与候选区域之间的共享体积内时，就会发生这种情况。

确定感兴趣区域所包含的质量后，我们只需知道其体积即可计算积分密度对比。对于球形区域，可根据其半径轻松确定体积。然而，爆米花空洞的体积并没有简单直接的计算方法，这正是我们接下来要解决的问题。

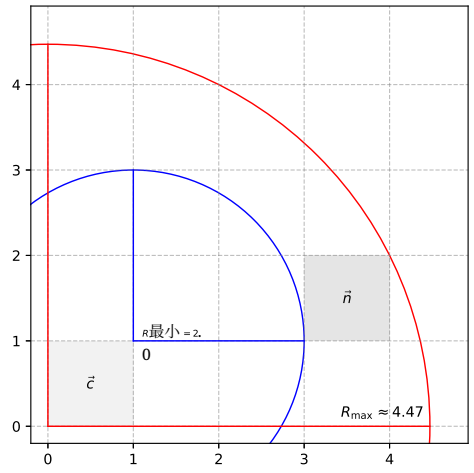

图1：邻域重叠算法的二维表示。给定一个相邻单元 $\vec{n}$；R_{min} 和 R_{max} 表示与源自中心单元的 $\vec{n}$ 接触的最近-最小和最远-最大球体。

3.2. 爆米花空洞的体积

爆米花空洞通常由许多部分重叠球体构成。因此，计算这类结构的体积需要求多个球形区域的并集体积。对于涉及两个或三个球体的简单构型，可以通过累加各球体的体积并减去重叠区域（球冠）来得到解析表达式 [45, 46]。然而，对于更大的系统，这些表达式很快就变得不切实际。对于包含许多球体的一般情况，我们采用 ARVO 软件包 [47, 48]中介绍的方法。

该技术以数值方式计算 $N > 2$ 个球体的并集体积，

$$M = \bigcup_{j=1}^{N} S_j,$$

通过将体积分转换为球面边界上的一组曲面积分，其中仅保留在计入重叠后仍暴露的球面边界。形式化地，

$$\begin{aligned} V(M) &= \iiint_{V(M)} dx\,dy\,dz \\ &= \iint_{B(M)} z\,dx\,dy \\ &= \sum_{i=1}^{N} \iint_{B_i(M)} z\,dx\,dy, \end{aligned}$$

其中，$B(M)$ 表示球体并集的边界，而$B_i(M)$ 表示球体S_i 的表面中未被任何其他球体遮挡的部分。

为评估所采用体积计算方法的数值表现，我们设计了一种由五个重叠球体组成的受控配置，使任意位置最多只有两个球体发生重叠。在此情形下，精确并集体积

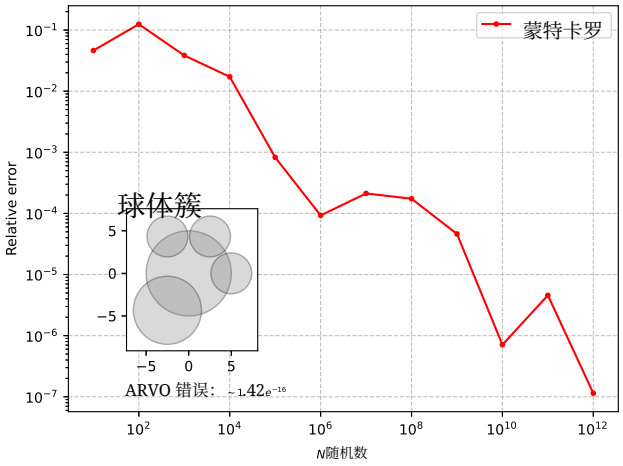

图2：左下角面板中计算由五个重叠球体组成的簇的体积时，蒙特卡罗误差与随机数数量的关系。

可以通过将各个球体的体积相加，再减去相应的成对球冠贡献来计算。该数值作为参考，用于将 ARVO 与采用 $N = 10^k$、其中 $k = 1$、2、……、12 个随机点的蒙特卡罗估计量进行比较。在图2中可以看到，蒙特卡罗估计值逐步改善，在使用 $10^{-7}N = 10^{12}$ 个随机点时，误差达到 10^{-7} 量级。另一方面，ARVO 达到了 10^{-16} 量级的相对误差。这种精度，加上算法的确定性，使 ARVO 成为计算爆米花空洞体积的有效方法。

4. 启发式优化

到目前为止，我们介绍了采用标准通用优化的基准方案。这些传统技术严格遵循数学正确性，并能对所有一般输入保持一致的可靠性。然而，宇宙学结构具有独特的统计特征，因此可以进一步提高效率。本节将介绍两种实用优化方法，它们利用目标宇宙学数据集的统计属性，在特定条件下采用一种宽松的一致性模型，以换取显著的性能提升。

4.1. 提前终止增长

随着爆米花空洞从低密度球体中生长出来，最初由多个球形空洞表示的碎片化区域，最终往往会形成数量相同的等效爆米花空洞。尽管这些冗余会在最终清理阶段消除，但在此之前，可能已经耗费大量计算来生长这些最终会被舍弃的结构。为解决这一问题，在初版论文发表和首次发布之后，开发版本对旧版代码进行了优化测试，方法是在 Popcorn 算法中实现生长提前终止。

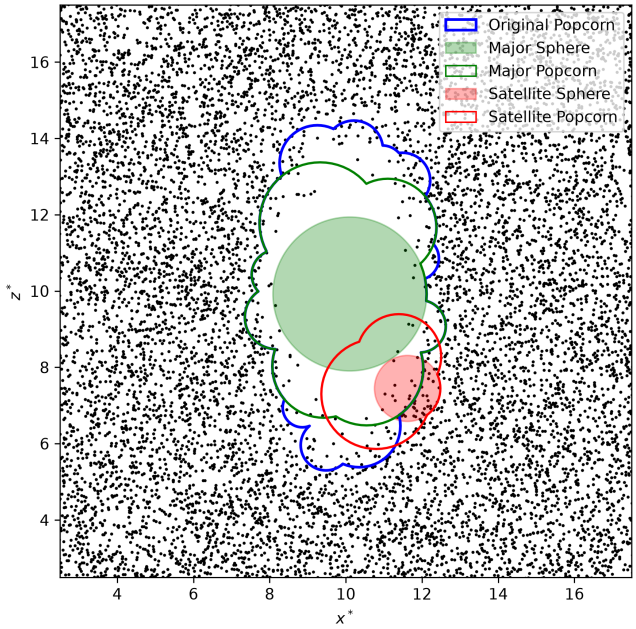

图3：合成示例中爆米花提前终止准则的示意图，原始爆米花空洞以蓝色轮廓表示。绿色表示在该区域中发现的最大（主要）球形空洞，以及采用此启发式方法得到的爆米花空洞。同样，红色表示其中一个卫星球体及其对应的爆米花空洞。

令V_n 表示添加第n个球形成员后的爆米花体积。我们定义相对体积增量：

$$\delta_n = \frac{V_n - V_{n-1}}{V_n}, \tag{5}$$

它衡量新加入球体对当前爆米花体积的相对贡献。在增长过程中，终止准则要求序列 $\{\delta_n\}$ 单调递减。每当新加入的球体带来的相对增量大于先前已接受成员所带来的增量时，就提前终止扩展。

作为第 2.4 小节所述重叠爆米花过滤流程的补充，这一判据的依据源于球形空洞查找器对低密度区域的采样方式。由于构造方式所限，球形目录由最大的互不重叠球体组成，因此，典型空洞由少数占据其中心区域的大球体，以及若干逐渐变小、填充至空洞边界的球体构成。因此，从这些中心球体之一开始扩展爆米花空洞时，随着结构逐渐适应周围几何结构，每个新加入的成员预计对总体积的贡献会逐渐减小。相反，对于较小的卫星球体，在低密度区域中逐步向中心靠近时，最初几步扩展通常会产生越来越大的成员。

我们可以在图3中观察到这一准则的实际效果：一个具有细长几何形状、难以用球形空洞表示的欠密度区域。在此示例中，球形空洞查找器

识别出七个球体，但为了便于可视化，我们重点关注其中两个：一个主球体及其一个卫星球体。在正常情况下，算法会将所有球体扩展成非常相似的延展形状，体积差最大为 4.3%。采用启发式方法后，差异变得更加显著：主球体变成了一个由十个成员组成的 Popcorn，而卫星球体仅扩展到三个成员；总体而言，体积减少了 75.9%。然而，这一变化意外地导致了回归：最终保留下来的爆米花空洞明显小于原始空洞（体积减少了 16.0%），这表明该单调准则会对清理阶段后实际存留的空洞的增长产生负面影响。

尽管空洞的尺寸有所减小，这种启发式方法仍大幅减少了构建空洞所需的工作量。为解决这一问题，我们扩展了该准则，允许严格递减的序列中出现少量“违规”，而这一序列是在旧版准则中引入的。关键在于根据根球的尺寸来定义违规次数，使该准则能够积极剔除较小的空洞，同时宽松地允许较大的空洞继续生长。总体而言，对于图3所示的合成示例，这种自适应准则能够恢复原始Popcorn算法中观察到的生长行为，同时保留提前终止带来的计算优势。不过，在某些配置下，它仍然容易导致复杂结构的生长过早终止。第 6 节将进一步评估由此产生的计算性能与统计属性之间的权衡。

4.2. NNS-候选项解耦

如前所述，*PVF* 的核心组件是空间网格，它既充当加速结构，也构成球形空洞的初始候选集。这种双重作用带来了可能相互冲突的设计要求。候选单元必须足够精细，才能对搜索空间进行采样而不错过潜在空洞；同时又必须足够粗糙，以避免生成冗余候选项。相比之下，*NNS* 单元的尺寸应当通过平衡每个单元中的示踪粒子数量与局部搜索优化的效果，来最大限度地提高查询效率。为解决这一潜在问题，我们探索了一种替代设计，其中

NNS 网格将独立于候选中心。在此，我们将保留原始候选分辨率，并根据输入条件搜索最优的 *NNS*网格参数。

NNS 分辨率可以通过每个单元的目标平均粒子数确定，记为 N_0。该参数与每个维度上的网格单元数有关，关系如下：

$$N_{\text{grid}} \approx \text{ceil}\left[\left(\frac{N_{\text{part}}}{N_0}\right)^{1/3}\right], \tag{6}$$

其中，N_{grid} 是网格每个维度上的单元数，N_{part} 是粒子总数。

选择合适的N_0 值至关重要：单元过小会增加索引开销，而单元过大则会降低邻居搜索局部性。为了确定

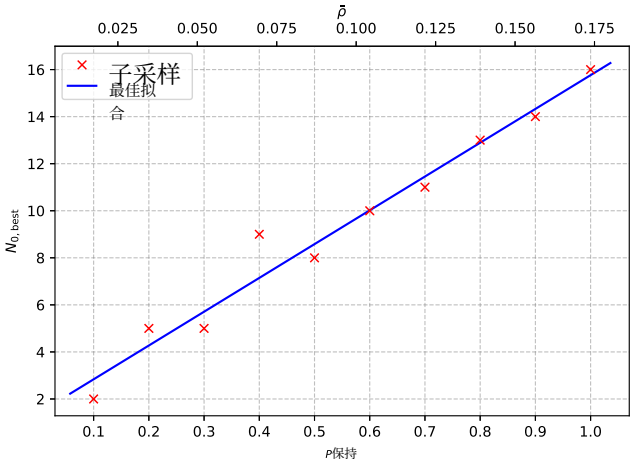

图 4：模拟随机子样本中，平均粒子密度 $\bar{\rho}$ 与 $N0,\ _{\text{best}}$ 即 P_{keep} = {1.0、0.9、...、0.1} 的粒子被剔除后，*NNS* 网格结构中最优的每个单元平均粒子数之间的关系。

这些合适的值，我们使用一个立方体积为 $1000h^{-1}\text{Mpc}^{\ 2}$、边长为 $1000h^{-1}\text{Mpc}^{\ 2}$ 且包含约 1.74 亿个粒子的模拟，对随机子样本进行了实证研究。每个样本均通过以随机概率P_{keep} = {1删除粒子生成，概率取值为 0、0.9、.....、0.1}。针对每种情况，我们评估了N_0 的多个离散取值，并选择执行时间最短的配置。所得关于平均密度 $\bar{\rho}$的实证关系如图 4 所示，可近似表示为：

$$N_0 = a\bar{\rho} + b \quad a \approx 82.656 \quad b \approx 1.4. \tag{7}$$

然而，这一经验关系的适用范围有限。当示踪粒子密度足够高时，继续提高网格分辨率最终会适得其反，因为开启和处理各个单元的成本会被单元内部的距离计算所主导。相反，在密度非常低时，最优值会低于每个单元平均一个示踪体，使N_0 不再适合作为搜索分辨率的参数化方式。因此，不应将拟合关系视为普适的准则，而应将其看作对我们模拟所探索范围的经验性刻画。

尽管存在这些局限，该实验仍表明，将候选采样和邻居搜索视为彼此独立的设计问题是有用的。特别是，实验显示，最优邻居搜索分辨率在很大程度上取决于示踪体密度，且不必与采样候选空洞中心所需的分辨率一致。因此，当前实现为实现可独立调节的空间离散化迈出了第一步，而剩余的局限则促使未来进一步以更通用的方式处理这些尺度。

[2]在天文学中，宇宙学距离通常以百万秒差距表示，其中 1 Mpc = 10^6 pc = 3.08568×10^{19} km [49]。此外，使用h^{-1}Mpc 也是一种常见做法，其中h 参数用于确定当前哈勃参数的值：$H_0 = 100h$ km s^{-1}Mpc^{-1}。这一惯例沿袭自早期的星系红移巡天 [50, 51]。

5. 软件架构

Popcorn 空洞查找器流程由三个可执行单元组成：

球形空洞查找器 (pvf.svf**)** 使用离散示踪体数据集识别球形空洞。

Popcorn 空洞查找器 (pvf.popcorn**)** 使用离散示踪体数据集将球形空洞扩展为爆米花空洞。

清理 (pvf.cleanpop**)** 识别重叠的爆米花空洞对，并移除冗余空洞。

所有可执行程序均通过单个参数文件运行，该文件指定输入示踪粒子数据集、输出空洞目录，以及球形空洞和爆米花空洞算法的查找参数。我们支持多种输入数据集格式，这些数据集来自一些最重要的宇宙学模拟 [52, 53, 54, 55]。

***PVF*格式** 包括简单的空格分隔 ASCII 格式，以及从旧版代码继承而来的打包二进制格式。

粒子目录GADGET2 和 GADGET4 二进制格式、Gadget4 HDF5 格式以及 SWIFT HDF5 格式。

群组目录GADGET4 子晕和 FoF 群组、SWIFT FoF 群组；均为 HDF5 格式。

HDF5 支持默认处于禁用状态，因为它并非查找过程的必要条件。将这些功能置于条件编译之后，即可在没有 HDF5 库的情况下构建代码。如果未启用 HDF5 支持的构建版本接收到上述格式之一的输入，程序会退出并显示错误信息，说明如何启用该功能。原始代码库中没有此功能；引入它是为了改善用户体验。因此，新代码库只需要支持 C++20 或更高版本并支持 OpenMP 的编译器，以及 MPI 库 [56, 57]。由于并行执行对于大型数据集中的识别至关重要，接下来我们将介绍这些库在*PVF*中的作用。

5.1. 并行执行模型

在空洞识别过程中，输入数据集被视为不可变数据，即在评估空洞候选区域时，不会对粒子分布进行任何修改。这使得我们能够在多个层级上采用简单的并行策略。

该实现采用混合并行化模型，将通过 MPI 实现的分布式内存并行与通过 OpenMP（下文简称 OMP）实现的共享内存多线程相结合。在最高层级，输入空间被划分为均匀区块，并分配给各个 MPI 秩。在球形空洞查找器（pvf.svf）中，这对应于候选单元的子集；而在爆米花阶段（pvf.popcorn）中，工作负载则分配到初始球形空洞上。

在每个 MPI 进程内，分配的工作负载会进一步划分给多个 OpenMP 线程，各线程独立评估各自子集中的候选空洞。这种分层方法使算法能够扩展到多个计算节点，同时在每个节点内保持高效的线程级并行。

SVF 中的清理阶段引入了全局同步约束：只有在识别过程完成、所有候选空洞都已在本地可用后，才能执行清理。

从运行时角度来看，该实现支持灵活配置 MPI 秩和 OMP 线程，从而能够在分布式和共享内存架构上实现混合并行执行。然而，当前设计继承了旧版 *PVF* 的一个重要限制：每个 MPI 秩都会维护一份完整的示踪体目录副本。虽然这种方法简化了数据访问模式，并避免了检测阶段的通信开销，但由于各秩之间存在冗余存储，也增加了内存消耗。因此，该实现倾向于采用较少数量的 MPI 秩，并在每个节点上使用更高程度的 OMP 多线程。我们有意保留了这一设计选择，以确保与 *PVF* 先前版本在方法上的连续性，并便于开展我们在第 7 节中讨论的未来实验。

5.2. 可复现性

除复现科学结果外，确定性执行在软件开发过程中也尤为重要，因为它能让人们将修改后的代码版本与已知的参考输出进行比较，从而简化调试。

在 *PVF* 中，球形空洞识别依赖随机版采样，而采样通过多层并行执行，因此随机数序列取决于运行时条件。原始实现并未处理这一问题，而是接受了这样一个事实：即使输入相同，多次运行也会产生不同的输出；由于球形空洞尺寸对所选中心高度敏感，这种差异往往十分明显。尽管这种随机性不会影响结果的科学有效性，却妨碍了代码开发过程中对结果进行清晰、无歧义的验证。因此，我们自然引入了确定性执行模式，以控制随机性来源和执行顺序，确保在固定运行时条件下获得可复现的结果。

首先，我们用 Xoshiro128* 浮点数生成器 [58], 取代了基础的 C 伪随机数生成器；该生成器以性能良好、能够高质量地生成通用随机数而闻名。更重要的是，它可以拆分为多个独立流。在 MPI 层级上，每个进程都通过确定性变换从全局种子派生出各自独立的种子，因此所有进程都能生成彼此不同且不重叠的随机序列，同时保证多次运行之间的可复现性。在每个 MPI 进程内部，OpenMP 线程通过序列拆分策略创建各自独立的随机流并为其设定种子。

除了数字生成之外，OMP 调度策略还引入了另一个随机性来源：采用动态调度时，不同执行过程中迭代可能会被分配给不同的线程，从而改变线程本地随机数生成器的使用顺序。因此，即使使用相同的种子，也无法再保证可复现性。静态调度策略通过将相同的

迭代分配给相同的线程来消除这一影响。然而，由于不同候选者在识别阶段的计算成本可能存在差异，静态调度可能容易导致负载不均衡。为兼顾这两种情况，我们改用 OMP 的运行时调度。这样，用户便可根据需求进行选择：静态调度用于保证可复现性，动态调度用于提升性能。

空洞最终排序是造成差异的最后一个来源。原始实现仅按半径对空洞排序，然后写入输出文件。然而，这会留下一个细微的边界情况：两个半径相同的空洞在不同运行中的顺序可能不同。虽然这种情况很少见，但我们在测试期间确实遇到了。因此，我们引入原始候选单元 ID 作为次级排序键，以打破平局、确保排序严格有序，并彻底解决可复现性问题。

这种方法有一个局限：只有在进程数、线程数和调度策略固定时，才能保证可复现性。这一局限在高性能计算代码中很常见（例如 Sparkling、GADGET；25、53），也是一种典型的权衡：在不牺牲性能或大幅改动实现的前提下，实现确定性执行。

6. 结果与比较

为了评估这项工作的影响，我们比较了两种实现的三个方面：静态分析，使用软件度量指标来刻画代码的结构复杂度和可维护性；统计效率性，评估真实宇宙学场景的物理结果；以及性能，测量代码的运行时行为。本次基准测试将旧版 Popcorn 代码的最新公开发行版 [3] (rev 63cbabe8) 与我们实现的两个带标签修订版进行比较：用于统计和性能测量的 cpc-benchmarks，以及用于静态分析的 cpc-final [4]。

6.1. 静态分析

静态分析提供了软件结构和实现复杂性的定量视角。虽然这并不直接对应性能或功能正确性，但它为可维护性、可读性和整体设计质量提供了有用的指标。鉴于代码以数据为中心，且不存在类层次结构和多态，该分析主要关注用于刻画代码规模和控制流的度量指标。

本次分析使用从代码行数（LOC）[59]，核心工具的角度对代码库进行了分类概览。我们对所有项目源代码运行了 Tokei，包括头文件、文档、实现代码和测试套件；所得度量指标汇总于图5。最显著的变化是总行数减少，降幅约为

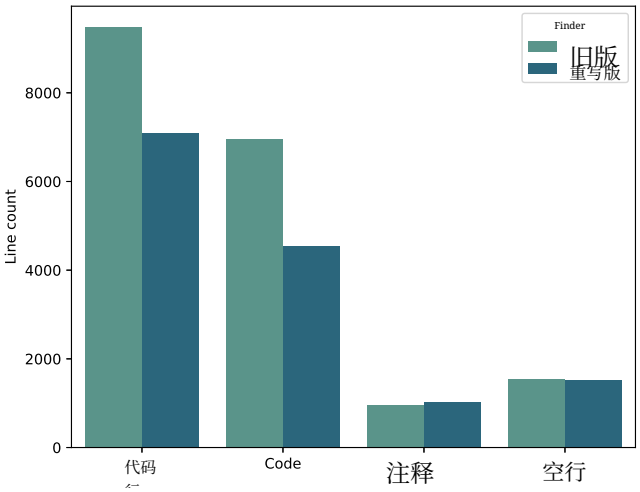

图5：不同 Finder 实现的代码行数（LOC）类别分布。

25%。进一步来看，减少的几乎全是代码片段，而注释和空行的数量基本保持不变。这出乎意料，因为与旧版不同，新版本为几乎所有公共接口都提供了详细文档。

虽然这两种实现使用等效算法来解决同一问题，但其内部结构存在显著差异。新版进行了大量重新设计，而简单的代码行数度量指标所提供的高层概述并未充分体现这些工作。为了进一步了解这一再工程过程的影响，我们使用了*Lizard* [60], 这一专注于通过圈复杂度数（CCN）指标分析控制流复杂度的工具。主要结果见表1。

表1：Lizard 得出的代码行数（LOC）和圈复杂度数（CCN）的汇总度量指标。

Finder	警告	函数	$\overline{\text{LOC}}$	$\overline{\text{CCN}}$
旧版	11	198	30.1	5.3
重写版	0	206	14.9	2.6

虽然两个实现包含的函数数量大致相当，但新版本将平均代码行数和圈复杂度降低了约一半，这表明其复杂性有所降低，代码也更加简洁。此外，Lizard 在旧版实现中发出了 11 条警告，且全部涉及圈复杂度为 ≥ 15 的函数；这通常被认为是结构较为复杂、具有中等风险的情况。其中一些警告指向死代码，另一些则涉及识别过程中的基础函数。这些诊断结果以及与其新对应实现的比较见表2。

最显著的案例出现在 inflate_fast 中，其 CCN 和 NLOC 均减少了约 90%。在这里，

[3]https://gitlab.com/dante.paz/popcorn_空洞_Finder [4]这些修订版仅在配置文件格式上有所不同，对性能或统计结果的影响可以忽略不计。

表2：原始实现中因圈复杂度（≥ 15）较高而被标记的函数，与新代码库中的对应实现进行比较。

	NLOC		CCN	
函数	Old	New	Old	New
inflate _fast	333	37	65	6
compute_intersecs	163	53	24	6
svf_work	124	80	23	12
popcorn_work	122	84	18	13

该功能被拆分到三个新的子进程中：

- 使用下界 Δ 近似值（NLOC = 48/CCN = 9）离散地膨胀一个种子。
- 计算具有固定半径的种子的精确 Δ（NLOC = 50/圈复杂度 = 8）。
- 在有界范围 [$R_{\inf}$、$R_{\sup}$] 内查找精确的空洞边界粒子（NLOC = 80/圈复杂度 = 13）。

总体而言，这种算法分解与关注点分离的应用降低了新版本的算法复杂度，因而也降低了风险。

Lizard 还提供了用于检测重复代码的扩展，可通过 --Eduplicate 标志启用。使用此功能后，分析显示旧版代码库中有 29.91% 由重复代码片段组成。其中大多数重复代码集中在三个主要区域：Finder 算法、输入/输出（IO）子系统以及解析逻辑。Finder 算法中的一部分重复代码对应于死代码。因此，我们将重点放在后两个区域上。

在 IO 组件中，许多重复代码来源于以下几类共享逻辑：读取分散在多个文件中的模拟快照、处理不同 Gadget 二进制格式的重复实现，以及与 HDF5 输入例程相关的通用样板代码。在新版本中，我们通过更通用的抽象处理了这些模式。现在，多文件快照读取通过一个可复用的调度例程实现；该例程接受一个负责读取单个文件的可调用对象，同时管理资源分配和执行流程。同样，我们实现了统一的 Gadget 读取器，通过一个参数化函数支持所有二进制格式，使格式特定的行为能够在不重复外围逻辑的情况下得到处理；之所以能够做到这一点，是因为这些格式的差异仅在于数据块中是否存在标头。引入这一抽象并采用 HDF5 Lite API 后，IO 层中的样板代码显著减少。

旧版代码中的其他重复内容源于主可执行程序源文件中反复出现的配置解析逻辑。这会带来隐患，因为每次修改参数，都必须在所有文件中反复进行一致的调整。在新版本中，配置解析被集中到一个专用模块中，以确保行为一致并便于未来扩展，例如引入带有详细错误说明的验证步骤。经过这些重构工作，新版本中的重复代码率从 29.91% 降至 2.68%。

6.2. 统计效率

现在，我们转而考察代码的运行时行为：在一个真实的宇宙学数据集上运行完整流程，以比较两种实现的统计结果和计算性能。该模拟使用 GADGET2 [52], 完成，模拟的是一个平坦的 ΛCDM 宇宙，其中包含一个边长为 $500h^{-1}$Mpc 的盒体，内有 512^3 个暗物质示踪粒子；采用的宇宙学参数如下：

$$\Omega_{m,0} = 0.258 \quad \Omega_{\Lambda,0} = 0.742 \quad H_0 = 0.796. \tag{8}$$

我们为此测试设定的目标是空洞尺寸函数（*VSF*），它统计宇宙中单位体积内特定半径的空洞数量。由于 *PVF* 的主要动机是相较球形空洞寻找器进行几何细化，这项测试集中展示了爆米花空洞群体的主要特点。

为了找到具有统计显著性的空洞目录，我们必须确定一组合适的参数。为 Δ_v和$R_{\max}$ 选择取值通常很直接。由于球形空洞寻找器与*PVF*密切相关，我们倾向于沿用球形理论中的Δ_v；该理论预测，空洞是扩展结构，其体积中有 .8 的部分符合 $\Delta_v \approx -0$。另一方面，$R_{\max}$ 可以根据实证证据推导出来：利用示踪体类型或平均示踪体密度等信息，我们可以预期空洞不会大于某个上限。本次实验选择 $R_{\max} = 45h^{-1}$Mpc，其数值最终远大于所找到的最大球体；但即使情况相反，其影响仍会在结果中显现。

选择合适的$R_{\min}$ 值并不那么直接，因为在小尺度下，球形空洞查找器（SVF）可能产生偏离预期物理分布的空洞丰度。空洞层级理论预测，*VSF*遵循 Schechter 型形式 [1, 61], , 描述物理相关尺度范围内的空洞丰度。然而在实践中，这种行为不一定能一直恢复到任意小的半径。这些尺度主要受到离散示踪粒子分布所提供的有限空间信息、球形空洞碎裂以及其他探测器特有属性的影响；这些因素共同产生了一个人为的区间，我们将其统称为**散粒噪声**。

为确定这些区间之间的转换，我们首先根据示踪粒子间的平均间距，选取一个有意设置得较为保守的临时最小半径 $\hat{}R_{\min}$，

$$\langle r \rangle = (\bar{\rho})^{-1/3}. \tag{9}$$

在我们的模拟中，$\langle r \rangle = 512/500 \approx 1h^{-1}$Mpc，因此我们采用 $\hat{}R_{\min} = 2.5h^{-1}$Mpc；不过，实际使用中，最大约为平均间距三倍的取值也同样是稳妥的初始选择。随后，我们使用 Δ_v、$R_{\max}$ 和 $\hat{}R_{\min}$ 执行球形空洞查找器，使生成的目录

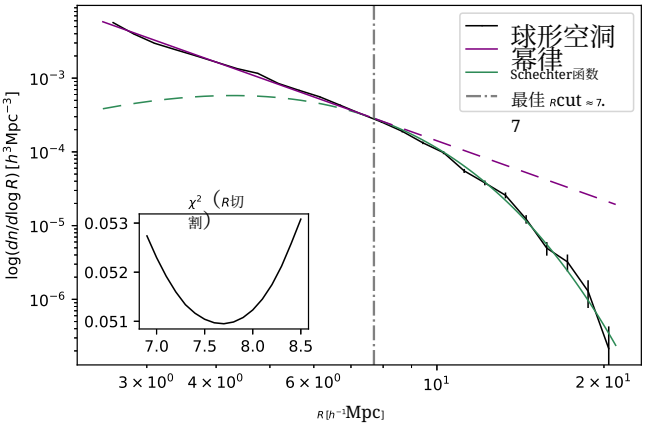

图6：一组球形空洞的经验空洞尺度函数 及由截断点R_{cut}分割的两条曲线的分段拟合。主图显示使拟合损失最小的R_{cut}。左下角的小图显示了邻近候选截断点的损失。

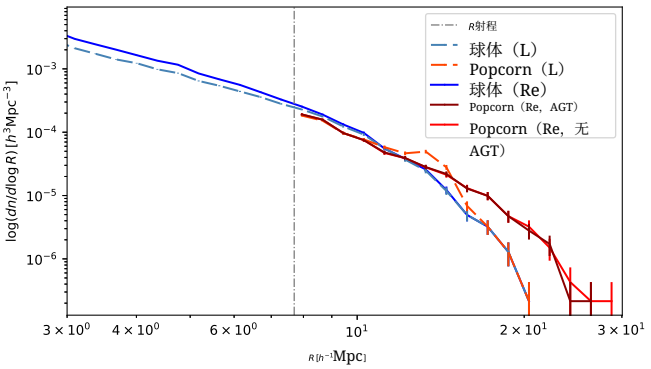

图7：不同配置的空洞搜寻器以及*PVF*实现的经验空洞尺度函数。球形算法使用蓝色系，爆米花变体使用红色系。虚线表示旧版实现（L），实线表示重写实现（Re）。“AGT”爆米花变体采用了有损的“自适应增长终止”准则，以牺牲不断增大的空洞为代价来提前终止，从而提升性能。

能够保留空洞，直至散粒噪声效应可能开始显著影响结果的尺度。

由此得到的空洞尺度函数 呈现出两个明显的区间。在半径足够大时，测得的丰度遵循预期的谢克特行为；而在较小半径下，它会系统性地偏离这一形式；在当前实现中，经验上发现其遵循幂律 [5]。因此，我们将空洞尺度函数 建模为连续分段函数：小半径处为幂律，大半径处为谢克特函数，两者由散粒噪声半径R_{shot}分隔：

$$\frac{dn}{d\log R} = \begin{cases} bR^a & R < R_{\mathrm{shot}} \\ \phi^*\left(\frac{R}{R^*}\right)^{\alpha} e^{-R/R^*} & R \geq R_{\mathrm{shot}}. \end{cases} \tag{10}$$

为了拟合连续模型，我们调整这两个区间，使其在点（R_{shot},y_{shot}）处相接，从而对五个相互依赖的参数进行耦合重参数化：R_{shot}、y_{shot}、a,α和R^*。然后，我们选取一组候选截断半径，并对曲线的其余四个参数进行拟合。我们将使 χ^2 拟合评分最小的截断半径确定为散粒噪声半径R_{shot}，并将该值用作爆米花阶段采用的最小半径，

$$R_{\mathrm{min}} = R_{\mathrm{shot}}. \tag{11}$$

图6展示了这一过程，其中我们使用球形空洞半径的30个对数分箱拟合了一个空洞尺度函数。

确定合适的参数后，我们通过对先前获得的球形空洞目录运行爆米花阶段来完成整个流程。为了评估重新实现的质量，我们分别使用两个代码库运行这些流程，此外还运行了一个禁用自适应增长终止准则（第4.1小节中讨论）的变体

[5]在我们采用约束较少的随机游走作为重新定位策略时，情况并非如此。我们将在第 7 节说明当前方法的优势。

。所有这些运行都得到了图7所示的空洞尺度函数曲线，其中爆米花空洞按其有效半径进行分箱：

$$R_{\mathrm{eff}} = \left(\frac{3V}{4\pi}\right)^{1/3}. \tag{12}$$

在R_{shot}的右侧，我们可以立即看到，球形空洞曲线在$R = 20h^{-1}\mathrm{Mpc}$附近趋于一致，而一些爆米花曲线则达到略低于$R = 30h^{-1}\mathrm{Mpc}$的位置。这意味着，爆米花算法能够有效地将碎片化或不完整的球形欠密度区扩展为更大的空洞结构。然而，其中一条爆米花曲线落后了，结果接近球形空洞。这发生在旧版实现初次发布后的版本中，原因是其采用了过于激进的增长终止准则。相当令人意外的是，我们早期使用旧版代码处理暗物质晕的测试，并未呈现出此处使用粒子时观察到的这种紧密的球形—爆米花曲线对齐。相比之下，新的自适应终止准则能够在最后三个分箱之前紧密复现“原始”行为，而最后三个分箱中的空洞总共只有四个。

在R_{shot}的左侧，我们可以看到球形空洞曲线之间的另一处分歧，这由其重叠清理阶段中的一个缺陷所导致。正如我们在第 2.2 小节中所解释的那样，“干净”的球形空洞配置对重叠消解的顺序十分敏感。这一特性严重限制了并行解的适用范围，因为并行处理必须谨慎避免同时处理相互重叠的链。相反，旧版实现采用了两阶段流程：首先使用随机初步过滤器，将大型集合划分为多个区块；当最大相对半径差低于 5% 时，这些区块便可并行处理；随后再通过确定性顺序遍历处理区块中剩余的部分。尽管这一任意阈值旨在将统计上可察觉的影响降至最低，但它仍然引入了明显偏差

在这一特定情况下。这种影响在R_{shot}的较低端更为突出，并非巧合，因为那里包含了绝大多数空洞；但我们不能依赖这一行为在$R \gg R_{\text{shot}}$上保持一致，这也是我们在重新实现中从头重新实现重叠清理流程的原因。

6.3. 性能

通过统计评估，我们成功识别并缓解了旧版代码库中的多项缺陷。为验证这些改进的整体影响，最后我们进行了严格的性能比较，展示这些结构性变更如何影响执行速度。性能分析评估了重写实现的三个方面：整体流水线运行时间、球形空洞重叠移除的耗时，以及确定性调度的影响。用于这些度量指标的系统是科尔多瓦国立大学超级计算部门的 serafin 集群“UNC Supercómputo”[6]，每个节点配备 2× AMD EPYC 7532 CPU 和 256GiB 内存。代码使用 gcc 12.3.0 和 openmpi 4.1.6 编译。图 8 展示了这项分析的完整总结。

流水线运行时间：我们首先比较完整识别流水线的实际运行时间。由于旧版实现与重写实现的可执行阶段和后处理不同，因此我们将主要识别阶段与后续处理阶段合并，计算每种配置的总运行时间。对于球形空洞查找器，重写实现将运行时间从旧版的 8767.8 s 降至 5420.9 s，对应 1.62× 倍加速。这一改进显著加快了后续爆米花形成阶段所需输入目录的生成。

然而，Popcorn 结果说明了考虑每个配置所执行的工作量的重要性。旧版实现用时 360.8 秒完成整个流程，而不采用提前终止生长的重写实现（Popcorn（Re，No AGT））则需要 8773.4 秒。这一差异并不表示重写算法本身出现了性能退化，而是揭示了旧版终止启发式方法所避免的大量冗余生长工作。启用自适应增长终止准则后，重写实现仅需 139.4 秒即可完成，相较于不受约束的重写版本减少了 62.9× 。这一结果还凸显了

负载分布对重写后的 Popcorn 实现的重要性：在无约束配置中，实验处理的 3408 个空洞里，约 99.7% 的运行时都花在处理最后八个空洞上。因此，终止准则不仅减少了总工作量，还能防止少数病态增长主导执行时间。

球形空洞清理：我们的下一项实验是评估球形空洞重叠移除算法的性能。由于不同实现每次运行时处理的

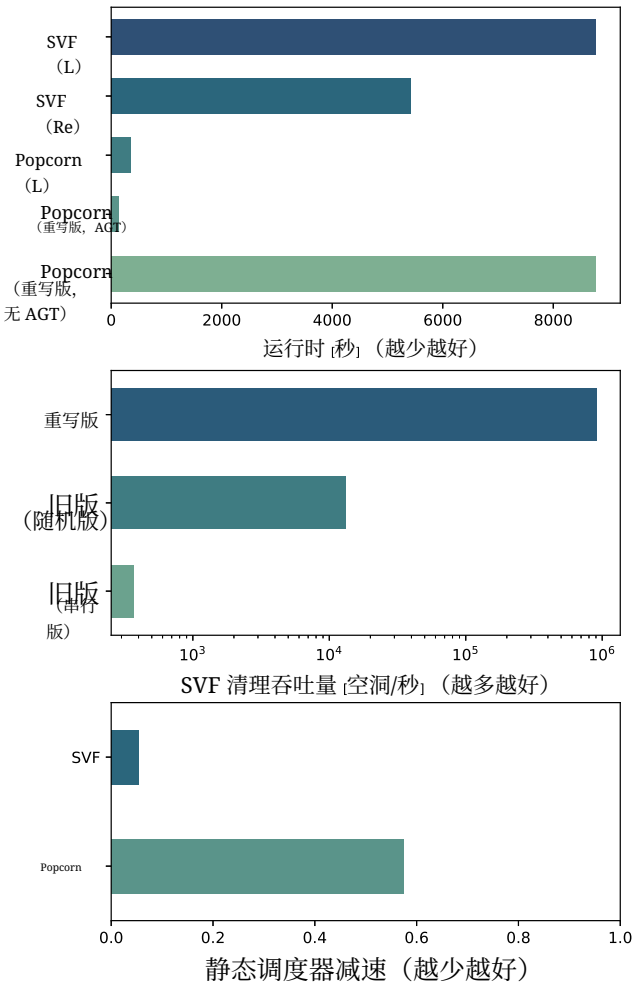

图 8：*PVF* 性能分析。上图展示了 SVF 和 Popcorn 的旧版（L）与重写版（Re）的墙钟时间，以及采用自适应增长终止准则（AGT）的另一种 Popcorn 版本。中图展示了不同重叠球形空洞清理策略的处理速度，其中我们将旧版策略细分为并行随机版本和纯串行版本。下图展示了静态调度器造成的减速，该调度器可实现确定性执行。

空洞数量不同，因此直接比较运行时长没有意义。相反，我们测量了清理吞吐量，即每秒已处理空洞的数量。

旧版程序用 480.5 秒清理了 630 万个空洞，吞吐量为 1.33×10^4 空洞/秒。相比之下，重写实现用 7.3 秒清理了 680 万个空洞，吞吐量约为 9.18×10^5 空洞/秒。总体而言，这相当于提升了约 69×。

为完整起见，我们还单独测量了旧版实现的串行部分，并排除了随机预处理阶段的耗时。在原有的 630 万个空洞中，最终仅剩 18 万个，处理耗时 458.9 秒，吞吐量约为 373 个空洞/秒。这说明旧版重叠移除策略带来了显著开销：如果我们只是移除预处理，以消除图7所示的偏差，清理工作也将耗时近 5 小时，远远超过发现这些空洞所需的时间。重写后的

[6] https://supercomputo.unc.edu.ar/

因此，该流程实现了效率显著更高的方案，同时消除了旧版预处理阶段引入的随机性。

调度开销：在本工作中，我们为 SVF 引入了确定性模式，从而实现完整的流水线可复现性。这需要新增一种静态调度策略，以确保所有并行线程按相同顺序处理输入。最后一项测试评估了在空洞识别中，相较于原始动态调度器，采用这一静态调度器所产生的成本。在 SVF 中，我们仅观察到静态调度下运行时增加了约 5%。

尽管 Popcorn 本身不受调度策略和随机种子影响而具有确定性，但我们仍将其纳入本实验，以刻画其增长工作负载中潜在的调度开销。在这种情况下，影响明显更大，运行时增加了约 58%。这些结果表明，静态调度可能引入可测量的负载均衡成本，尤其是对于高度不规则的 popcorn 工作负载而言。尽管如此，在 SVF 中的开销可以忽略不计，而 SVF 是唯一与这种调度选择相关的步骤。

7. 总结与结论

在本研究中，我们重新审视了 Popcorn 空洞查找器的实现，旨在提升其性能、可维护性、可复现性和统计可靠性。我们没有将现有实现视为黑箱，而是重建并记录了完整的空洞识别过程，从最初的球形候选体到爆米花空洞的构建与筛选。这不仅详细阐述了该算法及其底层设计选择，也使我们得以识别和评估原始实现中若干计算与统计行为的来源。我们重新审视并实现了球形空洞和爆米花空洞识别算法，着重将其概念阶段分离，并减少不必要的复杂性。除算法本身外，重写后的实现还提升了*PVF*的实际易用性。HDF5 支持改为可选，使得无需使用基于 HDF5 的输入格式的配置可以在不依赖 HDF5 的情况下编译。此外，我们通过明确指定随机种子和采用确定性调度，为球形查找器引入了可复现性，使其能够在受控条件下重复运行。

对代码库进行静态分析显示，源代码总规模、单个函数规模、圈复杂度和重复代码均有所减少。这些变化提供了经验证据，表明重写实现不仅在选定的工作负载下更加高效，而且也更易于维护和扩展。统计分析表明，重写实现为统计分析提供了更加一致的基础，同时保留了爆米花空洞表示非球形结构的能力。性能评估显示，新的球形空洞查找器实现了 1.62× 倍加速，而重写的 Popcorn 实现结合新的自适应增长终止策略，在统计质量略有下降的情况下达到了 62.9× 倍加速，这些结果强调，Popcorn 的性能很大程度上取决于所执行工作的数量和分布，而不仅仅取决于单个算法操作的成本。

本研究还揭示了空洞大小函数在小尺度上的行为存在显著差异。先前的实现 [28, 25]采用了随机游走重新居中过程，其参数（例如步长和步数）会直接影响所得空洞尺度函数的小尺度形状。因此，向散粒噪声区间过渡的位置和强度可能会随用户无法直接调整的内部算法参数而变化，导致散粒噪声的判定不够稳健。本实现采用的单元内重新居中策略则产生了定性上不同且更规则的行为。小尺度区间可用幂律进行经验描述，而在更大半径处则呈现预期的 Schechter 形式，因此可以直接根据这两个区间之间的过渡来定义散粒噪声。重要的是，这种刻画不会引入对隐藏参数的隐性依赖，从而为选择最小可靠空洞尺度提供了更具可重复性的依据。

除了 *PVF*，我们还发布了全新的开源 Python API*pvfpy* [7]，能够读取和处理输出识别文件。具体而言，空洞可视化（图3）、散粒噪声判定（图6）以及空洞大小函数（图7）均由该 API 生成。近期新增的一项功能提供了另一种表示爆米花空洞的方式：将其表示为长椭球，这些椭球根据成员球体的空间分布得出。这种表示以主轴尺寸和方向为参数，简洁地描述其非球形几何结构，便于表征形状并比较不同空洞。完整的软件栈向后兼容，因此旧版 *PVF* 的配置文件和输出文件均可用于更新后的代码库。总体而言，这项工作展现了重新审视原始实现的价值，而不仅仅着眼于其最初的科学成果。通过提升其可维护性和可复现性，并系统评估其算法选择，我们为爆米花空洞未来的发展和统计应用奠定了更可靠的基础。

8. CRediT 作者贡献声明

A. R. Schachner： 撰写——初稿，软件，调查，形式分析。 **D. J. Paz：** 撰写——审阅与编辑，软件，监督，方法学，验证。 **J. B. Cabral：** 撰写——审阅与编辑，监督，方法学，验证。 **J. A. Rico Gallego：** 监督，资金获取。 **C. M. Correa：** 撰写——审阅与编辑，监督。 **A.N.Ruiz：** 撰写——审阅与编辑，软件，监督。 **S. R. Gualpa：** 软件。

[7] https://gitlab.com/void-research-labs/pvfpy

9. 关于在稿件准备过程中使用生成式人工智能和人工智能辅助技术的声明

在准备本项工作期间，作者使用了 OpenAI–ChatGPT、Google–Gemini 和 Anthropic–Claude，以提高文本的清晰度和准确性。在使用该工具/服务后，作者根据需要对内容进行了审阅和编辑，并对已发表文章的内容承担全部责任。

10. 竞争性利益声明

Alvaro Roy Schachner 获得了 Sistema Extremeño de Ciencia, Tecnología e Innovación (SECTI) 提供的财务支持。如有其他作者，他们声明，除已知情况外，他们不存在可能被认为影响本文所报告工作的竞争性财务利益或个人关系。

11. 致谢

本研究得到 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) 的支持，并由欧盟通过欧洲区域发展基金 (ERDF) 共同资助，资助渠道为“Consejería de Educación, Ciencia y Formación Profesional de la Junta de Extremadura”（参考编号：GR24208）。本研究使用了 UNC Supercómputo (CCAD) [8]，的计算资源，该机构隶属于阿根廷的 SNCAD。

12. 代码可用性

Popcorn 空洞查找器的源代码已公开发布于 https://gitlab.com/void-research-labs/popcorn-void-finder，采用 BSD三条款许可证。

13. 数据可用性

支持本文所呈现统计结果的数据已存入 Mendeley Data，审稿人可通过投稿时提供的私有数据集链接访问。相关论文发表后，该数据集将向公众开放。

参考文献

[1]R. K. Sheth、R. van de Weygaert，《空洞层级结构：为虚无大惊小怪》，皇家天文学会月刊350 (2) (2004) 517–538。arXiv:astro-ph/0311260，DOI:10.1111/j.1365-2966.2004.07661.x。[2] S. Contarini、G. Verza、A. Pisani，《利用空洞开展精密宇宙学的时代》，《天文学与天体物理学》评论34 (1) (2026) 1。arXiv:2601.14362，DOI:10. 1007/s00159-026-00166-x。

[3] Y.-C. Cai、N. Padilla、B. Li，《利用宇宙空洞检验引力》，皇家天文学会月刊451 (1) (2015) 1036–1055。arXiv:1410.1510，DOI:10.1093/mnras/stv777。

[4] C. M. Correa、D. J. Paz、N. D. Padilla、A. N. Ruiz、R. E. Angulo、A. G. Sánchez，《来自空洞周围几何与动力学畸变的非基准宇宙学检验》，皇家天文学会月刊 485 (4) (2019) 5761–5772。arXiv:1811.12251，DOI:10.1093/mnras/stz821。

[5] M. Aubert、M.-C. Cousinou、S. Escoffier、A. J. Hawken、S. Nadathur、S. Alam、J. Bautista、E. Burtin、C.-H. Chuang、A. de la Macorra、A. de Mattia、H. Gil-Marín、J. Hou、E. Jullo、J.-P. Kneib、R. Neveux、G. Rossi、D. Schneider、A. Smith、A. Tamone、M. Vargas Magaña、C. Zhao，《SDSS-IV扩展重子声学振荡光谱巡天的完成：从宇宙空洞测量结构增长率》，皇家天文学会月刊 513 (1) (2022) 186–203。arXiv:2007.09013，DOI:10.1093/mnras/stac828。

[6] N. Hamaus, M. Aubert, A. Pisani, S. Contarini, G. Verza, M.-C. Cousinou, S. Escoffier, A. Hawken, G. Lavaux, G. Pollina, B. D. Wandelt, J. Weller, M. Bonici, C. Carbone, L. Guzzo, A. Kovacs, F. Marulli, E. Massara, L. Moscardini, P. Ntelis, W. J. Percival, S. Radinovi´c, M. Sahlén, Z. Sakr, A. G. Sánchez, H. A. Winther, N. Auricchio, S. Awan, R. Bender, C. Bodendorf, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, V. Capobianco, J. Carretero, F. J. Castander, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, L. Conversi, Y. Copin, L. Corcione, M. Cropper, A. Da Silva, H. Degaudenzi, M. Douspis, F. Dubath, C. A. J. Duncan, X. Dupac, S. Dusini, A. Ealet, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, P. Franzetti, M. Fumana, B. Garilli, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, S. V. H. Haugan, W. Holmes, F. Hormuth, K. Jahnke, S. Kermiche, A. Kiessling, M. Kilbinger, T. Kitching, M. Kümmel, M. Kunz, H. Kurki-Suonio, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Marggraf, K. Markovic, R. Massey, S. Maurogordato, M. Melchior, M. Meneghetti, G. Meylan, M. Moresco, E. Munari, S. M. Niemi, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, V. Pettorino, S. Pires, M. Poncet, L. Popa, L. Pozzetti, R. Rebolo, J. Rhodes, H. Rix, M. Roncarelli, E. Rossetti, R. Saglia, P. Schneider, A. Secroun, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, J.-L. Starck, P. Tallada-Crespí, D. Tavagnacco, A. N. Taylor, I. Tereno, R. Toledo-Moreo, F. Torradeflot, E. A. Valentijn, L. Valenziano, Y. Wang, N. Weli-kala, G. Zamorani, J. Zoubian, S. Andreon, M. Baldi, S. Camera, S. Mei, C. Neissner, E. Romelli, 欧几里得：利用红移空间畸变和宇宙空洞的 Alcock-Paczynski 检验进行预测，《天文学与天体物理学》658（2022）A20。arXiv:2108.10347，DOI:10.1051/0004-6361/202142073。

[8]https://supercomputo.unc.edu.ar

[7] S. Contarini, G. Verza, A. Pisani, N. Hamaus, M. Sahlén, C. Carbone, S. Dusini, F. Marulli, L. Moscardini, A. Renzi, C. Sirignano, L. Stanco, M. Aubert, M. Bonici, G. Castignani, H. M. Courtois, S. Escoffier, D. Guinet, A. Kovacs, G. Lavaux, E. Massara, S. Nadathur, G. Pollina, T. Ronconi, F. Ruppin, Z. Sakr, A. Veropalumbo, B. D. Wandelt, A. A mara, N. Auricchio, M. Baldi, D. Bonino, E. Branchini, M. Brescia, J. Brinchmann, S. Camera, V. Capobianco, J. Carretero, M. Castellano, S. Cavuoti, R. Cledassou, G. Congedo, C. J. Conselice, L. Conversi, Y. Copin, L. Corcione, F. Courbin, M. Cropper, A. Da Silva, H. Degaudenzi, F. Dubath, C. A. J. Duncan, X. Dupac, S. Ealet, S. Farrens, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, B. Garilli, W. Gillard, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, L. Guzzo, S. Haugan, W. Holmes, F. Hormuth, K. Jahnke, M. Kü mmel, S. Kermiche, A. Kiessling, M. Kilbinger, M. Kunz, H. Kurki-Suonio, R. Laureijs, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Mansutti, O. Marggraf, K. Markovic, R. Massey, M. Melchior, M. Meneghetti, G. Meylan, M. Moresco, E. Munari, S. M. Niemi, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, W. J. Percival, V. Pettorino, S. Pires, G. Polenta, M. Poncet, L. Popa, L. Pozzetti, F. Raison, J. Rhodes, E. Rossetti, R. Saglia, B. Sartoris, P. Schneider, A. Secroun, G. Seidel, G. Sirri, C. Surace, P. Tallada-Crespí, A. N. Taylor, I. Tereno, R. Toledo-Moreo, F. Torradeflot, E. A. Valentijn, L. Valenziano, Y. Wang, J. Weller, G. Zamorani, J. Zoubian, S. Andreon, D. Maino, S. Mei, 欧几里得：基于空洞大小函数的宇宙学预测，《天文学与天体物理学》667（2022）A162。arXiv:2205.11525，DOI：10.1051/0004-6361/202244095。

[8] S. Radinovi´c, S. Nadathur, H. A. Winther, W. J. Perci- val, A. Woodfinden, E. Massara, E. Paillas, S. Contarini, N. Hamaus, A. Kovacs, A. Pisani, G. Verza, M. Aubert, A. Amara, N. Auricchio, M. Baldi, D. Bonino, E. Bran- chini, M. Brescia, S. Camera, V. Capobianco, C. Carbone, V. F. Cardone, J. Carretero, M. Castellano, S. Cavuoti, A. Cimatti, R. Cledassou, G. Congedo, L. Conversi, Y. Copin, L. Corcione, F. Courbin, A. Da Silva, M. Dous- pis, F.Dubath, X. Dupac, S. Farrens, S. Ferriol, P. Fosalba, M. Frailis, E. Franceschi, M. Fumana, S. Galeotta, B. Gar- illi, W. Gillard, B. Gillis, C. Giocoli, A. Grazian, F. Grupp, S. V. H. Haugan, W. Holmes, A. Hornstrup, K. Jahnke, M. Kümmel, A. Kiessling, M. Kilbinger, T. Kitch- ing, H. Kurki-Suonio, S. Ligori, P. B. Lilje, I. Lloro, E. Maiorano, O. Mansutti, O. Marggraf, K. Markovic, F. Marulli, R. Massey, S. Mei, M. Melchior, Y. Mel- lier, M. Meneghetti, E. Merlin, G. Meylan, M. Moresco, L. Moscardini, S. M. Niemi, J. W. Nightingale, T. Nutma, C. Padilla, S. Paltani, F. Pasian, K. Pedersen, V. Pettorino, S. Pires, G. Polenta, M. Poncet, L. A. Popa, L. Pozzetti, F. Raison, A. Renzi, J. Rhodes, G. Riccio, E. Romelli, M. Roncarelli, C. Rosset, R. Saglia, D. Sapone, B. Sartoris, P. Schneider, A. Secroun, G. Seidel, S. Serrano, C. Sirignano, G. Sirri, L. Stanco, J. L. Starck, C. Surace, P. Tallada-Crespí, I. Tereno, R. Toledo-Moreo, F. Torradeflot, I. Tutusaus, E. A. Valentijn, L. Valenziano, T. Vas- sallo, Y. Wang, J. Weller, G. Zamorani, J. Zoubian, V. Scottez, 欧几里得：利用重建的空洞—星系交叉相关函数进行宇宙学预报，《天文学与天体物理学》677（2023）A78。arXiv:2302. 05302，DOI:10.1051/0004-6361/202346121。

[9] A. M. Rodríguez Medrano、D. J. Paz、F. A. Stasyszyn、A. N.Ruiz，宇宙空洞演化在星系晕内部重子过程中的印记，《皇家天文学会月刊》511 (2) (2022) 2688–2701。arXiv:2109.13378，DOI:10.1093/mnras/stac127。

[10] C. M. Correa，作为宇宙学实验室的宇宙空洞，《阿根廷天文学协会公报》，拉普拉塔，阿根廷，64 (2023) 159–165。arXiv:2210.17459，DOI:10.48550/arXiv.2210.17459。

[11] C. M. Correa、D. J. Paz、A. G. Sánchez、A. N. Ruiz、N. D. Padilla、R. E. Angulo，空洞中的红移空间效应及其对宇宙学检验的影响。第一部分：空洞大小函数，《皇家天文学会月刊》500 (1) (2021) 911–925。arXiv:2007.12064，DOI:10.1093/mnras/staa3252。

[12] C. M. Correa, D. J. Paz, N. D. Padilla, A. G. Sánchez, A. N. Ruiz, R. E. Angulo，《空洞中的红移空间效应及其对宇宙学检验的影响——II：空洞-星系交叉相关函数》，《皇家天文学会月刊》509 (2) (2022) 1871–1884。arXiv:2107.01314，doi:10.1093/mnras/stab3070。

[13] E. Jennings, Y. Li, W. Hu，《空洞丰度与游走集形式主义》，《皇家天文学会月刊》434 (3) (2013) 2167–2181。arXiv: 1304.6087，doi:10.1093/mnras/stt1169。

[14] D. Paz, M. Lares, L. Ceccarelli, N. Padilla, D. G. Lam- bas，《空洞演化的线索 II：测量斯隆数字化巡天星系红移空间畸变中的密度和速度剖面》，《皇家天文学会月刊》436 (4) (2013) 3480–3491。arXiv:1306.5799，doi: 10.1093/mnras/stt1836。

[15] N. Hamaus、P. M. Sutter、G. Lavaux、B. D. Wandelt，利用空洞周围的红移空间畸变探测宇宙学与引力，《宇宙学与天体粒子物理学杂志》2015（11）（2015）036–036。arXiv:1507.04363，doi:10.1088/1475-7516/2015/11/036。

[16] E. Massara, W. J. Percival, N. Dalal, S. Nadathur, S. Radinović, H. A. Winther, A. Woodfinden，《空洞周围物质和有偏示踪体的速度剖面》，《皇家天文学会月刊》517 (3) (2022) 4458–4471。arXiv:2206.14120，DOI:10.1093/mnras/stac2892。

[17] J. M. Colberg, F. Pearce, C. Foster, E. Platen, R. Brunino, M. Neyrinck, S. Basilakos, A. Fairall, H. Feldman, S. Gottlöber, O. Hahn, F. Hoyle, V. Müller, L. Nelson, M. Plionis, C. Porciani, S. Shandarin, M. S. Vogeley, R. van de Weygaert, 《Aspen–Amsterdam 空洞 Finder 比较项目》, 《皇家天文学会月刊》387 (2) (2008) 933–944。arXiv:0803.0918, DOI:10.1111/j.1365-2966.2008.13307.x。

[18] M.Cautun, E.Paillas, Y.-C. Cai, S.Bose, J.Armijo, B.Li, N. Padilla, 《Santiago-Harvard-Edinburgh-Durham空洞比较》I：揭示手征引力测试中的SHEDding light，皇家天文学会月刊476（3）（2018）3195–3217。arXiv:1710.01730, DOI:10.1093/mnras/sty463。

[19]E. Paillas, M. Cautun, B. Li, Y.-C. Cai, N. Padilla, J. Armijo, S. Bose, 《Santiago-Harvard-Edinburgh-Durham空洞比较》II：利用弱引力透镜揭示Vainshtein屏蔽，皇家天文学会月刊484(1)（2019）1149–1165。arXiv:1810.02864, DOI:10.1093/mnras/stz022。

[20]E. Platen, R. Van De Weygaert, B. J. T. Jones, 《宇宙分水岭：WVF 空洞检测技术》, 皇家天文学会月刊380 (2) (2007) 551–570。arXiv:https: // academic.oup.com/mnras/article-pdf/380/2/551/18422647/mnras0380-0551.pdf, DOI:10.1111/j.1365-2966.2007.12125.x。URL https://doi.org/10.1111/j.1365-2966. 2007.12125.x

[21] M. C. Neyrinck，ZOBOV：一种无需参数的空洞查找算法，皇家天文学会月刊 386 (4) (2008) 2101–2109。arXiv:0712.3049，DOI:10.1111/j.1365-2966.2008.13180.x。

[22] P. M. Sutter, G. Lavaux, N. Hamaus, A. Pisani, B. D. Wandelt, M. Warren, F. Villaescusa-Navarro, P. Zivick, Q.Mao, B.B. Thompson, 《VIDE：空洞识别与检验工具包》, 《天文学与计算》9 (2015) 1–9。arXiv:1406.1191，DOI:10.1016/j.ascom.2014.10.002。

[23] H. El-Ad, T. Piran, 《大尺度结构中的空洞》, 《天体物理学杂志》491 (2)(1997) 421–435。arXiv: astro-ph/9702135，DOI:10.1086/304973。

[24] F. Hoyle, M. S. Vogeley, 《点源目录巡天和更新版茨维基星表中的空洞》, 《天体物理学杂志》566 (2) (2002) 641–651。arXiv: astro-ph/0109357，DOI:10.1086/338340。

[25] A. N. Ruiz, D. J. Paz, M. Lares, H. E. Luparello, L. Ceccarelli, D. G. Lambas, 《空洞演化线索——III：空洞壳层中的结构与动态调度》, 皇家天文学会月刊448 (2) (2015) 1471–1482。arXiv:1501.02120，DOI:10.1093/mnras/stv019。

[26] A. N. Ruiz, I. G. Alfaro, D. Garcia Lambas, 《低密度区域中的结构与动态调度：宇宙空洞内的星系—星系相关性》, 皇家天文学会月刊483 (3) (2019) 4070–4079。arXiv:1812.05532，DOI:10.1093/mnras/sty3443。

[27] A. Banerjee、N. Dalal，使用大质量中微子模拟非线性宇宙学结构形成，《宇宙学与天体粒子物理学杂志》2016（11）（2016）015。arXiv:1606.06167, DOI:10.1088/1475-7516/2016/11/015。

[28] D. J. Paz、C. M. Correa、S. R. Gualpa、A. N. Ruiz、C. S. Bederián、R. D. Graña、N. D. Padilla，通过孔洞大小猜测奶酪口味：使用爆米花空洞丰度进行的宇宙学检验，《皇家天文学会月刊》522（2）（2023）2553–2569。arXiv:2212.06849，DOI:10.1093/mnras/stad1146。

[29]N. Findlay, S. Nadathur, 《VERSUS：一种受 exc ursion-set 启发、面向第四阶段时代的空洞 Finder》, 《皇家天文学会月刊》551 (2) (2026) stag1425。DOI: 10.1093/mnras/stag1425。

[30]G. B. Berriman, S. L. Groom, 《天文学档案如何应对数据洪流？》, arXiv 电子预印本（2011 年 10 月）。arXiv：1111.0075，DOI：10.48550/arXiv.1111.0075。

[31] M. Garofalo, A. Botta, G. Ventre, 《天体物理学与大数据：挑战、方法和工具》, 《国际天文学联合会论文集》12 (03 2017)。DOI：10.1017/S1743921316012813。

[32] J. Buchner, 《天文学领域高影响力科学软件及其创作者》, 载于《美国天文学会通报》, 第 57 卷，2025 年，第 2025i029 页。arXiv：2511.12195，DOI：10.3847/25c2cfeb.fe9c1f84。

[33] E.Tollerud, A.Smith, A. Price-Whelan, K.Cruz, D. Norman, G. Narayan, S. Mumford, A. Allen, C.-k. Chan, B. Cherinka, A. Drlica-Wagner, D. Foreman-Mackey, A. Ginsburg, A. Gradvhol, J. Harrington, D. Hogg, J. Jar- taltepe, J. Kinney, N. Merchant, I. Momcheva, N. Mur- phy, J. Peek, M. S. Peeples, T. Pickering, D. Rodriguez, L. Shamir, M. Sinha, B. Sip˝ocz, J. Sobeck, M. Sosey, H. Stevance, P. Teuben, D. Vohl, B. Weiner, T. Ald- croft, A. Allen, M. Alpaslan, L. Anderson, G. Barentsen, D. Bektesevic, J. Benavides, B. Berriman, M. Blanton, J. Bosch, D. Bouquin, L. Bradley, G. Bryan, D. Burke, K. Burns, D. Buzasi, J. B. Cabral, J. V. d. M. Car- doso, B. Chen, W. Clarkson, M. Collins, L. Corrales, M. Craig, S. Crawford, S. Domagal-Goldman, C. Dong, M. Durbin, J. K. Faherty, W. Farr, L. Forschini, V. Z. Golkhou, H. M. Günther, H. Hafok, C. Hahn, N. Hathi, C. Hedges, S. Huang, C. Hummels, E. Hunt, D. Hup- penkothen, S. Juneau, M. van Kerkwijk, W. Kerzen- dorf, I. Laginja, C. Law, J. de Leon, T. Li, P. L. Lim, A. I. Malz, Y.-Y. Mao, P. Melchior, B. Merin, B. Miller,

M. Modjaz, T. Morton, S. Mullally, R. Ogando, J. K. Parejko, D. Paz, S. Pearson, K. Pontoppidan, B. Pope, D. Rapetti, M. Rawls, J. Read, T. Robitaille, G. Rud- nick, S. Sharma, S. Sharma, D. Shupe, J. Speagle, T. Starkenburg, F. Stasyszyn, O. Streicher, G. Tremblay, F. Villaescusa-Navarro, J. M. Vos, B. A. Weaver, A. Welt- man, A. Wetzel, P. K. G. Williams, B. Winkel, 《为天文学提供由社区驱动的软件支持：2020 年代》，载于：《美国天文学会通报》，第 51 卷，2019 年，第 180 页。

[34] A. Hunt，D. Thomas，《程序员修炼之道：从小工到专家》，培生教育，1999 年。URL https://books.google.com.ar/books?id=5wBQEp6ruIAC

[35] L. Ceccarelli，A. N. Ruiz，M. Lares，D. J. Paz，V. E. Maldonado，H. E. Luparello，D. Garcia Lambas，《璀璨宇宙：宇宙空洞运动的情景》，《皇家天文学会月刊》461（4）（2016）4013–4021。arXiv：1511.06741，DOI：10.1093/mnras/stw1524。

[36] M. Lares，H. E. Luparello，V. Maldonado，A. N. Ruiz，D. J. Paz，L. Ceccarelli，D. Garcia La mbas，《空洞与超结构：相关性及诱导的大尺度速度流》，《皇家天文学会月刊》470（1）（2017）85–94。arXiv：1705.06541，DOI：10.1093/mnras/stx1227。

[37] I. G. Alfaro, F. Rodriguez, A. N. Ruiz, D. G. Lambas，《低密度环境中的星系如何分布于晕中：宇宙空洞中晕占据分布的分析》，《天文学与天体物理学》638 (2020) A60。arXiv:2003.06255, DOI:10.1051/0004-6361/ 201937431。

[38] I. G. Alfaro, F. Rodriguez, A. N. Ruiz, H. E. Lupar- ello, D. G. Lambas，《极端密度环境中的星系如何分布于晕中：斯隆数字化巡天中晕占据分布的分析》，《天文学与天体物理学》665 (2022) A44。arXiv:2203.07526, DOI:10. 1051/0004-6361/202243542。

[39] A. M. Rodríguez-Medrano, D. J. Paz, F. A. Stasyszyn, F. Rodríguez, A. N. Ruiz, M. Merchán，《局部和大尺度效应对空洞星系天体物理学的影响》，皇家天文学会月刊 521 (1) (2023) 916–925。arXiv:2212.10594，DOI:10.1093/mnras/ stad623。

[40]J. Ridley，向日葵花盘中的堆积效率，数学生命科学 58 (1) (1982) 129–139。DOI：10.1016/0025-5564(82)90056-6。网址：https://www.sciencedirect.com/science/ article/pii/0025556482900566

[41] Á. González，《使用斐波那契网格和经纬度网格测量球面面积》，数学地球科学 42 (1) (2010) 49–64。arXiv:0912.4540, DOI:10.1007/s11004-009-9257-x。[42] P. J. E. Peebles，《宇宙大尺度结构》，1980。[43] R. W. Hockney、J. W. Eastwood，《使用粒子的计算机模拟》，1988。

[44]M. P. Allen、D. J. Tildesley，《液体的计算机模拟》，牛津大学出版社，2017。DOI：10.1093/oso/9780198803195.001.0001。网址：https://doi.org/10.1093/oso/ 9780198803195.001.0001

[45] A. Bondi，范德瓦耳斯体积和半径，《物理化学杂志》68 (3) (1964) 441–451。DOI：10. 1021/j100785a001。

[46]R. Pavanï, G. Ranghino，计算分子体积的一种方法，《计算机与化学》6 (3) (1982) 133–135。doi:https: // doi.org/10.1016/0097-8485(82)80006-5。URL https://www.sciencedirect.com/science/ article/pii/0097848582800065

[47]J. Buša, J. Džurina, E. Hayryan, S. Hayryan, C.-K. Hu, J. Plavka, I. Pokorný, J. Skˇrivánek, M.-C. Wu, ARVO：一种 Fortran 软件包，可通过解析方程计算重叠球体的溶剂可及表面积和排除体积，《计算物理学通讯》165 (1) (2005) 59–96。doi:10.1016/j.cpc.2004.08.002。URL https://www.sciencedirect.com/science/ article/pii/S0010465504004151

[48] J. Buša Jr, S. Hayryan, M.-C. Wu, J. Buša, C.-K. Hu，Arvo-cl：Arvo 软件包的 OpenCL 版本——通过解析方程高效计算蛋白质的可及表面积和排除体积的工具，《计算机物理通讯》183（2012）2494–2497。DOI:10.1016/j.cpc.2012.04.019。

[49] A. N. Cox，《艾伦天体物理量》，2000。

[50] M. Davis，P. J. E. Peebles，《星系红移巡天。V. 两点位置与速度相关函数》，《天体物理学杂志》267（1983）465–482。DOI:10.1086/160884。

[51] D. L. Tucker，A. Oemler, Jr.，R. P. Kirshner，H. Lin，S. A. Shectman，S. D. Landy，P. L. Schechter，V. Muller，S. Gottlober，J. Einasto，《拉斯坎帕纳斯红移巡天的星系-星系自相关函数》，《皇家天文学会月刊》285（1）（1997）L5–L9。arXiv:astro-ph/9611206, DOI:10.1093/mnras/285.1.L5。

[52] V. Springel，《宇宙学模拟代码 GADGET- 2》，《皇家天文学会月报》364 (4) (2005) 1105–1134。arXiv:astro-ph/

0505010，DOI:10.1111/j.1365-2966.2005.09655. x。

[53] V. Springel、R. Pakmor、O. Zier、M. Reinecke，使用 GADGET-4 代码模拟宇宙结构形成，《皇家天文学会月刊》506 (2) (2021) 2871–2949。arXiv:2010.03567，DOI:10.1093/mnras/stab1855。

[54]M. Schaller、J. Borrow、P. W. Draper、M. Ivkovic、S. McAlpine、B. Vandenbroucke、Y. Bahé、E. Chaikin、A. B. G. Chalk、T. K. Chan、C. Correa、M. van Daalen、W. Elbers、P. Gonnet、L. Hausammann、J. Helly、F. Huško、J. A. Kegerreis、F. S. J. Nobels、S. Ploeckinger、Y. Revaz、W. J. Roper、S. Ruiz-Bonilla、T. D. Sandnes、Y. Uyttenhove、J. S. Willis、Z. Xiang，SWIFT：用于天体物理学和宇宙学应用的现代高并行引力与光滑粒子流体动力学求解器，《皇家天文学会月刊》530 (2) (2024) 2378–2419。arXiv:2305.13380，DOI:10.1093/mnras/stae922。

[55] HDF Group，《分层数据格式》，第 5 版。网址：https://github.com/HDFGroup/hdf5

[56] L. Dagum、R. Menon，《OpenMP：共享内存编程的行业标准 API》，Computational Science & Engineering，IEEE 5 (1) (1998) 46–55。[57]消息传递接口论坛，《MPI：消息传递接口标准》第 5.0 版（2025 年 6 月）。网址：https://www.mpi-forum.org/docs/mpi-5.0/mpi50-report.pdf[58]D. Blackman、S. Vigna，《扰乱线性伪随机数生成器》，arXiv 电子预印本（2018），arXiv:1805.01407arXiv:1805.01407，DOI：10.48550/ arXiv.1805.01407。[59] XAMPPRocky 及贡献者，《Tokei：快速统计你的代码》（2026 年 9 月）。网址：https://github.com/XAMPPRocky/tokei[60] teryyin 及贡献者，《Lizard：简单的代码复杂度分析器》（2026 年 11 月）。网址：https://github.com/terryyin/lizard[61] P. Schechter，《星系光度函数的解析表达式》，《天体物理学杂志》203（1976）297–306。DOI：10.1086/154079。