

Lyman- α 森林及其与高红移星系在场水平有效场论中的交叉相关

罗杰·德·贝伦斯,^{1, 2} * 米哈伊尔·M·伊万诺夫,^{1, 2} † 詹姆斯·M·沙利文,^{1, 2} ‡ 陈士凡,^{3, 4} 和 秋津和之⁵ 麻省理工学院莱因韦伯理论物理中心, 美国马萨诸塞州剑桥市, 邮编 02139 2 美国国家科学基金会人工智能与基础相互作用 AI 研究所, 马萨诸塞州剑桥市, 邮编 02139 3 哥伦比亚大学物理系, 美国纽约州纽约市, 邮编 10027 4 美国国家航空航天局哈勃奖学金计划, 爱因斯坦研究员 5 高能加速器研究机构 (KEK) 粒子与核研究所理论中心, 日本茨城县筑波市, 邮编 305-0801

我们提出了一个用于 Lyman- α (Ly- α) 森林通量减量的场级微扰前向模型。我们在两个模拟套件上验证了该模型: 大体积的 AbacusSummit N 体模拟 (在暗物质场上绘制 Ly- α 森林) 以及 Sherwood 流体动力学模拟。在模拟的红移范围内 ($z = 2.0 - 3.2$), 模型的三维和一维功率谱与模拟的 Ly- α 场在 $k \approx 0.3 (1.0) \text{ h Mpc}^{-1}$ 以内的匹配度分别达到 1% (5%), 在与大质量暗物质晕的交叉相关方面表现相似。计数-单元统计在单元半径达到 $2 \text{ h}^{-1} \text{ Mpc}$ 时仍显示出良好的一致性。借助宇宙方差消除, 该模型能够实现 Ly- α 偏差参数的精确测量, 并稳健地检测到完整的二次视线偏差算子集合, 这与有效场论 (EFT) 中自然性的概念一致。我们量化了 Ly- α 森林的随机性 (类似于单晕项), 并发现其在大尺度上是白噪声 (尺度和方向无关的), 与有效场理论 (EFT) 的预测一致。我们进一步发现, 基于线性理论功率谱调制的表象通量功率谱模型, 即使在准线性尺度上也无法在场层面上有效描述。针对当前观测的暗能量光谱仪 (DESI), 我们生成了 Ly- α 森林的大尺度聚类模拟, 以验证宇宙学参数推断管线。展望其继任者 DESI-II, 我们生成了代表性样本的大体积莱曼断裂星系 (LBGs) 和 Ly- α 发射星系 (LAEs) 模拟, 这些模拟经过 Astrid 流体动力学模拟校准, 并与 $z = 3$ 的观测数据匹配, 从而能够对 Ly- α 森林和高红移星系数据进行联合分析。

目录

一、引言与主要结果	2
II. 构建场地级模型	4
三、模拟数据	8
四、结果	9
A. Ly- α 森林前向模型	9
B. Ly- α 森林与大质量晕的互相关	12
C. 误差功率谱	12
D. 互相关系数	13
E. 一维功率谱	16
V. 传递函数拟合	18
A. 高红移星系传递函数	21
VI. 微扰理论中的传递函数	21
VII. 大规模集群模拟	25
A. 验证程序: 全形状与 BAO 拟合	28
B. 在固定红移下的 Ly- α 森林模拟	29
C. Ly- α 森林与高红移星系之间的交叉相关	33

*belsunce@mit.edu †ivanov99@mit.edu ‡
jms3@mit.edu; 布林森奖学金研究员

八、总结与结论	34
致谢	39
A. 来自 Sherwood 模拟的传递函数	39
B. 不同 Abacus FPGA 模型上的字段级拟合	39
参考文献	42

一. 简介与主要结果

Lyman- α ($\text{Ly-}\alpha$) 森林由一系列出现在遥远类星体光谱中的吸收特征组成, 这些特征是由视线方向上的中性氢引起的。自20世纪90年代以来, 利用高分辨率仪器如高分辨率计量光谱仪 (HIRES; 1, 2) 和紫外及可见光计量光谱仪 (UVES; 3, 4) 进行的高分辨率观测, 使得对中性氢密度的小尺度波动进行了精确测量。这些观测确定了吸收气体存在于低密度、高度电离的星际介质 (IGM) 中, 从而在中性氢分布与潜在暗物质场之间提供了一个透明的联系。反过来, 这种联系使得对 $\text{Ly-}\alpha$ 森林进行精密模拟成为可能。由于中性气体与近似均匀的紫外背景处于光电离平衡, $\text{Ly-}\alpha$ 森林可以利用地基观测在广泛的红移范围 ($2 \leq z \leq 5$) 内探测从宇宙学尺度到Mpc尺度及以下的密度波动。高分辨率光谱使得通过测量视线方向 (或一维) 功率谱[5 - 12]能够深入分析小尺度区域 ($k_{\text{max}} < \sim 10 \text{ h Mpc}^{-1}$)。这些测量对广泛的基础物理非常敏感, 包括中微子性质[5, 8, 13 - 17]、原初黑洞[18 - 20]、暗物质模型[21 - 31]、电离IGM的热历史[32 - 43]、非最简宇宙学模型[44 - 46]以及谱指数的漂移[15, 47]。

在过去的二十年里, $\text{Ly-}\alpha$ 森林调查在光谱分辨率和所探测的总体宇宙体积方面都显著扩展。来自扩展的重子声波振荡光谱调查 (eBOSS; 48) 的大量中分辨率类星体光谱样本, 尤其是来自正在进行的暗能量光谱仪 (DESI; 49 - 52) 的样本, 现在能够在许多独立视线之间进行交叉相关分析。这些多光标图描绘了我们宇宙的大尺度结构, 包含的信息类似于高红移光谱星系和类星体调查提供的信息, 但具有显著更宽的动态范围[49, 53, 54]。

对 $\text{Ly-}\alpha$ 森林的观测通过测量重子声学振荡 (BAO) 特征[51, 55 - 58]以及三维相关函数的宽带形状[53, 56, 59 - 61]来约束宇宙的膨胀历史。中性氢与潜在暗物质之间的映射促使人们开发高保真度的 $\text{Ly-}\alpha$ 森林数值模拟[62 - 72]。这些模拟通常用于校准基于线性理论的模型, 并辅以考虑非线性增长、压力平滑和视线速度展宽的现象学拟合函数[73 - 75]。

在由类星体视线的有限采样所设定的当前精度下 (即在银河的射击噪声限制区域), 此类建模方法已经产生了稳健的宇宙学约束 [50, 51, 61, 76]。然而, 预计 DESI 在其运行期间将观测高达一百万条类星体光谱, 预测显示在结合所有示踪器和红移区间时, 累计精度将低于 0.2% [49]。实现这一精度水平需要对理论系统误差进行精细控制。事实上, 最近的研究表明, 现象学的 $\text{Ly-}\alpha$ 森林模型会使推断的 BAO 缩放参数产生偏差——这是使用 DESI 约束宇宙膨胀历史的主要观测量——偏差水平约为 0.3% [77, 78]。这些发现表明, 现有的建模框架正接近其极限, 这推动了为当前的 DESI 数据以及即将到来的调查 (如 DESI-II、WEAVE-QSO 调查[79]、主焦点光谱仪[PFS; 80]和 4MOST[81]) 开发更准确和更稳健的理论描述。

通过大尺度结构有效场论 (EFT) 的框架, 可以消除对 BAO 推断的偏差, 最近该框架已扩展至 $\text{Ly-}\alpha$ 森林 [77, 85 - 87]。EFT 形式主义通过仅纳入与示踪物相关的对称性, 对大尺度动力学提供微扰描述 [88 - 91]。在 $\text{Ly-}\alpha$ 森林的情况下, 这些对称性包括等效原理和沿视线方向 \hat{z} 的旋转不变性, 对应于 $\text{SO}(2)$ 群 [15, 85, 86, 92 - 95]。虽然这为通过单环功率谱直接约束宇宙学参数铺平了道路 (在低红移星系调查的背景下, 参见例如 [96 - 101]), 但对宇宙学分析来说, 一个关键挑战是所涉及的尺度范围很大, 需要大体积、高分辨率的模拟来验证推断流程。

高分辨率的流体动力学模拟能够在小到中等尺度上准确捕捉所涉及的物理过程, 但在捕捉的同时运行一系列涵盖广泛宇宙学和天体物理参数的模拟

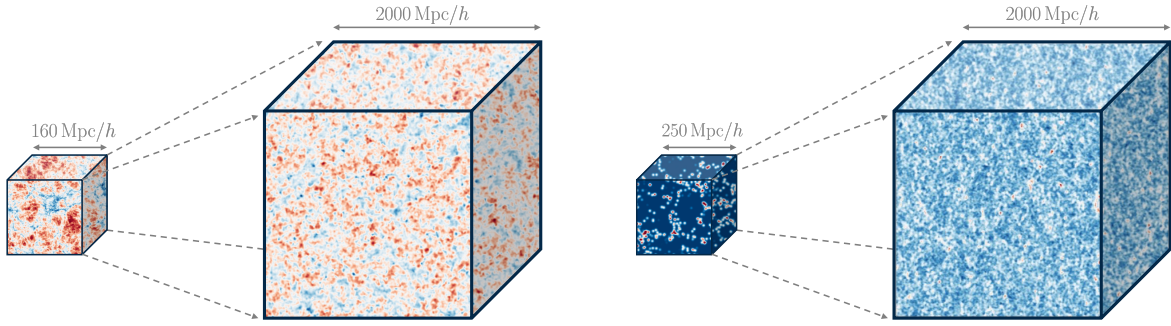


图 1. 结果摘要：我们在场域层面将解析的、微扰的前向模型拟合到高分辨率流体动力学模拟，并使用它生成大体积实现。在左图中，我们拟合了小体积 ($V = (160 \text{ h}^{-1} \text{ Mpc})^3$) 的 Sherwood 氢森林流体动力学模拟 [82]。在右图中，类似的框架被应用于高红移星系模拟（如 Lyman- α 发射体和 Lyman 跃迁断裂星系），来自 Astrid 的 ($250 \text{ h}^{-1} \text{ Mpc})^3$ 体积 [83, 84]。这些校准后的模型能够生成大体积模拟（此处： $V = (2000 \text{ h}^{-1} \text{ Mpc})^3$ ），在图中显示为更大的三维盒子。红色（蓝色）区域对应过密（欠密）区域。当前向模型与输入的流体动力学模拟使用相同的初始条件 (ICs) 时，场域层面的拟合能够受益于宇宙方差的抵消。巧合的是，两个使用相同初始条件在不同模拟上校准的模拟可以用于 Ly- α 森林和高红移星系的交叉相关分析——这是 DESI-II 的一个关键科学驱动因素。

宇宙学体积的计算成本高得令人望而却步。为减轻这一问题，已经开发了几种方法，将观测到的通量与大规模聚类模拟的基础物质场联系起来 [78, 102 – 109]。特别是，基于深度学习的重建方法 [110 – 114] 在中到小尺度上取得了良好效果，但无法完全捕捉长波长的准线性模式。另一种途径是直接预测如功率谱等统计摘要的仿真器 [115 – 118]。虽然这些方法在广泛的尺度范围内能够达到百分比级的准确度，但它们依赖于 (i) 小体积流体动力学模拟的功率谱拟合，而这些模拟受到宇宙方差的影响；(ii) 并且仅限于功率谱。虽然配对固定 (paired-fixed) 模拟可以降低宇宙方差的限制 [119]，但由此得到的仿真器不能直接推广到其他统计摘要。然而，场级的前向模型需要匹配所有傅里叶模态的所有振幅和相位——这比比较摘要统计量对理论框架的检验更严格——同时还要产生场的高阶矩。

为了实现来自 DESI 和 DESI-II 的 Ly- α 森林的宇宙学分析，并纳入二点及最终的三点统计，我们提出了一个计算上高效的框架，用于生成基于高保真流体动力学模拟校准的大体积模拟。这项工作为文献 [87] 提供了理论背景，该文献首次将该技术引入 Ly- α 森林，并扩展了早期在微扰、场级建模方面的发展 [120 – 125]。相同的框架还可以推广到高红移星系的模拟，使得验证 Ly- α 森林与星系位置之间的交叉相关测量成为可能。这种联合分析有助于打破增长率 f 与其他很难约束的速度梯度偏差之间的简并性 [126, 127]。更广泛地说，这里开发的模拟为验证端到端的全形状推断流程以及量化关键物理系统效应提供了一个受控环境，包括 BAO 特征的位移 [77, 128]。

我们使用两组互补的模拟来验证我们的微扰前向模型。首先，我们采用 AbacusSummit 系列的大体积 N 体模拟，并在其上绘制了 Ly- α 森林，覆盖体积为 $V = 23 (\text{h}^{-1} \text{ Gpc})^3$ [102]。其次，我们使用两组 Sherwood 流体动力学模拟，覆盖体积分别为 $V = 160^3 (\text{h}^{-1} \text{ Mpc})^3$ 和 $V = 80^3 (\text{h}^{-1} \text{ Mpc})^3$ [82]。进行场级拟合的一个关键优势是能够抵消宇宙方差，即使对于小体积模拟，也可以对宇宙学参数和干扰参数进行严格约束。

在这项工作中，我们提出了两个主要结果。首先，我们直接在场层次上使用微扰偏差展开来拟合和预测偏差传递函数的形状。其次，利用得到的场层次拟合结果，我们生成大规模聚类模拟。这些模拟的示例如图 1 所示，左面板显示了 Ly- α 森林，右面板显示了高红移星系。这里开发的方法对于即将到来的宇宙学分析至关重要，因为现有的大规模聚类模拟 [129, 130] 在针对当前 Ly- α 宽带形状分析基准 $r \approx 25 \text{ h}^{-1} \text{ Mpc}$ 小一个数量级的尺度时已经接近其极限 [61]。特别是由于 Ly- α 森林是探测高红移宇宙 ($2 \lesssim z \lesssim 5$) 的唯一途径之一，为未来的下一代测量如 DESI-II 和 Spec-S5 [131] 做准备。

本文提供了在参考文献 [87] 的一篇快报中提出的 Ly- α 前向模型的理论基础，组织结构如下：我们在第 II 节回顾了 Ly- α 森林和晕（作为高红移星系的代理）在红移空间中的微扰前向模型。在第 III 节中，我们介绍了使用的合成数据。我们在第 IV 节评估了微扰前向模型的性能，并研究了从我们场级拟合中获得的传递函数。

第五节。在第六节中，我们将测量的传递函数与理论上预期的传递函数进行比较，并在第七节中使用这些传递函数创建涵盖宇宙卷的大片聚类模拟。我们在第八节总结并讨论未来的工作。

二. 建立田间级模型

宇宙学 Ly- α 森林相关性的描述从线性理论模型 [73, 92] 开始，

$$\delta_F^{\text{lin}}(\mathbf{k}, z) = (b_1 - b_\eta f \mu^2) \delta_1(\mathbf{k}, z), \quad (1)$$

其中 b_1, b_η 是线性偏置参数， $f = d \ln D_+ / d \ln \mu$ 是线性密度场，可以重写为

$$\delta_1(\mathbf{k}, z) = \delta_{1,0}(\mathbf{k}) D_+(z), \quad (2)$$

其中 $\delta_{1,0}(\mathbf{k})$ 是外推到红移为零的初始条件场， $D_+(z)$ 是线性增长因子。在模拟中， $\delta_{1,0}(\mathbf{k})$ 是从初始条件的线性物质功率谱生成的随机标量场。为了简洁起见，我们将省略 δ_1 的时间依赖性，隐含假设该量总是在模拟红移 z 处进行评估。方程 (1) 中的模型重现了众所周知的通量功率谱线性理论模型，

$$\begin{aligned} \langle \delta_F(\mathbf{k}) \delta_F(\mathbf{k}') \rangle &= (2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') \langle \delta_F(\mathbf{k}) \delta_F(\mathbf{k}') \rangle' = (2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') P(k, \mu), \\ P_{11}(k) &\underbrace{=}_{\text{linear}} (b_1 - f b_\eta \mu^2)^2 P_{\text{lin}}(k), \end{aligned} \quad (3)$$

其中 $P_{\text{lin}}(k)$ 是线性 matter power spectrum (evaluated at redshift z) where we use $\int_{\mathbf{k}} = \int d^3k / (2\pi)^3$. The 上述系数 $(b_1 - f b_\eta \mu^2)^2$ 是众所周知的星系 Kaiser 因子的推广 [73, 92, 132]，通过设定 $b_\eta = -1$ 可以恢复。由此讨论可以清楚地看出，在线性近似下，Ly- α 场的所有相位都被场 $\delta_1(\mathbf{k})$ 捕获。因此，将方程 (1) 的理论模型拆分为显式显示振幅和相位依赖的部分是很方便的。将 Kaiser 因子吸收到方程 (1) 中作为动量依赖的传递函数会得到

$$\delta_F^{\text{lin}}(\mathbf{k}, z) = \beta_1(k, \mu) \delta_1(\mathbf{k}, z). \quad (4)$$

这种方法的优势将在非线性层面显现出来，此时转移函数 β_1 也将考虑一环修正。

线性理论模型在描述模拟 Ly- α 场 δ_{truth}^F 方面的成功 可以使用误差来估计 功率谱

$$P_{\text{err}}(\mathbf{k}) = \langle |\delta_{\text{truth}}^F(\mathbf{k}) - \delta_F^{\text{lin}}(\mathbf{k})|^2 \rangle'. \quad (5)$$

在非常大尺度上（即在 $k \rightarrow 0$ 的极限下），我们模型的非线性修正预计很小，因此模型中唯一预期的误差来源应当是随机场 ϵ ，

$$\delta_F(\mathbf{k}) = \delta_F^{\text{lin}}(\mathbf{k}) + \epsilon(\mathbf{k}), \quad (6)$$

从定义上讲，这与 δ_{lin}^F 无关 由宇宙学涨落产生。一般来说， $\epsilon(\mathbf{k})$ 是 Ly- α 由与大尺度初始条件无关的小尺度过程产生的通量减量分量。在星系和晕的背景下，场 $\epsilon(\mathbf{k})$ 捕捉到由追踪体离散性引起的射击噪声。使用这个类比，Ly- α 森林的随机场可以被认为源自吸收线的离散性。由于吸收线的数量众多，预计其非常小。然而，这一逻辑并不完全正确，因为它假设离散性是在吸收线的层面产生的。可以认为 Ly- α 吸收是由中性氢云产生的，而这些氢云的分布也具有随机成分。Ly- α 云的这种离散性可以作为理解森林随机性的第一个代理。

在 LSS 的 EFT 框架内 [89 - 91, 94, 133 - 135]，人们使用微扰泰勒展开来参数化未知函数。在这种方法中，误差功率谱在 $k \rightarrow 0$ 极限下具有以下展开式：

$$P_{\text{err}}(k, \mu) = n_0(1 + \alpha_1 k^2 + \alpha_2 k^2 \mu^2 + \dots), \quad (7)$$

其中 n_0 、 $\alpha_{1,2}$ 是由 Ly- α 物理决定的维度常数。对于星系， n_0 可以在领先阶近似为 $1/\bar{n}$ ，其中 $\bar{n} = N/V$ 是星系的数密度（ N 是共动体积 V 中星系的数量）。在这种情况下，相关的距离尺度是各个星系之间的平均间距 $R = (V/N)^{1/3}$ 。将相同的推理应用到 Ly- α 林，并为了简化假设单个示踪物是一个中性氢云，其与相邻云的间隔约为 $0.5 h^{-1} \text{Mpc}$ ，我们得到一个估计

$$n_0 \sim R^3 \sim 0.1 [h^{-1} \text{Mpc}]^3 \quad \alpha_{1,2} \sim R^2 \sim 0.3 [h^{-1} \text{Mpc}]^2 . \quad (8)$$

正如我们将看到的，这一天 $\delta_{\text{lin}}(k_{\text{NL}}, z) = 1$ ， $\Rightarrow k_{\text{NL}} \approx 5 h \text{Mpc}^{-1}$ at $z = 2.8$ 。不同于微扰偏差展开中出现的尺度。后者可以被估计为非线性尺度，在该尺度上物质密度波动达到数量级一 [136]。

$$\frac{k_{\text{NL}}^3}{2\pi^2} P_{\text{lin}}(k_{\text{NL}}, z) = 1, \quad \Rightarrow \quad k_{\text{NL}} \approx 5 h \text{Mpc}^{-1} \quad \text{at} \quad z = 2.8. \quad (9)$$

如果理论模型是准确的，我们期望在大尺度上恢复由方程 (7) 提示的尺度依赖性。如果对天真的线性模型 (4) 存在显著修正，这将产生方程 (7) 没有捕捉到的明显尺度和方向依赖性。我们预计会有哪些修正？

首先，我们模拟中的暗物质粒子位移很大，必须以非微扰方式处理。如果不加以考虑，它们会对重子声学振荡产生较大扭曲，从而导致线性模型与密度场的实际相位不匹配。这种效应是可以理解的，并且可以通过使用由泽尔多维奇位移移动的线性密度场 $\tilde{\delta}_1$ 代替 δ_1 来加以修正 [137]。在下文中， $\tilde{\delta}_1$ 将被称为移位线性场。

$$\tilde{\delta}_1(\mathbf{k}) = \int d^3 \mathbf{q} \delta_1(\mathbf{q}) e^{-i\mathbf{k} \cdot (\mathbf{q} + \psi_1(\mathbf{q}) + f\hat{z}(\psi_1(\mathbf{q}) \cdot \hat{z}))}, \quad (10)$$

其中 \mathbf{q} 表示拉格朗日空间（初始）坐标， ψ_1 是泽尔多维奇位移 $\psi_1(\mathbf{q}) = \int d^3 k e^{i\mathbf{q} \cdot \mathbf{k}} \frac{i\mathbf{k}}{k^2} \delta_1(\mathbf{k})$ 。

$$\psi_1(\mathbf{q}) = \int d^3 k e^{i\mathbf{q} \cdot \mathbf{k}} \frac{i\mathbf{k}}{k^2} \delta_1(\mathbf{k}). \quad (11)$$

请注意，上述 $\tilde{\delta}_1$ 在线性场 δ_1 中具有无限的泰勒展开。在微扰理论中，可以将其写为：

$$\tilde{\delta}_1 = \sum_{n=1}^3 \sum_{\substack{\mathbf{k}_1, \dots, \mathbf{k}_n \\ \mathbf{k}_1 + \dots + \mathbf{k}_n = \mathbf{k}}} \delta_1(\mathbf{k}_i) + (2\pi)^3 \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{1\dots n}) \tilde{K}_n(\mathbf{k}_1, \dots, \mathbf{k}_n), \quad (12)$$

其中 $\tilde{K}_1(\mathbf{k}) = 1$ 并且

$$\tilde{K}_2(k_1, k_2) = \mathbf{k} \cdot \mathbf{k}_1 \frac{k_2}{2k_2^2} + \frac{(f\mu k)}{2} \left(\frac{k_{1z}}{k_1^2} + \frac{k_{2z}}{k_2^2} \right), \quad \text{etc.} \quad (13)$$

其中 $k_{iz} = (\mathbf{k}_i \cdot \hat{z})$ ，且 $\mathbf{k} \equiv \mathbf{k}_1 + \dots + \mathbf{k}_n$ 对于第 n 个核。这一展开与 Zel'dovich 物质密度场的展开非常相似 [137 – 140]

$$\delta_Z(\mathbf{k}) = \int d^3 \mathbf{q} e^{-i\mathbf{k} \cdot (\mathbf{q} + \psi_1(\mathbf{q}) + f\hat{z}(\psi_1(\mathbf{q}) \cdot \hat{z}))} = \sum_{n=1}^{\infty} \sum_{\substack{\mathbf{k}_1, \dots, \mathbf{k}_n \\ \mathbf{k}_1 + \dots + \mathbf{k}_n = \mathbf{k}}} \delta_1(\mathbf{k}_i) (2\pi)^3 \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_{1\dots n}) F_n^{\text{ZA}}(\mathbf{k}_1, \dots, \mathbf{k}_n), \quad (14)$$

$F_n^{\text{ZA}} = 1 + f(\mathbf{k} \cdot \hat{z})^2$ ， $F_2^{\text{ZA}} = 1/2$

$$\frac{(\mathbf{k} \cdot \mathbf{k}_1)(\mathbf{k} \cdot \mathbf{k}_2)}{(k_1^2 k_2^2)} + \frac{f(k_1 \cdot \hat{z})k_{2z}(k_{1z})k_2}{2k_1 k_2^2} + \frac{f(k_2 \cdot \hat{z})k_{1z}(k_{2z})k_1}{2k_1 k_2^2} + \frac{f^2(\mu k)^2}{2k^2 k_1^2 k_2^2} k_{1z} k_{2z}, \quad \text{etc.} \quad (15)$$

这些构建模块将在我们未来的讨论中非常有用。Zel'dovich 位移引入的变化会产生更高阶的非线性效应。在接下来的内容中，我们将更系统地讨论这些效应。

偏置展开中的非线性是第二个重要的修正来源。在微扰欧拉偏置中这些由……给出

$$\delta_F(\mathbf{k}) = \sum_{n=1}^{\infty} \delta_F^{(n)} = \sum_{n=1}^{\infty} \left[\prod_{j=1}^n \int \frac{d^3 \mathbf{k}_j}{(2\pi)^3} \delta_1(\mathbf{k}_j) \right] K_n(\mathbf{k}_1, \dots, \mathbf{k}_n) (2\pi)^3 \delta_D^{(3)}(\mathbf{k} - \mathbf{k}_1 - \dots - \mathbf{k}_n), \quad (16)$$