

Lyman- α Forest and its Cross-Correlation with High-Redshift Galaxies in Effective Field Theory at the Field Level

Roger de Belsunce,^{1,2,*} Mikhail M. Ivanov,^{1,2,†} James M. Sullivan,^{1,2,‡} Shi-Fan Chen,^{3,4} and Kazuyuki Akitsu⁵

¹Center for Theoretical Physics – a *Leinweber Institute*, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, MA 02139, USA

³Department of Physics, Columbia University, New York, NY 10027, USA

⁴NASA Hubble Fellowship Program, Einstein Fellow

⁵Theory Center, Institute of Particle and Nuclear Studies,

High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan

We present a field-level perturbative forward model for the Lyman- α (Ly- α) forest flux decrement. We validate it on two simulation suites: large-volume ABACUS SUMMIT N -body simulations with the Ly- α forest painted onto the dark matter field, and the Sherwood hydrodynamic simulations. Across the redshift range of the simulations ($z = 2.0$ – 3.2), the 3D and 1D power spectra of the model match the simulated Ly- α fields at the 1% (5%) level up to $k \approx 0.3$ (1.0) $h \text{ Mpc}^{-1}$, with similar performance for the cross-correlation with massive dark matter halos. The counts-in-cells statistic shows excellent agreement down to cell radii of $2 h^{-1} \text{ Mpc}$. Leveraging cosmic variance cancellation, the model enables precision measurements of Ly- α bias parameters and robustly detects the full set of quadratic line-of-sight bias operators, consistent with the notion of naturalness in effective field theory (EFT). We quantify the stochasticity of the Ly- α forest (the analog to the one-halo term), and find it to be white (scale- and orientation-independent) on large scales, matching EFT predictions. We further find that phenomenological flux power spectrum models, based on modulations of the linear-theory power spectrum, fail at the field level even on quasi-linear scales. For the currently observing Dark Energy Spectroscopic Instrument (DESI), we generate large-scale clustering mocks of the Ly- α forest to validate cosmological parameter inference pipelines. Looking ahead to its successor, DESI-II, we produce large-volume mocks of representative samples of Lyman-break galaxies (LBGs) and Ly- α emitters (LAEs), calibrated on Astrid hydrodynamic simulations and matched to observations at $z = 3$, enabling joint analyses of Ly- α forest and high-redshift galaxy data.

CONTENTS

I. Introduction and key results	2
II. Building the field-level model	4
III. Simulation data	8
IV. Results	9
A. Ly- α forest forward model	9
B. Cross-correlation of the Ly- α forest with massive halos	12
C. The error power spectrum	12
D. Cross-correlation coefficient	13
E. One-dimensional power spectrum	16
V. Transfer function fits	18
A. High-redshift galaxy transfer functions	21
VI. Transfer functions in perturbation theory	21
VII. Large-scale clustering mocks	25
A. Validation procedure: Full-shape & BAO fits	28
B. Ly- α forest mocks at fixed redshift	29
C. Cross-correlations between the Ly- α forest and high-redshift galaxies	33

* belsunce@mit.edu

† ivanov99@mit.edu

‡ jms3@mit.edu; Brinson Prize Fellow

Lyman- α 森林及其与高红移星系在场层次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 NASA 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莱曼- 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 AbacusSummit N 体模拟 (在暗物质场上绘制 Ly- α 森林) 和 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) $h \text{ Mpc}^{-1}$ 处分别匹配至 1% (5%) 水平，与大质量暗物质晕的交叉相关也表现出类似的性能。计数进单元统计在单元半径为 $2 h^{-1} \text{ Mpc}$ 时表现出极好的匹配。利用宇宙方差抵消，该模型能够精确测量 Ly- α 偏差参数，并稳健地检测完整的一组二次视线偏差算子，这与有效场论 (EFT) 中的自然性概念一致。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，也无法在场层面上准确描述。对于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
二、构建田间水平模型	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. 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

VIII. Summary and Conclusions

34

Acknowledgments

39

A. Transfer functions from Sherwood simulations

39

B. Field-level fits on different Abacus FGPA mocks

39

References

42

I. INTRODUCTION AND KEY RESULTS

The Lyman- α (Ly- α) forest consists of a series of absorption features in the spectra of distant quasars, produced by intervening neutral hydrogen along the line of sight. Since the 1990s, high-resolution observations with instruments such as the High Resolution Echelle Spectrometer [HIRES; 1, 2] and the Ultraviolet and Visual Echelle Spectrograph [UVES; 3, 4] have enabled precise measurements of small-scale fluctuations in the neutral hydrogen density. These observations established that the absorbing gas resides in the low-density, highly ionized intergalactic medium (IGM), providing a transparent link between the neutral hydrogen distribution and the underlying dark matter field. This connection, in turn, enables precision simulations of the Ly- α forest. Because the neutral gas is in photoionization equilibrium with an approximately uniform ultraviolet background, the Ly- α forest probes density fluctuations from cosmological down to Mpc scales and below over a wide redshift range ($2 \leq z \leq 5$) using ground-based observations.

High-resolution spectra have enabled analyses deep into the small-scale regime ($k_{\max} \lesssim 10 h \text{ Mpc}^{-1}$) through measurements of the line-of-sight (or one-dimensional) power spectrum [5–12]. These measurements are sensitive to a broad range of fundamental physics, including neutrino properties [5, 8, 13–17], primordial black holes [18–20], dark matter models [21–31], the thermal history of the ionized IGM [32–43], non-minimal cosmological models [44–46], and the running of the spectral index [15, 47].

Over the past two decades, Ly- α forest surveys have expanded dramatically in both spectral resolution and the total cosmological volume probed. Large samples of medium-resolution quasar spectra from the extended Baryon Oscillation Spectroscopic Survey [eBOSS; 48] and, in particular, from the ongoing Dark Energy Spectroscopic Instrument [DESI; 49–52] now enable cross-correlation analyses across many independent lines of sight. These multi-skewer maps trace the large-scale structure of our Universe, containing information similar to that provided by spectroscopic galaxy and quasar surveys at high redshift, but with a significantly wider dynamic range [49, 53, 54].

Observations of the Ly- α forest constrain the expansion history of the Universe through measurements of the baryon acoustic oscillation (BAO) feature [51, 55–58] and through the broadband shape of the three-dimensional correlation function [53, 56, 59–61]. The mapping between neutral hydrogen and underlying dark matter has motivated the development of high-fidelity numerical simulations of the Ly- α forest [62–72]. These simulations are commonly used to calibrate linear-theory-based models augmented by phenomenological fitting functions that account for nonlinear growth, pressure smoothing, and line-of-sight velocity broadening [73–75].

At the current precision set by the finite sampling of quasar sightlines (i.e., the shot-noise limited regime for galaxies), such modeling approaches have yielded robust cosmological constraints [50, 51, 61, 76]. However, DESI is expected to observe up to one million quasar spectra over its lifetime, with forecasts indicating a cumulative precision below the 0.2% level when combining all tracers and redshift bins [49]. Achieving this level of precision requires exquisite control over theoretical systematics. Indeed, recent studies have demonstrated that phenomenological Ly- α forest models bias the inferred BAO scaling parameters – the primary observables for constraining the cosmic expansion history with DESI – at the 0.3% level [77, 78]. These findings indicate that existing modeling frameworks are approaching their limits, motivating the development of more accurate and robust theoretical descriptions for current DESI data and forthcoming surveys such as DESI-II, the WEAVE-QSO survey [79], the Prime Focus Spectrograph [PFS; 80] and 4MOST [81].

One can remove the bias on the BAO inference within the framework of the effective field theory (EFT) of large-scale structure, which has recently been extended to the Ly- α forest [77, 85–87]. The EFT formalism provides a perturbative description of large-scale dynamics by incorporating only the symmetries relevant to the tracer [88–91]. In the case of the Ly- α forest, these include the equivalence principle and rotational invariance around the line-of-sight direction \hat{z} , corresponding to the $SO(2)$ group [15, 85, 86, 92–95]. Whilst this paves the way to directly constrain cosmological parameters through the one-loop power spectrum (in the context of low-redshift galaxy surveys, see, e.g., [96–101]) a key challenge for cosmological analyses is that the large range of scales involved require large-volume, high-resolution simulations to validate inference pipelines.

High-resolution hydrodynamic simulations accurately capture the involved physics on small to intermediate scales but running a set of simulations covering a wide range of cosmological and astrophysical parameters, whilst capturing

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 FGPA 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 (Ly- α) 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Ly- α 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，Ly- α 森林可以利用地基观测在广泛的红移范围内 ($2 \leq z \leq 5$) 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 ($k_{\max} < \sim 10 h \text{ Mpc}^{-1}$)。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]、非最小宇宙学模型[44–46]以及谱指数的漂移[15, 47]。

在过去的二十年里，Ly- α 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) 的大量样本，现在使得跨多个独立视线的交叉相关分析成为可能。这些多光束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下光谱学星系和类星体调查提供的信息，但动态范围明显更广 [49, 53, 54]。

对Ly- α 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Ly- α 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、压力平滑和视线方向速度展宽的经验拟合函数[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 Ly- α 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 Ly- α 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 Ly- α 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 $SO(2)$ 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大范围尺度需要大体积、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，但在覆盖广泛的宇宙学和天体物理参数范围的同时进行一系列模拟

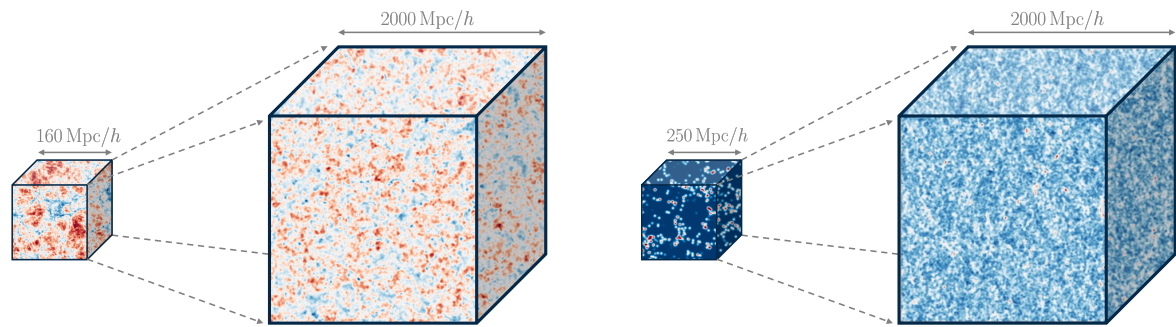


FIG. 1. **Summary of Results:** We fit an analytic, perturbative forward model at the field level to high-resolution hydrodynamic simulations, and use it to generate large-volume realizations. In the left panel, we fit small-volume ($V = (160 h^{-1} \text{Mpc})^3$) Sherwood hydrodynamic simulations of the Ly- α forest [82]. In the right panel, a similar framework is applied to high-redshift galaxy simulations (such as Lyman- α emitters and Lyman-break galaxies) from Astrid in a $(250 h^{-1} \text{Mpc})^3$ volume [83, 84]. These calibrated models enable the generation of large-volume simulations (here: $V = (2000 h^{-1} \text{Mpc})^3$) here depicted as larger 3D boxes. Red (blue) regions correspond to over- (under-)dense regions. The field-level fits benefit from cosmic variance cancellation when using the same set of initial conditions (ICs) for the forward model as for the input hydrodynamic simulation. Coincidentally, two simulations calibrated on different simulations using the same ICs can be used for cross-correlation analyses of the Ly- α forest and high-redshift galaxies – a key science driver for DESI-II.

cosmological volumes is prohibitively expensive. To mitigate this, several approaches have been developed to connect the observed flux to the underlying matter field for large-scale clustering simulations [78, 102–109]. In particular, deep learning-based reconstruction methods [110–114] yield promising results at intermediate to small scales but do not fully capture the long-wavelength quasi-linear modes. An alternative avenue are emulators that directly predict a summary statistic such as the power spectrum [115–118]. Whilst these offer per cent level accuracy for a wide range of scales they depend on (i) fits to power spectra of small-volume hydrodynamic simulations suffering from cosmic variance; and (ii) are restricted to the power spectrum. Whilst paired-fixed simulations reduce cosmic variance limitations [119], the resulting emulators cannot directly be generalized to other summary statistics. A forward model at the field level, however, needs to match *all* the amplitudes and phases of *all* Fourier modes – a more stringent test of the theoretical framework than comparing summary statistics – all whilst yielding higher-order moments of the field.

To enable cosmological analyses of the Ly- α forest from DESI and DESI-II that incorporate two- and, eventually, three-point statistics, we present a computationally efficient framework for generating large-volume simulations calibrated on high-fidelity hydrodynamic simulations. This work provides the theoretical background for Ref. [87], where this technique was first introduced for the Ly- α forest, and extends earlier developments in perturbative, field-level modeling [120–125]. The same framework generalizes to simulations of high-redshift galaxies, enabling validation of cross-correlation measurements between the Ly- α forest and galaxy positions. Such joint analyses help break the degeneracy between the growth rate f and the otherwise poorly constrained velocity-gradient bias [126, 127]. More broadly, the simulations developed here provide a controlled environment for validating end-to-end full-shape inference pipelines and for quantifying key physical systematics, including shifts of the BAO feature [77, 128].

We validate our perturbative forward model using two complementary sets of simulations. First, we employ the large-volume N -body simulations from the ABACUSSUMMIT suite, onto which the Ly- α forest is painted, spanning a volume of $V = 2^3 (h^{-1} \text{Gpc})^3$ [102]. Second, we use two sets of the Sherwood hydrodynamic simulations, which cover a volume of $V = 160^3 (h^{-1} \text{Mpc})^3$ and $V = 80^3 (h^{-1} \text{Mpc})^3$ [82]. A key advantage of performing field-level fits is the resulting cancellation of cosmic variance, which enables tight constraints on both cosmological and nuisance parameters even for small-volume simulations.

In this work, we present two main results. First, we fit and *predict* the shapes of the bias transfer functions using a perturbative bias expansion directly at the field level. Second, using the resulting field-level fits, we generate large-scale clustering simulations. Examples of these simulations are shown in Fig. 1, with the Ly- α forest displayed in the left panel and high-redshift galaxies in the right panel. The methodology developed here is critical for forthcoming cosmological analyses, as existing large-scale clustering mocks [129, 130] are approaching their limits when targeting scales an order of magnitude smaller than the current baseline for analyses of the Ly- α broadband shape of $r \approx 25 h^{-1} \text{Mpc}$ [61]. Especially, since the Ly- α forest is one of the only ways of probing the high-redshift Universe ($2 \lesssim z \lesssim 5$) ahead of next-generation surveys such as DESI-II and Spec-S5 [131].

This paper provides the theoretical foundation of the Ly- α forward model presented in a *Letter* in Ref. [87] and is organized as follows: We review the perturbative forward model of the Ly- α forest and halos (as proxies for high-redshift galaxies) in redshift space in Sec. II. In Sec. III we present the used synthetic data. We assess the performance of our perturbative forward model in Sec. IV and investigate the obtained transfer functions from our field-level fits in

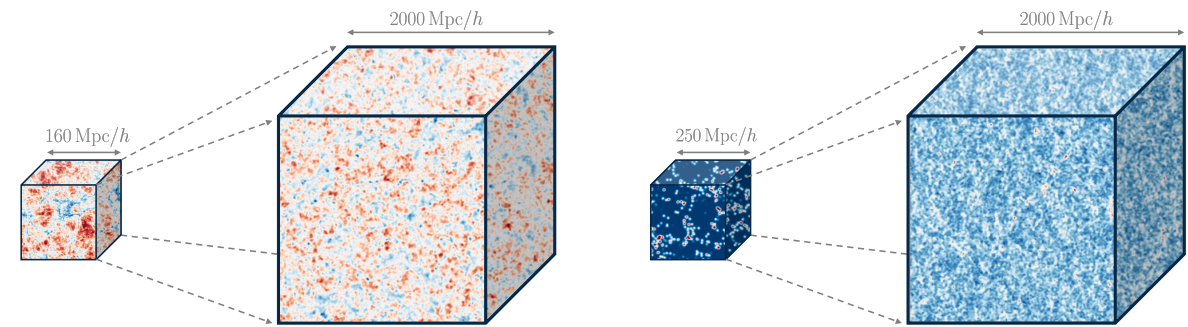


图 1. 结果摘要：我们在场域层面将解析的、微扰的前向模型拟合到高分辨率流体动力学模拟，并使用它生成大体积实现。在左图中，我们拟合了小体积 ($V = (160 h^{-1} \text{Mpc})^3$) 的 Sherwood 氢森林流体动力学模拟 [82]。在右图中，类似的框架被应用于高红移星系模拟（如 Lyman- α 发射体和 Lyman 跃迁断裂星系），来自 Astrid 的 $(250 h^{-1} \text{Mpc})^3$ 体积 [83, 84]。这些校准后的模型能够生成大体积模拟（此处： $V = (2000 h^{-1} \text{Mpc})^3$ ），在图中显示为更大的三维盒子。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。

为了实现对来自 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 $V = 2^3 (h^{-1} \text{Gpc})^3$ [102]。其次，我们使用两组 Sherwood 流体动力学模拟，覆盖体积分别为 $V = 160^3 (h^{-1} \text{Mpc})^3$ 和 $V = 80^3 (h^{-1} \text{Mpc})^3$ [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] 在针对比当前 Ly- α 宽带形状分析基准 $r \approx 25 h^{-1} \text{Mpc}$ 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，我们介绍了所使用的合成数据。我们在第 IV 节评估了微扰前向模型的性能，并研究了从我们的场级拟合中获得的传递函数。

Sec. V. In Sec. VI we compare the measured transfer functions to theoretically expected ones and use these transfer functions to create large-scale clustering mocks encompassing cosmological volumes in Sec. VII. We conclude and discuss future work in Sec. VIII.

II. BUILDING THE FIELD-LEVEL MODEL

The description of the cosmological Ly- α forest correlations begins with the linear theory model [73, 92],

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

where b_1, b_η are linear bias parameters, $f = d \ln D_+ / d \ln a$ is the logarithmic growth factor, $\mu \equiv k_{\parallel} / k = \hat{\mathbf{k}} \cdot \hat{\mathbf{z}}$ is the cosine of the angle to the line-of-sight, $\hat{\mathbf{z}}$, and δ_1 the linear density field, which can be rewritten as

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

where $\delta_{1,0}(\mathbf{k})$ is the initial condition field extrapolated to redshift zero and $D_+(z)$ is the linear growth factor. In simulations, $\delta_{1,0}(\mathbf{k})$ is the random scalar field generated from the linear matter power spectrum of initial conditions. For brevity, we will suppress the time dependence of δ_1 , implicitly assuming that this quantity is always evaluated at the simulation redshift z . The model in Eq. (1) reproduces the well-known linear theory model for the flux power spectrum,

$$\langle \delta_F(\mathbf{k}) \delta_F(\mathbf{k}') \rangle = (2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') \langle \delta_F(\mathbf{k}) \delta_F(\mathbf{k}') \rangle' = (2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') P(k, \mu), \quad (3)$$

$$P_{11}(k) \underbrace{=}_{\text{linear}} (b_1 - f b_\eta \mu^2)^2 P_{\text{lin}}(k),$$

where $P_{\text{lin}}(k)$ is the linear matter power spectrum (evaluated at redshift z) where we use $\int_{\mathbf{k}} = \int d^3k / (2\pi)^3$. The above coefficient $(b_1 - f b_\eta \mu^2)^2$ is the well-known generalization of the Kaiser factor for galaxies [73, 92, 132], which can be recovered by setting $b_\eta = -1$.

From this discussion it is clear that all phases of the Ly- α field are captured by the field $\delta_1(\mathbf{k})$ in the linear approximation. It is thus convenient to split the theory model from Eq. (1) into parts that make the amplitude and phase dependence manifest. Absorbing the Kaiser factor into Eq. (1) into a momentum-dependent transfer function yields

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

The advantage of this approach will become evident at the non-linear level, where the transfer function β_1 will also account for the one-loop corrections.

The success of the linear theory model in describing the simulated Ly- α field δ_F^{truth} can be estimated using the error power spectrum,

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

On very large scales (i.e. in the limit $k \rightarrow 0$), the non-linear corrections to our model are expected to be small, so the only expected source of error in the model should be the stochastic field ϵ ,

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

which by definition does not correlate with δ_F^{lin} generated by cosmological fluctuations. In general, $\epsilon(\mathbf{k})$ is the Ly- α flux decrement component produced by small-scale processes unrelated to large-scale initial conditions. In the context of galaxies and halos, the field $\epsilon(\mathbf{k})$ captures the shot noise which arises due to the discreteness of tracers. Using this analogy, the stochastic field of the Ly- α forest could be thought to originate from the discreteness of absorption lines. Due to their large numbers it is expected to be extremely small. This logic, however, is not fully correct because it assumes that the discreteness is generated at the level of absorption lines. One can consider that Ly- α absorption is produced by neutral hydrogen clouds whose distribution also has a stochastic component. This discreteness of the Ly- α clouds can be used as a first proxy to understand the stochasticity of the forest.

Within the EFT for LSS framework [89–91, 94, 133–135], one uses perturbative Taylor expansions to parameterize unknown functions. In this approach, the error power spectrum assumes the following expansion valid in the $k \rightarrow 0$ limit:

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

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学体积的大规模聚类模拟。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， $f = d \ln D_+ / d \ln a$ 是逻辑增长因子， $\mu \equiv k_{\parallel} / k = \hat{\mathbf{k}} \cdot \hat{\mathbf{z}}$ 是视线方向的余弦， $\hat{\mathbf{z}}$ 是视线方向， δ_1 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，

$$\langle \delta_F(\mathbf{k}) \delta_F(\mathbf{k}') \rangle = (2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') \langle \delta_F(\mathbf{k}) \delta_F(\mathbf{k}') \rangle' = (2\pi)^3 \delta_D^{(3)}(\mathbf{k} + \mathbf{k}') P(k, \mu), \quad (3)$$

$$P_{11}(k) \underbrace{=}_{\text{linear}} (b_1 - f b_\eta \mu^2)^2 P_{\text{lin}}(k),$$

其中 $P_{\text{lin}}(k)$ 是线性物质功率谱（在红移 z 处计算），我们使用 $\int_{\mathbf{k}} = \int d^3k / (2\pi)^3$ 。The above coefficient $(b_1 - f b_\eta \mu^2)^2$ 是众所周知的星系凯泽因子的推广 [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，Ly- α 场的所有相位在线性近似下都被 $\delta_1(\mathbf{k})$ 场捕捉到。因此，将方程 (1) 的理论模型拆分为能够显现幅度和相位依赖的部分是方便的。将凯泽因子吸收到方程 (1) 中，形成一个动量依赖的传递函数，则得到

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

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_F^{truth} 方面的成功可以使用误差来估计功率谱

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

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_F^{lin} 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水平上产生的。人们可以认为 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一种初步代理。

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：

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

where $n_0, \alpha_{1,2}$ are dimensional constants whose values are set by the Ly- α physics. For galaxies, n_0 can be estimated at leading order as $1/\bar{n}$, where $\bar{n} = N/V$ is the galaxy number density (N is the number of galaxies in the comoving volume V). The relevant distance scale in this case is the mean separation $R = (V/N)^{1/3}$ between the individual galaxies. Applying the same argument to the Ly- α forest and assuming for simplicity that the individual tracer is a neutral hydrogen cloud whose separation to its neighboring cloud is about $0.5 h^{-1}\text{Mpc}$, we get an estimate

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

As we shall see, this naive estimate will turn out to be quite accurate for the actual Ly- α simulations. Importantly, the stochasticity expansion features scales that are in general different from those that appear in the perturbative bias expansion. The latter can be estimated as a non-linear scale, where matter density fluctuations become of order one [136],

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

If the theory model is accurate, we expect to recover the scale-dependence suggested by Eq. (7) on large scales. If there is a significant correction to the naive linear model (4), this will generate a noticeable scale and orientation dependence not captured by Eq. (7). What corrections do we expect?

First, the displacements of dark matter particles in our simulations are large, and have to be treated non-perturbatively. If unaccounted for, they produce large distortions of baryon acoustic oscillations, which show up as a mismatch between the linear model and actual phases of the density field. This effect is well understood, and can be corrected for by using the linear density field $\tilde{\delta}_1$ shifted by the Zel'dovich displacement in lieu of δ_1 [137]. In what follows $\tilde{\delta}_1$ will be referred to as the shifted linear field.

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

where \mathbf{q} denotes Lagrangian space (initial) coordinates, and $\boldsymbol{\psi}_1$ is the Zel'dovich displacement

$$\boldsymbol{\psi}_1(\mathbf{q}) = \int d^3\mathbf{k} e^{i\mathbf{q}\cdot\mathbf{k}} \frac{i\mathbf{k}}{k^2} \delta_1(\mathbf{k}). \quad (11)$$

Note that $\tilde{\delta}_1$ above has an infinite Taylor expansion in the linear field δ_1 . In perturbation theory one can write this as:

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

where $\tilde{K}_1(\mathbf{k}) = 1$ and

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

where $k_{iz} = (\mathbf{k}_i \cdot \hat{z})$, and $\mathbf{k} \equiv \mathbf{k}_1 + \dots + \mathbf{k}_n$ for the n 'th kernel. This expansion is very similar to that of the Zel'dovich matter density field [137–140]

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

with $F_1^{\text{ZA}} = 1 + f(\hat{\mathbf{k}} \cdot \hat{z})^2$, and

$$F_2^{\text{ZA}} = \frac{1}{2} \frac{(\mathbf{k}\cdot\mathbf{k}_1)(\mathbf{k}\cdot\mathbf{k}_2)}{k_1^2 k_2^2} + \frac{f}{2} \frac{(\mathbf{k}_1 \cdot \mathbf{k}) k_{2z}(k\mu)}{k_1^2 k_2^2} + \frac{f}{2} \frac{(\mathbf{k}_2 \cdot \mathbf{k}) k_{1z}(k\mu)}{k_1^2 k_2^2} + \frac{f^2(\mu k)^2}{2k_1^2 k_2^2} k_{1z} k_{2z}, \quad \text{etc.} \quad (15)$$

These building blocks will be useful in our future discussion. The shifts implemented by the Zel'dovich displacement introduce higher order non-linear effects. In the following, we will discuss these effects more systematically.

Nonlinearities in the bias expansion are a second important source of corrections. In the perturbative Eulerian bias formulation these are given by

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

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 $R = (V/N)^{1/3}$ 。将相同的推理应用于 Ly- α 森林，并简单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 n_0 的估计将证明是相当准确的。重要的是，随机性展开中的尺度与扰动性偏差展开中出现的尺度不同。后者可以被估算为非线性尺度，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。

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

其中 \mathbf{q} 表示拉格朗日空间 (初始) 坐标， $\boldsymbol{\psi}_1$ 是泽尔多维奇位移

$$\boldsymbol{\psi}_1(\mathbf{q}) = \int d^3\mathbf{k} e^{i\mathbf{q}\cdot\mathbf{k}} \frac{i\mathbf{k}}{k^2} \delta_1(\mathbf{k}). \quad (11)$$

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：

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

其中 $\tilde{K}_1(\mathbf{k}) = 1$ 并且

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

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]

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

$F_1^{\text{ZA}} = 1 + f(\hat{\mathbf{k}} \cdot \hat{z})^2$ ， $F_2^{\text{ZA}} = 1/2$

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

这些构建模块将在后续讨论中非常有用。泽尔多维奇位移会引入更高阶的非线性效应。

下面我们将系统讨论这些效应。偏差展开中的非线性是第二个重要的修正来源。

在微扰偏差展开中，这些修正由下式给出：

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