arXiv:2611.32086v1 [cs.CC] 24 Nov 2026

适用于任意常数熵率的渐近最优双源提取器

Xin Li*

摘要

我们构造了适用于任意常数熵率、具有渐近最优参数的显式双源提取器。具体而言，对于定义在 n 比特上、最小熵为 $k=\delta n$ 的源，任意常数$\delta \in (0,1]$, 且误差为 $\varepsilon \geq 2^{-b_\delta n}$，其中 b_δ 是一个依赖于 δ的常数），我们的提取器输出$2k-O(\log(1/\varepsilon)+1)$ 比特，且与均匀分布的距离至多为 ε。特别地，对于任意常数熵率，这给出了输出比特数为 $\Omega(n)$、误差为 $2^{-\Omega(n)}$ 的双源提取器。此前，已知具有这种误差的显式双源提取器仅适用于熵率约为 $4/9$ 的情形（Bourgain 2005，Lewko 2019）。我们还将构造推广为适用于任意常数熵率、具有渐近最优参数的（强）不可篡改双源提取器和和集提取器。

我们的提取器可以扩展到处理略低于常数的熵率，但误差不是最优的。通过采用不同的构造和分析，我们给出了显式的（不可篡改）双源提取器，其熵为 $k \geq Cn/\log n$（其中某个常数 $C>1$），输出比特数为 $\Omega(k)$，且误差接近 $2^{-\Omega(\sqrt{n})}$。

Doron 和 Goodman 的一项独立工作为熵 $k \geq n^{1-\alpha}$ 构造了误差为 2^{-k^α}、适用于某个常数 $\alpha>0$的显式双源提取器。他们的提取器适用于比我们更低的熵，而对于任意常数熵率，我们的提取器具有更好的误差和输出长度。我们针对熵 $k \geq Cn/\log n$的提取器是在尚未得知他们的结果之前独立获得的；而我们针对具有渐近最优参数的常数熵率的提取器，则是在西蒙斯研究所了解到他们未发表的结果之后，采用不同的思路和技术获得的。

*计算机科学系 约翰斯·霍普金斯大学， lixint s@cs.jhu.edu。

目录

1 引言 **1**

1.1 我们的结果 .31.2 技术概述 .4

2 预备知识与常用工具 16

2.1 源强提取与熵松弛转移 .172.2 初等概率、傅里叶分析与秩 .182.3 奇次幂特征与稠密张成空间 .212.4 近似对偶性及其秩敏感形式.242.5 已发表的带种子原语与熵高效提取 .252.6 常数种子强凝聚器 .272.7 关于行数增长的一致定量结果 .272.8 任意公共特征的加权近似对偶性 .282.9 一致的共享特征正则化定理 .29

3 小误差下的常数率提取 32

3.1 普通双源提取 .323.2 和集提取 .383.3 固定-C 不可篡改双源提取 .55

4 常数率下的熵高效输出 79

4.1 具有加性熵损失的双源输出 .804.2 具有加性熵损失的源强不可篡改输出.854.3 推论：具有加性熵损失的普通源强输出.944.4 具有加性熵损失的和集输出 .94

5 递减熵率 **100**

5.1 增长行构件及其证明 .1005.2 短输出不可篡改性：构造与证明 .1095.3 具有加性熵损失的源强输出 .112

1 引言

随机性提取理论研究的是从弱随机源中提取近乎均匀的随机比特的过程。弱随机源是存在各种缺陷和偏差的概率分布。经过数十年的研究，人们设计出了许多不同类型的随机性提取器，即能够执行随机性提取的显式构造对象。这些对象后来已成为伪随机性领域的基础对象，并广泛应用于密码学、复杂性理论、图论等其他领域。本文使用标准的最小熵来衡量弱随机源中的随机性大小。具体而言，随机变量 X 的最小熵定义为 $H_\infty(X) = \min_{x \in \mathrm{supp}(X)} \log_2(1/\Pr[X = x])$。若一个源 X 的支撑集包含于 $\{0,1\}^n$，则称其为一个 $(n, H_\infty(X))$-源，其熵率为 $H_\infty(X)/n$。

众所周知，一般而言，仅具有最小熵保证的单个任意弱源无法进行确定性提取，因此促使人们研究具有结构的模型。

定义 1.1. 设 $\mathcal{X}$ 是 $\{0,1\}^n$ 上的一族分布。函数 $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ 是 $\mathcal{X}$ 的确定性提取器，误差为 ε，如果对于每个分布 $X \in \mathcal{X}$，都有

$$\Delta(\mathrm{Ext}(X), U_m) \leqslant \varepsilon,$$

其中 U_m 表示 $\{0,1\}^m$ 上的均匀分布，$\Delta()$ 表示统计距离。有时我们也写作 $\mathrm{Ext}(X) \approx_\varepsilon U_m$。如果 Ext 可由一个多项式时间算法计算，我们就称 Ext 是显式的。

此外，人们还研究了许多这类模型。其中研究最为深入的模型包括：双源提取器，其输入由两个独立的 (n, k) 源 X, Y 组成；仿射提取器，其输入源是在某个未知仿射子空间上的均匀分布；和集提取器，其输入是两个独立的 (n, k) 源 X, Y 的 $X + Y$（这里 $+$ 表示在 $\mathbb{F}_2$ 上的逐位加法）；以及不可篡改双源提取器，它除了提取功能外，还能保证输入源遭到篡改时的安全性。本文研究双源提取器、和集提取器和不可篡改双源提取器。下面我们将简要回顾这些提取器的发展历程。

双源提取器。Chor 和 Goldreich [20] 开创了对弱随机源和显式双源提取器的形式化研究。他们的构造给出了一个熵率高于 $1/2$ 且误差呈指数级小的显式双源提取器，而概率方法表明，当每个源的熵略大于 $\log n + 2\log(1/\varepsilon) + O(1)$ 时，存在误差为 ε、输出长度为 $m = 2k - 2\log(1/\varepsilon) - O(1)$ 的双源提取器。特别地，这意味着当 $k \geq c \log n$ 时，对于某个常数 $c > 1$，存在输出长度为 $\Omega(k)$ 且误差为 $2^{-\Omega(k)}$ 的双源提取器。Dodis、Elbaz、Oliveira 和 Raz [28] 改进了 [20], 中的提取器的输出长度，但熵率仍需高于 $1/2$。Raz [56] 构造了一个显式双源提取器，其中一个源的熵率需高于 $1/2$，而另一个源可以具有对数级熵。此后，突破熵率 $1/2$ 障碍一直是双源提取器研究中的一个重大开放问题。

Bourgain 于 2005年 的一项工作 [11] 首次突破了熵率 $1/2$的障碍，采用了和积现象以及指数和的估计。然而，Bourgain 的提取器仅适用于略低于二分之一的熵率。随后，Lewko [39] 重新研究了这一方法，并将最小熵率提高到任意接近 $4/9$ 的水平。这两个提取器的误差均为 $2^{-\Omega(n)}$。

另一方面，大量研究工作 [4, 5, 56, 11, 54, 6, 40, 42, 44, 43, 45, 22, 24, 17, 46, 27, 14, 23, 7, 25, 26, 47, 48, 49] 研究了由多个独立源组成的更一般模型，

并逐步减少源的数量。这些工作最终促成了 Chattopadhyay 和 Zuckerman 的突破性成果 [17], ，他们针对熵为 $k \geqslant \log^C n$ （其中某个常数 $C > 1$ ）构造出了具有多项式级小误差的显式单比特双源提取器。后续研究进一步改进了这些构造，并最终在 Li 于 2023年 的工作中达到顶峰 [49], ，该工作针对 $O(\log n)$ 的渐近最优熵构造出了显式双源提取器，具有 1 比特输出和常数误差。这些提取器还给出了目前已知的 ***Ramsey***图 的最佳显式构造。

然而，近期按照 [17] 的框架构造的双源提取器，最多只能达到多项式级小误差；这不足以满足密码学任务的要求，因为这类任务通常需要可忽略误差。鉴于随机性提取器的一项主要应用是密码学，寻找小误差双源提取器仍是一个重要的开放问题。此前已知的最佳构造由 Lewko 提出 [39] ，但仍要求熵率约为 $4/9$。

和集提取器。Chattopadhyay 和 Li [15] 引入并开创了对和集源提取器的研究，这一模型统一了此前研究过的若干模型，例如独立源、仿射源和小空间源。与双源提取器的情形不同，简单应用概率方法无法证明存在适用于小熵（例如 $O(\log n)$）的这类提取器。尽管如此， [15] 构造出了针对常数个独立源之和的显式提取器，其熵为多对数级，并达到了多项式级小误差。随后，Chattopadhyay 和 Liao [16] 将其改进为适用于两个独立源之和的提取器：熵为 $k = \text{polylog}(n)$ 时达到多项式级小误差；熵为 $k = \log n (\log\log n)^{1+o(1)}$ 时达到常数误差。之后，Li [49] 获得了适用于两个独立源之和的和集提取器，其熵达到渐近最优的 $O(\log n)$，且误差为常数。

然而，如果要求误差可忽略，由于和集提取器比双源提取器更强，前者的熵要求至少与后者相同。一种已知构造是 [20], 中的佩利图提取器，尽管其中的求和是在奇特征域上进行的。该提取器要求一个源的熵率高于 $1/2$ ，而另一个源的熵可以小至 $O(\log n)$ k，且提取器的误差为 $2^{-\Omega(k)}$。因此，对于 $k = \omega(\log n)$，误差可忽略。其分析基于 Karatsuba 的特征和估计 [37, 38]。

不可篡改提取器。许多标准随机性提取器模型也可以扩展为一种更强的版本，称为不可篡改提取器。这一概念源于密码学应用，例如在存在主动对手时进行隐私放大 [30] 以及不可篡改码 [34]。由于这类提取器具有密码学应用，并且近期发现了它们与构造标准提取器之间令人惊讶的联系 [42, 17], ，在过去约 15 年里，这类提取器也一直受到广泛研究。在这种设定中，某种特定类型的对手会篡改提取器的输入，而不可篡改提取器保证，即使以篡改后的输出为条件，提取器在原始输入上的输出仍接近均匀分布。我们首先通过定义篡改函数来讨论不可篡改提取器。

定义 1.2 （篡改函数）。对于任意函数 $f : S \to S$，若 f 没有不动点，则称 $f(s) \neq s$ 对所有 $s \in S$均成立。对于任意 $n > 0$，令 $\mathcal{F}_n$ 表示所有函数 $f : \{0,1\}^n \to \{0,1\}^n$ 构成的集合。$\mathcal{F}_n$ 的任意子集都是一个篡改函数族。

与近期文献一样，对于不可篡改提取器，我们只考虑没有不动点的篡改函数，因为如果篡改是恒等函数，则不可篡改性是不可能实现的。例如， [18]中给出了更一般的定义。现在，我们可以将双源提取器扩展为其相应的不可篡改版本。

定义 1.3。 如果函数 $\mathsf{nmExt} : (\{0,1\}^n)^2 \to \{0,1\}^m$ 满足以下性质，则称其为一个（k, ε） C-不可篡改双源提取器：设 X, Y是两个独立的（n, k）源，并且{（f_i, g_i）}$^{i\in[C]}$ 是 C 任意的篡改函数对，使得对于每个 $i \in [C]$，f_i中的至少一个, g_i 没有不动点，则

$$(\mathsf{nmExt}(X, Y), (\mathsf{nmExt}(f_i(X), g_i(Y)))_{i=1}^{C}) \approx_\varepsilon (U_m, (\mathsf{nmExt}(f_i(X), g_i(Y)))_{i=1}^{C}),$$

其中 U_m 与 $(\mathsf{nmExt}(f_i(X)\ g_i(Y)))_{i=1}^{C}$独立。若上述性质在额外条件化于源 X 或 Y时仍成立，则称该不可篡改双源提取器是（源）强的。

注 1.4. 在 $C = 0$的特殊情形下，这就给出了一个标准双源提取器，而强版本则给出了一个强双源提取器。

概率方法可用于证明存在具有优异参数的此类不可篡改提取器。例如， [18] 证明了，对于 $C = 1$，当 $k \geq m + \frac{3}{2}\log(1/\varepsilon) + O(1)$ 且 $k \geq \log n + 3\log(1/\varepsilon) + O(1)$ 时，存在适用于 (n, k) 源的此类提取器。若误差较小，那么对于针对 C 个篡改函数的不可篡改双源提取器，一个必要的信息论界是 $k \geq (C + 1)m/2$。对于强版本，我们需要 $k \geq (C + 1)m$。

不出所料，由于不可篡改提取器具有强得多的性质，构造显式不可篡改提取器要比构造标准提取器困难得多。历史上，Chattopadhyay、Goyal 和 Li [13] 首次为熵为$n - n^{\Omega(1)}$的源构造了显式不可篡改双源提取器，其输出长度为多项式级，误差则为指数级小。他们还在同样的极高熵条件下，构造了适用于多次篡改的版本。后续工作 [14, 47, 48, 21, 49]从多个方面改进了这些构造，最终 Li [49], 给出了一种显式构造：该构造的不可篡改双源提取器要求一个源的熵为 $(2/3 + \gamma)n$，另一个源的熵为 $k \geqslant C \log n$，其中某个常数为 $C > 1$；输出长度为 $\Omega(k)$，误差为 $2^{-\Omega(k)}$。在小熵情形下，Li和Zhong [51] 也获得了与相应标准提取器性能相当的不可篡改双源提取器：针对多对数熵、误差为多项式级小误差的提取器，以及针对 $O(\log n)$ 熵、误差为常数误差的提取器。

总之，构造小误差（例如可忽略误差）的（不可篡改）双源相关提取器仍然是一个重要的开放问题。对于双源提取器和和集提取器，瓶颈至多略低于熵率 $1/2$；对于不可篡改双源提取器，瓶颈则处于熵率 $2/3$。

1.1 我们的结果

本文在小误差（不可篡改）双源提取器和和集提取器方面显著改进了此前的结果。我们先陈述双源提取器的主要结果。

定理 1.5. 存在绝对常数 $C_0, C_1 > 0$，使得对于每个常数 $0 < \delta \leqslant 1$，都存在常数 $b_\delta > 0$ 和 n_δ，满足以下性质。对于所有 $n \geqslant n_\delta$、整数 $\delta n \leqslant k \leqslant n$ 和 $2^{-b_\delta n} \leqslant \varepsilon \leqslant \frac{1}{4}$，都存在一个显式双源提取器 $\mathsf{TExt} : \{0,1\}^{\overline{n}} \times \{0,1\}^n \longrightarrow \{0,1\}^m$，使得对于任意两个独立的 (n, k) 源 X, Y，

$$\Delta(\mathrm{TExt}(X, Y), U_m) \leqslant \varepsilon, \qquad m \geqslant 2k - C_0 \log(1/\varepsilon) - C_1.$$

因此，我们的提取器对任意常数熵率都达到了渐近最优参数。特别是，这意味着对于任意常数熵率，都存在一个输出长度为 $\Omega(n)$、误差为 $2^{-\Omega(n)}$ 的双源提取器。

我们可以将我们的结果推广到和集提取器和不可篡改双源提取器。具体而言，我们有以下定理。

定理 1.6. 对于每个常数 $0 < \delta \leqslant 1$，都存在常数 $b_\delta > 0$ 和 n_δ，使得以下结论成立。对于所有 $n \geqslant n_\delta$、整数 $\delta n \leqslant k \leqslant n$ 和 $2^{-b_\delta n} \leqslant \varepsilon \leqslant \frac{1}{4}$，都存在一个显式和集提取器 SumExt：$\mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m$，使得对于任意两个独立的 (n, k) 源 X, Y，

$$\Delta(\mathrm{SumExt}(X+Y), U_m) \leqslant \varepsilon, \qquad m \geqslant k - 2\log(1/\varepsilon) - 10.$$

定理 1.7。 对于每个常数 $0 < \delta \leqslant 1$ 和整数 $C \geqslant 0$，存在常数 $b_{\delta,\, C} > 0$，使得对于充分大的 n、整数 $\delta n \leqslant k \leqslant n$以及 $2^{-b_{\delta,\, C} n} \leqslant \varepsilon \leqslant 1/4$，存在显式非篡改双源提取器 nmExt：$\{0,\ 1\}^n \times \{0,\ 1\}^n \to \{0,\ 1\}^m$，满足

$$m \geqslant \frac{k}{C+1} - O(\log(1/\varepsilon) + 1)$$

对于任意两个独立的 (n, k) 源 X, Y，以及任意 C 对篡改函数$\{(f_i,\ g_i)\}^{i\in[C]}$，使得对于每个 $i \in [C]$，f_i，g_i 中至少有一个没有不动点，

$$(\mathsf{nmExt}(X,Y), (\mathsf{nmExt}(f_i(X), g_i(Y)))_{i=1}^C, Y) \approx_\varepsilon (U_m, (\mathsf{nmExt}(f_i(X), g_i(Y)))_{i=1}^C, Y), \tag{1.1}$$

$$(\mathsf{nmExt}(X,Y), (\mathsf{nmExt}(f_i(X), g_i(Y)))_{i=1}^C, X) \approx_\varepsilon (U_m, (\mathsf{nmExt}(f_i(X), g_i(Y)))_{i=1}^C, X). \tag{1.2}$$

注 1.8。 注意，在上述定理中，$C = 0$ 的特殊情形也给出了一个标准强双源提取器，其输出长度为 $m \geqslant k - O(\log(1/\varepsilon) + 1)$。

我们可以将我们的结果扩展到略低于常数的熵率。具体来说，我们有以下定理。

定理 1.9（熵率 $C/\log n$）。固定一个整数 $t \geqslant 0$和 $0 < \vartheta < 1/2$。存在常数$C_{t,\ \vartheta} > 0$和 n_0，使得对于每个 $n \geqslant n_0$、整数 $C_{t,\ \vartheta} n/\log n \leqslant k \leqslant n$以及 $2^{-n^{1/2-\vartheta}} \leqslant \varepsilon \leqslant 1/4$，都存在一个显式的强 t-不可篡改双源提取器 nmExt：$\{0,\ 1\}^n \times \{0,\ 1\}^n \to \{0,\ 1\}^m$，满足

$$m \geqslant \frac{k}{t+1} - O(\log(1/\varepsilon) + 1).$$

1.2 技术概述

这里我们非正式地概述构造中用到的技术。接下来，我们将着重介绍标准双源提取器的构造，以阐明我们的主要思路。首先，我们描述用于任意常数熵率的双源提取器背后的思路，其输出长度为 $\Omega(n)$，误差为 $2^{-\Omega(n)}$。不失一般性，我们考虑平坦源，也就是定义在某些未知子集上的均匀分布。

我们的出发点是 Ben-Sasson 和 Ron-Zewi 的工作 [9], ，该工作展示了如何从一个足够好的仿射提取器构造双源提取器。我们不会在此采用这一具体方法，但会使用他们的一项重要观察：设 E：$\mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ 是一个函数，并以自然方式将其视为一个 $2^n \times 2^n$ 矩阵 M_E，其中（x，y）项对应于 E（x，y）。设 r是 M_E在 $\mathbb{F}_2$上的秩，则还存在函数 h_X、h_Y：$\mathbb{F}_2^n \to \mathbb{F}_2^r$，使得 E（x，y）$= h_X$（x）h_Y（y），其中，表示 $\mathbb{F}_2$上的内积。此外，如果 E是熵率为 ρ 且秩为 O（n）的双源弥散器，那么对于任意常数 $\gamma > 0$，E 也是熵率为 $\rho + \gamma$、误差为 $2^{-\Omega(n)}$的双源提取器。这里，双源弥散器是一种弱于双源提取器的对象，因为它只要求输出为非恒定的（在输出一比特的情况下）。证明使用了近似对偶性：如果该函数不是具有上述参数的双源提取器，那么就可以在矩阵 [9], 中找到一个相对较大的子矩形

其中的所有项都相同，这与扩张器性质相矛盾。因此，我们的目标归结为：对于任意常数熵率，构造一个秩为 $O(n)$ 的双源弥散器。得到它之后，我们可以利用近似对偶性将其提升为误差为 $2^{-\Omega(n)}$的提取器。为了获得 $\Omega(n)$ 的输出长度，我们证明实际上可以产生 $\Omega(n)$ 个输出比特，使得这些比特的任意非平凡子集的奇偶校验仍然 $2^{-\Omega(n)}$ 接近均匀分布。因此，根据标准的异或引理，这 $\Omega(n)$ 个比特本身 $2^{-\Omega(n)}$ 接近均匀分布。

现在我们解释如何利用这些要素得到这样的双源提取器。为此，我们首先使用标准技术，从每个输入弱源中得到一个某处高熵源。具体而言，我们使用 [56] 中的带种子压缩器，它具有常数种子长度，并且可以将任意熵率为 $\delta/2$ 的弱源压缩为熵率为 $1-\beta$、误差为 ε、输出长度为 $\Omega(n)$ 的源。这里 δ、β、ε 是任意固定常数，而种子长度 d 是依赖于这些常数的一个常数。通过枚举所有可能的种子，我们得到一个具有常数 $r = 2^d$ 行数的某处高熵源；当在常数种子空间上取平均时，每一行都接近于某个熵率至少为 $1-\beta$ 的分布。这与例如 [6, 58], 中的一些压缩器形成对比，后者只能保证至少有一行接近于具有高熵的分布，但误差却小到指数级。这里的误差 ε 虽然更大，但对于我们的目的——首先得到一个分散器——这已经足够了。对于 $\widetilde{\ }$ $\widetilde{\ }$ 输入源 X、Y，令这两个某处高熵源为 X、Y。在我们的分析中，接着证明存在两个子源 X^*、Y^*，使得将其分别限制于 X^*、Y^* 后，每个某处高熵源 $\widetilde{X^*}$、$\widetilde{Y^*}$ 至少有 $\frac{3r}{4}+1$ 行的熵率至少为 $1-\tau/2$，其中 $\tau > 0$ 是某个小常数。因此，令 $D = \frac{r}{2}$，则至少有 $D+1$ 行在两者中同时具有至少为 $1-\tau/2$ X^* $\widetilde{\ }$ Y^* X^* 的熵率。现在我们从 和 Y^* 中各删除任意一行，最终得到 $r-1$ 行，其中至少有 D 行在 X^* 和 Y^* 中同时具有至少为 $1-\tau/2$ 的熵率 $\widetilde{\ }$ $\widetilde{\ }$。这些就是我们将在分析中使用的优行；从现在起，我们将输入源视为 $X = X^*$、$Y = Y^*$。

此时，我们的构造是：对每一行 a 应用特定的双源提取器 TExt，以提取高 $\widetilde{\ }$ $\widetilde{\ }$ $\widetilde{\ }$ $\widetilde{\ }$熵源的信息X, Y，即计算 TExt(X_a, Y_a)。然后，我们使用一个合并器将这些 $r-1$ 输出合并起来。然而，为了实现我们先前的计划，我们需要仔细构造双源提取器 TExt 和合并器。乍一看，似乎直接使用内积作为 TExt 很有吸引力。它的秩确实为 $O(n)$，但之后如何将输出合并起来并不明确。为了解决这个问题，我们使用一个稍作修改的双源提取器。具体来说，设 M $\widetilde{\ }$ $\widetilde{\ }$为 X 中每一行的长度，并设 Y。令 $K = \mathbb{F}_{2^M}$，并将每一行视为有限域 K K中的一个元素。对于 $\widetilde{\ }$ $\widetilde{\ }$第 a行，记 $U_a(X) = X_a$ 和 $V_a(Y) = Y_a$。接下来，我们选择一个适当的 $\ell = \Theta(n)$ $< M$，使得 $M = L\ell$，并令 $F = \mathbb{F}_{2^\ell}$。最后，选择一个整数 $J \geqslant 1$ 并计算以下函数：

$$H_J(U_a(X), V_a(Y)) = \sum_{j=1}^{J} U_a(X)^{2j-1} V_a(Y)^{2j-1} = \langle \Psi_J(U_a(X)), \Psi_J(V_a(Y)) \rangle_K, \quad \Psi_J(z) = (z, z^3, \ldots, z^{2J-1}),$$

以及 $W_a(X,\ Y) = \text{TExt}(U_a(X)\ V_a(Y)) = \text{Tr}_{K/F}\, H_J(U_a(X)\ V_a(Y))$，其中将线性映射 $\text{Tr}^{K/F}$映射到较小的域 F 对于使该函数成为双源提取器是必要的。

不难证明，当每个输入的熵足够高时，TExt$(U_a(X)\ V_a(Y))$ 确实是一个双源提取器，并且输出比特的任意奇偶校验在 $\mathbb{F}_2$ 上的秩为 $O(n)$，因为函数 $\text{Tr}_{K/F}\, H_J(U_a(x)\ V_a(y))$ 在 F 上的秩只有常数。不过，如何将这些输出合并起来仍有待解决。为此，我们证明了：当限制在足够大的矩形 $A \times B$ 上时，输出比特的任意非平凡奇偶校验的秩有一个下界，其中 A 和 B 分别是 $U_a(X)$ 和 $V_a(Y)$ 的支撑集。对于一行优行 a，有 $|A|,\ |B| \geq 2^{(1-\tau)M}$。这一步关键地利用韦伊界来论证，编码 $\Psi_J(z)$ 的任意非平凡奇偶特征在 K 上的偏差都很小；编码中的奇数指数则用于保证韦伊界在特征为 2 的域中成立。

我们利用这一点证明，当 $|A|,\ |B| \geq 2^{(1-\tau)M}$ 时， $\Psi_J(A)$ 和 $\Psi_J(B)$ 的张成空间的维数都至少为 $JM - \tau M - 1$。事实上，在极端情形 $\tau = 0$ 下，两个张成空间的维数都达到全维 JM；否则就会存在一个固定不变的非平凡奇偶特征，这与韦伊界矛盾。现在，输出比特的任意非平凡奇偶校验都可以写成 $h_{J,\ \lambda} = \mathrm{Tr}_{K/\mathbb{F}_2}(\lambda H_J(U_a(X)\ V_a(Y)))$，其中 $\lambda \in K,\ \lambda \neq 0$。由于二元双线性形式

$$B_\lambda(p,q) = \mathrm{Tr}_{K/\mathbb{F}_2}\left(\lambda \sum_{j=1}^{J} p_j q_j\right) \qquad (p,q \in K^J)$$

是非退化的，将其参数限制在 $\Psi_J(A)$ 和 $\Psi_J(B)$ 的张成空间上，代价至多为它们的二元余维数之和。因此，对于所有 $|A|,\ |B| \geq 2^{(1-\tau)M}$，我们有

$$\boxed{\mathrm{rank}_{\mathbb{F}_2}[h_{J,\lambda}]_{A\times B} \geq (J-2\tau)M - 2, \qquad \mathrm{rank}_{\mathbb{F}_2}[h_{J,\lambda}]_{K\times K} \leq JM.} \tag{1.3}$$

这立即启发了一种区分优行的方法：我们可以为不同的行 J_a 选择不同的数值 a。这里我们采用了与以下文献中针对任意常数熵率构造仿射提取器相似的思想 [41],，即将这些数值选为等比数列： $J_a = 2^{a-1}$，从而所有 J_b 的求和为 $b < a$ $\kappa_a = \sum_{b<a} J_b = J_a - 1$。

我们利用这一性质证明，在 D 优行上，为了表述简洁，假设第 1 行到 D，联合分布 $(W_1(X,Y)$、$\cdots$ $W_D(X,Y))$ 的支撑集为 F^D 的满支撑。为此，我们实际上证明：将分布 $(W_1(X,Y)$、$\cdots$ $W_D(X,Y))$ 限制到X^*、Y^* 时，它与均匀分布的距离呈指数级接近，误差为 $< 2^{-D\ell}$，其中 ℓ 是每个 $W_i(X,Y)$ 的输出长度。这由一个混合论证得出：对于任意 $a \in [D]$，令转录为 $\Pi_a = (W_1(X,Y)$、$\cdots$ $W_{a-1}(X,Y))$，我们证明联合分布 $(W_a(X,Y)\ \Pi_a)$ $2^{-\Omega(M)}$ 接近 (U_F, Π_a)。为证明这一点，考虑第 a 行，并令 $U = U_a(X)\ V = V_a(Y)$。通过固定额外随机性，前面的 $a-1$ 行可以视为 (U, V) 的确定性函数，因此转录Π_a 也同样是 (U,V) 的确定性函数 Π_a。现在我们使用标准的异或引理，并证明对于任意 $\mu \in F$、$\mu \neq 0$ 以及任意 $\beta \in \mathbb{F}_2^{\ell(a-1)}$，奇偶特征

$$\mathrm{Tr}_{F/\mathbb{F}_2}(\mu W_a) + \beta \cdot \Pi_a = \mathrm{Tr}_{K/\mathbb{F}_2}(\mu H_{J_a}(U,V)) + \beta \cdot \Pi_a$$

具有偏差 $2^{-\Omega(M)}$。这是分析中的关键步骤之一。

证明由上述秩界以及再次应用近似对偶提升得到。具体而言，先前的秩下界和上界共同表明，当限制在 $A \times B$ 上时，每个这样的混合奇偶特征都是非恒定的，因为对于充分大的 M，奇偶特征的秩至少为 $(J_a - 2\tau)M - 2 - ((J_a - 1)M) = (1-2\tau)M - 2 > 1$。回顾我们的构造：我们确保在一行优行中，每个 $U_a(X)$ 和 $V_a(Y)$ 实际上都具有熵 $(1-\tau/2)M$；现在我们可以应用近似对偶性，将该特征的偏差提升至 $2^{-\Omega(M)}$。通过仔细选择参数，总误差为 $< 2^{-D\ell}$。这完成了我们分析的第一部分。

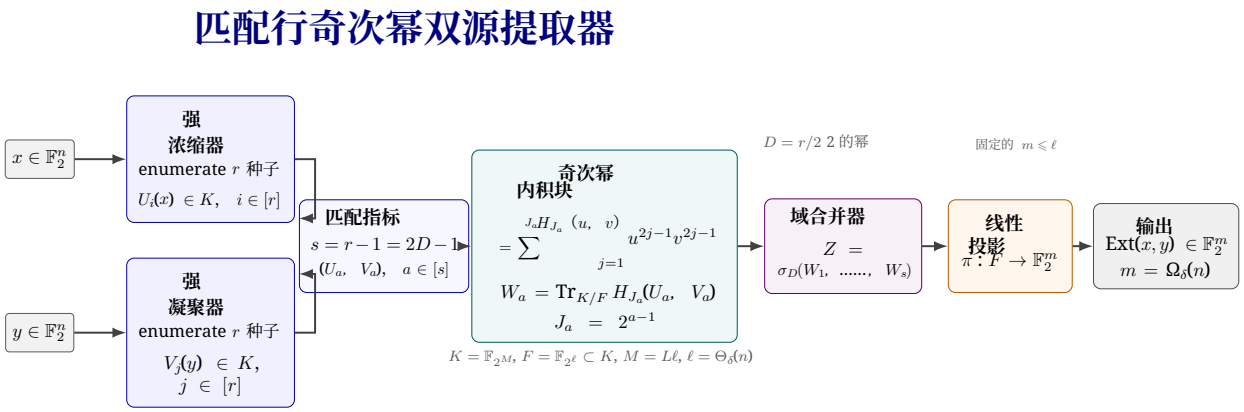
合并器。 现在我们描述最终的合并器，其输出 $\Omega(n)$ 个比特。它需要满足两个性质：首先，它需要将 $2D-1$ 行合并在一起，并确保输出比特的任何非平凡奇偶校验都是非恒定的。其次，它需要使任何此类奇偶校验的秩保持为 $O(n)$，从而我们可以再次应用近似对偶提升，以实现指数级小的误差。

第二个性质实际上更容易实现。由于我们是在大域 F 上进行运算，并且每个输出 $W_a(x,y) \in F$ 都可以写成 $W_a(x,y) = \sum_{j=1}^{R} A_j(x) B_j(y)$，其中 R 是一个常数，而每个 A_j, B_j 都是从 $\mathbb{F}_2^n$ 到 F 的函数，因此对 F 应用任意常次数多项式，就能将其作用于

$2D-1$ 个输出并进行奇偶校验；借助迹对偶基，可以使秩保持为 $O(n)$。为了实现第一个性质，我们按如下方式设计合并器：首先对 $s=2D-1$ 个输出应用以下对称多项式： $\sigma_D(z_1,\ldots,z_s)$ $=\sum_{S\subseteq[s]}\prod_{a\in S} z_a$，然后取一个固定的满射二元线性$|S|=D$
投影到 $m=\Omega(n)$ 比特，其中 m 足够小，因此我们可以应用异或引理，确保输出比特 $2^{-\Omega(n)}$ 接近均匀分布。

要理解其为何有效，回顾一下我们之前证明过，存在 D 个支撑集为 F^D 的优行。特别地，这意味着对于任意元素 $\alpha\in F$，所有 D 个优行都取值为 α 的概率为正。由于 $D=\frac{r}{2}=2^{d-1}$ 是 2 的幂，在特征为二的情况下，我们有 $(1+\alpha T)^D=1+\alpha^D T^D$。注意，生成函数$\prod_{a=1}^{2D-1}(1+W_a T)$ 中 T^D 的系数恰好是 $\sigma_D(W_1,\ldots\ldots,W_{2D-1})$。因此，只要 W_a 个 D 中有任意一个等于 α，就有 $\sigma_D(W_1,\ldots\ldots,W_{2D-1})=\alpha^D$，而与其余的 $D-1$ 个输入无关。等价地，这可以看作对输出取多数 。因此， σ_D 的输出在 F 中也具有满支撑，所以线性投影的任意非平凡奇偶校验都是非恒定的。最后，我们可以再次使用近似对偶提升，将误差降至指数级小，并再次使用异或引理，得出 $\Omega(n)$ 个输出比特 $2^{-\Omega(n)}$接近均匀分布。特别地，在分析中，我们两次使用近似对偶提升来降低误差。

下面是展示该构造的示意图：



仅求值以下操作：对每个源应用常数种子强凝聚器；仅求值$r-1$ 在 K上的匹配奇次幂内积；取 F的迹；次数为 D 的初等对称合并器；固定线性投影。这里 $J_a=2^{a-1}$、 $F=\mathbb{F}_{2^\ell}\subset K=\mathbb{F}_{2^M}$、 $\ell=\Theta_\delta(n)$，且 D 是 2 的幂。

和集提取器。我们的和集提取器构造大致与双源提取器具有相同的结构，但做了几处修改。首先，如果我们直接应用任意凝聚器并尝试每一种可能的种子，就会破坏和集结构，后续分析也将无法进行。相反，我们需要一种凝聚器，使得对于每个固定种子，其函数都是线性函数。我们构造了这样一种凝聚器来替代Raz的浓缩器：首先使用Li和Zhong的某处高熵线性浓缩器 [50], 然后如 [33, 32]中所示，使用曲线归并器将各行合并起来。

其次，由于现在我们得到的唯一输入是一个和 $X+Y$，因此无法直接应用函数$H_J(U_a(X)\ V_a(Y))$ $=\sum_{j=1}^{J} U_a(X)^{2j-1}V_a(Y)^{2j-1}$。相反，我们使用 $(U_a(X)\ +V_a(Y))^d$，其中$d=2^h-1,\quad h\geqslant 3$。再次令 $u=U_a(X)\ v=V_a(Y)$，在特征为二的情况下，Lucas定理意味着

d的所有二项式系数都是奇数，因此

$$\boxed{(u+v)^d = \sum_{j=0}^{d} u^j v^{d-j}.} \tag{1.4}$$

因此，对于每个非零 $\lambda \in K$，

$$\operatorname{rank}_2 \left[\operatorname{Tr}_{K/\mathbb{F}_2}(\lambda(u+v)^d) \right]_{u,v \in K} \leqslant (d+1)M = 2^h M. \tag{1.5}$$

只要 h 是一个常数，全局秩仍然是 $O_h(M) = O(n)$。随后，我们使用矩-迹码论证，并在命题3.20中证明，通过适当地选择 h ，可以确保高熵矩形 $A \times B$ 上的受限秩可以增长为M的任意大固定倍数。因此，我们可以再次采用之前的策略：为不同的行选择不同的 h ，以区分优行，其余论证可以照此进行。

不可篡改双源提取器。 从高层次上看，我们的 C 篡改不可篡改双源提取器遵循标准双源提取器的一般框架：对每个源应用常数种子强凝聚器，在每个对齐的行对上计算一个适用于高熵率的合适的不可篡改双源提取器，并使用 σ_D合并行输出。具体而言，我们将使用构造的秩性质和近似对偶提升。额外的困难在于，我们需要确保非可篡改性质。下面简要介绍新构造所需的几个新要素。

首先，为了证明不可篡改性，我们的策略是在输出比特上使用异或引理的不可篡改版本，例如 [29] 中的版本。这要求证明：任意混合二元特征——它对应于原始输入的输出比特中任意非空子集与篡改输入的输出比特中任意子集的线性组合——都具有指数级小的偏差。在对标准双源提取器的分析中，我们通过证明 D “良好输出”的联合分布具有 F^D 的满支撑集来实现这一点，然后利用 σ_D 表现得类似于多数函数这一性质。这意味着 σ_D 的输出具有 F 的满支撑集。因此，任意非平凡特征都不是常数，随后我们通过近似对偶提升将偏差降至指数级小。在这里，我们采用类似的策略来证明： D良好输出与其所有 C 篡改版本的联合分布具有 $F^{D(C+1)}$ 的满支撑集，从而应用 σ_D 后得到的联合输出具有 F^{C+1} 的满支撑集。

然而，一般而言这并不成立，因为有些篡改可能根本没有熵，有些篡改也可能彼此相同。为了解决这一问题，我们每次分析一个混合二元特征。如果某个活跃篡改对的任一侧存在重输入值，我们就将相应的原始源限制为该值，这不会造成太多熵损失。于是，该篡改对对该特征的贡献只取决于另一个原始源，因此我们进一步固定这一比特贡献，额外熵损失至多为一比特，并将这对篡改从该特征中移除。重复这一过程后，只剩下两个输入边缘分布都具有所需熵的篡改。对于相同的篡改副本，我们可以将这些输出上的奇偶校验合并为仅作用于其中一个副本的奇偶校验。因此，我们可以假设至多有 C 个彼此不同且熵都足够的篡改。

区分执行。与以往构造不可篡改提取器的方法一样，我们需要一种策略来区分原始输入及其篡改版本。这里我们采用的策略比以往研究中的建议生成器更简单。具体而言，我们将每个完整输入视为

x 一个多项式 P_x 的系数向量，该多项式定义在一个域上，且每个元素占用的比特数为一个较小的线性量。因此，该多项式的次数为某个常数 b。对于浓缩器输出中各行的每个索引 a，我们选择一个不同的固定求值点 t_a，并计算标签（或辅助信息）$(P_x(t_a)\ P_y(t_a))$。这种方法的优点是，每个 $(P_x(t_a)$ 或 $P_y(t_a))$ 分别是 X 和 Y 的确定性函数，因此，在固定这些标签的条件下进行分析很方便，同时还能保持 X 和 Y的独立性。

为了应用上述 σ_D 多数论证，我们希望证明，在提取器的每次篡改执行以及原始输入上的执行中，都有多数的 D 优行具有相同的索引。由于最终我们会对每一行应用一个高熵率不可篡改双源提取器，因此我们希望确保在这些 D 行中，所有输入都能通过标签加以区分。回想一下，我们之前假设每种篡改都不相同。直观地说，我们希望证明，由(x, y)计算出的两个不同的标签集合$\{(P_x(t_a)\ P_y(t_a))\}$ 和由(x', y')计算出的$\{(P_{x'}(t_a)\ P_{y'}(t_a))\}$ 最多只能在 b 个位置上相同（回想一下，b 是多项式 P_x, P_y的次数）。因此，标签发生碰撞的位置总共至多有 $\binom{C+1}{2}b$ 个。如果有必要，我们可以循环使用凝聚器种子来生成足够多的行，从而确保这些位置与凝聚器的坏种子合计只占所有可能行的 $< 1/2$ 比例。

然而，上述论证存在几个问题。首先，我们只知道任意一对篡改与原始输入对在所有可能的输入字符串上都不同，但两对不同的篡改仍可能在某些输入上相同。其次，即使两对篡改（或一对篡改与原始输入对）在所有可能的输入字符串上都不同，标签不同的位置也可能取决于具体的输入字符串。为了解决这些问题，我们固定相等/不等模式，并转而考察一个子源。具体而言，设原始输入为 $x_0 = x,\ y_0 = y$，篡改后的输入为$x_i = f_i(x)\ y_i = g_i(y)$，其中 $i \in [C]$。对于每个 $0 \le i < j \le C$，我们记录两个比特，分别表示 $x_i = x_j$ 和$y_i = y_j$。总共得到 $2\binom{C+1}{2}$ 比特；我们可以固定其中一种特定模式，使得两个源的熵损失都不超过 $\binom{C+1}{2}$ 比特，并且它们仍然相互独立。关键在于，在条件于这一固定模式后，对于每个 $i,\ j$，等式 $x_i = x_j$ 在保留下来的左支撑集上要么始终成立，要么始终不成立；对保留下来的右支撑集上的 $y_i = y_j$ 也同样如此。此外，除非 $i,\ j > 0$，这两个等式不可能同时成立。如果对于某个 $i,\ j$确实如此，那么在分析混合特征时，我们仍可将这两种篡改合并。因此，经过这一步，我们可以假设对于每个可能的输入字符串，篡改函数至少在一侧不同。这样便解决了第一个问题。

接下来，我们做同样的事，并固定标签相等示性函数。假设先前对凝聚器种子进行可能的循环后，总共有 R 行。对于每个不同的 i, j，在第 $a \in [R]$行，它们的标签分别为$(P_{x_i}(t_a)\ P_{y_i}(t_a))$和$(P_{x_j}(t_a)\ P_{y_j}(t_a))$。现在，我们记录两个比特，分别表示 $P_{x_i}(t_a) = P_{x_j}(t_a)$和 $P_{y_i}(t_a) = P_{y_j}(t_a)$是否成立。同样，这些比特至多有 $2R\binom{C+1}{2}$ 比特；我们可以固定这些比特的一种特定模式，使得两个源的熵损失都至多为 $R\binom{C+1}{2}$ 比特，并且它们仍然独立。现在，在以这种固定为条件后，对于任意 $i,\ j$，在整个保留的乘积支撑集中，所有输入字符串中发生标签碰撞的索引集合都相同 。这样就解决了第二个问题。

一种能够抵抗一次篡改的带标签高熵不可篡改提取器。接下来，我们将一种能够抵抗一次篡改的带标签高熵不可篡改提取器用作起始提取器，并将其应用于每一对行。稍后我们会将其改造为能够抵抗 C 次篡改的提取器。在 [49], Li 构造了熵率为 $1 - \delta$的显式不可篡改双源提取器；然而，我们所

需的秩性质并不能仅由不可篡改提取器的保证提供。因此，我们提出了一种专为此目的设计的新构造。

新型带标签的高熵不可篡改提取器是一种非对称提取器，专门设计用于满足秩性质。令 z，$w \in \mathbb{F}_2^M$ 表示完整的左行和右行，其中$M = 6n_0$。选择二元域 $K \subset F \subseteq L$，满足 $L = \mathbb{F}_{2^{n_0}}$ 和 $[F:\ K] = 2$。令 $\bar{x}$ 为 z 的前缀，包含 n_0 个比特，并将 $\bar{x}$ 视为 L 中的一个元素。将完整的右行解析为 $w =$（r，t，Y）$\in L^2 \times K^J$，其中 $J\log_2|K| = 4n_0$。对于系数向量 $d =$（d_0，d_1，...），记 P_d（T）$= \sum_j d_j T^j$。定义

$$\Gamma_c(\bar{x}) = (\bar{x}, \bar{x}^{1+2}, \ldots, \bar{x}^{1+2^c}), \tag{1.6}$$

$$S_\alpha(\bar{x}) = V_0(\bar{x}) + \alpha V_1(\bar{x}) = (u, v) \in K^2. \tag{1.7}$$

这里，$\alpha \in F$ 是我们上文所述的固定数值标签，我们通过使用固定的 K-基将 F 与 K^2 视为相同对象。函数 V_0、V_1 来自 Li 的近满熵仿射非篡改提取器 [49]。对于不同的固定标签 $\alpha \neq \alpha'$ 以及任意仿射映射 $h:\ L \to L$（包括恒等映射），即使给定 $S_{\alpha'}$（h（U_L）），诚实种子 S_α（U_L）也与均匀分布在指数意义下接近。标签仅通过 S_α（$\bar{x}$）的坐标（u，v）进入求值起始项。

在最终 输出缩短迹之前，提取器会计算以下取值于 K的函数 :

$$\mathcal{E}_\alpha(z, w) = E_\alpha(\bar{x}; r, t, Y) = \underbrace{\mathrm{Tr}_{L/K}\Big(tP_{\Gamma_c(\bar{x})}(r)\Big)}_{\text{Gold recognition}} + \underbrace{vP_Y(u)}_{\text{tagged extraction}}, \qquad (u, v) = S_\alpha(\bar{x}). \tag{1.8}$$

因此，提取器会忽略诚实左行中未使用的后缀，但任意左侧篡改都可能依赖于该后缀。$\bar{x}$ 的 Gold 坐标的二进制度至多为二，而第二项则在种子坐标 u 处求值右侧系数多项式，并将结果乘以 v。

这两个加数各有不同用途：*Gold*项用于识别任意左侧篡改何时表现得像仿射篡改，而带标签求值项则在找到这种仿射结构后排除常数混合特征。另一方面，右侧篡改可以是任意函数。

更具体地说，考虑两个分别在集合 A 和 B上均匀分布的独立源，以及两个篡改函数 f 和 g。我们将论证，对于不同标签 $\alpha \neq \alpha'$，混合二元特征

$$\chi(z, w) = \mathrm{Tr}_{K/\mathbb{F}_2}(\lambda\mathcal{E}_\alpha(z, w) + \mu\mathcal{E}_{\alpha'}(f(z), g(w))), \qquad \lambda \neq 0,$$

在 $A \times B$上不是常数，随后利用近似对偶提升，将偏差降至稍大积集上的指数级小值。$\mu = 0$ 的情形容易分析，因此我们重点考虑 $\mu \neq 0$的情形。为此，在分析中固定未使用的左侧后缀，此时左侧篡改函数 f 是 $\bar{x}$的确定性函数。稍微滥用一下记号，令篡改后左行的相应前缀为 $f(\bar{x})$，并假设 $\bar{x}$ 仍支撑在 A上。若 f 是仿射函数，那么 $\Gamma_c(f(\bar{x}))$也是关于 $\bar{x}$的比特的二次函数。因此，我们通过 *Gold* 缺陷 $d_c(f;\ A)$来衡量 f 偏离仿射函数的程度；其定义为：将 $\Gamma_c \circ f$ 限制在 A上后，其二元坐标函数在所有次数至多为 2 的布尔函数空间模下所张成空间的维数。

如果 $d_c(f;\ A)$ 很大，那么这意味着 $\chi(z, w)$ 中 Gold 项之和的秩已经很大，因为 $\Gamma_c(\bar{x})$ 关于 $\bar{x}$ 的比特是二次的，而 $d_c(f;\ A)$ 衡量了 $\Gamma_c(f(\bar{x}))$ 在二次布尔函数空间之外的维数。另一方面，带标签的提取部分之和的秩很小，因为其中每一项都展开为关于 K 的 J 分离乘积，并且二元矩阵秩至多为 $J\log_2|K|$。综合起来，这意味着 $\chi(z, w)$ 的二元矩阵秩为 > 1，因此它不是常数。

反过来，如果 $\chi(z,w)$ 是常数，那么这就迫使 $d_c(f;\,A)$ 很小。我们的摊还识别定理（定理 3.41）产生了一个大子集 $A_1 \subseteq A$，在该子集上，f 与仿射映射 h 一致。我们的定理的关键特征在于，Gold 缺陷 $\rho = d_c(f;\,A)$ 对仿射一致子集的熵亏损仅贡献 $O(\rho/c)$，而不是 $O(\rho)$。更准确地说，该定理给出了一个仿射一致子集 $A_1 \subseteq A$，其环境熵亏损为 $O(a+\rho/c+c)$，其中 $a = n_0 - \log_2|A|$。因此在我们的应用中，可以将 a 选为 n_0 的一个足够小的常数比例，并将 c 选为一个足够大的常数，从而使 A_1 的亏损成为 n_0 的任意小比例。尽管 A_1 不一定是仿射子空间，但环境均匀仿射子空间上的仿射非篡改提取器性质通过计数可知，存在一个值 z'，使得 $\{S_\alpha(\bar{x}) : \bar{x}\in A_1,\ S_{\alpha'}(h(\bar{x})) = z'\}$ 包含许多互不相同的值。随后，我们利用这一点和一个维数论证得出另一个矛盾。粗略地说，我们固定 $S_{\alpha'}(h(\bar{x})) = z'$，以及一对 (r,t) 和其篡改版本 (r',t')，使得右行仍具有很高的熵。现在，在混合特征 $\chi(z,w)$ 中，求和中的每一项都变成仅关于 $\bar{x}$ 或 w 的局部函数，原始输入上的带标签提取项 $vP_Y(u)$ 除外。利用 $vP_Y(u)$ 的双线性形式以及 $\chi(z,w)$ 为常数这一假设，我们可以论证，与 w 相关的两个子空间和 $\bar{x}$ 的一个编码彼此正交，这与它们的维数界相矛盾。

扩展到 C 种篡改。上述带标签提取器可抵抗一种篡改，并具有左源强性质。接下来，我们将其扩展为可处理 C 种篡改的提取器，采用类似交替提取的递归过程，在每个中间层将可处理的篡改数量加倍。不过，我们必须维持两种不同的不变量：运行递归所需的左源强提取器性质，以及有界域分离，后者是使全局二元特征秩与输入长度呈线性关系所必需的。具体而言，仅对二元矩阵秩加以限制并不足以满足第二个不变量，因为中间步骤会执行非线性运算，可能导致秩增加。

在第 $i+1$层，令 Z、$W \in \mathbb{F}_2^N$为两行输入，并令 Ext^i 为我们在前一层构造的左源强提取器。首先，我们使用贯穿整个构造的同一标签，将 Ext^i 应用于 Z和 W的两个短前缀，例如 Z'和 W'。假设 Ext^i 生成长度为 w_0的 $\mathsf{R} = \mathsf{R}(Z',\ W')$。令 $q = C+1$ 和 $v = w_0/(256q)$，设 P为 R 的一个前缀，包含 $64v$ 比特。计算

$$\mathsf{L} = H_{8v}(Z;P), \qquad \mathrm{Ext}^{i+1}(Z,W) = H_v(\mathsf{R};\mathsf{L}).$$

这里 H_t 是某种特定形式的标准强种子提取器：对于域 $\mathbb{F}_{2^{4t}}$上的种子 $(u,\ t_0)$，我们将源视为系数词 $d = (d_j)$，并且

$$H_t(d;(u,t_0)) = \mathrm{Tr}_{\mathbb{F}_{2^{4t}}/\mathbb{F}_{2^t}}\left(t_0\sum_j d_j u^j\right).$$

在 H_{8v} 和 H_v中，输入源的长度都是相关域次数的固定倍数，因此系数多项式的次数为 $O_C(1)$。

统计归纳。 假设 Ext^i 可以处理 C_i 次比较，我们将证明 Ext^{i+1} 可以处理$C_{i+1} = 2C_i$ 次篡改。为此，将这 $2C_i$ 次篡改分成两组 T，S 每组大小为 C_i。令 R_0 和 L_0 分别表示 R 和 L 在原始输入上的输出。我们将逐步固定一系列随机变量，同时保持 Z 和 W之间的独立性。

首先固定所有执行中的前缀 Z'，因此生成的所有 R 变量都是 W 的确定性函数；如果我们选择 Z' 足够短，Z 仍然具有足够的熵。由 Ext^i 能够处理 C_i 篡改这一事实可知，P_0 接近均匀分布，且独立于 $\mathbf{P}_S = (P_j)_{j\in S}$。因此，$H_{8v}$ 的强提取器性质保证了 L_0 接近均匀分布，且独立于

$\mathbf{L}_S = (\mathsf{L}_j)_{j \in S}$。此外，我们可以固定所有 P 变量，同时仍保持这一性质；此时 L 变量是 Z 的确定性函数。注意，在开始时，同样由 Ext^i 能够处理 C_i 篡改且具有左源强这一事实可知，R_0 接近均匀分布，且独立于篡改集合 T 中的所有 R 变量。因此，除了之前固定的所有 P 变量之外，我们还可以固定所有这些 R 变量。根据我们对长度参数的选择，R_0 仍然具有足够的熵。最后，H_v 的强提取器性质保证了 Ext^{i+1} 在原始输入上的新输出接近均匀分布，且独立于 S 中的所有篡改输出以及所有 L 变量。由于 T 中的所有 R 变量都已被固定，这意味着 Ext^{i+1} 在原始输入上的输出实际上接近均匀分布，且独立于 $S \cup T$ 中所有被篡改的输出，同时在左源上具有强性质。

实现秩性质。回顾标准双源提取器的分析，我们需要每一对行的输出满足以下两个秩性质：(1) 输出及篡改输出的每个混合特征中，只要诚实输出的系数非零，在限制到高熵矩形 $A \times B$ 上，对于每个给定常数 $K_* > 0$，其秩至少为 $K_* N$，这是将最终输出合并所必需的；(2) 输出中任意特征的全局秩由 $O(N)$ 上界约束，这是近似对偶提升最后一步所必需的。这正是我们使用上述特殊形式的强种子提取器 H 的地方。其思想是，我们将在合适的公共基域 $E = \mathbb{F}_{2^{\Theta_{C^{(N)}}}}$ 上工作，使得出现在这一固定深度递归中的每个域字，在 E 上仅具有 $O_C(1)$ 个坐标。我们在 E 上相容地选择坐标基，并且所有前缀和系数选择都使用完整的 E 坐标块。因此，这些选择以及相对迹都是 E-线性的，而乘法具有有界的坐标展开。对于一个取值于 E 的、关于两个 未条件化输入的函数，将其分离宽度定义为

$$\mathrm{sep}_E(U) := \min \left\{ r : \; U(z,w) = \sum_{i=1}^{r} a_i(z) b_i(w), \quad a_i, b_i \text{ are } E\text{-valued} \right\}.$$

关键在于，如果 E值函数具有常数分离宽度，那么输出比特的任意奇偶校验都具有 $O(N)$ 二元矩阵秩。特别地，上述 H 函数在 E上具有常数分离宽度。分离宽度具有以下基本闭包规则：

$$\begin{aligned} \mathrm{sep}_E(U+V) &\leqslant \mathrm{sep}_E(U) + \mathrm{sep}_E(V), \\ \mathrm{sep}_E(UV) &\leqslant \mathrm{sep}_E(U)\,\mathrm{sep}_E(V), \\ \mathrm{sep}_E(U^{2^j}) &\leqslant \mathrm{sep}_E(U). \end{aligned}$$

归纳应用这些规则，并注意到我们只进行常数次递归，因此可以将所有中间输出的分离宽度界定为 $O_C(1)$。

然而，我们还需要对限制在高熵矩形 $A \times B$ 上的二元秩给出一个下界。因此，我们改变<style id='2'>递归的最后一层，并将函数$H_v(\mathsf{R};\mathsf{L})$替换为

$$\mathrm{Tr}_{L/K}\left(bt(1+ur)^{J-1} \right), \qquad \mathsf{L} = (u,b) \in K^2, \quad \mathsf{R} = (r,t) \in L^2,$$

其中 J 是一个足够大的固定 2 的幂。在特征 2 下，恒等式 $(1+ur)^{J-1} = \sum_{i=0}^{J-1} (ur)^i$ 将其表示为已分离域字中的定宽多项式；因此，其全局二元特征秩仍为 $O_{C,\,J}(N)$。但是，在我们论证所得到的乘积矩形上，类似于标准双源提取器中的论证，诚实的 R_0 和 L_0 具有较大的支撑集，由此得到独立的受限秩下界

$$\mathrm{rank}_{\mathrm{restricted}} \geqslant 4Jv - \frac{w_0}{3} - 5.$$

因此，在熵参数确定之后再选择 J，可以使其大于任意指定常数乘以 N。

完成提取器。 构造的其余部分及分析与我们之前描述的标准双源提取器类似。最后一步是将常数数量的输出合并起来；在分析中，我们还要两次使用近似对偶提升，将误差降至指数级小。

实现最优输出长度。 我们使用输出长度为 $\Omega(n)$ 且误差为 $2^{-\Omega(n)}$的提取器，为给定误差 ε构造具有最优输出长度的改进的提取器。这里我们以能达到输出长度$2k - O(\log(1/\varepsilon) + 1)$ 的标准双源提取器为例，说明主要思想。我们之前的双源提取器 $G(X, Y)$ 的误差为 $2^{-\Omega(n)}$，实际上是一个强双源提取器，因此自然可以采用以下方法：首先使用该双源提取器生成一个短种子 $S = G(X, Y)$，其长度为 d 比特；然后使用两个具有最优参数 E_L, E_R 的强种子提取器，并输出

$$\mathrm{Ext}(X, Y) = \big(S, E_L(X, S), E_R(Y, S)\big).$$

强性质保证了 $(S, E_L(X, S))$ 或 $(S, E_R(Y, S))$ 接近均匀分布。然而，要证明 $(S, E_L(X, S)$ $E_R(Y, S))$ 接近均匀分布并不直接。

令 $s = \lceil \log_2(1/\varepsilon) \rceil$，我们非对称地选择两个输出长度： $E_L(X, S)$ 输出 $p_L = k - d - 2s - O(1)$ 个比特，而 $E_R(Y, S)$ 输出 $p_R = k - O(s + 1)$ 个比特。在分析中，主要问题在于 S 同时依赖于两个源。我们通过根据 Y 的实际边缘分布对第一个输出和种子的条件碰撞概率进行加权并加以界定，从而解决这一问题。利用柯西-施瓦茨不等式，可以将所得的条件碰撞概率界转换为统计距离。这自然表明，我们需要考虑在给定 Y 的条件下， $G(X, Y)$ 在 X 的不同副本上的输出，这可以通过更强的<style id='27'>单侧不可篡改性性质来刻画。也就是说，我们只需要 $G(X, Y)$ 对单边篡改具有不可篡改性。因此，例如，我们可以直接取 $G(X, Y)$ 作为上述描述的不可篡改双源提取器。采用第 4.1 节中的参数选择，可得

$$(Y, S, E_L(X, S)) \approx_{\varepsilon/4} (Y, U_d, U_{p_L}),$$

其中右侧的均匀分布变量彼此独立，且与 Y 独立。一旦具备这一性质，我们就可以使用强种子提取器E_R，安全地附加 $E_R(Y, S)$ 的第二个输出，并且输出长度和误差也随之得到。

对于抵抗 C 次篡改的强不可篡改提取，我们采用对称构造

$$\mathsf{nmExt}(X, Y) = \big(S, E(X, S) \oplus E(Y, S)\big),$$

其中， E 是一个强种子提取器。这里的种子必须能够抵抗 $2C + 1$ 次篡改，因为碰撞论证会为每个原始输入和篡改后的输入引入第二个输入。幸运的是，我们的输出长度为 $\Omega(n)$ 且具有 $2^{-\Omega(n)}$ 的多重篡改不可篡改双源提取器仍然能够应对这一情况。

对于和集提取器，我们使用一个种子 $S = G(Z)$，它可由 $Z = X + Y$计算得到，方法是结合先前的和集提取器与线性种子提取器LExt。与之前一样，我们需要和集提取器具有单侧不可篡改性，因此采用了稍作修改的构造，以实现这一性质，见定理3.33。该构造如下：

$$\mathrm{SumExt}(Z) = \big(S, \mathrm{LExt}(Z, S)\big).$$

实现 Cn/对数 n **熵**。上述构造可以直接扩展，以处理略低于常数的熵率。然而，要实现 Cn/对数 n 熵，我们使用某处高熵压缩器，例如 [6, 58, 50], 其中的压缩器输出多对数数量的行，每行长度为n/多对数 n，且至少有一行与具有高熵的分布指数接近。假设我们从每个源得到 $r = 2^d$ 行，那么我们尝试两个源的行的所有可能配对，得到 r^2 对，其中至少有一对在两个源中都具有高熵。

从这里开始，我们只需将每一行视为 F^t 中的向量，其中$F = \mathbb{F}_{2^\ell}$是适当选择的域，并在 F上应用标准的内积双源提取器。我们将证明，通过适当选择参数， r^2 输出的联合分布与以下内容的凸组合在指数意义下十分接近：仿射某处随机源。后者是在 r^2 行上的仿射源，并且其中某一行实际上服从均匀分布。

这项工作受到 Aggarwal、Dodis 和 Lovett [1] 关于分割状态不可篡改码的研究启发。在该研究中，他们证明：如果 L 和 R 是 $\mathbb{F}_p^n$上的两个独立均匀随机字符串，其中 p为某个素数，并且 f、$g : \mathbb{F}_p^n \to \mathbb{F}_p^n$ 是定义在 L 和 R上的两个任意函数，那么联合分布 $(L, R, f(L)\ g(R))$ 与“仿射分布” $\{(U, aU + b)|a, b \in \mathbb{F}_p\}$ 的凸组合“接近”，其中 U是 $\mathbb{F}_p$上的均匀分布。在我们的情形中，不同的是，L和 R 是两个独立的高熵源，它们来自对 X 和 Y应用某处高熵压缩器后的输出；通过固定额外的随机性，我们从压缩器得到的其他输出可以被视为定义在 L 和 R上的确定性篡改函数。此外，在我们的应用中，我们有 $r^2 - 1$ 对篡改函数，而不是仅仅一对。还有一个技术问题：现在我们是在域$\mathbb{F}_{2^\ell}$ 上工作，而不是在素域 $\mathbb{F}_p$上。

不过，对于高最小熵源，我们确实可以证明该结果的二元仿射多输出类比形式 [1]。证明再次使用了 [8]中的近似对偶表述，大致如下。首先，我们不使用具有非线性映射的任意某处高熵压缩器，而是像构造和集提取器时那样，使用 [50] 中的线性凝聚器。这样做的优点是，在分析中，内积输出的每个二元奇偶特征实际上都是原始输入比特上的双线性函数。即使我们考虑 r^2 输出的联合二元奇偶特征，这一点仍然成立。因此，组合后的特征维数仅为 $2n$，与成对数量 r^2无关；在应用近似对偶性时，这有助于我们获得更好的参数。另一方面，使用非线性的某处高熵压缩器可能会导致特征维数大得多，并且依赖于 r^2。现在，在分析中，我们考察 r^2 输出的联合分布，并考虑其支撑集的仿射包。如果该分布已经接近于仿射子空间上的均匀分布，那么基本上就完成了，因为内积保证至少有一个输出接近均匀分布。否则，在该仿射子空间上存在某个非平凡特征，使得它在该分布下具有较大偏差；于是我们可以调用近似对偶性，找到一个大矩形，在其上该特征为常数。随后，我们利用这个矩形将原始输入空间划分为互不相交的矩形（从而保持源的独立性），并继续这一过程。通过使用势函数，我们证明该过程会在有限步内终止；除去一小部分概率质量后，其余矩形上的输出分布都接近于某个仿射子空间上的均匀分布。此外，我们还证明在这一过程中不会损失太多熵。因此，在任意足够大的矩形上，根据凝聚器和内积双源提取器的性质，输出仿射子空间的某个优行投影具有满支撑。因此，该仿射子空间上的均匀分布具有一个完全均匀的诚实行。现在我们可以得出结论：原始联合分布接近于仿射某处随机源的凸组合，其中包含 r^2 行。

只要行数 r^2 相对于每行的长度较小，就有非常优秀的已知提取器，例如 Rao [55] 利用重复条件化的方法以及后续研究所提出的提取器

使用仿射相关性破坏器 [12, 49] ，可以从这类仿射某处随机源中提取误差呈指数级小的随机比特。对于不可篡改提取器，这里还需要进行一些额外工作，最终我们会应用基于建议的相关性破坏器。

然而，上述方法只能得到误差为 $2^{-k^{\Omega(1)}}$ 而非 $2^{-\Omega(k)}$ 的双源提取器，即使熵率为常数也是如此。事实上，误差来自上述构造中的两个步骤：将源转换为仿射某处随机源，以及随后从这种源中提取。若要通过第二步将误差降至 $2^{-\Omega(k)}$ ，即使行数为常数，也需要行长度达到 $\Omega(k)$；这会迫使我们在第一步中选取大小为 $2^{\Omega(k)}$ 的域，因此内积中的块数只能是常数。在这种情况下，近似对偶性的分析会失效。我们的正则化定理（定理 2.30 和 5.5）中的熵损失界更具体地揭示了这一局限。对于常数行数和常数熵率，若每个域上内积的输出有 ℓ 比特，那么在秩阈值为 R 时，正则化熵损失的两个主要部分大致为：来自高秩近似对偶性的 $O\left(\frac{k\ell}{R}\right)$，以及来自低秩精确消息步骤的 $O(\ell R)$。平衡这两项可得 $R = \Theta(\sqrt{k})$，合并后的熵损失为 $O(\ell\sqrt{k})$。这迫使我们选择足够小的常数因子，使 $\ell = O(\sqrt{k})$ 成立。因此，通过这种方法，我们只能达到输出长度和误差指数的平方根尺度。可以结合强种子提取器，将输出长度提升至 $\Omega(k)$，但误差仍为 $2^{-\Omega(\sqrt{k})}$。当熵率为 $\Theta(1/\log n)$ 时，凝聚器得到的行数会进一步增加至多对数级，而输出扩展损失会导致误差范围为 $2^{-n^{1/2-\vartheta}}$ ，对每个固定的 $\vartheta > 0$ 都如此。毫不意外的是，我们针对常数熵率、具有渐近最优参数的提取器采用了截然不同的构造和技术。

独立研究与致谢。Doron 和 Goodman [31] 构造了显式双源提取器，可处理熵为 $k \geq n^{1-\alpha}$ 、误差为 $2^{-k^{\alpha}}$ 且常数为 $\alpha > 0$的情形。对于常数熵率，他们的误差仍为 $2^{-k^{\alpha}}$ 。因此，他们的提取器适用于比我们的提取器熵更低的源；而对于任意常数熵率，我们的提取器具有更好的误差和输出长度。他们的技术与我们针对熵为 $Cn/\log n$的提取器有一些相似之处：两种构造都先利用近似对偶性将两个源转化为仿射某处随机源的凸组合，然后再应用针对这类源的提取器。Doron 和 Goodman 更进一步，运用额外的新思路和技术将熵降低到$k \geq n^{1-\alpha}$。另一方面，我们针对常数熵率的渐近最优提取器也采用了与他们的结果无关的新思路和技术。

Doron 和 Goodman 于2025年12月底得到了初步结果。在补充细节并撰写结果的过程中，作者于2026年8月左右独立发现了本文中针对熵为 $Cn/\log n$ 的提取器，当时并不知道他们的结果。此后，作者一直沿着两个方向开展研究：降低熵要求和改进误差。然而，2026年9月参加西蒙斯研究所的“伪随机性与高维扩张”项目期间，作者了解到他们尚未发表的结果及其核心思路。随后，作者将重点转向降低提取器的误差，并最终获得了针对常数熵率、具有渐近最优参数的新提取器构造。作者感谢西蒙斯研究所提供的款待与支持。

人工智能披露。 作者在探索证明思路、开展文献调研以及起草稿件的过程中大量使用了 ChatGPT Astra。大部分高层次的想法都属于作者，而 ChatGPT 也为完整证明贡献了重要的技术思路。ChatGPT Astra

还用于协助完成选择适当参数这一繁琐工作。在与 ChatGPT 的多次交流中，许多中间方案均以失败告终，最终才形成了成功的构造。本文所展示的构造还包含了对先前更为复杂构造的简化。

论文结构。 第2节汇集了通用预备知识，包括源强提取器、秩敏感近似对偶性、浓缩器、带种子原语以及共享特征正则化定理。第 3 节构造了具有线性输出长度和指数级小误差的常数率方案：标准双源提取器、和集提取器，以及能够抵御任意固定数量篡改的非可塑双源提取器。第4 节将前述构造用作短种子，并通过熵回收论证，为上述提取器实现最优输出长度。最后，第 5 节给出了熵为 $k = \Omega(n/\log n)$ 的构造。

2 预备知识与通用工具

引用的定理仅在其所述假设下使用。内部引理及其证明按照构造所需的顺序排列。在不同小节中，同一符号可能表示局部域长度；原始输入长度仍记为 n。

记号与约定。对于非负整数 r，记 $[r] = \{1, \dots, r\}$，其中 $[0] = \varnothing$，并且 $(x)_+ = \max\{x, 0\}$。记号 $\mathbf{1}_E$ 表示事件或集合 E的示性函数。对于非空有限集 S， U_S 表示其均匀分布； U_m 是 $U_{\mathbb{F}_2^m}$的缩写。乘积 $U_S^{\otimes r}$由 r 个相互独立的均匀副本构成。在理想比较中，除非另有说明，新引入的均匀随机变量与保留变量相互独立。我们用 $\mathrm{L}(X)$ 表示 X的分布，用 $\operatorname{supp} X$ 表示其正概率支撑集；当 $\Delta(\mathrm{L}(X), \mathrm{L}(Y)) \leqslant \varepsilon$时，记作 $X \approx_\varepsilon Y$，其中使用下文定义的统计距离。随机元组之间的符号 $\perp$ 表示独立。对于有限字母表上的分布 P，其碰撞概率为 $\mathrm{cp}(P) = \sum_z P(z)^2$；我们也记 $\mathrm{cp}(X) = \mathrm{cp}(\mathrm{L}(X))$。对于子分布，采用相同的求和方式，但不对其质量进行归一化。

除非另有说明，向量空间张成、维数、余维数、正交性和矩阵秩均是在 $\mathbb{F}_2$上计算的。对于布尔核，记号 $\operatorname{rank}_2 h$ 表示其取 $0/1$ 值的矩阵的秩，而非其符号矩阵的实数秩。对于有限集 A， $\mathbb{F}_2^A$ 表示所有函数 $A \to \mathbb{F}_2$构成的空间。对于二元向量空间的非空子集 S 以及任意 $s_0 \in S$，其仿射包为

$$\mathrm{aff}(S) = s_0 + \mathrm{span}_{\mathbb{F}_2}\{s - s_0 : s \in S\}.$$

其维数即所示方向空间的维数。仿射源在一个仿射子空间上均匀分布；当一个元组的拼接具有此性质时，称其为联合仿射。对于核 Q，记号 $Q(A, B)$表示$\{Q(x, y) : x \in A, \ y \in B\}$，而非概率分布。除非另有说明，域的二元坐标表示之间的所有仿射映射均为 $\mathbb{F}_2$-仿射。

对于比特串 z，前缀$d(z)$表示其前 d 个坐标，其中 $0 \leqslant d \leqslant |z|$。比特串上的运算 $\oplus$是逐坐标的模二加法，而 $z \,\|\, z'$ 表示拼接。在向量空间或码上， $\oplus$ 表示直和。对于域 K， K^* 和 $K^\times$ 均表示 $K \setminus \{0\}$。

定义 2.1 （普通双源提取和和集提取）。 具有分量熵 k 和误差 ε 的普通双源提取器是一个映射 $E\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$，使得对于最小熵至少为 k 的每个独立对，都有 $E(X, Y) \approx_\varepsilon U_m$。和集提取器是一个映射 $S : \mathbb{F}_2^n \to \{0,1\}^m$，使得对于每个这样的独立对，都有 $S(X + Y) \approx_\varepsilon U_m$。仅和求值器的保证本身并不能断言在公开一个原始加数后源仍具有源强。

定义 2.2（不可篡改双源提取）。映射 $E\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ 是一个在熵 k 和误差 ε 下的 t-不可篡改双源提取器，如果对于熵至少为 k 的独立源 X, Y 以及满足 $\Pr[(f_j(X), g_j(Y)) = (X, Y)] = 0$ 的任意分离映射 f_j, g_j，都有

$$\Delta\big((E(X,Y), (E(f_j(X), g_j(Y)))_{j=1}^{t}), (U_m, (E(f_j(X), g_j(Y)))_{j=1}^{t})\big) \leqslant \varepsilon.$$

新鲜的均匀随机变量与实际的联合篡改输出分布相互独立。对篡改输入不作任何熵假设。对于 $t = 0$，比较元组为空，此时该定义即为普通双源提取。所有固定-C 的表述均采用此约定。

2.1 源强提取与熵松弛转移

定义 2.3（分离源强）。设 $Z = F(X, Y)$，对于一个由 t 个独立篡改组成的可容许元组，令 $T_j = F(f_j(X), g_j(Y))$，且 $\mathbf{T} = (T_1, \ldots, T_t)$。若一个 t-不可篡改函数 F 满足：对于所有相互独立的 k-源及所有此类篡改，

$$\Delta((Z, \mathbf{T}, X), (U_m, \mathbf{T}, X)) \leqslant \xi, \qquad \Delta((Z, \mathbf{T}, Y), (U_m, \mathbf{T}, Y)) \leqslant \xi.$$

这两项保证彼此独立；它们都不涉及 (X, Y) 的联合分布。对于普通源强提取，令 $t = 0$，并省略 $\mathbf{T}$。

引理 2.4（源强输出不能超过一个源的熵）。设 k 为满足$0 \leqslant k \leqslant n$的整数。若一个确定性双源函数在分量熵 k 下具有误差为 $\varepsilon < 1$ 的源强性质，且输出长度为 m，则

$$\varepsilon \geqslant 1 - 2^{k-m}, \qquad \text{hence} \qquad m \leqslant k + \log \frac{1}{1-\varepsilon}.$$

In 尤其对于 $\varepsilon < 1/2$，源强输出不能包含超过 k 比特 .

证明。取恰好在 2^k 个点上均匀分布的 Y，并任选一个独立的 k-源 X。对于每个固定值 $X = x$，条件诚实输出仅是 Y 支撑集中 2^k 个点的函数，因此其支撑大小至多为 2^k。它与在 2^m 个输出上均匀分布的统计距离至少为 $1 - 2^{k-m}$。对 x 求平均，即可得到强距离下界。重新整理后即得所述结论。 □

有限初始长度。当线性输出定理适用于每个正输入长度，但其域构造仅针对 $n \geqslant n_0$ 给出时，采用以下确定性补全即可。对于 $1 \leqslant n < n_0$，返回一个恒为零的比特。即使比较元组任意取值，且保留任一原始源，将该比特替换为独立均匀分布的比特，其统计距离也恰为 $1/2$。每个输出特征的二元真值表秩均为零。如果大长度情形下的输出和误差常数为 c，$a > 0$，则将其分别替换为 $\min\{c,\ 1/n_0\}$ 和 $\min\{a,\ 1/n_0\}$。这样，单比特补全在整个有限初始范围内同时满足 $m(n) \geqslant cn$，且误差至多为 2^{-an}。下文会明确采用此约定；它并未将加性损失定理扩展到其所述的非平凡误差范围之外。

引理 2.5（逐原子熵松弛强强化）。设 F 是一个在分量熵阈值 h下、输出长度为 m_0且误差为 γ 的 t不可篡改双源提取器，其中 $t = 0$ 表示普通提取。固定 $k \geqslant h$，令 $0 \leqslant r \leqslant m_0$，并令 F_r 为前 r 个输出比特。对于每一对独立的 k-源以及任意由 t 个篡改组成的可容许元组，关于 F_r 的两个源强比较的误差均至多为

$$2^{(t+1)r}\big(\gamma + 2^{h-k}\big). \tag{2.1}$$

如果用秩为 r 的固定二元线性满射替代坐标截断，并将所有实际的篡改输出通过同一映射投影，该陈述仍然成立。

证明。 截断和任何固定的满射线性投影都会根据数据处理性质保留初始 t-不可篡改误差，因此以下处理 r比特的输出。令$\mathbf{T}$ 表示实际的比较元组。固定一个（z，$\mathbf{v}$）原子，其中包含（$t+1$）r 输出比特；对于原始左支撑集中的每个数值 x，定义

$$D_{z,\mathbf{v}}(x) = \Pr[Z = z, \mathbf{T} = \mathbf{v} \mid X = x] - 2^{-r} \Pr[\mathbf{T} = \mathbf{v} \mid X = x].$$

记 $b = k - h$。若集合 $E = \{x : D_{z,\ \mathbf{v}}(x) > \gamma\}$ 的概率至少为 2^{-b}，则在 E上以 X 为条件后，最小熵将至少为 h （每个条件原子的概率至多为 $2^{-k}/\Pr[E]$）。独立源 Y 将保持不变，原有的篡改映射在较小的乘积支撑上仍然是可接受的。对于该源，这个单一联合原子的正超额将超过 γ，与不可篡改性相矛盾。因此$\Pr[E] < 2^{-b}$。由于$|D_{z,\ \mathbf{v}}(x)| \leqslant 1$,

$$\mathbb{E}_X[D_{z,\mathbf{v}}(X)]_+ \leqslant \gamma + 2^{-b}.$$

条件有符号原子差之和为零。因此，条件全变差等于其正部之和。对 $2^{(t+1)r}$ 原子求和即可证明（2.1）中的第一个比较。交换 X 与 Y 后重复该论证即可证明第二个比较。对于篡改图像，不假设其具有独立性或熵。

□

备注2.6（回收参数与条件熵）。对于源强输出扩展，$C \geqslant 0$ 统计最终的篡改输出，而处理 $R = 2C + 1$辅助篡改则需要短种子。整个第4.2节始终保持这一点区分。对于有限变量，记

$$\widetilde{H}_\infty(A \mid T) = -\log \sum_t \Pr[T = t] \max_a \Pr[A = a \mid T = t].$$

揭示一个至多有 2^b 个值的变量，在该平均条件最小熵中至多付出 b 的代价。随后由马尔可夫不等式可知，在质量 ξ之外，逐点熵至少为平均界减去 $\log(1/\xi)$。对于次分布，我们还将全变差距离的一半作为代价；在中心化回收证明中，会明确计入丢弃质量。条件源的独立性仅用于矩形通信叶，或用于一个明确独立的参考实验。

2.2 初等概率、傅里叶分析与秩

定义 2.7 （源与提取）。 对于有限随机变量 X，其最小熵为$H_\infty(X) = -\log \max_x \Pr[X = x]$。一个平坦 k-源，其中 k为整数，是在恰好包含 2^k个点的集合上均匀分布的。我们用统计距离 $\Delta(P, Q) = \frac{1}{2}\sum_z |P(z) - Q(z)|$ 表示有限分布P, Q之间的距离。

一个强种子（k, ε）提取器 E 满足

$$\Delta((U_d, E(X, U_d)), U_{d+m}) \leqslant \varepsilon$$

对于最小熵至少为 k 的每个源，其中 U_d 与 X 独立。普通双源提取器使用独立的 X，Y；分别强性意味着两种保证分别保留 X 或保留 Y。在侧阈值为 k 时，布尔双源分散器在每个支撑集大小至少为 2^k 的支撑集乘积上都不是常数。

引理 2.8（平坦分解与初等条件化）。 对于点概率至多为 $1/K$ 的每个分布（其中 $K \geqslant 1$ 为整数），它都是定义在 K 元集合上的均匀分布的凸组合。对这样的分布分别进行独立分解，会得到乘积分布的凸组合。如果事件 A 仅取决于 X，那么对于每个确定性视图 T，

$$H_\infty(T(X) \mid A) \geqslant H_\infty(T(X)) + \log \Pr[A].$$

独立源中的分离事件保持独立性。

证明。 多面体 $\{p : \sum_x p_x = 1,\ 0 \leqslant p_x \leqslant 1/K\}$ 的极点仅为定义在 K 元素集合上的均匀分布。事实上，如果两个坐标严格位于各自的界限之间，那么在它们之间转移足够小的量，就能将该点表示为两个不同点的中点。如果只有一个坐标是分数，那么由方程 $\sum p_x = 1$ 及 K 的整数性可知，它也必须是一个界限坐标。有限维多面体是其极点的凸包。将两个凸分解作张量积，即得独立性结论。条件化界由 $\Pr[T(X) = z \mid A] \leqslant \Pr[T(X) = z]/\Pr[A]$ 得出。分别对事件作条件化后，联合条件分布可直接分解。

□

平坦分解中的整数下取整至多损失一个熵比特；之后的每次使用都会预留一个固定的正线性熵裕量。

引理 2.9（源局部泄露）。 若 $H_\infty(X) \geqslant k$，且 $L(X)$ 至多取 2^v 个值，则除总概率至多为 2^{-t} 的标签外，

$$H_\infty(X \mid L(X) = l) \geqslant k - v - t.$$

证明。 丢弃概率低于 $2^{-(v+t)}$的标签。这样的标签至多有 2^v 个。对每个剩余标签，应用引理 2.8。

□

引理 2.9a（交互式转录泄露）。 设 X、Y 是相互独立的有限随机变量。设 T 是一个有限确定性二进制通信协议的转录：在每个内部节点，发言方由此前的转录决定，而下一比特是该发言方的输入和此前转录的函数。假设在每次执行中，左方和右方分别至多发送 b_L 和 b_R 比特。每个有支撑的转录叶节点都保持 X、Y 的独立性。对于每个 $u > 0$，在转录之外的概率至多为 2^{1-u}，并且对每个确定性视图 V（X） 和 W（Y） 均同时成立，

$$H_\infty(V(X) \mid T = \tau) \geqslant H_\infty(V(X)) - b_L - u,$$
$$H_\infty(W(Y) \mid T = \tau) \geqslant H_\infty(W(Y)) - b_R - u.$$

相同的良好叶节点适用于所有视图：对这样的叶节点进行条件化后，任何仅依赖于 X 的事件的概率至多增加 2^{b_L+u}，而任何仅依赖于 Y 的事件的概率至多增加2^{b_R+u}。此外，按照备注 *2.6*中的平均条件最小熵约定，

$$\widetilde{H}_\infty(V(X) \mid T) \geqslant H_\infty(V(X)) - b_L, \qquad \widetilde{H}_\infty(W(Y) \mid T) \geqslant H_\infty(W(Y)) - b_R.$$

证明。 每个受支持的转录都有一个矩形原像 $A_\tau \times B_\tau$。令 $p_\tau = \Pr[X \in A_\tau]$和$q_\tau = \Pr[Y \in B_\tau]$。由独立性可得 $\Pr[T = \tau] = p_\tau q_\tau$，而该叶节点上的条件分布是 $X \mid A_\tau$ 和 $Y \mid B_\tau$ 的乘积分布。

固定一个数值右输入 y。将协议树修剪到该输入后，只剩下左消息分支：每条右消息都由 y 和此前的转录决定。修剪后的二叉树沿一条路径至多有 b_L 个分支比特，因此至多有 2^{b_L} 个叶节点。对 y 取平均可得

$$\sum_{\tau} q_\tau = \mathbb{E}_Y \big|\{\tau : Y \in B_\tau\}\big| \leqslant 2^{b_L}.$$

对称论证给出 $\sum_\tau p_\tau \leqslant 2^{b_R}$。因此，

$$\Pr_T[p_T < 2^{-b_L-u}] = \sum_{\tau : p_\tau < 2^{-b_L-u}} p_\tau q_\tau \leqslant 2^{-b_L-u} \sum_\tau q_\tau \leqslant 2^{-u}.$$

类似地，$\Pr_T[q_T < 2^{-b_R-u}] \leqslant 2^{-u}$。在这两个异常集合并集之外的每个叶节点上，每个左侧局部事件 E 都满足

$$\Pr[E \mid T = \tau] = \Pr[E \mid X \in A_\tau] \leqslant \frac{\Pr[E]}{p_\tau} \leqslant 2^{b_L+u} \Pr[E].$$

这同时给出了对所有左侧局部视图的左侧事件界和最小熵界。相同的论证也证明了右侧界。没有对视图数量使用并集界。

最后，写成 $a_V = \max_v \Pr[V(X) = v]$。现在对所有受支撑的叶节点取平均，同样的条件原子界给出

$$\sum_\tau \Pr[T = \tau] \max_v \Pr[V(X) = v \mid T = \tau] \leqslant \sum_\tau p_\tau q_\tau \frac{a_V}{p_\tau} \leqslant 2^{b_L} a_V.$$

取负的二进制对数即可证明左侧的平均条件最小熵断言；右侧完全相同。该证明不要求任一源是平坦的，也不要求任何视图是单射的。

□

引理 2.10（条件代表元表）。对于任意有限对 $(U,\ X)$，其分布是分布 $(U,\ f(U))$ 的凸组合，其中确定性表 f 的随机选择与后续的 U 独立，并且每个分量都具有原始 U 边缘分布。对于乘积分布的两侧，该操作可以独立执行。

证明。对于每个概率为正的 u，独立地按照 $X \mid U = u$ 抽取 $f(u)$，然后从其原始分布中抽取 U。对所有表取平均即可重现该二元组。概率为零的条目可任意填充。如果 U 是 X 的函数，则每个抽取的代表元都位于其正确的纤维中。两个表的选择相互独立，因此可得最后一个断言。

□

引理 2.11（傅里叶距离）。设 $P,\ Q$ 是定义在 $\mathbb{F}_2^m$ 上的分布，并记 $\widehat{P}(\lambda) = \mathbb{E}_{Z\sim P}(-1)^{\lambda\cdot Z}$。则

$$\Delta(P,Q) \leqslant \frac{1}{2} \left(\sum_{\lambda \in \mathbb{F}_2^m} |\widehat{P}(\lambda) - \widehat{Q}(\lambda)|^2 \right)^{1/2}.$$

特别地，如果 P 的每个非零特征的偏差至多为 β，则 $\Delta(P,\ U_m) \leqslant 2^{m/2}\beta/2$。如果将 $(A,\ B)$ 与 $(U_a,\ B)$ 进行比较，则只有具有非零 A 系数的特征会产生贡献。

证明。由帕塞瓦尔恒等式可得 $2^m \sum_z (P(z) - Q(z))^2 = \sum_\lambda |\widehat{P}(\lambda) - \widehat{Q}(\lambda)|^2$。对 $\frac{1}{2}\sum_z |P(z) - Q(z)|$ 应用柯西–施瓦茨不等式。关于均匀边缘分布的断言由正交性得出。

□

统计距离在确定性映射和随机核下会收缩，并且在具有共同权重的混合下具有凸性。这些事实直接由定义中的 ℓ_1 和的三角不等式得出。

定义 2.12（分离与公共特征）。若 $H(x,\ y) = \sum_{i=1}^{R} a_i(x) b_i(y)$，则域值核在 F 上的分离长度至多为 R。对于布尔核，这就是其在$\mathbb{F}_2$上的真值表秩。若存在固定的 $u(x)\ v(y) \in \mathbb{F}_2^R$，使得每个输出特征都是关于同一对映射的双线性形式，则向量输出具有维数为 R 的公共二元特征 。

秩在加法下具有次可加性，并且在任意分离预组合下不会增加。分离长度在乘积下相乘。对于 $F = \mathbb{F}_{2^\ell}$，一个分离的 F-乘积的绝对迹秩至多为 ℓ。更一般地，次数为 a 的扩域 K/F 的迹对偶基给出

$$\mathrm{Tr}_{K/F}(xy) = \sum_{j=1}^{a} x_j y_j.$$

因此，相对迹对于 a 个分离乘积的代价至多为 a 。将有限多个域值行的所有左因子和右因子串接起来，可以为整个表生成一对公共二元特征，而不仅仅是分别针对各个输出特征的独立秩界。

显式构造的域。我们使用固定的二进制基，并采用固定特征下的确定性有限域构造，如 Shoup [57]所述。只要求相对扩域次数超过某个固定常数时，可以将其选为 2 的幂；随后，固定高度的二次塔便可提供 K/F 及其相对迹。所有此类算术运算相对于所显示的二进制域次数都是多项式时间的。域常数均独立于各个源固定。

2.3 奇次幂特征与稠密张成

引理 2.13（限制下的秩）。设 $H\colon\ V \times W \to \mathbb{F}_2$ 为双线性形式，且设 $V_0 \leqslant V$、 $W_0 \leqslant W$。则

$$\mathrm{rank}(H|_{V_0 \times W_0}) \geqslant \mathrm{rank}(H) - \mathrm{codim}\, V_0 - \mathrm{codim}\, W_0.$$

如果 $A \subseteq V$ 和 $B \subseteq W$ 分别张成 V_0 和 W_0，则矩阵 $[H(a,\ b)]_{a \in A,\ b \in B}$ 的秩恰好等于 $H|_{V_0 \times W_0}$。

证明。对一个参数施加限制，至多会从独立的行或列中删除其余维数对应数量的行或列。这证明了不等式。对于第二个断言，从 A 和 B中选取张成基。相应的子矩阵表示受限制的双线性形式，而其他每一行或列都是这些行或列的线性组合。 □

引理 2.14（奇次数加性特征界）。参考文献。 [53, 引理 1]（加性韦伊界的特征为 2 的正奇数次情形）。设 $K = \mathbb{F}_{2^N}$，且令 $P \in K[z]$ 具有正奇数次 d。则

$$\left|\mathbb{E}_{z \sim U_K} (-1)^{\mathrm{Tr}_{K/\mathbb{F}_2} P(z)}\right| \leqslant (d-1) 2^{-N/2}.$$

这是特征为二情况下的加性韦伊界；正的奇数最高次数排除了阿廷–施莱尔例外形式。我们在固定次数和充分大的 N 下使用它；见 [53, 引理1]。线性非恒定相位的偏差为零。

引理 2.15（小特征偏差迫使稠密特征张成空间）。令 $\Phi:\Omega\to\mathbb{F}_2^r$，并假设 $\Phi(U_\Omega)$ 的每个非零特征的绝对偏差至多为 β。如果 $A\subseteq\Omega$ 的密度至少为 $\alpha>0$ 且$\beta\leqslant\alpha/2$，则

$$\operatorname{codim}_{\mathbb{F}_2}\operatorname{span}\Phi(A)\leqslant\log(1/\alpha)+1.$$

证明。记作 $V=\mathbb{F}_2^r$，并将 $V^*=\mathrm{Hom}_{\mathbb{F}_2}(V,\mathbb{F}_2)$ 视为所有二元线性特征组成的空间，利用标准点积将 V^* 与 V 进行对应。令

$$S=\operatorname{span}_{\mathbb{F}_2}\{\Phi(\omega):\omega\in A\}\leqslant V,\qquad c=\operatorname{codim}_{\mathbb{F}_2}S.$$

S 的湮灭子为

$$S^\perp=\{a\in V:a\cdot s=0\text{ for all }s\in S\}.$$

由初等线性代数可得 $\dim S^\perp=c$，因此 $|S^\perp|=2^c$。我们首先证明对每个 $v\in V$均成立的特征正交恒等式，

$$\mathbf{1}[v\in S]=2^{-c}\sum_{a\in S^\perp}(-1)^{a\cdot v}. \tag{2.2}$$

若 $v\in S$，则每个加数均为 1。若 $v\notin S$，则双重湮灭子恒等式 $(S^\perp)^\perp=S$ 给出一个$a_0\in S^\perp$，使得 $a_0\cdot v=1$。以 a和 $a+a_0$ 为索引的项随后符号相反并相互抵消，因此总和为零。

令 U 在有限集 Ω 上服从均匀分布。每个 $\omega\in A$ 均满足 $\Phi(\omega)\in S$，即使 Φ 是非单射的。consequently,

$$\Pr[\Phi(U)\in S]\geqslant\Pr[U\in A]=\frac{|A|}{|\Omega|}\geqslant\alpha. \tag{2.3}$$

将 $v=\Phi(U)$代入（2.2）并对 U取平均。由于零特征的期望为 1，我们得到 精确的 恒等式

$$\begin{aligned}\Pr[\Phi(U)\in S]&=2^{-c}\sum_{a\in S^\perp}\mathbb{E}_U(-1)^{a\cdot\Phi(U)}\\&=2^{-c}+2^{-c}\sum_{a\in S^\perp\setminus\{0\}}\mathbb{E}_U(-1)^{a\cdot\Phi(U)}.\end{aligned} \tag{2.4}$$

该假设将每个非零特征的期望的绝对值限制为 β。此类项恰有 2^c-1 个，因此（2.3）和（2.4）推出

$$\alpha\leqslant 2^{-c}+(1-2^{-c})\beta=\beta+(1-\beta)2^{-c}.$$

因为 $\beta\leqslant\alpha/2$ 且 $\alpha\leqslant 1$，因此有 $\beta<1$。重新整理可得稍微更精确的界

$$2^{-c}\geqslant\frac{\alpha-\beta}{1-\beta}\geqslant\frac{\alpha}{2}.$$

现在取以 2 为底的对数，得到

$$c\leqslant\log_2(2/\alpha)=\log_2(1/\alpha)+1,$$

正如所述。证明仅使用 A 在 Ω 中的密度；并不假设 A 是仿射的，也不假设 Φ 是单射。

□

引理 2.16（全域奇次幂特征）。对于 $K=\mathbb{F}_{2^N}$ 和一个固定整数 $J\geqslant 1$，定义

$$\Psi_J(z)=(z,z^3,\dots,z^{2J-1})\in K^J.$$

$\Psi_J(U_K)$ 的每个非零二元特征的偏差至多为 $\beta_J=(2J-2)2^{-N/2}$。对于每个固定的$0<\tau<1/2$ 以及足够大的 N（取决于 J，τ），每个满足 $|A|\geqslant 2^{(1-\tau)N}$ 的 $A\subseteq K$ 都满足

$$JN \text{ 维二元空间 } K^J \text{ 中 } \mathbb{F}_2\text{span } \Psi_J(A) \text{ 的余维为 } \leqslant \tau N+1 \qquad .$$

证明。我们将 K^J 视为一个维数为 JN 的二元向量空间。由 K 上绝对迹配对的非退化性可知，每个 $\mathbb{F}_2$-线性泛函 $\ell\colon K\to\mathbb{F}_2$ 都唯一地具有如下形式：对某个 $a\in K$，$\ell(x)=\mathrm{Tr}^{K/\mathbb{F}_2}(ax)$。事实上，对于 $a\neq 0$，泛函 $x\mapsto\mathrm{Tr}(ax)$ 非零，因为乘以 a 是 K 上的一个置换，而绝对迹是一个非零的二元线性泛函。由此得到的从 K 到其二元对偶空间的单射，因维数相等而为同构。逐坐标应用这一事实，K^J 的每个二元特征都唯一地具有如下形式

$$\chi_{\mathbf{a}}(v_1,\dots,v_J)=(-1)^{\mathrm{Tr}_{K/\mathbb{F}_2}(\sum_{j=1}^{J}a_jv_j)},\qquad \mathbf{a}=(a_1,\dots,a_J)\in K^J. \tag{2.5}$$

当且仅当 $\mathbf{a}\neq 0$ 时，它作为二元线性特征非零。它在 $\Psi_J(z)$ 上的求值为

$$\chi_{\mathbf{a}}(\Psi_J(z))=(-1)^{\mathrm{Tr}_{K/\mathbb{F}_2}P_{\mathbf{a}}(z)},\qquad P_{\mathbf{a}}(z)=\sum_{j=1}^{J}a_jz^{2j-1}.$$

选取满足 $a_{j_0}\neq 0$ 的最大索引 j_0。指数 1、3、.....、$2J-1$ 互不相同，因此 $P_{\mathbf{a}}$ 具有正奇数次 $e=2j_0-1\leqslant 2J-1$ 和非零首项系数。

如果 $e=1$，那么 $P_{\mathbf{a}}(z)=a_1z$，其中 $a_1\neq 0$。其绝对迹是一个非零二元线性泛函，它在 0 和 1 这两个值上分别取值于 K 中恰好一半的元素；因此其特征期望恰好为零。如果 $e\geqslant 3$，引理 2.14 给出

$$\left|\mathbb{E}_{z\sim U_K}(-1)^{\mathrm{Tr}_{K/\mathbb{F}_2}P_{\mathbf{a}}(z)}\right|\leqslant(e-1)2^{-N/2}\leqslant(2J-2)2^{-N/2}.$$

奇数首项次数至关重要：形如 $R(z)^2+R(z)+c$ 的非恒定多项式具有偶数次，而这类阿廷–施莱尔相位正是特征为二时加性特征和中的例外情形。结合线性情形和高次数情形，即可证明第一个断言，并且该断言对整个 JN 比特特征向量的每个非零二元特征都均匀成立。

接下来固定 $0<\tau<1/2$。当 $J=1$ 时，前一个特征偏差为零，以下论证对任意 N 都成立。对于 $J\geqslant 2$，由于 J 和 τ 已固定，取足够大的 N，使得

$$(1/2-\tau)N\geqslant 1+\log_2(2J-2). \tag{2.6}$$

这等价于

$$\beta_J=(2J-2)2^{-N/2}\leqslant 2^{-\tau N-1}.$$

令 $A\subseteq K$ 满足 $|A|\geqslant 2^{(1-\tau)N}$。它在 K 中的密度至少为

$$\frac{|A|}{|K|}\geqslant 2^{-\tau N}=:\alpha_0.$$

取 $\Omega=K$ 和 $\Phi=\Psi_J$，并固定一个二进制坐标识别 $K^J\cong\mathbb{F}_2^{JN}$。刚刚证明的特征界表明，$\Phi(U_K)$ 的每个非零二元特征的偏差至多为 $\beta_J\leqslant\alpha_0/2$。引理 2.15 因此适用，并给出

$$\mathrm{codim}_{\mathbb{F}_2}\,\mathrm{span}_{\mathbb{F}_2}\,\Psi_J(A)\leqslant\log_2(1/\alpha_0)+1=\tau N+1.$$

这里的所有维数都是二进制维数，即在 K^J中；我们不取 K-线性张成空间 $\Psi_J(A)$。严格不等式 $\tau<1/2$ 使得对每个固定的2.6，条件 (J) 都可以达到。 □

2.4 近似对偶性及其秩敏感形式

引理 2.17 (近似对偶性的集合形式)。 参考文献。 [10, 引理 7.9], 使用多项式 Freiman–Ruzsa 定理 [35]。存在一个绝对常数 C ，使得以下结论成立。对于有限的A， $B \subseteq \mathbb{F}_2^R$ 满足 $|\mathbb{E}_{A\times B}(-1)^{\langle a,\ b\rangle}| \geqslant \varepsilon$，以及任意固定的正整数 t，都存在一个常值子矩形，其两侧的相对密度至少为 2^{-q} ，其中

$$q \leqslant C\left[\frac{R}{t} + 2^t \log\frac{2}{\varepsilon} + (t+1)\log(R+2)\right]. \tag{2.7}$$

这是 Bhowmick–Dvir–Lovett [10, 引理 7.9 及其证明], 在倍增参数为 $K = 2^{R/t}$时的特征为二的推论，其中使用了 Gowers–Green–Manners–Tao [35]证明的多项式 Freiman–Ruzsa 定理。所引用估计的证明分别界定两个相对密度，且适用于集合大小不相等的情形。其多项式分解如上所示转化为对数项。有界的 R 以及整数取整均并入 C。我们需要其秩敏感的带索引版本，因此显式移除了重复项，而不是通过附加所有原始输入标签来计入重复项。

引理 2.18 （有偏带索引矩阵）。 令 M 为一个由非空集合 A， B索引的二元矩阵，这些集合的大小至多为 2^n，且秩为 $r \geqslant 1$。设

$$\left|\mathbb{E}_{A\times B}(-1)^{M(x,y)}\right| \geqslant \varepsilon.$$

对于每个固定的正整数 L，存在一个单色矩形 $A'\times B'$，其两侧的相对密度均至少为 2^{-q}，其中

$$q \leqslant C\frac{r}{L} + C_L\left(\log\frac{2}{\varepsilon} + \log(n+2) + \log(r+2)\right). \tag{2.8}$$

这里的 C 是绝对的，而 C_L 仅取决于 L。

证明。 将 $M(x,y) = \langle a_x, b_y\rangle$ 分解为 $\mathbb{F}_2^r$。将 a_X的分布（其中 $X \in A$为均匀分布）按其非零原子概率划分为二进分层。由于这些概率至少为 2^{-n}，至多需要 $n+1$个分层。对 b_Y也进行同样的处理。

带符号平均值是至多$(n+1)^2$ 个类别对贡献之和。某个贡献的绝对值至少为 $\varepsilon/(n+1)^2$。该类别对的两个类别概率以及绝对条件偏差都至少为 $\varepsilon/(n+1)^2$。

在一个含有 t 个不同标签的类别中，条件概率之比至多为二。它们至多为 $2/t$，因此该分布是定义在大小为$\{1, \lfloor t/2\rfloor\}$的集合上的均匀分布的凸组合。这些集合的大小至少为 $t/3$ （单点集无妨）。两侧的独立分解会产生一对平坦标签集，其绝对偏差至少为 $\varepsilon/(n+1)^2$。

将（2.7）应用于该对。如果其单色标签子集的相对密度为 ρ，则每个选定子集在其整个二进分层中的条件质量至少为 $\rho/6$ ： 其中每个标签的概率至少为 $1/(2t)$。因此，提升到原始索引后，相对质量至少为

$$\frac{\varepsilon\rho}{6(n+1)^2}$$

两边。将 $\varepsilon/(n+1)^2$ 代入（2.7），并加入最后这个提升因子的对数。对于固定的 L，改变常数后，结果为（2.8）。 □

引理 2.19（非对称秩到偏差升级）。设两个输入长度 n_L，n_R 均介于 n 的固定正倍数之间。设一族布尔矩阵的秩至多为 An。假设每个矩阵在所有边长至少为 $2^{a_0 n_L}$ 和$2^{b_0 n_R}$的允许子矩形上均为非恒定，且允许的支撑集在分别限制下封闭。对于固定的 $0 < a_0 < a_1 \leqslant 1$和 $0 < b_0 < b_1 \leqslant 1$，存在 $\eta > 0$，使得该类中的每个矩阵在熵至少为 $a_1 n_L$，$b_1 n_R$ 的独立源上的绝对偏差至多为 $2^{-\eta n}$。

证明。利用引理取独立的平坦分量 2.8。若偏差至少为 $2^{-\eta n}$，引理2.18 将给出一个常值子矩形，其对数密度损失至多为

$$CAn/t + C_t(\eta n + O(\log(n+2))).$$

首先选取固定的 t，使第一项小于两个可用熵间隙各自的四分之一。然后选取 η，使第二项中 n 的系数也小于两个熵间隙各自的四分之一。当 n 足够大时，对数项和向下取整造成的一比特损失都能控制在剩余余量内。所得矩形超过了两个禁用阈值，矛盾。由于常数具有一致性，且绝对偏差具有凸性，因此可以对独立的平坦分解取平均。 □

推论 2.20 (线性秩离散器升级)。 对于固定的 A，ρ，$\xi > 0$，其中 $\rho + \xi \leqslant 1$，存在$\eta > 0$，使得每个秩至多为 An 的布尔 ρn-双源分散器，在边熵 $(\rho + \xi)n$ 下都是误差为 $2^{-\eta n}$的普通提取器。对于满足这两个假设的所有函数，该指数是一致的。

证明。 应用引理 2.19，令输入长度和熵间隙分别相等，然后利用单比特统计距离是其特征偏差的一半这一事实。低秩分散器到提取器原理源自 Ben-Sasson–Ron-Zewi [9,引理 2。17]其指数小误差的推论最初使用了多项式 Freiman–Ruzsa 猜想，该猜想现已由 Gowers–Green–Manners–Tao [35]证明。另请参见 Alrabiah–Goodman–Mosheiff–Ribeiro [2, 定理 9] 以了解这一无条件表述。 □

引理 2.21（低秩转录的统一容忍度）。 固定常数 $0 < \rho < \rho + \xi \leqslant 1$、$A$，$\kappa \geqslant 0$以及 $\omega > 0$。令 $U(x, y) \in \mathbb{F}_2^u$、$V(x, y) \in \mathbb{F}_2^v$为两个 N 比特输入的函数。假设对于 $\alpha \neq 0$和 β，相位 $\alpha \cdot U + \beta \cdot V$在每个边熵至少为 ρN 的矩形上，其全局二元秩至多为 AN，且受限秩大于 $(\kappa + \omega)N$，并且该条件是一致成立的。令 $T(x, y) \in \mathbb{F}_2^z$的每个奇偶校验的全局秩至多为 κN。存在仅取决于这些固定常数的 $\eta > 0$，使得对于充分大的 N、熵至少为 $(\rho + \xi)N$ 的独立熵源，以及 $u + v + z \leqslant \eta N$，

$$(U, V, T) \approx_{2^{-\eta N/2}} (U_u, V, T).$$

理想分布保留了 (V, T) 的实际联合边缘分布。

证明。 对于 $\alpha \neq 0$，添加转录奇偶校验后，受限秩至少为 $\omega N > 1$，全局秩至多为 $(A + \kappa)N$。推论 2.20将指数调低以吸收偏差/距离因子 2 后，得到均匀偏差 $2^{-\eta N}$。只有满足 $\alpha \neq 0$的特征才会对与 (U_u, V, T)的比较产生贡献。引理 2.11 给出的距离至多为 $2^{(u+v+z)/2-\eta N}/2 \leqslant 2^{-\eta N/2}$。没有对联合计算的转录的任何取值进行条件化。 □

2.5 已发表的带种子原语与熵高效提取

我们按照种子不计入输出长度的约定陈述带种子结果。如果某个源将种子计入其中，我们会明确删除其固定的种子坐标。

定理 2.22 (两种强种子参数范围)。 参考文献。 [36,定理 5.12 和 5.14 以及注 5。15]; 另见 [12, 引理 4。10]。存在绝对常数，使得以下性质成立。对于一个熵为 n、长度为 比特的源，以及误差 k，令 ζlog(s =)1/ζ。

(i) 强种子提取器的输出至少为 $k - 2s - O(1)$ 比特，种子长度为 $\log(n+2) + O(\log(k+2)\log((k+2)/\zeta))$。(ii) 当 $k \geqslant B(\log(n+2) + s + 1)$ 对某个充分大的绝对常数 B 成立时，强种子提取器的输出为 $\lfloor k/2 \rfloor$ 比特，种子长度为 $O(\log(n+2) + s + 1)$。

两种构造均可在关于 $n + s$ 的多项式时间内完成。对于端点 $k = n$，在调用针对 $k < n$ 表述的结论时，可以附加一个固定的输入比特来处理。

(i) 是 GUV [36], 的定理5.14和注5.15，删除种子坐标后即可得到。(ii) 可由其固定比例定理和强性注释推出；所示的熵下界吸收了种子坐标和所有加性常数。 [12, 引理 4.10] 中也以一种方便的半输出形式记录了这一结果。

引理 2.23 （近乎全量输出的强种子提取）。存在绝对常数 $A, B > 0$，使得对于每个 n、整数 $k \leqslant n$ 和 $0 < \zeta \leqslant 1/4$，都存在一个强种子提取器，其误差至多为 ζ，输出长度至少为 $\max\{0, k - A(s+1)\}$ （向下取整至整数），且种子长度至多为

$$B\left\lceil \log^2(n+2) + (s+1)\log\left(2 + \frac{n}{s+1}\right)\right\rceil, \qquad s = \log(1/\zeta). \tag{2.9}$$

它在 $n + s$ 中是确定性多项式时间的。忽略种子填充和输出截断不会影响结论。

证明。若 $s \leqslant \log(n+2)$，则使用定理 2.22的第 (i) 部分。其种子长度为 $O(\log^2(n+2))$，熵损失为 $O(s+1)$。若所述输出长度非正，则使用空输出。

现在假设 $s > \log(n+2)$。令 $T = \lceil \log_2(n+2) \rceil + 2$，将每个提取阶段的误差设为$\zeta_0 = \zeta/(8T)$，并令

$$t = \lceil s + \log(8T) \rceil + 1.$$

从累积输出长度 $M = 0$ 开始。在某一阶段定义 $h = k - M - t$。如果 $h < H_*$，则停止，其中 $H_* = B_*(s + \log(n+2) + \log(8T) + 1)$，且 B_* 是一个充分大的绝对常数。否则，使用新的独立种子，应用(ii)部分，阈值为 h、误差为 ζ_0，输出 $\lfloor h/2 \rfloor$ 比特，并将该长度加到 M中。

所有阈值和长度都是以下内容的确定性函数： n、 k、 ζ。固定所有先前种子的任意取值。原始源在公开其先前输出之前仍具有 k 最小熵。由于该输出恰好有 M 比特，引理2.9 表明，在条件概率至多为 2^{-t}的情况下例外，给定其先前输出的源具有至少 h的熵。对先前种子取平均后，仍得到相同的例外界。因此，当前阶段的强保证使其输出在误差至多为 $\zeta_0 + 2^{-t}$的意义下，独立于其种子以及实际的先前输出/种子转录，且服从均匀分布。

反复应用这一比较可证明，完整输出与所有种子联合起来后，与均匀分布的距离不超过各阶段误差之和。形式化的归纳首先比较实际的前缀/当前输出分布与实际前缀和一个新鲜均匀当前输出的乘积分布，然后应用先前的归纳假设。未来的种子相互独立，因此在整个过程中都可以将它们附加进来。无需在先前的混合之后替换源分布。

每个阶段都会将 h 替换为 $\lceil h/2 \rceil$，因此阶段数至多为 T，实际上为 $O(\log(2+n/(s+1)))$。它们的总误差至多为 ζ。终止时，$k - M \leqslant t + H_* + O(1) = O(s+1)$，因为 $s > \log(n+2)$。在此情形下，每个种子的长度为 $O(s+1)$，因此总长度为 $O((s+1)\log(2+n/(s+1)))$。这就证明了统一界。至多进行 $O(\log n)$ 次多项式时间带种子调用。 □

注记 2.24（为何该种子长度足够）。 加性损失输出证明要求，当 $s = \log(1/\varepsilon) \leqslant b_\delta n$ 时，生成的种子所含比特数不超过一个足够小的线性数量。函数（2.9）满足这一性质，因为 $b\log(2+1/b) \to 0$ 当 $b \to 0$ 时成立。这个从 GUV 得到的显式推导避免了采用更强的熵高效种子参数陈述。其内部种子稍长一些也无妨，因为该种子会保留在最终输出中。

2.6 常数种子强凝聚器

定理 2.25（常数种子强凝聚器）。 参考文献。 [58, 定义 2.10 和定理 4.3]。对于固定的 $0 < \rho \leqslant 1$ 和 $0 < \beta,\ \varepsilon < 1$，存在一个常数种子长度 d 以及一个高效映射

$$\mathrm{Cond} : \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^{m_0}, \qquad m_0 = \Omega(n),$$

使得，对于每个最小熵至少为 ρn的 X，

$$2^{-d} \sum_{i \in \{0,1\}^d} \min_{Z_i:\ H_\infty(Z_i) \geqslant (1-\beta)m_0} \Delta(\mathrm{Cond}(X,i), Z_i) \leqslant \varepsilon.$$

我们采用 Zuckerman [58, 定义 2.10 和定理 4.3 中的强凝聚器约定，该约定归功于 Raz]。这些映射不必是线性的。对于下文所用的较小目标缺陷，当 n充分大时，其输出长度至多可取为 $2n$，这是由均匀输入的支撑集大小界得出的。

2.7 随行数增长的均匀定量结果

接下来三个陈述中的均匀性至关重要。它们的常数都是绝对常数；其算法不要求熵率或行数固定。

引理 2.26（定量近似对偶性）。 参考文献： [8, 引理1.11], 及 PFR 定理 [35]。存在一个绝对常数 C_{AD}，使得下述结论成立。若 $A,\ B \subseteq \mathbb{F}_2^a$ 非空，且

$$|\mathbb{E}_{x \sim U_A, y \sim U_B} (-1)^{x \cdot y}| \geqslant 2^{-v},$$

那么，对每个整数 $u \geqslant 1$，都存在非空的 $A' \subseteq A$ 和 $B' \subseteq B$，使得 $x \cdot y$ 为常数，且二者的相对大小均至少为

$$2^{-C_{\mathrm{AD}}\{a/u + 2^u(v+1) + (u+1)\log(a+2)\}}. \tag{2.10}$$

这是 Ben-Sasson–Lovett–Ron-Zewi 的引理1.11 [8], ，其 PFR 假设由 [35]给出。在该引理中取参数 $K = 2^{a/u}$，使其参数 $t = a/\log K$ 恰好为 u。其两个密度界的负对数至多为 $a/u + 2^u(v+1) + u\log(4a) + \log a$的绝对常数倍。这就得到（2.10），较小的维数可通过增大 C_{AD}来涵盖。我们使用完整的权衡关系，而不仅仅是特殊情形 $-v \leqslant \sqrt{a}$。特别是， u 可以随 n增长。

定理 2.27（均匀线性某处凝聚器）。 参考文献。 [50, 定理31和算法 4],使用 [35]。存在绝对常数 $A \geqslant 1$, $C_c \geqslant 1$, $c_c > 0$, 以及 $\beta_0 > 0$ ，使得在下文所用的参数范围内，对于 $0 < \delta \leqslant 1/2$ ，存在可均匀显式构造的二元线性映射

$$C_1, \dots, C_J : \mathbb{F}_2^n \to \mathbb{F}_2^m$$

with

$$J \leqslant C_c \delta^{-A}, \qquad C_c^{-1} n \delta^A \leqslant m \leqslant n, \tag{2.11}$$

其在每个最小熵至少为 δn 的源上的元组，都与某个混合分布 $2^{-c_c m}$-接近；该混合分布具有一个指定行，其最小熵至少为（$1/2 + \beta_0$）m。多项式指数和常数均独立于 δ。

这是 Li–Zhong 线性凝聚器、 [50] 的一般弱源版本，即 $O(\log(1/\delta))$ 个阶段；每个阶段都具有常数分支因子和常数长度缩减，从而得到（2.11）中的绝对多项式界。一般弱源分析使用已证明的 PFR 定理。增大 A, C_c 即可处理整数舍入以及有限多个初始长度。我们仅在 m 按 n 除以某个固定幂次的 $\log n$增长时使用这一表述，因此所有阶段和误差估计都处于大长度范围内。这比简单写成$O_\delta(1)$ 行和 $\Theta_\delta(n)$ 比特而不追踪对率的依赖性更强。

表格约定。 由 s 比特的 T 行构成的初等 *somewhere-uniform* 分布，至少有一个行边缘分布恰为均匀分布；其余行可以与该行任意相关。某处均匀分布是初等分布的正混合，其中优良索引可以在不同分量间变化。仿射某处随机 分布是在某个仿射子空间上均匀分布，且该子空间投影到某一行后覆盖整个 $\mathbb{F}_2^s$。这两种约定都不假设各行之间相互独立。

定理 2.28（行数增长时的仿射某处随机提取）。 参考文献。 [52,引理 3.6(iii)]；比较 [50, 定理 39]。存在一个绝对常数整数 $A_0 \geqslant 1$ ，使得

$$H(T) = (2T)^{A_0 \log(2T)}, \qquad a_s = s/H(T),$$

当 $a_s \geqslant (2T)^4$ 时，以下结论成立。存在一个定义一致显式的映射，作用于由 s 比特构成的 T 行，可从任意均匀仿射源中提取至少 $\lfloor a_s \rfloor$ 比特，误差至多为 2^{-a_s}；该源投影到至少一行后覆盖整个 $\mathbb{F}_2^s$。在此参数范围内，求值可在多项式时间内完成。

我们采用 Li–Zhong [52, 引理 3.6(iii)] 的定量形式中的 Li 的仿射某处随机提取器；另见 [50, 定理 39]。原始构造包含 $O(\log T)$ 个递归阶段，并有损失 $T^{O(\log T)}$，其中指数为绝对常数。行之间不需要具备内部独立性。对于$T = (\log n)^{O(1)}$，有

$$H(T) = 2^{O((\log\log n)^2)} = n^{o(1)}. \tag{2.12}$$

仅针对固定 T、且常数未明确给出的 T 相关定理是不够的。在我们的应用中，不需要对这个最终的仿射提取器给出次数界。

2.8 任意公共特征的加权近似对偶性

对于递减率双源应用，特征是原始输入比特（及其源局部篡改副本）。我们对任意独立特征映射陈述分析，因为正则化定理同样适用于非线性预合成。

引理 2.29（加权秩投影引理）。设 X, Y 是定义在 $\mathbb{F}_2^n$ 的子集上的独立平坦源。假设一个布尔相位关于两侧各至多 P 比特的独立特征变量是仿射双线性的，且其在当前两个特征仿射包上的双线性部分的秩为 r。若其偏差大于 2^{-H}，则对每个整数 $u \geqslant 1$，都存在非空原始子集 A', B'，使相位在这些子集上为常数，且两者的相对概率均至少为 $2^{-g(r,H,u)}$，其中

$$g(r, H, u) = C_w \left\{ \frac{r+2}{u} + 2^u (H+1) + (u+1)\log(n+P+2) \right\}. \tag{2.13}$$

常数 C_w 是绝对常数。特征映射无需是单射、关于原始输入呈线性，也无需服从均匀分布。

证明。在两个特征包的仿射坐标中，将相位写作

$$f(x, y) = L(x) \cdot R(y) + a(x) + b(y) + c,$$

其中秩因子有 r 个坐标。这里，所有函数都与各自的原始特征映射复合。独立投影

$$Z = (L(X), a(X), 1), \qquad W = (R(Y), 1, b(Y) + c)$$

属于 $\mathbb{F}_2^{r+2}$，并满足 $f(X,\ Y) = Z \cdot W$。它们的正原子概率是$1/|\operatorname{supp}\ X|$和 $1/|\operatorname{supp}\ Y|$的整数倍，因此至少为 2^{-n}。

令 $\varepsilon = 2^{-H}$，并选取一个符号，使带符号相关大于 ε。将每个投影支撑集划分为至多 $n+1$ 个二进概率桶。若 p_i，q_j 是桶质量，c_{ij} 是它们的条件带符号相关，则 $\sum p_i q_j c_{ij} > \varepsilon$。满足 $c_{ij} > \varepsilon/2$ 的桶对总权重大于 $\varepsilon/2$：否则总和至多为 $\varepsilon/2 + (1-\varepsilon/2)\varepsilon/2 < \varepsilon$。某个这样的桶对满足

$$p_i q_j > \frac{\varepsilon}{2(n+1)^2}. \tag{2.14}$$

每个桶的质量也至少为这个量。

在大小为 N_1的桶中，每个条件原子的取值都介于 $1/(2N_1)$ 和 $2/N_1$之间。令$m_1 = \max\{1, \lfloor N_1/2 \rfloor\}$。这些原子的概率至多为 $1/m_1$，因此引理 2.8 将条件分布分解为定义在 m_1 个点上的平坦分布，其中 $m_1 \geqslant N_1/3$。在右侧独立地进行同样的操作。某一对平坦分量的带符号相关高于 $\varepsilon/2$。将引理 2.26 应用于这对分量，维数为 $r+2$，指数为 $H+1$。如果单色集的相对平坦密度至少为 2^{-b}，那么在它们各自原始桶下的条件概率至少为 $2^{-b}/6$ 。取它们的整个原始原像，并利用（2.14），可得相对原始概率至少为

$$\frac{\varepsilon}{12(n+1)^2} 2^{-C_{\mathrm{AD}}\{(r+2)/u + 2^u(H+2) + (u+1)\log(r+4)\}}.$$

取负对数并使用 $r \leqslant P$，即可在增大一个绝对常数后证明（2.13）。拉回单色集会保留相位恒等式。这并不要求投影后的源在进行桶选择之前已经是平坦的。

□

2.9 Auniform 共享特征正则化定理

Let

$$\mathcal{F} : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^D.$$

假设固定映射 u_L， $u_R : \mathbb{F}_2^n \to \mathbb{F}_2^P$ 用双线性形式表示每个特征：

$$\lambda \cdot \mathcal{F}(x,y) = u_L(x)^{\mathsf{T}} M_\lambda u_R(y) \qquad (\lambda \in \mathbb{F}_2^D). \tag{2.15}$$

用零坐标填充特征，可以得到统一的长度 P。同一组映射必须适用于所有 λ。

定理 2.30 （一致正仿射正则化）。 存在一个绝对常数 C_0 ，使得以下结论成立。对于独立平坦输入、整数 $1 \leqslant r_* \leqslant P$、 $u \geqslant 1$以及实数 $H \geqslant 0$， $t \geqslant 1$，存在一个有限划分，将其划分为原始乘积矩形。在每个叶节点上，条件输出与其精确仿射包上的均匀分布相差不超过$2^{D/2-H-1}$ 。除总叶节点质量至多为 2^{1-t}的部分外，每个源的最小熵损失至多为

$$\mathcal{L} = C_0 \Big[\frac{P}{u} + \frac{P}{r_*}\big(2^u(H+1) + (u+1)\log(n+P+2)\big) + D(r_*+2) + t \Big] \tag{2.16}$$

比特。该比较结果是均匀仿射分布的正混合。此划分仅用于证明，无需计算。

2.9.1 进展规则

在当前矩形 $A \times B$ 上，令 U、V分别为 $u_L(A)$ $u_R(B)$ 的仿射包，并令 W 为 $\mathcal{F}(A, B)$ 的仿射包。如果在 W 上非常值的每个特征的偏差都至多为 2^{-H}，则终止。否则，选取一个违规特征，并令 r 为其双线性部分在 U、V 的方向上的秩。

当 $r > r_*$时，使用引理 2.29 得到一个单色子矩形 $A' \times B'$。若 U'， V'是其特征仿射包，则

$$\operatorname{codim}_U U' + \operatorname{codim}_V V' \geqslant r. \tag{2.17}$$

事实上，常值仿射双线性相位的四点差分为零。因此，其双线性部分在 U', V' 的两个方向空间上为零。施加余维数 c_1, c_2 的限制至多会使矩阵秩降低$c_1 + c_2$ ，从而证明了（2.17）。在一个原始输入上查询某个特征仿射包原像的成员资格，该特征仿射包的余维至少为 $r/2$。其肯定概率至少为 $2^{-g(r,H,u)}$。由于被查询的特征仿射包是适当的，其概率小于 1，且两种结果仍然都是原始乘积矩形。

当 $r \leqslant r_*$时，将特征在当前仿射坐标中写作 $L(x) \cdot R(y) + a(x) + b(y) + c$。揭示 $(L(x)\ a(x))$，至多需要 $r+1$ 比特，然后揭示右侧比特 $L(x) \cdot R(y) + b(y) + c$。这样便使特征成为常值，最多使用 $r_* + 2$ 个已声明的消息比特。由于该特征在 W 上并非常值，每个由此得到的非空输出包的维数至少小 1。

每个矩形的所有选择都可以固定下来，从而构成一棵确定性有限树。每次完成操作都会使每个分支上的原始矩形基数严格减小。在高秩情况下，两种归属结果都是真子集；在低秩情况下，相位值单元不可能成为非常数相位的整个支撑集。尽管没有断言其深度存在有效界，这棵树仍然是有限的。

2.9.2 一个共享势控制每个高秩特征

两个当前特征仿射包维数之和起初至多为 $2P$，且始终不会增加。每当高秩判断为是时，该和至少减少 $r/2$。因此，逐路径地，

$$\sum_{\text{high yes}} r_j \leqslant 4P, \qquad N_+ \leqslant 4P/r_*. \tag{2.18}$$

每次低秩操作都会使输出包的维数减少，因此这类操作至多有 D 次。它们声明的消息总长度至多为

$$B_{\mathrm{ex}} \leqslant D(r_* + 2). \tag{2.19}$$

由于高秩至少为二， $r + 2 \leqslant 2r$ 因此，方程（2.13）和（2.18）给出了肯定信息的确定性界：

$$I_+ \leqslant C_1 \left[\frac{P}{u} + \frac{P}{r_*} \big(2^u (H+1) + (u+1) \log(n + P + 2) \big) \right]. \tag{2.20}$$

仅凭不同的特征分解无法证明这一界；正是固定特征对（2.15）使得该势函数对整个路径保持一致。

2.9.3 两个惊奇超鞅

令 I_- 表示高秩否定回答的总信息。在肯定概率为 p的查询处， $2^{I_-/2 - N_+}$ 的条件乘数的均值为

$$(1-p)(1-p)^{-1/2} + p/2 = \sqrt{1-p} + p/2 \leqslant 1.$$

其他操作不会改变它。它是一个从一开始取值为一的非负超鞅，并且该树是有限的。马尔可夫不等式意味着

$$\Pr[I_- > 2N_+ + 2t] \leqslant 2^{-t}. \tag{2.21}$$

这使我们无需对被拒绝的查询进行计数，就能控制任意多个此类查询。

令 I_{ex} 表示所有低秩精确消息的信息，并令 B 表示它们累积的声明比特长度。对于一个长度为 a 比特、结果概率为正的消息 p_z， $2^{I_{\mathrm{ex}} - B}$ 的条件乘数的期望为

$$\sum_{z : p_z > 0} p_z \, (2^a p_z)^{-1} = \frac{|\{z : p_z > 0\}|}{2^a} \leqslant 1.$$

在高秩查询处，该过程保持不变。方程 (2.19) 和马尔可夫不等式给出

$$\Pr[I_{\mathrm{ex}} > D(r_* + 2) + t] \leqslant 2^{-t}. \tag{2.22}$$

这两个估计都不假设转录比特是独立的或无偏的。

2.9.4 熵与正仿射补全

对于初始平坦矩形 $A' \times B'$ 内的叶 $A \times B$，其概率为 $(|A'|/|A|)(|B'|/|B|)$。回答概率会逐项相消，因此其负对数恰好为 $I_+ + I_- + I_{\mathrm{ex}}$。每个源的熵损失均非负，且不超过该总和。在 (2.21) 和 (2.22) 的并集之外，结合 (2.20)、 $2N_+ \leqslant 8P/r_*$ 和消息界可得 (2.16)，其中 C_0 为绝对常数。

在每个终端叶上，其仿射包的所有非恒定特征的偏差都至多为 2^{-H}， W 。在 W 上的仿射坐标中，由帕塞瓦尔恒等式和柯西–施瓦茨不等式可得

$$\Delta(\mathcal{F}(X,Y) \mid \mathrm{leaf}, U_W) \leqslant \tfrac{1}{2} \sqrt{2^{\dim W} - 1} \, 2^{-H} \leqslant 2^{D/2 - H - 1}.$$

用原始叶概率对这些均匀分布取平均，得到一个正混合，并由此证明定理 2.30。对于双线性原始映射， $u_L(x) = x, u_R(y) = y$ 给出 $P = n$。这是逆对数双源构造所利用的特殊简化。

3 小误差下的常数率提取

本节中的全部三种构造都具有线性输出，且误差相对于原始输入长度呈指数级小。固定-C 定理涵盖任意固定数量的独立篡改。对于$\delta n \leqslant k \leqslant n$，输出长度 $c_\delta n$ 尤其至少为 $c_\delta k$。在全文中，对于任意域 F和 $1 \leqslant D \leqslant s$，定义

$$\sigma_D(z_1, \dots, z_s) := \sum_{\substack{I \subseteq \{1,\dots,s\} \\ |I| = D}} \prod_{i \in I} z_i, \qquad z_1, \dots, z_s \in F.$$

因此， σ_D 是次数为 D 的初等对称多项式；等价地，它是 $\prod_{i=1}^{s}(1 + z_i T)$ 中T^D的系数。

阅读构造框。 每个编号的构造框都以其所使用的先前已建立原语开头，随后将<style id='3'>公开设置 与对输入执行的步骤分开。公共常数、域表示、行顺序、投影和子程序参数均根据给定的长度、熵阈值、误差目标和篡改次数确定，而绝不根据源支持集或篡改映射确定。证明稍后可以证明某个足够小或足够大的公共常数有效；求值不会针对输入搜索该常数。除带种子子程序所显示的种子参数外，不提供任何外部随机性。优行、子源、条件代表元表和近似对偶划分都是分析对象，而不是隐藏的算法步骤。构造框使用单独的计数器，因此所有现有的定理编号和公式编号均予以保留。

求值公共域合并器。 每次调用 $\sigma_D(w_1, \dots, w_s)$ 都可以通过同一个固定递推式求值：初始化 $e_0 = 1$和 $e_1 = \cdots = e_D = 0$；对于 $i = 1, \dots, s$，按递减顺序更新$e_j \leftarrow e_j + w_i e_{j-1}$，即从 $j = \min\{i, D\} \dots,$ 1；返回 e_D。递减的更新顺序使用之前的多项式系数，因此在第 i 行之后， e_j是 $\prod_{b=1}^{i}(1 + w_b T)$ 中 T^j 的系数，其中 $j \leqslant D$。这是一条求值规则，而不是证明中对优行的选择或强制规则。

3.1 普通双源提取

3.1.1 In成分与证明：奇次幂受限秩与稠密良好行 s

Fix

$$\tau = 2^{-10}. \tag{3.1}$$

对于 $K = \mathbb{F}_{2^M}$，核 $H_J(u, v) = \sum_{j=1}^{J} u^{2j-1} v^{2j-1}$是域内积 $H_J(u, v) = \langle \Psi_J(u)\ \Psi_J(v) \rangle_K$。

引理 3.1（奇次幂受限秩）。对于每个固定的 $J \geqslant 1$、充分大的 M 和 $\lambda \in K^\times$，令 $h_\lambda(u, v) = \mathrm{Tr}_{K/\mathbb{F}_2}(\lambda H_J(u, v))$。其全局二元矩阵秩至多为 JM。对于任意的$A, B \subseteq K$，其中 $|A|, |B| \geqslant 2^{(1-\tau)M}$，

$$\operatorname{rank}_2 [h_\lambda]_{A \times B} \geqslant (J - 2\tau)M - 2. \tag{3.2}$$

亏损 τ 与 J独立。

证明。 迹配对 $\mathrm{Tr}^{K/\mathbb{F}_2}(\lambda\sum_j p_j q_j)$ 在 K^J 上是非退化的，其二元维数为 JM。根据引理 2.16， $\Psi_J(A)$ 和 $\Psi_J(B)$ 的张成空间的余维分别至多为 $\tau M+1$。引理 2.13至多损失它们的和，从而得到下界。展开 J 分离的域乘积，并使用迹对偶基，即可得到全局上界。

□

引理 3.2 （指定对新鲜性）。 固定 $J \geqslant 1$ 和 $0 \leqslant \kappa < J - 2\tau$。存在 $\eta > 0$，使得对于充分大的 M，任意转录 $T\colon\ K^2 \to \mathbb{F}_2^v$ （其奇偶性的秩至多为 κM ）、任意满射二进制线性映射 $\pi\colon K \to \mathbb{F}_2^u$，以及熵至少为 $(1-\tau/2)M$ 的独立源 U, V，均满足

$$(\pi H_J(U,V), T(U,V)) \approx_{2^{-\eta M/2}} (U_u, T(U,V)), \qquad u+v \leqslant \eta M. \tag{3.3}$$

指数关于 T、 π 和这些源是一致的。

证明。 每个非零投影特征都是具有非零系数的绝对迹特征。由引理 3.1，其全局秩至多为 JM，受限秩至少为 $(J-2\tau)M-2$。取$\omega = (J-2\tau-\kappa)/2 > 0$ 并提高长度阈值。应用引理 2.21 ，其中当前比较坐标为空， $\rho = 1-\tau$，且 $\xi = \tau/2$。由此得证。

□

3.1.2 选择包含许多优行的子源

对于分布 P on $\{0,1\}^{m_0}$ and $0 \leqslant h \leqslant m_0$，定义

$$e_h(P) = \sum_z (P(z) - 2^{-h})_+.$$

这等于它到最近的 h-源的距离。事实上，每个这样的源都必须移除这部分超额质量；反过来，将其截断到上限 2^{-h} ，再把移除的质量分配到未使用的容量中，即可达到这一距离。

引理 3.3 （实际子源上的许多优行）。 假设确定性行 $C_i(X) \in \{0,1\}^{m_0}$, $1 \leqslant i \leqslant r$ 满足 $\sum_i e_h(C_i(X)) \leqslant \varepsilon_c r$，其中 $\varepsilon_c < 1/8$。存在一个分布 X^*，它通过在一个概率至少为 $(1-8\varepsilon_c)2^{-r}$ 的事件上对 X 和辅助随机性进行条件化而得到；并且存在一个指标集 I，满足 $|I| \geqslant 7r/8$，使得

$$H_\infty(C_i(X^*)) \geqslant h - r - \log\frac{1}{1-8\varepsilon_c} \qquad (i \in I).$$

特别地，$\mathrm{supp}\, X^* \subseteq \mathrm{supp}\ X$。

证明。 令 $p_i(z) = \mathbb{P}[C_i(X) = z]$。在给定 $X = x$ 的条件下，抽取接受指示变量 A_i ，其中

$$\mathbb{P}[A_i = 1 \mid X = x] = \min\{1, 2^{-h}/p_i(C_i(x))\}.$$

只抽取具有正概率的值，因此该比值有定义。在给定 X 的条件下，这些示性函数可以独立抽取。于是 $\mathbb{P}[A_i = 0] = e_h(C_i(X))$。对于 $N_{\mathrm{rej}} = \sum_i (1-A_i)$，有$\mathbb{E}N_{\mathrm{rej}} \leqslant \varepsilon_c r$；因此由马尔可夫不等式可得

$$\mathbb{P}[N_{\mathrm{rej}} > r/8] \leqslant 8\varepsilon_c.$$

在补事件上，至少有 $7r/8$ 次接受。这是关于整数计数的陈述，无需对 r 作出整除性假设。在至多 2^r 个精确接受

模式中，某个至少有 $7r/8$ 次接受的模式具有概率 $p \geqslant (1-8\varepsilon_c)2^{-r}$。以此模式为条件，令 I 为其接受索引。由此得到的 X 的边缘分布为 X^*。

对于 $i \in I$，条件事件蕴含 $A_i = 1$，因此

$$\mathbb{P}[C_i(X^*) = z] \leqslant \frac{\mathbb{P}[C_i(X) = z, A_i = 1]}{p} = \frac{\min\{p_i(z), 2^{-h}\}}{p} \leqslant \frac{2^{-h}}{p}.$$

取对数即可证明熵下界。条件化不会增加 X 的支撑集中的点。

□

辅助随机性不会为提取器带来随机性。进行此条件化仅仅是为了证明，两个原始输入支撑集中都包含一个乘积子源，使得所需事件在该子源上发生。分别将引理应用于独立的 X 和 Y，可保持 X^* 和 Y^* 的独立性。

3.1.3 主定理与构造接口

定理 3.4（线性输出与指数误差）。固定 $0 < \delta \leqslant 1$。存在正常数 $c_\delta, a_\delta, C_\delta$，以及一个确定性多项式时间族

$$\mathrm{Ext}_n : \{0,1\}^n \times \{0,1\}^n \longrightarrow \mathbb{F}_2^{m(n)}, \qquad m(n) \geqslant c_\delta n,$$

使得满足 $H_\infty(X)\ H_\infty(Y) \geqslant \delta n$ 的独立源 X, Y 满足

$$\Delta(\mathrm{Ext}_n(X,Y), U_{m(n)}) \leqslant 2^{-a_\delta n}.$$

对于每个 $\beta \in \mathbb{F}_2^{m(n)} \setminus \{0\}$，

$$\mathrm{rank}_2\big((x,y) \mapsto \beta \cdot \mathrm{Ext}_n(x,y)\big) \leqslant C_\delta n.$$

此外，存在一个域 $F_n = \mathbb{F}_{2^{\ell(n)}}$，其具有 $\ell(n) = \Theta_\delta(n)$，一个整数 $R = O_\delta(1)$，显式映射G_L, $G_R\colon \{0,1\}^n \to F_n^R$，以及一个满射二进制线性映射 $\pi_n\colon F_n \to \mathbb{F}_2^{m(n)}$，使得

$$\mathrm{Ext}_n(x,y) = \pi_n(\langle G_L(x), G_R(y)\rangle_{F_n}). \tag{3.4}$$

因此，每个编码都有 $R\ell(n) = O_\delta(n)$ 比特。推论 3.9还证明了一个单独的源强线性输出伴随构造，其秩形式相同。

这是一个独立双输入构造：它不是通过对 $x+y$ 应用和集提取器得到的。在两个输入上枚举相同的 r 强浓缩器种子，其中 $r \geqslant 4$ 是 2 的幂。仅使用相应的左/右行，并删除固定的最后一个索引，留下$s = r-1 = 2D-1$ 列，其中 $D = r/2$。在第 a 列，将每个输入的第 a 行视为公共域 K 中的一个元素，并使用

$$H_{J_a}(u,v) = \sum_{j=1}^{J_a} u^{2j-1} v^{2j-1}, \qquad W_a = \mathrm{Tr}_{K/F}\, H_{J_a}(u,v), \qquad J_a = 2^{a-1}. \tag{3.5}$$

返回 $Z = \sigma_D(W_1$、……、$W_s)$ 的固定二元线性投影。公共输出域 F有$\Theta_\delta(n)$ 比特。依赖源的好行、子源和条件代表元表仅在分析中出现。

3.1.4 完整的确定性求值器：参数顺序与域

构造 3.1 普通常数率双源提取器

前文中的要素。定理 2.25中的强浓缩器；引理 3.1–3.2中的奇次幂核和新鲜度界；第 2.2节中的有限域表示和相对迹。多项式 σ_D 在本节开头定义。

公开参数和固定设置。 固定 $0 < \delta \leqslant 1$ 和 $\tau = 2^{-10}$。将凝聚器的率设为$\delta/2$，目标亏损设为 $\tau/4$，误差设为 $1/100$。按固定顺序在其种子中填充无效比特，使其种子空间有$r = 2^d \geqslant 4$ 个元素。令 $m_0 = m_0(n)$ 为其输出长度。按以下顺序选择其余常数：

$$s = r - 1 = 2D - 1, \qquad D = r/2. \tag{3.6}$$

对于 $a = 1, \ldots, s$，设

$$J_a = 2^{a-1}, \qquad \kappa_a = \sum_{b<a} J_b = J_a - 1. \tag{3.7}$$

令 $\eta_a > 0$ 为 th 引理 3.2 为 (J_a, κ_a)给出的指数，令 $\eta =$ 最小值$_a\, \eta_a$，并固定 t 的一个幂 wo

$$L \geqslant \max\{2, 4s/\eta\}. \tag{3.8}$$

对于每个充分大的输入长度 n，设

$$\ell = \lfloor m_0/L \rfloor, \qquad M = L\ell, \qquad K = \mathbb{F}_{2^M}, \quad F = \mathbb{F}_{2^\ell} \subseteq K. \tag{3.9}$$

固定二进制基和一个满射二进制线性映射 π_n： $F \to \mathbb{F}_2^{m(n)}$，其中 $m(n)$ = 最小$\{\ell,\ \lfloor \nu_\delta n \rfloor\}$ 且 $\nu_\delta > 0$ 是第 3.1.7 节中认证的固定充分小常数。该选择与源无关。

输入。 一个有序对 $(x, y) \in \{0,1\}^n \times \{0,1\}^n$。

求值步骤。

1. 凝聚每个输入。 对于每个种子索引 $i \in [r]$，在 x 和 y上求值凝聚器。保留每个结果的前 M 比特，并将它们与 $U_i(x)$ $V_i(y) \in K$对应起来。

2. 使用固定的匹配列。 保留 $a = 1, \ldots, s$对应的配对 $(U_a(x)$ $V_a(y))$。两侧都省略索引 r；不要根据熵来选择列。

3. 求值每个指定的行核。 计算

$$W_a(x, y) = \mathrm{Tr}_{K/F} \sum_{j=1}^{J_a} U_a(x)^{2j-1} V_a(y)^{2j-1} \quad (a \in [s]). \tag{3.10}$$

4. 在域中合并，然后投影。计算

$$Z_n(x, y) = \sigma_D(W_1(x, y), \ldots, W_s(x, y)), \tag{3.11}$$

并返回 $\pi_n(Z_n(x,\ y))$.

输出。 $m(n)$ $=\Omega_\delta(n)$ 比特，保证如定理 3.4所述。对于有限初始范围，使用第 $x_1 + y_1$ 节中的单比特补全 3.1.7。

未评估。接受模式、高熵子源、代表性表格和单色矩形。

3.1.5 实际子源上的完整前缀新鲜性

从熵至少为$(\delta/2)n$的独立熵源开始。分别在两侧应用引理3.3 ，其中 $h=(1-\tau/4)m_0$。在得到的乘积分布$(X^*,\ Y^*)$上，存在接受集合 $I_X,\ I_Y\subseteq[r]$ 其大小至少为 $7r/8$。由于 $\varepsilon_c=1/100$ 给出$\log(1/(1-8\varepsilon_c))<1$，每个 $i\in I_X$经过截断后都有，

$$H_\infty(U_i(X^*))\geqslant h-r-1-(m_0-M)\geqslant(1-\tau/2)M, \tag{3.12}$$

并且当 $i\in I_Y$ 时， $V_i(Y^*)$ 也满足类似的界，对充分大的 n 均成立。最后一个不等式用到了 $m_0-M<L$，其中 $L,\ r$ 是常数，而 $M=\Theta_\delta(n)$。当$a\in I_X\cap I_Y$时，保留列 a 是良好列。因此

$$\begin{aligned}|(I_X\cap I_Y)\cap[r-1]|&\geqslant|I_X|+|I_Y|-r-1\\&\geqslant 3r/4-1\geqslant r/2=D,\end{aligned} \tag{3.13}$$

其中最后一个不等式用到了 $r\geqslant 4$。因此，固定删除无需从源中选择索引，即可保留所需的多数。

引理 3.5 （每个良好块都相对于其完整前缀是新鲜的）。 在同一个选定的乘积分布上，每个良好列 a 都满足

$$(W_a,W_{<a})\approx_{2^{-\eta M/2}}(U_F,W_{<a}). \tag{3.14}$$

证明。 对原始左侧输入在 U_a条件下应用引理 2.10 ，并对右侧输入在 V_a条件下独立地应用。固定这两个代表表，恰好保留了这些良好边缘分布。现在，每个先前特征都是其自身行核的独立前复合，其秩至多为 J_bM；每个前缀奇偶校验的秩至多为 κ_aM。因此，在指定的行空间上，当前特征与前缀奇偶校验之和的受限秩至少为

$$(J_a-2\tau)M-2-(J_a-1)M=(1-2\tau)M-2>1.$$

相对迹是满射的。其输出与完整前缀的总长度为 $a\ell\leqslant s\ell\leqslant\eta M/4\ \leqslant\ \eta_aM$。引理 3.2 适用，对独立表取平均可证明（3.14）。前缀本身从未被固定。不同的良好列可以使用不同的表表示；所有结论都针对同一个原始选定分布。代表表论证和秩相减都不要求浓缩器索引存在非对角线索引对。特别是，在两个输入上使用相同的确定性种子索引并不会损害它们的独立性。

□

推论 3.6 （所选良好块的联合均匀性）。 对于任意一组 I 良好列，$\Delta(W_I,\ U_F^{\otimes|I|})\leqslant|I|2^{-\eta M/2}$。

证明。 将 I 按递增顺序排列，并将每个完整前缀估计投影到前面所选的坐标上。归纳和统计距离的收缩性给出了误差之和。 □

3.1.6 原始乘积支集上的多数强制

引理 3.7 （强制每个域值）。 设 D 是 2的幂，设 $s=2D-1$，且设 F 的特征为二。如果 $z_1,\ldots,z_s$ 中任意指定的 D 等于 α，那么

$$\sigma_D(z_1,\ldots,z_s)=\alpha^D \tag{3.15}$$

对于任意 其他坐标的任意取值。在布尔输入上，σ_D 是严格多数。

证明。 多项式 σ_D 是 $\prod_i(1+z_i t)$ 中 t^D 的系数。设 $H \subseteq [2D-1]$ 为指定的 D 个索引的集合，其对应条目等于 α，固定指定坐标可得

$$\prod_i (1+z_i t) = (1+\alpha t)^D \prod_{i \notin H} (1+z_i t) = (1+\alpha^D t^D) \prod_{i \notin H} (1+z_i t).$$

剩余乘积的次数至多为 $D-1$，从而证明了（3.15）。对于重量为$w < 2D$的布尔输入，其值模 $\binom{w}{D}$ 为二。 t^D 在 $(1+t)^w$ 中的系数是 w 在 D次幂处的二进制位，因此当且仅当 $w \geqslant D$时，它等于一。

□

命题 3.8（每个大型原始矩形上的满像）。 对于充分大的 n， Z_n（$A \times B$）$=F$，只要 A， $B \subseteq \{0、1\}^n$ 的大小至少为 $2^{(\delta/2)n}$。

证明。 从 U_A、U_B 开始，并在上述实际乘积子源中选择 D 个良好列。对于每个 $\alpha \in F$，它们的同时取值（α，……，α）的概率至少为 $2^{-D\ell} - D2^{-\eta M/2} > 0$，根据（3.8）。引理 3.7 强制 $Z_n = \alpha^D$ 在该处取值，而与其他列无关。由于 D 是 2 的幂， $\alpha \mapsto \alpha^D$ 置换 F。每个见证都是 $A \times B$的一个实际点。

□

3.1.7 原始输入秩与最终提取步骤

取 K 的迹对偶 F-基，可将每个分离的 K 乘积展开为至多 $L = [K$ 个 $F]$ 分离的F-乘积。因此，计入所有凝聚器预处理后，第 a 列的分离长度为

$$R_a \leqslant L J_a. \tag{3.16}$$

展开固定多项式 σ_D 后，在原始 n-比特输入定义域上逐点有：

$$Z_n(x,y) = \sum_{t=1}^{R} A_t(x) B_t(y), \qquad R \leqslant \sum_{|I|=D} \prod_{a \in I} L J_a \leqslant \prod_{a=1}^{s} (1 + L J_a) = O_\delta(1). \tag{3.17}$$

因此，每个绝对迹特征的二元秩至多为 $R\ell = O_\delta(n)$。根据命题 3.8，它在原始矩形上的非恒定性成立；这并非声称受限秩下界在任意凝聚器预处理后仍然成立。

推论 2.20，取原始熵间隙 $\delta n/2$，给出一个共同的 $\theta_\delta > 0$ ，使得每个非零特征在熵为 δn的源上的偏差至多为 $2^{-\theta_\delta n}$ ；这里需降低指数并提高长度阈值。选取一个固定的正有理数 $\nu_\delta \leqslant \theta_\delta$ ，并令

$$m(n) = \min\{\ell, \lfloor \nu_\delta n \rfloor\}, \tag{3.18}$$

$$\mathrm{Ext}_n(x,y) = \pi_n(Z_n(x,y)). \tag{3.19}$$

其中 π_n 是任意一个秩为此值的固定满射二进制线性映射。根据引理2.11，

$$\Delta(\mathrm{Ext}_n(X,Y), U_{m(n)}) \leqslant \tfrac{1}{2} 2^{m(n)/2 - \theta_\delta n} \leqslant 2^{-\theta_\delta n/2}.$$

相同的特征恒等式保持了原始输入秩的界。取 $G_L = (A_t)_t$， $G_R = (B_t)_t$即可证明（3.4）。该算法使用常数个固定幂运算和域运算，以及一个显式凝聚器；它不会寻找证明中的子源。

对于低于固定截断值的有限多个长度 N_δ，使用一个比特 $x_1 + y_1$。其误差（包括分别条件化的误差）至多为 $1/2$，且其秩至多为二。它具有如下域配对表示$(x_1,\ \mathbf{1}) \cdot (1,\ y_1)$ ，定义在 $\mathbb{F}_{2^n}$上，随后进行坐标投影。将所有编码填充至同一常数长度，并将输出和误差常数降至不超过 $1/N_\delta$。这就完成了在每个正长度下的定理3.4 的证明。

3.1.8 源强线性输出伴随构造

推论 3.9（恒定熵率下的分别源强）。 对于每个固定的 $\delta > 0$，存在c'_δ、 $a'_\delta > 0$ 以及一个显式确定性多项式时间族 $\mathrm{Ext}^{\mathrm{ss}}_n$：$\{0、1\}^n \times \{0、1\}^n \to \{0、1\}^{r_n}$，$r_n \geqslant c'_\delta n$，使得对于最小熵至少为 δn 的独立源，

$$\Delta((\mathrm{Ext}^{\mathrm{ss}}_n(X,Y),X),(U_{r_n},X)), \quad \Delta((\mathrm{Ext}^{\mathrm{ss}}_n(X,Y),Y),(U_{r_n},Y)) \leqslant 2^{-a'_\delta n}.$$

其输出是上述域值内积计算器的一个固定线性投影，并且每个非零输出奇偶校验的二进制真值表秩仍为 $O_\delta(n)$。

构造 3.2 源强普通线性输出伴随构造

前文部分的构成要素。定理 3.4（构造3.1）中的普通提取器，以及熵余量强强化引理2.5 （零比较）。

公开参数和固定设置。固定 $\delta > 0$，并令 $h := \delta n/4$。将普通族实例化为率 $\delta/4$，等价地，分量熵阈值为 h；记其输出和误差常数为c_0，$a_0 > 0$。固定 $0 < b_\delta < \min\{c_0/2,\ a_0/4,\ \delta/16\}$，并令 $r_n = \lfloor b_\delta n \rfloor$ 取充分大的值。

输入。 $(x,\ y) \in \{0,\ 1\}^n \times \{0,\ 1\}^n$。

求值步骤。

1. 在 $(x、\ y)$ 上求值率为 $\delta/4$ 的普通提取器。**2.** 返回其前 r_n 个输出比特。在有限初始范围内，返回基础族的单比特补全$x_1 + y_1$。

输出。 推论 3.9的伴随函数 $^{\mathrm{ss}}_n(x、\ y)$；该 函数 分别对任一源都是强的。不会执行额外的种子提取器调用或条件化步骤。

证明。 对所有充分大的 n，令熵为 $h = \delta n/4$ 时的基础构造具有至少$c_0 n$ 个输出比特，且误差至多为 $2^{-a_0 n}$。取 $0 < b_\delta < \min\{c_0/2、\ a_0/4、\ \delta/16\}$。对于输入熵$k \geqslant \delta n$，松弛量 $k - h$ 至少为 $3\delta n/4$。引理 2.5 结合 $t = 0$ 将两个源强误差均界定为

$$2^{r_n}\left(2^{-a_0 n} + 2^{-3\delta n/4}\right) \leqslant 2^{-\Omega_\delta(n)}.$$

因此，对于某个固定整数 $N_\delta \geqslant 1$ 和 $a_* > 0$，误差至多为 $2^{-a_* n}$，并且对于$n \geqslant N_\delta$误差至多为 $r_n \geqslant b_\delta n/2$。在 N_δ 以下，单比特补全的两个强误差均至多为 $1/2$。取

$$c'_\delta = \min\{b_\delta/2, 1/N_\delta\}, \qquad a'_\delta = \min\{a_*, 1/N_\delta\}$$

证明了对每个 $n \geqslant 1$ 都成立的所示界。固定输出投影保留了秩和分离域表示，从而证明了最后的断言。□

3.2 和集提取

3.2.1 In梯度及其证明：线性强凝聚器与矩秩

我们回顾两个针对和集的已引用定理以及特征和引理；秩敏感的近似对偶性升级已在预备知识中证明。基于这些输入的归约如下所示。

定理 3.10（线性某处凝聚器）。 参考文献。 [50, 定理 31 和引理 27], , 其中多项式 Freiman–Ruzsa 定理由 [35]提供。对于每个固定的 $\rho > 0$，存在固定的 $t \geqslant 1$， $\beta \in (0, 1/4]$， c， $c' > 0$，以及显式的二元线性映射

$$P_1, \dots, P_t : \mathbb{F}_2^n \to \mathbb{F}_2^m, \qquad m = \Theta_\rho(n),$$

使得，对于每个 $H_\infty(X) \geqslant \rho n$，元组 $(P_i X)_{i \leqslant t}$ 与某个正混合的分布相距不超过 $2^{-c'n}$，其中这些分布都至少有一个坐标的最小熵为 $(1/2 + \beta)m$。

这是 Li–Zhong 定理 31 中一般弱源构造的固定率推论 [50], , 由第 4 节中的基本增率引理 27 和常深度迭代得到。多项式 Freiman–Ruzsa 输入由 [35] 提供。下面处理通过固定数量的比特进行填充或截断的情况。这些坐标是原始输入的联合二元线性函数，而不是非线性源凝聚器。

定理 3.11（常种子曲线合并器）。 参考文献。 [33] 和 [32, 定理 18]。令 $\mathbb{E} = \mathbb{F}_q$，选取互异的 γ_1，……，$\gamma_t \in \mathbb{E}$，并令 $c_i(U)$ 为它们的次数至多为 $t - 1$ 的拉格朗日基多项式。对于任意数量 v 的 $\mathbb{E}$ 坐标，定义

$$\mathrm{CM}((z_i)_{i \leqslant t}, u) = \sum_{i=1}^{t} c_i(u) z_i. \tag{3.20}$$

如果 $(Z_i)_{i \leqslant t}$ 是在 $\mathbb{E}^v$ 中含有一个均匀坐标的混合分布，且 U 在 $\mathbb{E}$上服从独立均匀分布，那么对于 $0 < \vartheta$， $\varepsilon < 1$ 以及

$$q \geqslant \max\{t, (2t/\varepsilon)^{1/\vartheta}\},$$

$\mathrm{CM}(Z, U)$ 与某个最小熵至少为 $(1 - \vartheta)v \log q$ 的分布之间的统计距离为 ε。

这是 Dvir–Wigderson 曲线构造 [33] 结合 Dvir–Kopparty–Saraf–Sudan 的定理18 [32]得到的结果。关键在于，关于 q 的下界独立于 v。每个固定种子映射对于整个表是 $\mathbb{E}$-线性的，并且与其坐标向量上的每个 $\mathbb{E}$-线性映射交换。我们证明的是高熵情形的结论，而不是均匀行情形的结论，因此无需对后者作假设。

奇次数加性特征估计是引理 2.14。

我们使用推论 2.20中的一致线性秩离散器升级。对于单比特特征，如有必要则降低其误差指数，以吸收偏差与统计距离之间的两倍因子。

相容的有限域及其运算可在固定特征 [57] 下用确定性多项式时间构造。最终求值器使用的所有扩域比率均为固定常数。

对于域 F，函数 $B(x, y)$ 的分离长度至多为 r，条件是 $B(x, y) = \sum_{j=1}^{r} a_j(x) b_j(y)$。分别进行前复合不会增加此长度，求和会使其相加，求乘积则会使其相乘。若 $F = \mathbb{F}_{2^\ell}$，则此类函数的每个绝对迹特征的二元秩至多为 $r\ell$，这由 $F/\mathbb{F}_2$ 的迹对偶基可得。对来自次数为 a 的扩域 K/F 的分离乘积取迹，至多需要 a 个分离 F 乘积。

3.2.2 从线性某处凝聚器到多个良好线性视图 s

对于有限字母表 Ω 和 $h \leqslant \log |\Omega|$，令 有限域 K_h 表示所有原子概率均至多为 2^{-h} 的分布，并定义

$$e_h(P) = \sum_{z \in \Omega} (P(z) - 2^{-h})_+.$$

引理 3.12（到熵上限的距离）。 $e_h(P) = \min_{Q \in \mathcal{K}_h} \Delta(P, Q)$。任何受 2^{-h} 限制的子分布都可以扩展为具有该上限的概率分布。

证明。 每个原子上限为 2^{-h} 的比较分布都必须移除所显示的超额质量。将 P 限制在该值即可实现这一移除，并且其缺失质量能够容纳在剩余容量中，因为 $|\Omega|2^{-h} \geqslant 1$。将该质量重新分配到那里。

□

引理 3.13 （通过函数提升封顶比较）。 令 Z 取值于 Ω，并令$A : \Omega \to \Omega'$ 为满射。若 AZ 与 有限域 K_h中的某个分布的 ε相近，则 Z 也与 有限域 K_h中的某个分布的 ε相近。

证明。 通过在 A 的每个非空实际纤维上对 Z 的条件分布（在空纤维上任意指定）来提升 AZ 的封顶比较分布 Q 。每个提升后的原子至多为其像的质量$Q(w) \leqslant 2^{-h}$，并且在各纤维上求和全变差会保持该距离。

□

引理 3.14 （仅用于证明的线性投影）。 对于最小熵至少为 h 的 $Z \in \mathbb{F}_q^v$ 和一个整数$r \leqslant v$，存在一个 $\mathbb{F}_q$-线性映射 $A\colon \mathbb{F}_q^v \to \mathbb{F}_q^r$，使得

$$\Delta(AZ, U_{\mathbb{F}_q^r}) \leqslant \tfrac{1}{2}\sqrt{q^r 2^{-h}}.$$

如果该界小于 $1/2$，则所选映射为满射。

证明。 对于均匀随机的 $r \times v$ 线性矩阵 A，

$$\mathbb{E}_A \operatorname{cp}(AZ) = q^{-r} + (1 - q^{-r}) \operatorname{cp}(Z).$$

柯西–施瓦茨不等式和碰撞概率到距离的界表明，到 $U_{\mathbb{F}_q^r}$ 的平均距离至多为 $\tfrac{1}{2}\sqrt{q^r \operatorname{cp}(Z)} \leqslant \tfrac{1}{2}\sqrt{q^r 2^{-h}}$。选择一个达到此平均界的矩阵。非满射矩阵的输出支撑集至多包含 q^{r-1} 个值，与均匀分布的距离为 $\geqslant 1 - 1/q \geqslant 1/2$。

□

引理 3.15 （曲线合并器的高熵输入）。 假设 $(Z_1, \dots, Z_t) \in (\mathbb{F}_q^v)^t$ 是一个正混合，其中某一行的最小熵至少为 h。固定 $0 < \tau < 1/2$ 并选取

$$r = \left\lfloor \frac{h - 2\log(1/\tau)}{\log q} \right\rfloor \geqslant 1.$$

在满足 定理 3.11 中的域大小条件下，$\mathrm{CM}(Z, U_q)$ 与 有限域 $K_{(1-\vartheta)r\log q}$ 相差不超过 $(\tau + \varepsilon)$。

证明。 考虑一个含有高熵行的混合分量 Z_g。引理 3.14 给出一个满射A，使得 AZ_g 与均匀分布相差不超过 $\tau/2$。将该边缘分布替换为均匀分布，同时保留其他投影行的条件分布。投影后的表至多改变 τ，并变为某处随机；曲线合并器使其输出与所述熵上限相差不超过 $\tau + \varepsilon$。由于

$$A \operatorname{CM}(Z, u) = \operatorname{CM}((AZ_i)_i, u),$$

引理 3.13 将这个带熵上限的比较转移回未投影的输出。投影可能因分量而异：它仅用于证明。对各分量取平均是可行的，因为熵上限类是凸的，且目标熵相同。

□

引理 3.16 （公开常数种子）。 设 S 在大小为 Q 的集合上均匀分布，与 Z 具有任意联合分布，并假设 $e_h(Z) \leqslant \varepsilon$。那么

$$\frac{1}{Q}\sum_{s} e_{h-\log Q}(Z \mid S = s) \leqslant \varepsilon.$$

证明。令 $p_{s,z}=\Pr[S=s,Z=z]$ 和 $c=2^{-h}$。由 S 的均匀性，左侧为 $\sum_{s,z}(p_{s,z}-c)_+$。对于非负的 a_s，$\sum_s(a_s-c)_+\leqslant(\sum_s a_s-c)_+$，因此

$$\sum_{s,z}(p_{s,z}-c)_+\leqslant\sum_z\left(\sum_s p_{s,z}-c\right)_+=e_h(Z).$$

使用引理 3.12 将超额质量与距离对应起来。 □

定理 3.17（半数以上常数种子线性强浓缩器）。固定 $\rho>0$ 和一个常数$\varepsilon_*>0$。存在常数 $\beta>0$、一个 2 的幂 $Q\geqslant 16$，以及显式的二元线性映射

$$L_u:\mathbb{F}_2^n\to\mathbb{F}_2^M,\quad u\in\mathbb{F}_Q,\qquad M=\Theta_\rho(n),$$

使得每个 $H_\infty(X)\geqslant\rho n$ 都满足

$$\frac{1}{Q}\sum_{u\in\mathbb{F}_Q}\min_{Z_u:H_\infty(Z_u)\geqslant(1/2+\beta/2)M}\Delta(L_uX,Z_u)\leqslant\varepsilon_*.\tag{3.21}$$

所有常数也可以依赖于 ε_*。对于每个固定的种子，这些映射都是线性的，而不仅仅是低次多项式。

构造 3.3 常数种子线性强浓缩器

前文的构件。定理 3.10 中的线性某处凝聚器，以及定理 3.11中的曲线合并器。引理 3.15 和 3.16 提供了高熵分析和保留种子分析，并不增加额外的求值步骤。

公开参数和固定设置。给定固定速率 $\rho>0$ 和常数误差 $\varepsilon_*>0$，应用定理 3.10 得到 $P_1,\dots,P_t$、行长度 m 和 $0<\beta\leqslant 1/4$。令 $\vartheta=\beta/4$ 和 $\varepsilon=$为 $\min\{\varepsilon_*/4,1/8\}$。固定一个 2 的幂 $Q=2^s\geqslant$，使其 $\max\{16,t,(2t/\varepsilon)^{1/\vartheta}\}$，以及互不相同的 $\gamma_1,\dots,\gamma_t\in\mathbb{F}_Q$和相应的拉格朗日基多项式 $c_i(U)$。令 $M=s\lceil m/s\rceil$。

输入。 $z\in\mathbb{F}_2^n$ 和一个数值种子 $u\in\mathbb{F}_Q$。

求值步骤。

1. 对所有 $i\in[t]$ 求值 $P_i(z)$，并在末尾补零，使每行恰好包含 M 比特。将每行解析为由 $\mathbb{F}_Q$中的 M/s 个元素组成的向量。**2.** 计算逐坐标的域上的和 $L_u(z)=\sum_{i=1}^t c_i(u)P_i(z)$，并返回其 M 个二进制坐标。

输出。一个 M比特线性视图。要获得完整的确定性表，请按固定域顺序，对每个 $u\in\mathbb{F}_Q$ 运行相同的过程；在这种用法中不采样种子。保证由定理 3.17给出。

未评估。引理 3.14中的依赖源的线性投影；它不属于 L_u。

证明。使用定理 3.10，及其 t，m，β并在必要时减小 β以确保 $\beta\leqslant 1/4$。选取$\vartheta=\beta/4$ 和 $\tau=\varepsilon=\min\{\varepsilon_*/4,\ 1/8\}$。选取一个满足以下条件的固定 2 的幂 $Q=2^s\geqslant 16$：

$$Q\geqslant\max\{t,(2t/\varepsilon)^{1/\vartheta}\}.$$

将每个 P_i 填充少于 s 个零比特，使其长度达到 $M=s\lceil m/s\rceil$，并将其视为 $\mathbb{F}_Q^{M/s}$ 中的一个向量。实际映射定义如下：

$$L_u(z)=\sum_{i=1}^t c_i(u)P_i(z),\qquad u\in\mathbb{F}_Q,\tag{3.22}$$

其中填充方式如上所述。乘以 $\mathbb{F}_Q$ 的一个固定元素是二元线性的，因此所有这些映射都是二元线性的。

输入比较表的行熵为 $h = (1/2+\beta)m$。引理 3.15 使得使用独立均匀种子的输出接近熵

$$H = (1-\vartheta)s\left\lfloor \frac{h - 2\log(1/\tau)}{s} \right\rfloor.$$

总误差至多为 $2^{-c'n} + \tau + \varepsilon \leqslant \varepsilon_*$（当 n 足够大时）。应用引理 3.16：到熵的平均固定种子距离 $H - s$ 具有相同的界。最后，

$$H - s \geqslant (1-\beta/4)(1/2+\beta)m - O_{\rho,\varepsilon_*}(1) \geqslant (1/2+\beta/2)M$$

当 n 足够大时。最后一个不等式具有正线性余量，因为 $(1-\beta/4)\ (1/2+\beta)\ -(1/2+\beta/2)$ $= 3\beta/8 - \beta^2/4 > 0$。这证明了（3.21）。 □

备注 3.18（需要哪种误差）。种子空间 $\mathbb{F}_Q$ 对于常数 ρ 是固定的，ε_*。该定理并未声称在种子长度为常数时具有指数小的强凝缩器误差。弥散证明中只会使用（3.21）中的小常数误差。

3.2.3 具有共同多数的一个实际乘积子源

引理 3.19（截顶模式选择）。设 Q 边缘视图 L_iX 满足 $Q^{-1}\sum_i e_h(L_iX) \leqslant \varepsilon_*$，其中 $\varepsilon_* \leqslant 1/32$。存在一个实际的源局部条件分布 X^*，其支撑集包含于 $\operatorname{supp} X$ 中，并且存在一个集合 G_X，其中至少包含 $15Q/16$个索引，使得

$$H_\infty(L_iX^*) \geqslant h - Q - 1 \qquad (i \in G_X).$$

所选的共 同条件事件（允许使用私有辅助随机币）的概率至少为 2^{-Q-} [1].

证明。给定一个行值 v，以概率 $\min\{1, 2^{-h}/p_i(v)\}$独立地接受索引 i，其中p_i 是 L_iX的分布。拒绝的期望数量至多为 $Q\varepsilon_*$。由马尔可夫不等式可知，接受至少 $15Q/16$ 个索引的概率至少为 $1 - 16\varepsilon_* \geqslant 1/2$。至多有 2^Q 种接受模式，其中一种模式的概率为 $p \geqslant 2^{-Q-1}$。对其进行条件化。在任何被接受的索引处，

$$\Pr[L_iX = v \mid \text{pattern}] \leqslant \frac{\Pr[L_iX = v,\ i \text{ accepted}]}{p} \leqslant 2^{-h}/p \leqslant 2^{-h+Q+1}.$$

该模式是源局部的，条件化不会增加支撑点。 □

将此引理分别应用于独立的 X，Y。私有随机币和选择事件各自位于对应的一侧，因此 X^*，Y^* 仍然独立。至少有 $7Q/8$ 个索引属于 $G_X \cap G_Y$。从算法的列表中删除一个固定索引，并记

$$T = Q - 1 = 2D - 1, \qquad D = Q/2. \tag{3.23}$$

由于 $Q \geqslant 16$，至少有 $7Q/8 - 1 \geqslant D$ 个保留列对两个源都很好。在给定列上，每个加数都使用相同的映射；不存在左侧和右侧行索引之间的未知配对。

稍后我们将映射填充至长度 M to $N = a\lceil M/a \rceil$，其中 a 是一个固定常数。对于$h = (1/2+\beta/2)$ M，当 n充分大时，好的观察源的最小熵至少为

$$\alpha_1 N, \qquad \alpha_1 = 1/2 + \beta/4. \tag{3.24}$$

事实上，它们的熵至少为 $h-Q-1$，而填充增加的比特数少于 a。率固定后，所有这些损失都是常数。我们只需要每个原始乘积源的支撑集中有一个权重为正的乘积分布；不会用该分布对原始乘积源进行统计替换。

3.2.4 矩—迹码受限秩界

记 $K=\mathbb{F}_{2^N}$，令 $h\geqslant 3$ 和 $d=2^h-1$ 为固定值，并假设 $N>2h$ 足够大。本小节中的所有迹均为绝对迹。

命题 3.20（矩—迹码受限秩）。设 $p\geqslant 1$、$2p\leqslant 2^{h-2}$、$\lambda\in K^\times$、$A,\ B\subseteq K$、$|A|\geqslant p$ 和 $|B|\geqslant 2(d-1)2^{N/2}$。则

$$\operatorname{rank}_2[\operatorname{Tr}(\lambda(x+y)^d)]_{A\times B}\geqslant p\log|A|-p\log p+\log|B|-N-1. \tag{3.25}$$

完整二元矩阵的秩至多为 $(d+1)N=2^hN$。

证明。迹码限制损失。令 $\mathcal{C}_d$ 为函数构成的二元空间

$$f(z)=\varepsilon+\operatorname{Tr}\left(\sum_{\substack{1\leqslant j\leqslant d\\ j\text{ odd}}}a_jz^j\right),\qquad \varepsilon\in\mathbb{F}_2,\ a_j\in K.$$

次数至多为 d 的每个多项式迹都属于此空间：将每个正指数写成 2^vj，其中 j 为奇数，并利用 $\operatorname{Tr}(az^{2^vj})=\operatorname{Tr}(a^{2^{N-v}}z^j)$。对于非零系数元组，最大的非消失奇数次项的次数为正奇数，且至多为 d。奇数次特征和界（引理 2.14）表明，当 N足够大时，其绝对偏差严格小于 1。因此，系数（包括 ε）是唯一的，并且 $\mathcal{C}_d$ 的每个非常数成员的偏差至多为 $(d-1)2^{-N/2}$。

令 r 为限制 $\mathcal{C}_d\to\mathbb{F}_2^B$的核的维数。该核不包含常值函数一，且其中所有元素都在 B上消失。由正交性可得

$$\frac{|B|}{2^N}\leqslant 2^{-r}\sum_{f\text{ in the kernel}}\mathbb{E}_z(-1)^{f(z)}\leqslant 2^{-r}+(d-1)2^{-N/2}.$$

因此，关于 $|B|$ 的下界意味着

$$r\leqslant N-\log|B|+1. \tag{3.26}$$

非受限行张成空间中的矩。根据 Lucas定理，$(x+y)^d$中的所有二项式系数模二均为一。对于每个奇数 $j>d/2$，$\operatorname{Tr}(\lambda(x+y)^d)$的奇指数迹表示中 y^j 的系数恰好为 λx^{d-j}：没有更高的偶指数能够化简为 j，因为$2j>d$。取 $j=d-2v$ 为 $0\leqslant v<2^{h-2}$，然后除以 λ 并应用逆弗罗贝尼乌斯映射，即可恢复矩向量

$$(1,x,x^2,\dots,x^{2^{h-2}-1})$$

通过在不受限制的行函数空间上进行二元线性运算。对于 A 的两个不同的 p 元子集，如果这些向量之和相等，就会在至多 $2p$ 个不同的范德蒙德列之间产生非平凡的二元关系。前 $2p$ 个坐标排除了这种关系。因此，全定义域行张成空间的维数至少为

$$\log\binom{|A|}{p}\geqslant p\log|A|-p\log p.$$

限制与全局秩。 所有行函数都属于$\mathcal{C}_d$。将右参数限制为 B至多会损失（3.26）中的核维数，从而得到（3.25）。全局上界既可以通过将 $(x+y)^d$ 展开为 $d+1$ 分离的 K-乘积来得到，也可以通过单乘积实现（3.29）和迹对偶基来得到。

□

推论 3.21（固定速率下的任意掩码裕量）。固定 $1/2 < \alpha_0 < \alpha_1 < 1$ 和 $\kappa \geqslant 0$。选取

$$p = \left\lceil \frac{\kappa + 4}{\alpha_0} \right\rceil, \qquad h \geqslant 3, \quad 2^{h-2} \geqslant 2p.$$

存在 $\theta > 0$，它在 N、 $\lambda \in K^\times$ 以及每个布尔核 T 上均匀分布： $K^2 \to \mathbb{F}_2$ 的全局秩至多为 κN；对于独立的 U， V，其熵至少为 $\alpha_1 N$，

$$\left| \mathbb{E}(-1)^{\mathrm{Tr}(\lambda(U+V)^{2^h-1}) + T(U,V)} \right| \leqslant 2^{-\theta N} \tag{3.27}$$

当 N 充分大时。

证明。对于边熵至少为 $\alpha_0 N$ 的矩形， B 上的规模条件在 N 充分大时成立，$\alpha_0 > 1/2$ 因为 $\alpha_0 > 1/2$。命题 3.20 给出的秩至少为

$$(p\alpha_0 + \alpha_0 - 1)N - p\log p - 1 > (\kappa + 2)N.$$

加上 T 后，秩仍大于一，因此该相位在该阈值下是一个分散器。其全局秩至多为 $(2^h + \kappa)N$。推论 2.20 给出的熵增量为 $\alpha_1 - \alpha_0$，降低其指数后即可得到偏差估计。所引用引理的均匀性保证了对所有掩码和非零系数的均匀性。允许的熵率在所要求的裕量 κ 之前p、 h、 θ 就已固定；只有它们依赖于该裕量。

□

3.2.5 定理

所有对数均以 2 为底。熵是二进制最小熵。统计距离为 $\Delta(P,\ Q) = \frac{1}{2}\sum_z |P(z) - Q(z)|$。原始输入长度为 n；和集提取器仅能看到 $z = x + y \in \mathbb{F}_2^n$。

定理 3.22（常速率和集提取）。 对于每个固定的 $0 < \delta \leqslant 1$，存在正常数 c_δ， e_δ， C_δ，使得对于所有充分大的 n，一个可由一致多项式时间计算的映射

$$\mathrm{Ext}_n : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^{m_n}, \qquad m_n \geqslant c_\delta n,$$

满足

$$\Delta(\mathrm{Ext}_n(X+Y), U_{m_n}) \leqslant 2^{-e_\delta n}$$

对于每个满足 $H_\infty(X)\ H_\infty(Y) \geqslant \delta n$ 的独立对。对于每个非零输出泛函 λ，矩阵

$$[\lambda \cdot \mathrm{Ext}_n(x+y)]_{x,y \in \mathbb{F}_2^n}$$

在 $\mathbb{F}_2$ 上的秩至多为 $C_\delta n$。更确切地说，输出是一个域值表达式的二元线性投影，该表达式在一个二元维数为$\Theta_\delta(n)$的域上具有常数数量的分离乘积。因此，对于其提升特征，可以获得总二进制长度为 $O_\delta(n)$的显式分离编码。

这里的“一致多项式时间”是指对每个固定速率而言：算法中的常数固定，不依赖于源分布，也不依赖于 n。该构造并非搜索能成功的系数。

3.2.6 完整的仅求和构造

求值器的形式为

$$\mathrm{Ext}_n(z) = \pi\,\sigma_D\left(\mathrm{Tr}_{K/F}((L_1 z)^{2^{h_1}-1}), \ldots, \mathrm{Tr}_{K/F}((L_{2D-1} z)^{2^{h_{2D-1}}-1})\right). \tag{3.28}$$

这里， $L_i:\ \mathbb{F}_2^n \to K$ 是固定的二元线性映射， D 是固定的2的幂， $K = \mathbb{F}_{2^N}$、 $F = \mathbb{F}_{2^\ell}$ 满足 $N = a\ell = \Theta_\delta(n)$ 并具有固定的 a，而 σ_D 是次数为 D的初等对称多项式。矩指数 h_i 根据固定的秩层级递增。映射 π 是固定的二元线性输出投影。

每个块使用一个大域乘积。仅为进行秩分析，写作 $z = x + y$ 并令$d_i = 2^{h_i} - 1$。取一个次数为 $d_i + 1$的扩域 E_i/K，其迹对偶 K基为（$e_{i,\,j}$）$_{0\leqslant j\leqslant d_i}$ 和（e_{i}^{*}、$_j$）$_{0\leqslant j\leqslant d_i}$。定义源局部编码

$$A_i(x) = \sum_{j=0}^{d_i} (L_i x)^j e_{i,j}, \qquad B_i(y) = \sum_{j=0}^{d_i} (L_i y)^{d_i - j} e^*_{i,j}.$$

由于每一个 $(x+y)^{\,2^{h_i}-1}$ 的二项式系数都是奇数，迹对偶性给出了精确的恒等 ty

$$\mathrm{Tr}_{K/F}\big((L_i(x+y))^{d_i}\big) = \mathrm{Tr}_{E_i/F}\big(A_i(x)B_i(y)\big). \tag{3.29}$$

域次数 $[E_i:\ F] = a2^{h_i}$ 一旦熵率和列确定后便为常数。这是一种代数表示，并未改变仅和求值器。特别是，其二元特征的全局秩至多为 $2^{h_i}N$。

新的输入端步骤是一个显式线性强凝聚器，具有常数大小的种子空间，且输出率高于二分之一。较小的常数误差已足够：它用于选取一个实际乘积子源，以进行支撑论证，而不会计入最终统计误差。求值器会枚举该凝聚器的所有种子。

3.2.7 参数调度、原生域与实际行算法

构造 3.4 常数率和集提取器

前文中的组成部分。定理3.17 （构造3.3）中的线性强凝聚器；推论3.21中的矩掩码裕度；多项式 σ_D 及其强制恒等式，引理3.7；第2.2节中的相对迹。

公开参数与固定设置。 固定 $\delta > 0$。在 $\rho = \delta/2$ 和 $\varepsilon_* = 2^{-12}$ 处实例化凝聚器，得到 Q 个长度为 $M = M(n)$ 的线性视图，以及常数 $\beta > 0$。将种子枚举一次，省略固定的最后一个种子，并将其余映射称为 $L_1, \dots, L_T$，其中 $T = Q - 1 = 2D - 1$ 和 $D = Q/2$。令 $\alpha_0 = 1/2 + \beta/8$ 和 $\alpha_1 = 1/2 + \beta/4$。按列顺序递增，定义

$$\kappa_i = \sum_{j<i} 2^{h_j}, \qquad p_i = \left\lceil \frac{\kappa_i + 4}{\alpha_0} \right\rceil, \tag{3.30}$$

并取最小的 $h_i \geqslant 3$，使得 $2^{h_i - 2} \geqslant 2p_i$。令 $\theta_i > 0$ 为推论 3.21 中得到的指数，并令 $\theta = \min_i \theta_i$。固定一个整数

$$a \geqslant \max\{2, 4T/\theta\}. \tag{3.31}$$

设 $\ell = \lceil M/a \rceil$，$N = a\ell$，并固定相容域 $F = \mathbb{F}_{2^\ell} \subseteq K = \mathbb{F}_{2^N}$ 和二进制基。固定一个满射二进制线性输出投影 $\pi : F \to \mathbb{F}_2^{m_n}$，满足 $m_n = \text{最小值}\{\ell, \lfloor \zeta_\delta n \rfloor\}$，其中偏差常数 ζ_δ 已在定理 3.22 的证明中得到认证。

输入。 仅有 $z \in \mathbb{F}_2^n$，在定理中解释为 $X + Y$。

求值步骤。

1. 对每个 $i \in [T]$，求值 $L_i(z)$，追加 $N - M$ 个零比特，并将结果视为 K 的一个元素。

2. 对于每个保留列，计算

$$W_i(z) = \mathrm{Tr}_{K/F}\big((L_i z)^{2^{h_i - 1}}\big) \in F. \tag{3.32}$$

3. 在 F 中计算 $Z_n(z) = \sigma_D(W_1(z) \dots, W_T(z))$，并返回 $\pi(Z_n(z))$。

输出。 $m_n = \Omega_\delta(n)$ 比特，并满足定理 3.22的保证。所有预处理映射均为二元线性；没有向求值器提供 z 作为 $x + y$ 的分解。

未评估。 优列选择、封顶模式、代表表，以及分离乘积证明中的扩张域 E_i 。

用于线性映射的系数域 $\mathbb{F}_Q$ 不必是 K的子域：它用于指定二元线性预处理。迹域 $F \subseteq K$ 具有所示的相容表示。

对于 $W_j(x + y)$的每个输出特征，迹传递性和命题 3.20 给出全局二元秩界 $2^{h_j} N$。因此，完整的先前前缀的每个奇偶校验

$$\Pi_i = (W_1(X + Y), \dots, W_{i-1}(X + Y))$$

秩至多为 $\kappa_i N$。此界与之后对 a 的选择无关：行输出被缩短了，但本征域核心仍作用于 N 比特。局部定理中的所有渐近估计对于这些域次数均成立。

3.2.8 实际乘积源上的新鲜性

我们首先介绍用于转移局部秩估计的基本代表表方法，而无需固定联合前缀。

引理 3.23（分离代表表）。设 X，Y 相互独立，且 $U = L(X)$、$V = R(Y)$。存在相互独立的随机确定性映射 T，S，且它们与 U，V 独立，使得

$$(\mathsf{T}(U), \mathsf{S}(V)) \overset{\text{distribution}}{=} (X, Y), \qquad L(\mathsf{T}(u)) = u, \quad R(\mathsf{S}(v)) = v$$

在支撑集中的取值上成立。分别对每个固定映射对进行前复合，不会增加任何原始核的二元秩。

证明。先从条件分布 $X \mid U = u$ 中采样表中每个条目 $\mathsf{T}(u)$，再独立抽取 U；这样，$\mathsf{T}(U)$ 就会复现 X，且 $L(\mathsf{T}(u)) = u$ 在其支撑集上成立。对右侧重复此过程，且彼此独立。对于固定的表，分别进行前复合相当于选取或重复矩阵的行和列，因此不会增加二元秩。 □

引理 3.24（完整前缀新鲜性）。设 X^*，Y^* 是实际的独立乘积分布，其 i 阶观测边缘分布的熵至少为 $\alpha_1 N$。那么，行（3.32）满足

$$\Delta((W_i(X^* + Y^*), \Pi_i), (U_F, \Pi_i)) \leqslant \frac{1}{2}\, 2^{i\ell/2 - \theta_i N} \leqslant 2^{-\theta N/2} \tag{3.33}$$

for 足够大的 n。不对更早的源的熵作任何假设 ws.

证明。 固定当前域输出的一个非零特征，以及前缀的任意一个二元特征。在原生域中，它们的组合相位为

$$\mathrm{Tr}_{K/\mathbb{F}_2}(\lambda(L_i x + L_i y)^{2^{h_i}-1}) + T_{\mathrm{pre}}(x, y), \qquad \lambda \in F^\times \subset K^\times.$$

原始掩码的秩至多为 $\kappa_i N$。对当前的两个视图使用引理 3.23 。对于每一对固定的代表表，它仍然是在 $\kappa_i N$ 上的秩至多为 K^2的掩码，而两个视图的分布仍保持独立，并保留其熵。推论 3.21 将相位偏差界定为 $2^{-\theta_i N}$。对代表表对取平均；这与原始的联合输出/前缀分布相同。

将该分布与一个独立于实际前缀的均匀当前域值进行比较。当前特征为零的傅里叶系数会完全抵消。至多剩下 $(2^\ell - 1)2^{(i-1)\ell}$ 个系数。帕塞瓦尔恒等式和柯西–施瓦茨不等式给出（3.33）中的第一个界。由于 $a \geqslant 4T/\theta$，有$i\ell/2 \leqslant \theta N/8$，这意味着第二个较弱的界。我们没有对两个源施加条件Π_i；前缀的新鲜性是通过独立源对上的混合特征得到证明的。 □

3.2.9 多数强制与最终误差界

引理中的强制恒等式3.7 同样适用于 σ_D 和集输出域 F。

命题 3.25（每个低速率矩形上的完整输出支撑）。 对于上述固定构造，令

$$Z_n(z) = \sigma_D(W_1(z), \ldots, W_{2D-1}(z)) \in F.$$

对于任意 $AB \subseteq \mathbb{F}_2^n$ 与 $|A|$，$|B| \geqslant 2^{(\delta/2)n}$，对于充分大的 n，有 $Z_n(A + B) = F$.

证明。 将强线性凝聚器应用于 U_A 和 U_B。分别使用引理 3.19 ，得到一个实际乘积分布 X^*, Y^* 其支撑集包含于 $A \times B$，其中至少有 D 个共同保留的良好列。固定一组 H 个 D 这样的索引，并按算法顺序排列。它们的观察熵至少为 $\alpha_1 N$ ，由（3.24），因此引理3.24 适用于每个索引。将每个较早的前缀投影到较早选出的索引上，并使用逐次替换。这给出

$$\Delta((W_i(X^* + Y^*))_{i\in H}, U_F^{\otimes D}) \leqslant D2^{-\theta N/2}.$$

这里的初等归纳首先针对实际的前置选定边缘分布替换当前坐标，然后利用前一步的归纳界将该前置边缘分布替换为均匀分布。它不要求坏列或后续列之间具有独立性。

由（3.31）， $D\ell \leqslant \theta N/4$。因此，当 n 足够大时，

$$D2^{-\theta N/2} < 2^{-D\ell} = |F|^{-D}.$$

对于每个 $u \in F$，所有选定列都等于 u 这一事件在实际乘积分布下因此具有正概率。引理 3.7 使得合并值 u^D 在该事件上成立。映射$u \mapsto u^D$ 是 F上的弗罗贝尼乌斯自同构，因此每个域值都会出现。所有这些对都属于$A \times B$。

□

引理 3.26 （实际仅求和复合的秩）。每个 $W_i(x + y)$ 在 F 上的分离长度至多为 $a2^{h_i}$。因此， $Z_n(x + y)$ 的分离长度至多为

$$R_* := \sum_{\substack{I\subseteq[2D-1]\\|I|=D}} \prod_{i\in I} a2^{h_i}, \tag{3.34}$$

一个仅取决于固定速率的常数。 $Z_n(x + y)$ 的每个绝对迹特征的二元秩至多为 $R_*\ell = O_\delta(n)$。

证明。 应用精确域乘积恒等式（3.29）于 W_i（$x + y$）。在 E_i/F 上的迹对偶性至多产生 $[E_i : F] = a2^{h_i}$ 个分离的 F-乘积。固定的行乘积会将它们的分离长度相乘，而求和会将这些长度相加；因此展开 σ_D 即可得到（3.34）。最后，每个分离的 F-乘积在取绝对迹后至多贡献 ℓ 二元秩。所有恒等式都对实际的未条件化的仅求和求值器成立。

□

定理 *3.22*的证明。 对于每个非零 $\lambda \in F$，布尔函数

$$f_\lambda(x, y) = \mathrm{Tr}_{F/\mathbb{F}_2}(\lambda Z_n(x + y))$$

在每个边熵为 （$\delta/2$）n的矩形上都不是常函数，这是由命题 3.25得出的。引理 3.26 给出了均匀的 O_δ（n）秩界。现在将推论 2.20应用于原始的 n-比特加数，其熵增量为 $\delta/2$，可得到某个 $\zeta_\delta > 0$ ，使得每个这样的特征在所有相互独立的熵-δn 输入上偏差至多为 $2^{-\zeta_\delta n}$ 。取任意一个满射二元线性投影到

$$m_n = \min\{\ell, \lfloor\zeta_\delta n\rfloor\} = \Omega_\delta(n)$$

比特。非零输出奇偶性是 F 的非零绝对迹特征。因此

$$\Delta(\pi Z_n(X + Y), U_{m_n}) \leqslant \tfrac{1}{2}\sqrt{2^{m_n} - 1}\, 2^{-\zeta_\delta n} \leqslant \tfrac{1}{2}\, 2^{-\zeta_\delta n/2}.$$

这给出了误差和输出方面的结论，并保留了秩界。

线性映射、幂、域和对称多项式均独立于源而固定。对于固定特征，有限域构造和算术运算均可在多项式时间内完成；视图数量、矩阶、域扩张比和合并器次数均为固定常数。对称多项式也可以通过其初等系数递推进行求值。求值器不会计算引理 3.15 中使用的优列选择、哈希投影、接受模式、代表表或近似对偶性矩形。所有长度取整在相应阶段只产生常数开销，对于充分大的 n，这些开销均可吸收。这证明了一致显式性。

□

3.2.10 单侧不可篡改和集种子的组成与证明

引理 3.27 （秩与傅里叶约定）。 分离前复合不会增加函数矩阵的二元秩。秩具有次可加性。若 $F=\mathbb{F}_{2^\ell}$ 与 $H(x,\ y)=\sum_{j=1}^{r}a_j(x)b_j(y)$ 在F上，则 H 的每个绝对迹特征的二元秩至多为 $r\ell$。固定次数的相对迹在分离乘积中至多付出其次数的代价。

若 $(Z,\ T)\ \in\mathbb{F}_2^q\times\mathbb{F}_2^v$，且在 Z 上具有非零系数的每个混合特征的偏差至多为 ω，则

$$\Delta((Z,T),(U_q,T))\leqslant\tfrac{1}{2}\sqrt{(2^q-1)2^v}\,\omega.$$

对于一对 q 比特的输出，取 $v=q$ 时结论成立。

证明。 秩的断言是第 2.2 节中关于分离与迹对偶性的常见事实；距离界是引理 2.11 ，其中仅保留含诚实系数的特征。 □

引理 3.28 （实际条件代表元）。 对于确定性映射 $U=T(X)$，可以将 $(X,\ U)$ 的联合分布写成 $(R(U),\ U)$ 的混合，其中 R 是一个与 U 独立的随机确定性表，并且对于所有正概率的 u，都有 $T(R(u))=u$。对于独立源，分别采用这类表示，在固定这些表后仍保持独立性，并且不会增加任何分别提升后的矩阵秩。

证明。 应用引理 2.10。其代表元位于各自的实际纤维中，而分别进行前复合不会增加矩阵秩。

□

3.2.11 针对单侧篡改的矩截面

固定 $h\geqslant 3$、$d=2^h-1$、$K=\mathbb{F}_{2^N}$ 和 $N>2h$，令 N 在下文中足够大。置 $L=2^{h-2}$.

引理 3.29 （迹码限制核）。 设

$$\mathcal{C}_d=\left\{\varepsilon+\mathrm{Tr}_{K/\mathbb{F}_2}\sum_{\substack{1\leqslant j\leqslant d\\ j\ \mathrm{odd}}}c_jy^j\right\}.$$

其显示系数是唯一的，并且每个非常数成员的偏差至多为 $(d-1)2^{-N/2}$。如果 $B\subseteq K$ 的大小至少为 $2(d-1)2^{N/2}$，则

$$\dim\ker(\mathcal{C}_d\to\mathbb{F}_2^B)\leqslant N-\log|B|+1.$$

证明。 这些正是命题3.20证明的第一部分中确立的唯一系数结论和限制核结论；方程（3.26）给出了所示的维数界。该论证只使用右侧集合 B ，且同样适用于割线行函数。

□

引理 3.30（矩割线森林）。设 A 是特征为二的域 K 的有限子集，且 $f\colon A\to K$ 满足 $f(x)\neq x$。对于非零的 $a,\ b\in K$，令

$$v_L(x)=(1,x,\dots,x^{L-1}),\qquad w(x)=av_L(x)+bv_L(f(x)).$$

如果 $4p\leqslant L$ 和 $|A|\geqslant \max\{12,\ 4p\}$，则

$$\dim_{\mathbb{F}_2}\operatorname{span}\{w(x):x\in A\}\geqslant \log\binom{\lfloor |A|/3\rfloor}{p}\geqslant p\log|A|-p\log(4p).$$

证明。对于每个 $x\in A$，绘制无向边 $\{x,\ f(x)\}$。不存在自环；一条无向边至多被表示两次，因此至少有 $|A|/2$ 条不同的边。该有向图的出度至多为一（A 之外的顶点出度为零），所以每个无向连通分量至多包含一个环。合并重复的互反边后，每个环的长度至少为三。删除每个环中的一条边。剩余的森林至少包含不同边数的三分之二，因此至少有 $|A|/3$条边。为 $\lfloor |A|/3\rfloor$ 条保留的边各选取一个定义域代表元。

对这些代表元的不同 p元子集，$w(x)$的和彼此不同。否则，它们的对称差是一个至多含有 $2p$ 条边和至多含有 $4p$ 个顶点的非空森林。在一个叶顶点处，其在所提出关系中的系数为 a 或 b，且非零。因此，与之对应的至多 $4p\leqslant L$ 个范德蒙德列之间的关系是非平凡的，矛盾。二元张成空间中有 $\binom{\lfloor |A|/3\rfloor}{p}$ 个不同的和。最后，对于 $|A|\geqslant 12$有 $\lfloor |A|/3\rfloor\geqslant |A|/4$，且$\binom{u}{p}\geqslant (u/p)^p$。

□

命题 3.31（混合限制秩界）。对于 $\lambda\in K^\times$、$\mu\in K$和无不动点$f\colon A\to K$，令

$$M(x,y)=\operatorname{Tr}_{K/\mathbb{F}_2}\bigl(\lambda(x+y)^d+\mu(f(x)+y)^d\bigr).$$

若 $4p\leqslant L,|A|\geqslant$最大$\{12,\ 4p\}$，且 $|B|\geqslant 2(d-1)2^{N/2}$，则

$$\operatorname{rank}_2 M[A,B]\geqslant p\log|A|-p\log(4p)+\log|B|-N-1. \tag{3.35}$$

对于 f的任意扩域，完整定义域矩阵的秩至多为 $2^{h+1}N$。

证明。所有行函数都属于$\mathcal{C}_d$。对于 $0\leqslant t<L$，指数 $d-2t$ 为奇数且大于 $d/2$。在二项式展开中，没有其他指数能在迹运算下化为它，因此其系数恰好为

$$\lambda x^{2t}+\mu f(x)^{2t}.$$

系数提取后进行逆平方运算是二元线性的，并得到 $\sqrt{\lambda}\,v_L(x)+\sqrt{\mu}\,v_L(f(x))$。若 $\mu\neq 0$，引理 3.30 给出了完整行张成空间的下界。若 $\mu=0$，对不同的 p-子集 A求和所得的结果根据通常的范德蒙德线性无关性彼此不同；因此同样得到较弱的界。引理 3.29界定了限制到 B时损失的维数。

两个二项式展开中的每一个都具有 $d+1=2^h$ 相互分离的 K-乘积；引理 3.27 给出了全局界。单输出特征界为 2^hN。 □

推论 3.32（任意指定线性秩的固定速率掩码）。固定 $1/2<\alpha_0<\alpha_1<1$ 和一个常数 $\kappa>0$。可以选取一个固定的 h，使得对于所有充分大的 N，每个无不动点$f\colon K\to K$、每个 $\lambda\neq 0$、每个 μ以及每个全局秩至多为 κN 的布尔掩码 T均满足

$$\left|\mathbb{E}(-1)^{\operatorname{Tr}(\lambda(U+V)^{2^h-1}+\mu(f(U)+V)^{2^h-1})+T(U,V)}\right|\leqslant 2^{-\theta N}$$

对于最小熵至少为 α_1N 的独立 $U,\ V$，其中 $\theta>0$ 在 $f,\ \lambda,\ \mu,\ T$ 上服从均匀分布。对于 $\mu=0$，相同的结论无需对 f 作任何条件限制。

证明。 取 p 使得 $(p+1)\alpha_0 - 1 > \kappa + 2$，再取 h 使得 $4p \leqslant 2^{h-2}$。在每个边-$\alpha_0 N$ 矩形上，命题 3.31 表明，加上掩码后秩仍大于一。满秩至多为$(2^{h+1}+\kappa)N$。推论 2.20结合熵增量 $\alpha_1 - \alpha_0$，即可证明结论。已知仅在实际左支撑集上为无不动点的映射，可以通过非零平移将其扩展到该集合之外；应用中的所有分布均不变。 □

3.2.12 单侧不可篡改和集种子：主定理

同一源兼容架构还有一个强化版本，用于第 和集 熵回收定理的 4.4 节。只有第一个被加数受到篡改。这并不意味着和的某个函数具有双侧不可篡改性。同一种子还提供了第4.1节普通熵回收包装器所用的单侧接口。第3.1节中独立的普通常数率构造不使用此种子。

定理 3.33 （单侧不可篡改和集种子）。 对于每个固定的 $0 < \sigma \leqslant 1$，都存在c_σ，$e_\sigma > 0$ 以及一个确定性多项式时间映射 $G_n \colon \mathbb{F}_2^n \to \mathbb{F}_2^{q_n}$, $q_n \geqslant c_\sigma n$，使得对于最小熵至少为 σn 的独立源，以及每个满足在 $\operatorname{supp} X$ 上没有不动点的 f，均有

$$(G_n(X+Y), G_n(f(X)+Y)) \approx_{2^{-e_\sigma n}} (U_{q_n}, G_n(f(X)+Y)). \tag{3.36}$$

每个输出特征的全局提升二元秩 $O_\sigma(n)$，每个单比较混合特征也如此。任选一个非零平移 f 并投影诚实输出，即可得到普通提取。该族对每个正整数 n都有定义，并采用下文指定的有限长度补全。

构造。 通过对原始输入应用多项式求值指纹，增强线性强凝聚器视图。枚举一个常数大小的求值集，使每个非零原始差在几乎每个指纹中都能保留下来。对每个增广视图应用一个在固定秩层级中选定阶数的矩幂，取迹到公共域，并与相同次数为-D 的初等对称多项式合并。更大的矩阶保护含诚实系数的割线系数，使其免受每个更早前缀特征的影响。最终的二元线性投影提供种子。该求值器仅读取求和输入；篡改映射仅出现在证明中。

证明。 我们先建立割线秩估计，然后给出完整的参数与支撑集强制论证。

3.2.13 完整的单侧构造及其证明

构造 3.5 单侧不可篡改和集种子

前文中的工具。 定理3.17中的线性强凝聚器；推论 3.32中的割线掩码界；引理 3.7中的多数多项式；固定域表示与相对迹。

公开参数和固定设置。 固定目标分量率 $\sigma>0$。以率 $\nu=\sigma/64$ 和误差 2^{-16} 实例化凝聚器，得到映射 L_u、$u\in\mathbb{F}_Q$、行长度 M 以及常数 β。令 $r=\lfloor\beta M/16\rfloor$、$E=\mathbb{F}_{2^r}$ 和 $b_n=\lceil n/r\rceil$。固定一个常数 b_*，使其在大长度范围内界定 b_n；固定一个 2 的幂 $R\geqslant 64b_*$，以及互不相同的 v_1、……、$v_R\in E$。按字典序排列 $\mathbb{F}_Q\times[R]$ 并删除其最后一对。保留的数量为 $T=QR-1=2D-1$、$D=QR/2$。

令 $\alpha_0=1/2+\beta/8$ 和 $\alpha_1=1/2+\beta/4$。对于每个保留的索引 i，按顺序令 $\kappa_i=1+2\sum_{j<i}2^{h_j}$，并根据推论 3.32 选择 h_i。具体而言，其证明允许取一个整数 p_i，满足 $(p_i+1)\alpha_0-1>\kappa_i+2$，随后取 $h_i\geqslant 3$，满足 $2^{h_i-2}\geqslant 4p_i$。令 $\theta=\min_i\theta_i>0$ 为所给出的偏差指数，并固定 $a\geqslant\max\{2,16T/\theta\}$。令 $\ell=\lceil(M+r)/a\rceil$、$N=a\ell$ 和 $F=\mathbb{F}_{2^\ell}\subseteq K=\mathbb{F}_{2^N}$。固定一个到 $q_n=\min\{\ell,\lfloor cn/2\rfloor\}$ 比特的满射投影，使用下方最终全局提升所认证的常数 $c>0$。

输入。 仅 $z\in\mathbb{F}_2^n$。

求值步骤。

1. 将 z 填充至 b_nr 比特，将其解析为 z_0、……、$z_{b_n-1}\in E$，并构造 $P_z(V)=\sum_{j<b_n}z_jV^j$。2. 对于每个保留的有序对 $(u,\ v_j)$，连接 T_u、$v_j(z)=(L_uz,\ P_z(v_j))$。追加零以达到 N 比特；按照固定列顺序，将结果识别为 $T_i(z)\in K$。3. 对所有 $i\in[T]$，计算 $W_i(z)=\mathrm{Tr}^{K/F}(T_i(z)^{2^{h_i}-1})$。4. 计算 $Z_n(z)=\sigma_D(W_1(z)\text{……},\ W_T(z))$，并返回其固定的 q_n 比特投影 $G_n(z)$。

输出。 定理 3.33中的种子。其不可篡改性允许仅篡改第一个被加数，而第二个保持不变。对于有限初始范围，按照证明之后的规定，返回一个零比特。

未评估。 映射 f、其函数图和森林、接受模式，以及比较像是否具有重原子的判定，都不会影响已评估的视图或指数列表。

线性指纹可防止局部不动点 在原始阈值 νn 处使用定理 3.17，其中$\nu=\sigma/64$ 且误差为 $\varepsilon_c=2^{-16}$。令 M、Q、β 为其参数。置

$$r=\lfloor\beta M/16\rfloor,\qquad E=\mathbb{F}_{2^r},\qquad b_n=\lceil n/r\rceil.$$

对于固定的 b_*，当 n、$b_n\leqslant b_*$ 足够大时，将 z 填充为 b_n 个域系数，并写作 $P_z(V)=\sum_{j<b_n}z_jV^j$。选择一个2的幂 $R\geqslant 64b_*$，以及固定且互不相同的 v_1、……、$v_R\in E$。定义所有增广线性视图

$$T_{u,v}(z)=(L_uz,P_z(v))\in\mathbb{F}_2^{M+r}.\tag{3.37}$$

根据多项式根界，每个不同的原始对在这些索引中除 b_n-1 个之外的所有索引处都具有不同的指纹。因此，至少有 $63R/64$ 个指纹逐点检测 $x\neq f(x)$，其中 x。

实际公共良好视图 设 X、$f(X)$ Y的熵均至少为 νn。令 $K_t=(1/2+\beta/2)M$。对于每个 L_uX和 $L_uf(X)$，引入一个源局部接受位；以概率 z接受值 $\min\{1,2^{-K_t}/P(z)\}$。拒绝概率等于超过熵上限的部分，且至多为定理 3.17中的距离。具有任一左侧

拒绝的索引的期望比例至多为 $2\varepsilon_c$。因此，以至少 $1-32\varepsilon_c>1/2$ 的概率，至多有 $Q/16$ 个索引处于左侧不良状态。

另记录 R 确定性指纹差异位。完整左模式至多有 2^{2Q+R} 个。满足拒绝界的模式，其概率至少为 $2^{-2Q-R-1}$。固定该模式。 X 和 $f(X)$ 的每个接受行边缘分布都不超过 $2^{-K_t+2Q+R+1}$。检测指纹索引的集合已固定，且大小至少为 $63R/64$。在右侧独立地，一个正模式使 Y 中至少 $15Q/16$ 行的熵不超过 $K_t-O(Q)$。将由此得到的独立实际源分布记为 $X^*,\ Y^*$。

至少有 $7Q/8$ 个索引 u 对所有三个边缘分布都是好的。将它们与检测指纹索引配对。在这些 QR 增广视图中，超过 $3QR/4$ 个视图的诚实与比较左视图以及右视图的熵都至少为 $K_t-O(Q+R)$，并且满足

$$T_{u,v}(x)\neq T_{u,v}(f(x))\quad\text{on the selected left support.}\tag{3.38}$$

删除一个固定的网格索引，并记

$$T=QR-1=2D-1,\qquad D=QR/2.$$

于是， D 是 2 的幂，且至少有 D 列为共同良好。对于单次执行，使用相同的论证，但不设比较上限或差异位：当 X,Y 的熵至少为 νn时，超过一半的列都是良好的。

Let

$$\alpha_0=1/2+\beta/8,\qquad \alpha_1=1/2+\beta/4.$$

在填充任意固定数量的比特至本征长度 $N=M+r+O(1)$ 后，对于充分大的 n，每个共同良好视图的熵都至少为 $\alpha_1 N$。事实上，$(1/2+\beta/2)>(1/2+\beta/4)(1+\beta/16)$，且具有固定的正裕量；所有模式和填充造成的损失都是常数。

层级、实际前缀和域长度 对保留列排序一次。递归地设定

$$\kappa_i=1+2\sum_{j<i}2^{h_j}$$

并选择 h_i 由推论 3.32 在该掩码预算和固定速率 α_0,α_1. 令 $\theta_i>0$ 为其指数，并令 $\theta=$最小值$_i\,\theta_i>0$。取一个固定整数 $a\geqslant$最大值$\{2,16T/\theta\}$，然后令

$$\ell=\lceil (M+r)/a\rceil,\qquad N=a\ell,\qquad F=\mathbb{F}_{2^\ell}\subseteq K=\mathbb{F}_{2^N}.$$

填充每个 少于 a 个零比特，并将其与 K对应。实际的域行是 e

$$W_i(z)=\mathrm{Tr}_{K/F}\big(T_i(z)^{2^{h_i}-1}\big).\tag{3.39}$$

每个特征的秩至多为 $2^{h_i}N$。完整对前缀

$$\Pi_i=(W_j(X^*+Y^*),W_j(f(X^*)+Y^*):j<i)$$

在原始原始定义域上具有每个秩低于 $\kappa_i N$ 的奇偶校验。

引理 3.34（公共良好列的成对新鲜性）。 对于上文选定的那个实际乘积分布，每个公共良好列都满足

$$\Delta((W_i,W_i',\Pi_i),(U_F,U_F',\Pi_i))\leqslant 2^{-\theta N/2}$$

当 n 足够大时，其中 $W\prime_i=W_i$（f（X^*） $+Y^*$）。对于单次执行，类似的前缀分布成立，且误差不会更大。

证明。 取 $U = T_i X^*$ 和 $V = T_i Y^*$。由引理 3.28，为两个原始输入固定独立的条件代表元表。在取值 u支撑集上，比较视图是一个确定性映射 $f_i(u)$，且不同于 u，根据（3.38）。将 f_i 扩展到取值支撑集之外，并确保没有不动点。经过这次单独替换后，每个前缀奇偶校验的秩仍低于 $\kappa_i N$。原生诚实核心为 $(U+V)^{2^{h_i}-1}$；迹传递性将每个非零 F系数映射为非零原生系数。

应用推论 3.32 并对代表表取平均。引理 3.27给出

$$\Delta((W_i, W_i', \Pi_i), (U_F, W_i', \Pi_i)) \leqslant \tfrac{1}{2} 2^{i\ell - \theta_i N}.$$

现在改为根据比较视图来表示原始左侧输入。该视图具有相同的高熵，诚实视图在逐点意义上不同，而右侧源保持不变。同一界使 W_i'相对于（W_i，Π_i）新鲜。这两个断言都涉及原始联合分布，而非混合源。将每个当前坐标替换为独立均匀随机量的操作彼此可交换，且会收缩统计距离。这些操作的总代价至多为 $2^{i\ell-\theta_i N} \leqslant 2^{-\theta N/2}$。这里没有对联合前缀进行条件化。单次执行证明使用 $\mu = 0$。 □

强制、重比较与全局提升 定义

$$Z_n(z) = \sigma_D(W_1(z), \ldots, W_{2D-1}(z)). \tag{3.40}$$

若 D 个条目等于 u，则无论其他项为何，次数为-D 的初等对称多项式都等于 u^D：生成多项式含有因子 $(1+ut)^D = 1 + u^D t^D$，而剩余因子的次数至多为 $D-1$。

对于一组 D 公共良好指标，引理3.34 以及连续替换给出的距离至多为 $D2^{-\theta N/2}$，相对于 $U_F^{\otimes 2D}$。当 n足够大时，该值小于 $|F|^{-2D}$，因为 $2D\ell \leqslant \theta N/8$。因此，规定所有诚实选定条目为 u 、所有比较选定条目为 v 的每个事件都有正概率。它强制$(Z_n,\ Z_n') = (u^D,\ v^D)$。由于 D 是2的幂，这一对具有满支撑 F^2。单次执行论证表明，在侧熵至少为 νn的每个原始乘积支撑上都有满支撑。

首先只需处理全局无不动点的 f。仅在原始左支撑集上没有不动点的映射，可以通过非零平移扩展到该支撑集之外，同时不改变任一经评估的分布。为证明侧熵为（$\sigma/2$）时的混合弥散n，在任意相应支撑集上取均匀分布的 X，Y，并取一个诚实系数非零的拟议常数混合特征。如果 $f(X)$ 具有质量大于 $2^{-\sigma n/8}$的原子，则将其固定在左侧。这使诚实左熵至少为 $3\sigma n/8$。此时比较特征仅是 Y的布尔函数；固定其取值的代价至多为一比特。独立的诚实源仍然超过 νn，因此单次执行的满支撑性与其为常数相矛盾。如果不存在这样的原子，则 X，$f(X)$ Y均达到凝聚器阈值。公共良好构造和成对强制给出矛盾。比较系数为零的情形已经属于单次执行弥散。

(3.39) 中的每一行的域分离至多为 $a2^{h_i}$，因为 T_i 是线性的。(3.40) 的分离至多为

$$R_* := \sum_{|I|=D} \prod_{i \in I} a 2^{h_i} = O_\sigma(1).$$

$Z_n(x+y)$ $Z_n(f(x)+y)$ 的混合特征的秩至多为 $2R_*\ell = O_\sigma(n)$。推论 2.20 现在应用于原始的 n 比特源，将阈值从 $\sigma n/2$ 提高到 σn ，并给出所有含诚实输出的偏差均至多为 2^{-cn}，且对 f 一致。投影到 $q_n = \min\{\ell, \lfloor cn/2 \rfloor\}$ 比特。引理 3.27 给出的单侧不可篡改性误差至多为 $2^{q_n - 1 - cn} = 2^{-\Omega_\sigma(n)}$。这就证明了对所有充分大的 n，定理 3.33 均成立。所有参数选择都在获知源之前确定。

对于剩余的有限个正长度，返回一个零比特，并按照 c_σ、 e_σ 的方式减小第 2 节中的有限长度约定。这保留了单侧不可篡改性、普通提取和提升的秩界，并完成了对每个正长度的定理证明。

注记 3.35（仅篡改一个加数）。 定理 3.33 不允许同时平移两个加数。这种平移会保持 $X+Y$不变，并使诚实输出与比较输出相等，这与普通不可篡改性不相容。当第二个加数保持不变且 f 没有不动点时，则不会产生这种矛盾。

3.3 固定-C 不可篡改双源提取

3.3.1 带标签的仿射种子与单侧 Gold 识别

下面的起始器使用较短的左侧前缀和较长的右侧行。其逆向论证仅能识别左侧篡改。外层构造通过对原始输入进行源局部指纹处理，提供彼此不同的数值标签。

定理 3.36 （近满熵仿射非篡改性）。 参考文献： [49,定理 1.12 和第 3] 节。存在绝对常数 δ_A, c_A, $e_A>0$ 以及一个显式映射 V_N，其定义域为 N 比特，输出比特至少为 c_AN，它是一个熵为 $(1-\delta_A)N$、误差为 2^{-e_AN} 的仿射不可篡改提取器。它在该阈值下还具有普通仿射提取保证。

这是 Li 的定理1.12及其在 [49]第 3 节中的构造。根据下面的初等扩展引理，无不动点约定可用于活动仿射支撑集上。

引理 3.37 （将仿射比较从其活动支撑集扩展出去）。 如果 $A\subseteq\mathbb{F}_2^N$ 是一个非空仿射空间，且 f 是仿射的，并且对于每个 $x\in A$ 都有 $f(x)\neq x$，那么存在一个全局无不动点的仿射$\sim f$，其在 A 上与 f 一致。

证明。 选取一个仿射收缩 $P\colon\ \mathbb{F}_2^N\to A$，并令 $\widetilde{f}(x)=f(Px)+x+Px$。于是，对于每个 x，有 $\widetilde{f}(x)+x=f(Px)+Px\neq 0$，并且在 A 上有 $Px=x$。 □

引理 3.38 （等式划分）。 对于均匀仿射 $Z\in\mathbb{F}_2^N$、仿射映射 f 和整数$t\geqslant 1$，存在一个仿射划分，除去至多为 2^{-t} 的概率，在该划分的各部分上， $f(Z)=Z$ 恒等成立，或 $f(Z)\neq Z$ 处处成立。每个部分使 Z 的熵降低至多 t。

证明。 $f(Z)$ 的像 $+Z$ 是仿射的。如果它不包含零，则无需划分。否则，它是一个维数为 r 的线性空间，且像在该空间上均匀分布。如果 $r\leqslant t$，则固定完整的像。如果$r>t$，则固定其一个满射的 t 比特线性投影，并舍弃零值。所有非零值都能保证不等式成立。每种保留的固定至多使用 t 个独立仿射方程；后一种情况下被舍弃的概率为 2^{-t}。 □

引理 3.39 （仿射标签种子族）。 存在一个绝对常数整数 $A_*\geqslant 1$，使得对于$N=A_*s$ 和 $F=\mathbb{F}_{2^s}$，可以显式计算两个源局部函数 V_0, $V_1\colon\mathbb{F}_2^N\to F$，满足

$$S_a(z)=V_0(z)+aV_1(z)\qquad(a\in F) \tag{3.41}$$

对每个仿射 f以及互不相同的 a, $a'\in F$，均满足

$$(S_a(U_N),S_{a'}(f(U_N)))\approx_{2^{-2s}}(U_F,S_{a'}(f(U_N))). \tag{3.42}$$

每个域元素都是允许的标签。参数 A_* 独立于所有后续矩宽度和秩裕度。

证明。选取 A_* 足够大，使得 $4s \leqslant \delta_A N/2$，$c_A N \geqslant 2s$，且 $e_A N \geqslant 8s$。将 V_N 截断为 $2s$比特，并将其视为 $(V_0,\ V_1) \in F^2$。应用引理 3.38，令 $t = 4s$取为 $U_N,\ f(U_N)$。每个保留部分上的诚实熵至少为 $N - 4s \geqslant (1-\delta_A)N$。

在非复制部分上，根据引理 3.37 扩展其比较。定理 3.36 使得 V_0、V_1) 相对于比较向量独立均匀，误差为 2^{-8s}。其在 a 处的非零线性求值因此相对于比较求值独立均匀。在复制部分上，普通仿射提取使系数对以相同的误差均匀分布。两个求值向量 (1、a)、(1、a') 彼此独立，因此在该比较分布下，它们的两个值恰好独立均匀。恢复实际的比较边缘分布至多再付出 2^{-8s}。对各部分取平均并计入丢弃质量后，对于充分大的 $2 \cdot 2^{-8s} + 2^{-4s} \leqslant 2^{-2s}$ 至多为 s。

□

我们将 $\mathbb{F}_2^N$与 $L = \mathbb{F}_{2^N}$ 对应，并保留 $F \subseteq L$。在本小节及接下来的小节中，S_a 上的下标表示域标签，而不是提供给最终求值器的均匀种子。公式（3.41）将说明实际的输入相关建议，而不对 V_0、V_1施加任何次数界。

Gold 缺陷与图能量。 对于 $z \in L$，令 $\Gamma_c(z) = (z,\ z^{1+2},\ \ldots\ldots,\ z^{1+2^c})$。对于二元向量空间中的有限子集 S，记 $E(S) = |\{(z_1,\ z_2,\ z_3,\ z_4) \in S^4 :\ z_1 + z_2 + z_3 + z_4 = 0\}|$。令 $A \subseteq V = \mathbb{F}_2^N$，并令 $f : A \to W = \mathbb{F}_2^M$ 为任意映射。将 f 在 A 上的限制的图定义为

$$\mathrm{Graph}(f|_A) := \{(x, f(x)) : x \in A\} \subseteq V \times W.$$

这里 f 的定义域已经是 A；为与后续关于图能量的陈述保持一致，仍保留限制记号。将其归一化图能量定义为

$$\varepsilon_f = 2^{-3N} E(\mathrm{Graph}(f|_A)).$$

以下所有概率均针对整个环境空间中的均匀分布随机变量；属于 A 是事件的一部分。

引理 3.40（受限同态测试）。参考文献。[3, 引理 3.3]。存在绝对常数 $D_0,\ d_0 \geqslant 1$，使得对于所有正整数 $N,\ M$、每个 $A \subseteq \mathbb{F}_2^N$、每个映射 $f\colon A \to \mathbb{F}_2^M$ 以及 $0 < \rho \leqslant 1$,

$$E(\mathrm{Graph}(f|_A)) \geqslant \rho 2^{3N} \quad \Longrightarrow \quad |\{x \in A : f(x) = \ell(x)\}| \geqslant D_0^{-1} \rho^{d_0} 2^N \tag{3.43}$$

对于某个仿射映射 $\ell : \mathbb{F}_2^N \to \mathbb{F}_2^M$。

这是 [3, 引理 3.3]。这里只使用其存在性断言；其算法复杂度在此处无关紧要。

固定二进制识别 $L \cong \mathbb{F}_2^N$。Gold 编码的二进制坐标张成空间是定义在 L上的以下函数空间:

$$\mathcal{C}_c = \left\{ z \longmapsto \mathrm{Tr}_{L/\mathbb{F}_2}\left(a_0 z + \sum_{j=1}^{c} a_j z^{1+2^j} \right) : a_0, \ldots, a_c \in L \right\}.$$

根据迹配对的非退化性，这与 Γ_c 的二进制坐标函数的张成空间一致。对于 $A \subseteq L$ 和 $f\colon\ A \to L$, 其拉回空间为

$$(\mathcal{C}_c \circ f)|_A := \{x \mapsto \varphi(f(x)) : \varphi \in \mathcal{C}_c\} \leqslant \mathbb{F}_2^A.$$

这里 $\circ$ 表示函数复合。空间 $\mathrm{RM}(2,N)|_A$ 由限制到 A 的、关于 L 的固定二进制坐标且次数至多为二的布尔多项式构成；其中包括常函数。定义完整的 Gold 缺陷为

$$d_c(f;A)=\dim_{\mathbb{F}_2}\left(\frac{(\mathcal{C}_c\circ f)|_A+\mathrm{RM}(2,N)|_A}{\mathrm{RM}(2,N)|_A}\right). \tag{3.44}$$

Put

$$\mathcal{U}_f:=(\mathcal{C}_c\circ f)|_A,\qquad \mathcal{Q}_A:=\mathrm{RM}(2,N)|_A.$$

二者都是 $\mathbb{F}_2^A$ 的二进制向量子空间，但 $\mathcal{Q}_A$ 不一定包含于 $\mathcal{U}f$。符号 $+$ 在（3.44）中表示它们的向量空间和，因此分母$\mathcal{Q}_A$ 是分子的子空间。该分式表示商向量空间：当两个函数在A 上的差$\mathcal{Q}_A$属于其中时，就将它们视为相同。特别地，加上一个二次函数不会改变商类。第二同构定理给出

$$\frac{\mathcal{U}_f+\mathcal{Q}_A}{\mathcal{Q}_A}\cong\frac{\mathcal{U}_f}{\mathcal{U}_f\cap\mathcal{Q}_A},\qquad d_c(f;A)=\dim_{\mathbb{F}_2}\mathcal{U}_f-\dim_{\mathbb{F}_2}(\mathcal{U}_f\cap\mathcal{Q}_A). \tag{3.45}$$

因此，d_c（f；A）统计的是这样的独立 Gold 坐标组合：在将那些在 A 上与二次函数一致的组合视为相同之后，仍然保留下来的组合。不假设 f 具有单射性。特别地，如果 f 是二进制仿射的，则 d_c（f；A）$=0$，因为仿射映射的 Gold 坐标的二进制度至多为二。

等价的二次多项式加小误差表示。 商的定义有如下具体推论，这将在定理 3.41 的证明中使用。令 $B=L^{c+1}$，将其视为一个 N（$c+1$）维二元向量空间，并记 $r=d_c$（f；A）。存在一个向量值多项式映射 $Q_f:\ L\to B$，其二进制度至多为二；一个函数 $e_f:\ A\to B$；以及一个固定的二进制线性子空间 $W_f\leqslant B$，其维数为 r，使得

$$\Gamma_c(f(x))=Q_f(x)+e_f(x),\qquad e_f(x)\in W_f\quad(x\in A). \tag{3.46}$$

事实上，考虑二进制线性映射

$$T:B^*\longrightarrow\mathbb{F}_2^A/\mathcal{Q}_A,\qquad T(\lambda)=[x\longmapsto\lambda(\Gamma_c(f(x)))].$$

T的像恰好是 $(\mathrm{U}_f+\mathrm{Q}_A)/\mathrm{Q}_A$，因此秩为 $T=r$。选取 $\ker\ T$ 的一组基 $\lambda_1,\ldots,\lambda_{\dim B-r}$，并将其扩展为 B^* 的一组基。对于 $\ker\ T$ 中的每个基泛函，选取 L 上次数至多为二的布尔多项式 q_i，使其在 A 上与 $x\mapsto\lambda_i(\Gamma_c(f(x)))$ 一致；对于其余的 r 个基泛函，令 $q_i=0$。由于 λ_i 构成 B^* 的一组基，存在唯一的映射 $Q_f:\ L\to B$，满足对每个 i 都有 $\lambda_i(Q_f(x))=q_i(x)$。其二进制坐标是次数至多为二的多项式的线性组合，因此 Q_f 是二次的。现在在 A 上令 $e_f(x)=\Gamma_c(f(x))-Q_f(x)$。每个 $\lambda\in\ker\ T$ 都湮灭 $e_f(x)$，于是

$$e_f(A)\subseteq W_f:=(\ker T)^\perp:=\{z\in B:\lambda(z)=0\text{ for all }\lambda\in\ker T\},\qquad \dim_{\mathbb{F}_2}W_f=r.$$

这证明了（3.46）。特别地，e_f 至多取 2^r 个不同的值；该二次映射定义在 L 的全部范围上，而分解仅在 A 上成立。

对于有限二元向量空间 V 上的实值函数 g，记作

$$\|g\|_{U^3}^8:=\mathbb{E}_{x,h_1,h_2,h_3\sim U_V}\prod_{\omega\in\{0,1\}^3}g(x+\omega_1h_1+\omega_2h_2+\omega_3h_3),$$

其中，四个抽样向量相互独立。这就是第三个 Gowers一致范数（针对实值函数）。反复应用柯西–施瓦茨不等式可得 $|\mathbb{E}g| \leqslant \|g\|_{U^3}$。对于 $g = \mathbf{1}_D$，所显示的平均值恰好等于随机仿射三维立方体的全部八个顶点都位于 D 中的概率。方向向量不必线性独立，因此也包括退化立方体和重复顶点。在下一个证明中，$V = L^p$ 被视为一个二元向量空间。

定理 3.41（任意映射的摊销识别）。对于整数 $p \geqslant 1$、$c \geqslant 8p-1$、$N > 2c$、$|A| = 2^{N-a}$ 和 $r = d_c(f;\ A)$，存在某个仿射映射，使其在 A 的一个大小至少为 $2^{N-O(a+r/p+p)}$ 的子集上与 f 一致。该常数是绝对的。

证明。我们先给出论证中使用的两个代数事实和概率事实，然后解释取 p 个副本如何摊销缺陷 r。下文中，每个立方体均通过独立且均匀地选择其基点和三个方向来抽取；方向可以为零。

步骤 *1*：短 *Gold*关系具有较小的二元张成空间。设 $z_1, \ldots, z_m \in L$ 满足 y

$$\sum_{\nu=1}^{m} \Gamma_c(z_\nu) = 0, \qquad m \leqslant 8p \leqslant c+1. \tag{3.47}$$

令 $d = \dim_{\mathbb{F}_2}$ 张成空间$\{z_1, \ldots, z_m\}$，并在 L 上构造 $m \times m$ Moore矩阵，

$$\mathsf{M} = (z_\nu^{2^i})_{0 \leqslant i < m,\, 1 \leqslant \nu \leqslant m}.$$

对于 $i < j$，$\mathsf{M}\mathsf{M}^{\mathsf{T}}$ 的（i，j）项为

$$\sum_{\nu} z_\nu^{2^i + 2^j} = \left(\sum_{\nu} z_\nu^{1 + 2^{j-i}} \right)^{2^i} = 0,$$

因为 $1 \leqslant j - i \leqslant m - 1 \leqslant c$。含有 $i > j$ 的项因对称性而消失，而对角线项消失是因为

$$\sum_{\nu} z_\nu^{2^{i+1}} = \left(\sum_{\nu} z_\nu \right)^{2^{i+1}} = 0.$$

因此 $\mathsf{M}\mathsf{M}^{\mathsf{T}} = 0$。Moore行列式判据（包括其矩形秩形式）表明，$\operatorname{rank}_L \mathsf{M} = d$：秩等于标签张成空间的 $\mathbb{F}_2$维数，即使某些标签重复也是如此。由于 M^{T} 的像包含于 M 的核中，

$$d = \operatorname{rank} \mathsf{M} = \operatorname{rank} \mathsf{M}^{\mathsf{T}} \leqslant \dim \ker \mathsf{M} = m - d.$$

因此

$$\dim_{\mathbb{F}_2} \operatorname{span}\{z_1, \ldots, z_m\} \leqslant m/2. \tag{3.48}$$

这里没有使用 z_ν 互不相同这一假设。

步骤 *2*：一个三维立方体上的真关系是一种图能量检验。令 I 为 $\mathbb{F}_2^3$的任意非空真子集，并固定 $b \in L$。我们断言，均匀仿射三维立方体 $(x_\omega)_\omega$ 属于 L，满足

$$\Pr\bigl[x_\omega \in A\ (\forall \omega),\quad \sum_{\omega \in I} f(x_\omega) = b\bigr] \leqslant \varepsilon_f^{1/4}. \tag{3.49}$$

选取 $\omega_* \in I$ 和 $\omega_\circ \notin I$。它们在 $\mathbb{F}_2^3$ 中的两点仿射直线，是两个独立仿射布尔函数 β_1，β_2 的同时 1-纤维。再取一个仿射布尔函数β_3，将 β_1，β_2 补全为 $\mathbb{F}_2^3$ 上所有仿射函数的一组基。因此，均匀仿射立方体可写为

$$x_\omega = z_0 + \beta_1(\omega) U + \beta_2(\omega) V + \beta_3(\omega) z_3,$$

其中 z_0，z_3，U，V是 L中的独立均匀元素。固定 z_0，z_3并令 $F(U,\ V)$表示式 (3.49) 中事件的示性函数。若 $\rho = \mathbb{E}_{U,\ V} F(U,\ V)$，则由二变量柯西–施瓦茨不等式（盒子不等式）可得

$$\rho^4 \leqslant \mathbb{E}_{U,U',V,V'} F(U,V)F(U',V)F(U,V')F(U',V'),$$

对于独立均匀分布的 U、U'、V、$V\prime$。如果右侧的四个事件都成立，就将它们的四个输出方程相加。在特征为二的情形下，b 的四个副本相互抵消。对于任何不等于 ω_* 的顶点 $\omega \in I$，$\beta_1(\omega)$ $\beta_2(\omega)$ 中至少有一个为零，因此其贡献成对相同，也会相互抵消。剩下的项只有 f 在 ω_* 处的四个取值。它们的自变量为

$$z + U + V, \quad z + U' + V, \quad z + U + V', \quad z + U' + V', \qquad z = z_0 + \beta_3(\omega_*) z_3.$$

这四个自变量都属于 A，它们的和为零，其 f 值之和也为零。对于每个固定的 z，这四个自变量中有三个是 L 中的独立均匀分布元素；第四个是它们的和。因此，后一图平行四边形事件的概率恰好为

$$2^{-3N} E(\mathrm{Graph}(f|_A)) = \varepsilon_f.$$

因此，$\rho^4 \leqslant \varepsilon_f$ 对于每个固定的 z_0，z_3。取平均即可证明（3.49）。该界对所选集合 I 和常数 b均成立。

步骤 *3*：用一个 r维误差空间表示缺陷。应用在 d_c的定义之后紧接着建立的二次项加小误差表示（3.46(f；A)。记 $Q = Q_f$，$e = e_f$ 以及 $W = W_f \leqslant B = L^{c+1}$。于是

$$\Gamma_c(f(x)) = Q(x) + e(x), \qquad e(A) \subseteq W, \qquad \dim_{\mathbb{F}_2} W = r. \tag{3.50}$$

关键在于，误差 e 至多取 2^r 个值，而 Q 在整个环境输入空间上二进制度至多为二。不涉及 f 为单射或全局多项式的任何假设。

定义 E_p：$A^p \to W$，其中 $E_p(x_1, \ldots, x_p) = \sum_{i=1}^{p} e(x_i)$。某个纤维 $\mathcal{D} = E_p^{-1}(w)$ 的大小至少为$|A|^p/2^r$。作为环境空间 L^p的子集，其密度满足

$$\alpha = \frac{|\mathcal{D}|}{2^{Np}} \geqslant 2^{-pa-r}. \tag{3.51}$$

对非负示性函数使用 U^3 不等式可得

$$\text{均匀分布的三维立方体在 } L^p \text{ 中的全部八个顶点属于 } \mathcal{D}] = \|\mathbf{1}_{\mathcal{D}}\|_{U^3}^8 \geqslant \alpha^8 \geqslant 2^{-8(pa+r)}. \text{的概率} \tag{3.52}$$

对于此处计数的每个立方体，将（3.50）在其八个顶点上求和，并对 $i \in [p]$。八个相同的值 w 相互抵消。$\sum_i Q(x_i)$ 的各项值之和也为零，因为任意二元二次多项式的第三阶加法导数均为零。因此

$$\sum_{i=1}^{p} \sum_{\omega \in \mathbb{F}_2^3} \Gamma_c(f(x_{i,\omega})) = 0.$$

这个和式中有 $8p$ 个输出标签，而（3.48）意味着它们的二元张成空间维数至多为 $4p$。

步骤 *4*：提取许多真关系，并在各块上独立应用四角界。每次揭示一个块的 p 坐标立方体。若 d_i 是加入块 i 的八个标签时输出标签张成空间的二元维数增量，那么

$$0 \leqslant d_i \leqslant 8, \qquad \sum_{i=1}^{p} d_i \leqslant 4p.$$

至多有 $4p/7$ 个块具有 $d_i \geqslant 7$。因此至少有 $3p/7$ 个块具有 $d_i \leqslant 6$。对于每个这样的块，将八个标签模去所有更早标签的张成空间后，所得二进制关系空间的维数至少为二。我们可以选取一个支撑在非空真子集$I_i \subsetneq \mathbb{F}_2^3$上的非零关系：二维关系空间不可能只由零向量和全一向量组成。对于某些固定的二进制系数 $b_{i,\,h,\,\omega}$，该关系为

$$\sum_{\omega \in I_i} f(x_{i,\omega}) = \sum_{h<i} \sum_{\omega \in \mathbb{F}_2^3} b_{i,h,\omega} f(x_{h,\omega}). \tag{3.53}$$

所选块的集合、子集 I_i以及（3.53）中的系数向量，至多有 $2^{O(p^2)}$ 种可能的联合规约。该估计与 N 无关，因为所有系数均为二进制，且只出现 $8p$ 个标签。

固定一个包含至少 $3p/7$ 个所选块的规约。中的坐标立方体 L^p 是独立均匀分布的三维立方体。当我们依次揭示它们时，在每个所选块处，（3.53）右侧已经是 L的一个固定值。因此，（3.49）给出了其条件成功概率的界，包括属于当前八个顶点中的 A，该概率不超过 $\varepsilon_f^{1/4}$。反复应用条件化界，可得所有指定关系同时成立的概率不超过 $\varepsilon_f^{3p/28}$。（3.52）中的每个立方体都至少允许一个这样的规约。现在应用并集界可得

$$2^{-8(pa+r)} \leqslant 2^{O(p^2)} \varepsilon_f^{3p/28}. \tag{3.54}$$

取对数（或者在不等式不成立时单独处理 $\varepsilon_f = 0$）可得

$$\varepsilon_f \geqslant 2^{-O(a+r/p+p)}.$$

最后应用受限同态测试引理 3.40。该引理的常数是绝对常数，因此可得一个仿射映射 $\ell : L \to L$，它与 f 在 A的至少 $2^{N-O(a+r/p+p)}$ 个点上相等，证毕。 □

3.3.2 非对称标记启动器

引理 3.42（缩放多项式求值）。对于 $K = \mathbb{F}_{2^m}$，令 $\Phi_J(u, v) = v(1, u, \ldots, u^{J-1}) \in K^J$。若 $D \subseteq K^2$ 的密度为 2^{-d} 且满足 $(J-1)2^{-m} \leqslant 2^{-d-1}$，则 $\Phi_J(D)$ 的二元仿射张成的余维至多为 $d+1$。

证明。令 $q = |K| = 2^m$，并设 $\chi(z) = (-1)^{\mathrm{Tr}_{K/\mathbb{F}_2(z)}}$。$K^J$ 上的每个二进制线性泛函都具有如下形式

$$\Lambda(z_0, \ldots, z_{J-1}) = \mathrm{Tr}_{K/\mathbb{F}_2}\left(\sum_{i=0}^{J-1} a_i z_i \right)$$

根据迹配对的非退化性，存在唯一的向量（a_0、.....、a_{J-1}）$\in K^J$。如果 $\Lambda \neq 0$，则多项式 $P(T) = \sum_{i=0}^{J-1} a_i T^i$ 非零，且次数至多为 $J-1$。因此，对于独立

均匀分布的 $U,\ V\in K$，

$$\begin{aligned}\mathbb{E}_{U,V}(-1)^{\Lambda(\Phi_J(U,V))} &= \mathbb{E}_U\mathbb{E}_V\chi(VP(U))\\ &= \Pr_U[P(U)=0]\\ &\leqslant (J-1)q^{-1} =: \beta.\end{aligned}$$

中间的等式是精确的：若 $P(u)\neq 0$，则映射 $v\mapsto \mathrm{Tr}^{K/\mathbb{F}_2}(vP(u))$ 是一个非零二元线性泛函，其特征平均值为零；若 $P(u)=0$，则平均值为一。

令 $H=\mathrm{aff}^{\mathbb{F}_2}(\Phi_J(D))$，并将其在 K^J 中的二元余维数记为 b。若 $H=z_*+W$ and$W^\perp\leqslant (K^J)^*$ 是其方向空间的湮灭子，由特征正交性可得

$$\mathbf{1}_H(z)=2^{-b}\sum_{\Lambda\in W^\perp}(-1)^{\Lambda(z-z_*)}.$$

对 $z=\Phi_J(U,\ V)$ 取期望，将零特征上界为一，并将其余 2^b-1个特征上界为 β，可得

$$\Pr[\Phi_J(U,V)\in H]\leqslant 2^{-b}+(1-2^{-b})\beta\leqslant 2^{-b}+\beta.$$

但 $\Phi_J(D)\subseteq H$，因此左侧的概率至少为 $|D|/|K|^2=2^{-d}$。在假设$\beta\leqslant 2^{-d-1}$下，我们有

$$2^{-b}\geqslant 2^{-d}-\beta\geqslant 2^{-d-1}.$$

因此 $b\leqslant d+1$。请注意，结论针对的是二元仿射 张成空间；其差空间具有相同的维数，我们也会将此界用于差以及整个张成空间。 □

偏差与分离源强。 如果某个特征的诚实系数非零，则称该特征含诚实系数。如果每个此类特征的绝对偏差至多为 ϵ 对于阈值为k_0的所有独立源，则在阈值为 k_0+d 时，给定任一源后的平均绝对条件偏差至多为$\epsilon+2^{1-d}$。按其条件偏差的符号对该源进行划分。每个质量至少为 2^{-d}的类都保留该阈值和独立性；其余各类的总质量至多为 2^{1-d}。对于 q个、每个含 m 比特的输出，条件傅里叶反演和三角不等式将任一强误差界定为 $2^{qm-1}\,(\epsilon+2^{1-d})$：不含诚实系数的系数会与均匀分布乘以实际比较边缘分布相抵消。我们使用推论 2.20 的线性秩近似对偶提升，并固定线性熵裕量，从而由每个低阈值矩形上的非恒定性得到初始偏差。

兼容的 M比特行。 下文中，我们使用这一术语表示对 长度的算术限制，而不是对两行的分布施加限制，也不是对篡改映射作出保证。对于起始提取器，取引理 A_* 中的绝对整数3.39，以及固定的输出缩短因子 $D_1=12A_*t_*$，其中 t_* 是在证明中选定的固定正整数。起始长度 M 若可表示为兼容的，则称其为

$$M=6n_0,\qquad n_0=A_*s,\qquad 2t_*\mid s,$$

其中 s 足够大，以适用于固定常数 $c,\ J$，并且塔式表示是固定的

$$K=\mathbb{F}_{2^{s/2}}\subset F=\mathbb{F}_{2^s}\subset L=\mathbb{F}_{2^{n_0}}.$$

因此，右侧的行可解析为两个 L 字和 $J=8A_*$ K 字，因为 $2n_0+J(s/2)=6n_0=M$，而最终的迹可以落入次数为 $s/(2t_*)=M/D_1$的子域中。每个输入比特都有一个

在这些固定基中的指定坐标。“相容”一词并不暗含任何随机选择或额外的熵条件。可接受的起始长度包含固定整数 $12A_*t_*$的所有充分大的倍数。

引理 3.43（带标签起始提取器）。 对于每个如上所定义且充分大的兼容行长度 M，都存在一个固定标签提取器，可抵御一次带有不同标签的任意独立篡改，并且在左侧源上是强的，其熵为 $(1-\tau_1)M$，其中 $\tau_1 > 0$为绝对常数。其输出长度为 $\Omega(M)$，误差为 $2^{-\Omega(M)}$。此外，对于下文实际使用的辅助信息，它在兼容子域上允许常数域分离。

构造 3.6 非对称带标签起始提取器

组成材料 f　来自前文。 引理 3.39 中的标签仿射种子和域上的多项式求值 .

公开参数和固定设置。 设 $n_0 = A_*s$、 $K = \mathbb{F}_{2^{s/2}} \subset F = \mathbb{F}_{2^s} \subset L = \mathbb{F}_{2^{n_0}}$、 $J = 8A_*$和 $M = 6n_0$。使用证明中选定的固定宽度 c 。

输入。两个任意的任意的 M 比特行（具有兼容的公共长度）和固定的数值标签 $a \in F$.

求值步骤。

1. 将完整的左、右输入行分别记为 z， $w \in \mathbb{F}_2^M$。取 z 的前 n_0 比特作为前缀 $_{n_0}(z) \in L$，并将完整的右行解析为 $w = (r,\ t,\ Y) \in L^2 \times K^J$。记 $S_a(x) = (u,\ v) \in K^2$。

2. 对于系数向量 D，令 $P_D(T) = \sum_{i \geqslant 0} D_i T^i$ 为其从零开始的带索引坐标。计算

$$E_a = \mathrm{Tr}_{L/K}\big(tP_{\Gamma_c(x)}(r)\big) + vP_Y(u).$$

输出。 将 E_a 的相对迹返回到下文证明中指定的适当输出子域。输出域字的长度是 M 的固定正比例。

证明。 我们通过排除亏损足够小的每个矩形上存在常数混合特征，来证明固定标签命题。然后，误差和单独的左源保证可由秩到偏差定理及熵余量推出。在整个证明中，标签 $a \neq a'$ 均为固定的域元素；后续的外层构造负责固定其当前依赖输入的标签。

输入矩形和两个术语。固定两侧大小至少为 $2^{M-\alpha M}$ 的支撑集，其中$\alpha > 0$ 将在其他常数选定后确定。令 $z, w \in \mathbb{F}_2^M$表示完整的左右两行，并令 $T_L, T_R : \mathbb{F}_2^M \to \mathbb{F}_2^M$表示任意的独立比较映射。起始器仅使用 z的前 n_0 比特；将未使用的后缀固定为出现频率最高的值 s_0 ，并记为 $z = (x, s_0)$。由此得到的前缀支撑集 $A \subseteq L$ 满足 $|A| \geqslant 2^{n_0 - a_0}$ ，其中 $a_0 \leqslant \alpha M$。接着，比较中的左侧短前缀是确定性的、可能为非单射的函数

$$f(x) := \mathrm{prefix}_{n_0}(T_L(x, s_0)), \qquad x \in A.$$

篡改 T_L 是任意的，不必保留前缀。令 B 表示完整右行 w的支撑集，且 $|B| \geqslant 2^{6n_0 - b_0}$ 满足 $b_0 \leqslant \alpha M$。解析 $w = (r(w)\ t(w)\ Y(w))$及其被篡改后的值 $T_R(w) = (r'(w)\ t'(w)\ Y'(w))$； T_R 仍完全任意。因此，固定 s_0之后，受限乘积矩形由$(x, w) \in A \times B$参数化。从现在起， x 始终表示左侧短前缀，而 w 表示完整右行。

令 $m_K = s/2$，使得 $K = \mathbb{F}_{2^{m_K}}$ 以及 $Jm_K = 4n_0$。考虑两个未截断起始器输出的绝对迹特征，其中非零诚实系数为 $\lambda \in K^\times$ ，任意比较系数为 $\mu \in K$：

$$\chi_{s_0}(x, w) = \mathrm{Tr}_{K/\mathbb{F}_2}\big(\lambda E_a(x; r(w), t(w), Y(w)) + \mu E_{a'}(f(x); r'(w), t'(w), Y'(w))\big)\,.$$

假设它在 $A\times B$上为常数。这里的 E_a 是构造3.6中的未截断前缀级公式，而不是一个左参数为整个 M比特行的函数。对于诚实项，简写为

$$E_a(x;w)=G(x;w)+T_a(x;w),\qquad G(x;w)=\mathrm{Tr}_{L/K}(t(w)P_{\Gamma_c(x)}(r(w))),\qquad T_a(x;w)=vP_{Y(w)}(u),$$

带标签项是 $\Phi_J(u,\ v)$ 的 J坐标与 Y 的 J系数的配对。对两侧分别进行任意前复合后，其二元特征矩阵的秩至多为 $Jm_K=4n_0$。因此，两个带标签项之和（包括可能存在的常数矩阵）的秩至多为

$$k_G:=2Jm_K+1=8n_0+1. \tag{3.55}$$

完整混合特征为常数，意味着相应的混合 Gold 项限制在$A\times B$上时，其矩阵秩至多为 k_G。这里使用的是通常意义下的二元矩阵秩：对于函数 $h:\ A\times B\to\mathbb{F}_2$，其矩阵为 $(h(x,\ y))_{x\in A,\ y\in B}$；左侧函数与右侧函数的乘积贡献秩一。

两个 *Gold* 特征族的稠密张成空间。令 $D_G=(c+1)n_0$，为 L^{c+1} 的二元维数。对于 $\Gamma_c(Z)$ 的非零二元特征 Λ，其中 Z 在 L 上均匀分布，将其布尔相位记为

$$q(z)=\Lambda(\Gamma_c(z))=\mathrm{Tr}_{L/\mathbb{F}_2}\left(\gamma_0 z+\sum_{i=1}^{c}\gamma_i z^{1+2^i}\right),\qquad \gamma_i\in L.$$

若 $\gamma_1=\cdots=\gamma_c=0$，那么 q是一个非零二元线性函数，其特征偏差为零。否则，令 $j\leqslant c$ 为满足 $\gamma_j\neq 0$的最大索引。下面说明为何 q 的极化根基很小。相应的极化形式为

$$\begin{aligned}B_q(z,h)&:=q(z+h)+q(z)+q(h)\\&=\mathrm{Tr}_{L/\mathbb{F}_2}\left(\sum_{i=1}^{j}\gamma_i\left(zh^{2^i}+hz^{2^i}\right)\right).\end{aligned}$$

事实上，在特征为二的情况下，$(z+h)^{2^i}=z^{2^i}+h^{2^i}$，因此展开 $(z+h)^{1+2^i}$ 后，恰好只剩下所展示的两个交叉项；线性项 $\gamma_0 z$ 消失了。绝对迹具有弗罗贝尼乌斯不变性。因此，对任意 $w,\ z\in L$,

$$\mathrm{Tr}_{L/\mathbb{F}_2}(wz^{2^i})=\mathrm{Tr}_{L/\mathbb{F}_2}(w^{2^{-i}}z),\qquad w^{2^{-i}}:=w^{2^{n_0-i}},$$

其中，指数 2^{-i} 表示逆向的 i次弗罗贝尼乌斯自同构，而非有理指数。将此恒等式应用于 $w=\gamma_i h$，可将极化形式改写为

$$B_q(z,h)=\mathrm{Tr}_{L/\mathbb{F}_2}(zR(h)),\qquad R(h):=\sum_{i=1}^{j}\left(\gamma_i h^{2^i}+(\gamma_i h)^{2^{-i}}\right).$$

根据 L上的二元迹配对的非退化性，极化根基恰为

$$\mathrm{Rad}(q):=\{h\in L:B_q(z,h)=0\text{ for every }z\in L\}=\{h\in L:R(h)=0\}.$$

其中出现的所有弗罗贝尼乌斯映射都是 $\mathbb{F}_2$-线性的，因此 Rad(q) 是一个二元子空间。为界定其大小，将方程 $R(h)=0$ 两边提升到 2^j 次幂。由于弗罗贝尼乌斯自同构是双射，这等价于

$$0=R(h)^{2^j}=\sum_{i=1}^{j}\left(\gamma_i^{2^j}h^{2^{i+j}}+(\gamma_i h)^{2^{j-i}}\right).$$

这里， $n_0 > 2c \geqslant 2j$ 确保所有指数 $i+j$ 都严格小于 n_0，因此不会发生弗罗贝尼乌斯幂回绕。第一项求和中含有 $i=j$ 的项为 $\gamma_j^{2^j} h^{2^{2j}}$，其系数非零，且所有其他单项式的次数都严格更小。因此，这是关于 h 的一个非零普通多项式，其次数恰为 2^{2j}，且它在 L 中至多有 2^{2j} 个根。因此

$$|\operatorname{Rad}(q)| \leqslant 2^{2j}.$$

最后，令 $\beta = \mathbb{E}_{z\in L}(-1)^{q(z)}$。将偏差平方，并作变量替换 $(z, z') = (z, z+h)$，然后先对 z 取平均，得到

$$\begin{aligned}\beta^2 &= \mathbb{E}_{z,h\in L}(-1)^{q(z)+q(z+h)} \\ &= \mathbb{E}_{h\in L}(-1)^{q(h)}\mathbb{E}_{z\in L}(-1)^{B_q(z,h)} \\ &= \mathbb{E}_{h\in L}\big[\mathbf{1}_{\operatorname{Rad}(q)}(h)(-1)^{q(h)}\big] \leqslant \frac{|\operatorname{Rad}(q)|}{|L|} \leqslant 2^{2j-n_0}.\end{aligned}$$

倒数第二个等式再次用到了迹配对的非退化性：对 z 求 $(-1)^{B_{q(z,h)}}$ 的平均值，当且仅当 $R(h) = 0$ 时为 1，否则为 0。开平方得到

$$\left|\mathbb{E}_{Z\in L}(-1)^{\Lambda(\Gamma_c(Z))}\right| \leqslant 2^{j-n_0/2} \leqslant 2^{c-n_0/2} \tag{3.56}$$

对每个非零坐标特征 Λ。固定 c 并取足够大的 n_0，最后一个界至多为 2^{-a_0-1}。若 $\Gamma_c(A)$ 的二元仿射张成的余维为 u，则像引理 3.42 中那样，将该仿射包的示性函数按其湮灭子特征展开，得到

$$2^{-a_0} \leqslant 2^{-u} + 2^{c-n_0/2}.$$

因此 $u \leqslant a_0 + 1$。诚实 Gold 特征的坐标函数限制在 A 上后，其二元张成空间的维数至少为 $D_G - a_0 - 1$。

在右侧， B 到 $(r,t) \in L^2$ 的投影至少有 $2^{2n_0-b_0}$ 个元素：每对 (r,t) 至多有 $|K|^J = 2^{4n_0}$ 个原像。因此，该投影在 L^2 中的密度至少为 2^{-b_0} 。在3.42 中，针对 L 应用宽度为 $c+1$ 的引理。对于固定的 c 和充分大的 n_0，其根界假设成立。由此可得， $b_0 + 1$ 的仿射张成余维数至多为

$$\Phi_{c+1}(r,t) = t(1, r, \ldots, r^c) \in L^{c+1}.$$

如果比较 *Gold* 系数非零，则可识别左侧篡改。假设 $\mu \neq 0$。令 U 为二元特征矩阵，其行索引为 $x \in A$ ，并将 $\Gamma_c(x)$ 和 $\Gamma_c(f(x))$ 的坐标向量串接而成。类似地，令 V 将 B 中每个右输入对应的特征向量 $\Phi_{c+1}(r,t)$ 和 $\Phi_{c+1}(r',t')$ 串接而成，其中 (r',t') 是被篡改右输入的前两个 L-字。混合 Gold 特征矩阵为

$$\mathsf{U}\,\mathsf{P}\,\mathsf{V}^{\mathsf{T}},$$

其中，P 是定义在维数为 $2D_G$ 的空间上的非退化分块对角二元迹配对；其两个块均非退化，因为 λ, $\mu \neq 0$。特别地，P 可逆。

U 的诚实块的秩至少为 $D_G - a_0 - 1$。 $\Gamma_c(x)$ 的每个二进制坐标都是 x 的二次布尔函数，而 $\Gamma_c(f(x))$ 的坐标在模去 A 上所有二次函数后的张成空间维数为 $d_c(f;\, A)$。因此，加入篡改块后，秩至少增加这一缺陷量：

$$\operatorname{rank}\mathsf{U} \geqslant D_G - a_0 - 1 + d_c(f; A).$$

仅诚实块就给出秩 $\mathsf{V} \geqslant D_G - b_0 - 1$。将西尔维斯特秩不等式应用于上面的乘积，可得

$$\begin{aligned} k_G &\geqslant \operatorname{rank}(\mathsf{U}\mathsf{P}\mathsf{V}^{\mathsf{T}}) \\ &\geqslant \operatorname{rank}\mathsf{U} + \operatorname{rank}\mathsf{V} - 2D_G \\ &\geqslant d_c(f;A) - a_0 - b_0 - 2. \end{aligned}$$

因此

$$d_c(f;A) \leqslant k_G + a_0 + b_0 + 2. \tag{3.57}$$

根据定理 3.41，存在一个子集 $A_1 \subseteq A$，在该子集上 f 与一个仿射映射一致 $\ell\colon\ L \to L$ 且其环境亏损满足

$$a_1 := n_0 - \log_2 |A_1| \leqslant C_{\mathrm{rec}} \left(a_0 + \frac{k_G + a_0 + b_0 + 2}{p} + p + 1 \right) \tag{3.58}$$

对于某个绝对 C_{rec}。选取常数 p，使其足够大于 A_* 和 C_{rec}，然后令 $c = 8p$；只有在完成这些选择之后，才取 $\alpha > 0$ 足够小。由于 $k_G = 8n_0 + 1$ 和 $s = n_0/A_*$，(3.58) 的右侧对于所有充分大的相容长度至多为 $s/8$。这一选择还保证了 $n_0 > 2c$以及上面的张成空间估计。

当 $\mu = 0$ 时，比较 Gold 块不存在，因此无需进行识别。我们可以取$A_1 = A$，并在下述论证中使用对 S_a 的普通仿射提取；此情形保留的右侧熵至少与 $\mu \neq 0$ 情形一样多。

固定一个比较种子，同时不破坏诚实种子张成空间。 首先取 $\mu \neq 0$。在 A_1 上，我们有$f = \ell$，其中 ℓ 是仿射的。对于不同标签 a、a'，标签仿射种子引理在环境均匀分布 $Z \in L$ 上给出

$$(S_a(Z), S_{a'}(\ell(Z))) \approx_{2^{-2s}} (U_{K^2}, S_{a'}(\ell(Z))).$$

令 $\beta = |A_1|/|L| = 2^{-a_1}$。对于每个可能的比较种子值 z'，令 $T_{z'}$ 为对于满足 $S_{a'}(\ell(x)) = z'$ 的 $x \in A_1$，所出现的诚实种子值 $S_a(x)$ 的集合。那么

$$\beta \leqslant \Pr\left[S_a(Z) \in T_{S_{a'}(\ell(Z))}\right] \leqslant \max_{z'} \frac{|T_{z'}|}{2^s} + 2^{-2s}.$$

由于 $a_1 \leqslant s/8$，某个 z' 满足

$$|T_{z'}| \geqslant 2^{s - a_1 - 1} \tag{3.59}$$

当 s 足够大时。将左侧源限制为相应的非空纤维，并在必要时，为 $T_{z'}$中的每个诚实种子选择一个左侧输入。此时比较种子为常数，但诚实种子仍然在 K^2 内实现了至少为 2^{-a_1-1} 的密度。现在我们使用下面的估计来导出矛盾。对于 $\mu = 0$，无需固定比较种子。普通仿射提取给出 $S_a(U_L) \approx U_{K^2}$；将其应用于集合 $A_1 = A$，同样的计数论证可得到至少 2^{s-a_0-1} 个诚实种子值。通过使用不含篡改项的 $a_1 = a_0$，下面的估计同样适用于这种情况。

固定右侧的 *Gold* 系数并得到矛盾。 在右侧限制到诚实对 $(r,\ t) \in L^2$ 和篡改对 $(r',\ t') \in L^2$ 的最大纤维（如果 $\mu = 0$，则省略后者）。这四个域字至多泄露 $4n_0$ 比特，因此剩余的系数向量集合 $\mathcal{Y} \subseteq K^J$ 的大小至少为

$$|\mathcal{Y}| \geqslant 2^{6n_0 - b_0 - 4n_0} = 2^{2n_0 - b_0}. \tag{3.60}$$

这是对不同 Y 系数向量数量的界：对于 Y 的每个选择以及固定的 $(r,\ t)$，只有一个诚实右输入，因此投影在该纤维上是单射的。在所选矩形中，两个 Gold项仅依赖左侧输入，因为它们的系数对已被固定。篡改的带标签项仅依赖右输入，因为比较种子已固定。只有诚实的带标签项仍然真正同时依赖左右两侧。

为使抵消过程更加严谨，令$\mathcal{X}$ 表示所选比较种子纤维中保留的短左前缀集合，令$\mathcal{B}$ 表示所选系数纤维中保留的完整右侧行集合。在乘积$\mathcal{X}\ \times\mathcal{B}$上，根据前述固定条件，我们可以将所谓的常值混合布尔特征写成以下形式

$$\chi_{s_0}(x,w)=L_{\mathrm{loc}}(x)+R_{\mathrm{loc}}(w)+\mathrm{Tr}_{K/\mathbb{F}_2}(\lambda\,\Phi_J(u(x),v(x))\cdot Y(w))=c_\chi\in\mathbb{F}_2. \tag{3.61}$$

这里，$(u(x)\ v(x))\ =S_a(x),\ \ Y(w)$ 是完整右侧行 w 的诚实 K^J 系数向量，而$\cdot$ 是 K^J 上的普通 K-内积。函数 L_{loc} 包含两个 Gold 贡献项（其诚实和篡改的右侧系数对均已固定）； R_{loc} 包含篡改的带标签贡献项（其比较种子已固定）。如果 $\mu=0$，相应的比较项就不存在。因此，该分解无需对右侧篡改映射作任何假设。

任取 $x_1,\ x_2\in\mathcal{X}$和 $w_1,\ w_2\in\mathcal{B}$，并记

$$\begin{gathered}(u_i,v_i)=S_a(x_i),\qquad \Phi_i=\Phi_J(u_i,v_i),\qquad Y_j=Y(w_j),\\ L_i=L_{\mathrm{loc}}(x_i),\qquad R_j=R_{\mathrm{loc}}(w_j),\qquad H_{ij}=\mathrm{Tr}_{K/\mathbb{F}_2}(\lambda\Phi_i\cdot Y_j).\end{gathered}$$

χ_{s_0} 在整个乘积矩形上的常值性给出

$$\begin{aligned}&L_1+R_1+H_{11}=c_\chi,\quad L_1+R_2+H_{12}=c_\chi,\\ &L_2+R_1+H_{21}=c_\chi,\quad L_2+R_2+H_{22}=c_\chi.\end{aligned} \tag{3.62}$$

将这四个等式在 $\mathbb{F}_2$中相加。每个 L_i 均出现两次，一次与 w_1 一起出现，另一次与 w_2一起出现，因此仅在左侧的贡献相互抵消。同样，每个 R_j 也出现两次，一次与 x_1 一起出现，另一次与 x_2一起出现，并相互抵消。右侧的常数贡献为 $4c_\chi=0$。因此，

$$0=H_{11}+H_{12}+H_{21}+H_{22}=\mathrm{Tr}_{K/\mathbb{F}_2}(\lambda(\Phi_1\cdot Y_1+\Phi_1\cdot Y_2+\Phi_2\cdot Y_1+\Phi_2\cdot Y_2))\,.$$

由 K-内积的双线性性，四个混合项可以分解为，

$$\mathrm{Tr}_{K/\mathbb{F}_2}\left(\lambda\left[\Phi_J(u_1,v_1)+\Phi_J(u_2,v_2)\right]\cdot(Y_1+Y_2)\right)=0. \tag{3.63}$$

由于两个左侧前缀和两行右侧行向量是任意的，这表明每个诚实的缩放求值特征之差与每个保留的右侧系数向量之差之间都满足正交关系。诚实特征向量之差的二元张成空间的维数至少为 $Jm_K-a_1-2=4n_0-a_1-2$，这是由引理 3.42 应用于（3.59）中的种子值集合所得；乘以 $\lambda\neq0$ 保持该维数不变。另一方面，$\mathcal{Y}$中向量之差的二元张成空间的维数至少为$\log_2|\mathcal{Y}|\geqslant2n_0-b_0$。因此，在 K^J 上的非退化迹配对下，其湮灭子的维数至多为 $4n_0-\ (2n_0-b_0)\ =2n_0+b_0$。等式（3.63）将迫使

$$4n_0-a_1-2\leqslant2n_0+b_0,$$

这与 $a_1\leqslant s/8$、 $b_0\leqslant\alpha M$ 以及一个足够小的固定 α 矛盾。因此，没有任何含诚实成分的混合特征会在所声称的矩形上恒定。

指数误差、左源强度与前缀复用。 对于固定的数值标签，每个 Gold项都是 $(c+1)n_0$ 个分离二进制乘积之和，而每个带标签的项都是 Jm_K 个此类乘积之和。因此，每个混合特征的全局二元秩均为 $O(M)$， 其中常数仅取决于固定的起始器参数。将推论 2.20 的线性秩近似对偶提升应用于刚证明的非恒定性，并将 αM 的固定正比例作为熵余量。由此可得特征偏差 $2^{-\Omega(M)}$ ，且最小熵阈值略高。按照本引理之前的描述，依据条件特征偏差的符号对含诚实成分的左侧源进行划分，使得在保留整个源的情况下，平均意义下的界仍然成立；较小的符号类带来的代价至多为指数尾部。

最后，应用相对迹 $\mathrm{Tr}^{K/K'}$ 到子域 $K' = \mathbb{F}_{2^{M/D_1}} \subseteq K$，并选择足够大的固定除数$D_1 = 12A_*t_*$ 。K' 的每个非零特征都拉回为 K的非零特征，而较短的输出长度使傅里叶反演中的分解无关紧要。所得结果的输出长度为 $M/D_1 = \Omega(M)$，误差为 $2^{-\Omega(M)}$；当熵为 $(1-\tau_1)M$ 时，左源强度满足要求，其中 $\tau_1 > 0$ 为固定值（固定松弛常数后，可以取 $\tau_1 = \alpha/4$ ）。

同一基本方法也可应用于较长源词的前缀。为清楚起见，设 Z 为诚实左前缀， X 为完整左词。独立于 Z，抽取一个表 ，将每个前缀值 z 映射到一个从条件分布 $X \mid Z = z$中抽取的完整词。在抽取的 Z 处查询该表，即可重现 X的原始分布。对于每个固定的表，篡改前缀都是诚实前缀的确定性函数；两个原始侧仍各自拥有独立的表。针对前缀的左强保证会保留 Z 及其表，因此也保留由其生成的完整左词。对这些表取平均，即可证明较长输入在任意独立篡改下所需的结论。每个递归层级都会用到这一补全论证。

□

3.3.3 将比较次数加倍

普通强哈希。 对于输出长度 v，将源词解析为 $P \,\mathrm{over} K = \mathbb{F}_{2^{4v}}$ 的系数（用零填充），并将 $8v$ 比特种子解析为 $(u,\ t) \in K^2$。使用 $H_v = \mathrm{Tr}^{K/\mathbb{F}_{2^v}}(tP(u))$。对于任意两个长度固定且不同的字，在均匀种子下，碰撞概率至多为 $2^{-v} + d_{\max}2^{-4v}$，这是由根界得到的，其中 $d_{\max}$ 界定了所用次数。对于常数级的字长与域长比率，次数界为常数。对给定种子时输出的平均碰撞概率应用柯西–施瓦茨不等式，可知在最小熵为 k 时，强误差至多为 $\frac{1}{2}\sqrt{2^v(2^{-k} + O(2^{-4v}))}$。特别地， $k \geqslant 3v$ 足以达到误差 $O(2^{-v})$。

兼容的 N 比特行。 同样，在下文中，我们用这个术语表示对长度的算术限制。具体而言，一旦固定 C，包装器至多有 $h = \max\{1,\ \lceil \log_2 C \rceil\}$ 个层级。包装器某一层的长度 N 在兼容的 情况下满足

$$p = \frac{N}{8q} \in \mathbb{N}, \qquad w = \mathrm{outlen}(\mathcal{E}_{\mathrm{prev,p}}), \qquad v = \frac{w}{256q} \in \mathbb{N}, \quad q = C + 1,$$

其 p 比特前缀在前一层兼容，且当前层的固定塔和输出域具有指定的整数次数。特别是， $w = 256qv$ 使分拆

$$\mathsf{R} \in \mathbb{F}_{2^{w/2}}^2, \qquad \mathsf{L} \in \mathbb{F}_{2^{4v}}^2, \qquad \mathbb{F}_{2^{4v}} \subseteq \mathbb{F}_{2^{w/2}}$$

在最终层级有效；中间多项式哈希使用 $\mathbb{F}_{2^{4v}}$，而生成 L 的哈希使用 $\mathbb{F}_{2^{32v}}$ 和一个 $64v$ 比特的种子。一旦 C 固定，所有比率和层级数也就固定，因此我们可以为输入长度选定一个固定的整除格，并预先选好域嵌入。下文的整数 J 是在 此格确定后选取的，不会改变该格。

对于外层提取器，我们不假设任意的浓缩器行长度已经属于此格。相反，在确定其整数常数后，我们取长度为$N = A_C H j \ell \leqslant M$的兼容前缀，其中 $\ell = \lfloor M/(A_C H j) \rfloor$。因此 $0 \leqslant M - N < A_C H j = O_{\delta,\ C}(1)$，丢弃这有限数量的比特只会造成熵损失 $O_{\delta,\ C}(1)$。同样的选择也确保了外层构造中使用的嵌套域 $\mathbb{F}_{2^\ell} \subset \mathbb{F}_{2^{j\ell}} \subset \mathbb{F}_{2^{Hj\ell}}$。任何额外的固定标签域嵌入都作为这一公开的常数整除设置的一部分来安排。“兼容”绝不意味着输入数据具有某种特殊的代数形式。

定理 3.44（具有可调混合秩的带标签行）。 固定 $C \geqslant 1$。对于如上所定义的所有充分大的相容长度 N，存在一个亏损 $\tau_C > 0$ 和一个本征输出次数N/A_C，二者的比率均在选择裕量 $K_* > 0$ 之前固定。对于每个固定裕量，下面的求值器具有如下性质：在任意边长至少为 $2^{(1-\tau_C)N}$ 的矩形上，只要指定的固定标签与每个比较标签不同，该行的每个含诚实成分的混合特征以及至多 C 个独立比较都具有二元秩 $> K_* N$。比较标签可以相同。每个特征都具有全局秩 $O_{C,\ K_*}(N)$，并且在相容域上具有常数域分离，包括在下面实际进行指纹代入之后。

对于倍增递归，将前一层的输出记为 R，将新的左哈希字记为 L。这些是中间输出，而不是输入子集 $A,\ B$。它们的名称反映了下文证明中转录固定之后所确立的局部性。

固定数量比较的构造 3.7 标记行

前文中的构件。 构造 3.6， H_v 和引理 3.42。

公开参数和固定设置。 令 $q = C+1$, $h = \max\{1, \lceil \log_2 C \rceil\}$。从左强单比较起始构造开始。使用以下递归进行 h 个层级，仅在第 h 层使用最后一个公式。所有缩减比率仅取决于 C。在输入长度为 N 的层级，前一层的调用长度为 $p = N/(8q)$，并给出一个长度为 w 的字。令 $v = w/(256q)$。通过取相容长度来保证整除性。

输入。 两行 $x,\ y$ 以及起始域中的一个数值标签 a。

求值步骤。

1. 通过在两个 p 比特前缀上使用标签 a 运行前一层求值器，计算 R。令 $P = \mathrm{prefix}_{64v}(\mathrm{R})$ 和 $\mathrm{L} = H_{8v}(x;\ P)$。**2.** 在中间层，返回 $H_v(\mathrm{R};\ \mathrm{L})$。在最终层，将 $\mathrm{L} = (u,\ b) \in K^2$、$\mathrm{R} = (r,\ t) \in L^2$ 拆分，其中 $K = \mathbb{F}_{2^{4v}} \subseteq L = \mathbb{F}_{2^{w/2}}$。返回

$$\mathrm{Tr}_{L/K}\left(bt(1+ur)^{J-1}\right).$$

选取一个 2 的幂 J，使得 $4Jv - w/3 - 5 > K_* N$。

输出。 $4v = N/A_C$ 比特的本原域字。只有 J 取决于所要求的裕量。

证明。 我们区分在所有中间层使用的统计归纳 与仅在最终层级使用的受限秩论证。前者独立于 J；特别是，在递归过程中不会引入依赖于裕量的熵要求。我们使用以下记号：执行 0 是诚实对，而执行 $i \geqslant 1$ 是任意的独立篡改。它们的数值标签固定，且诚实标签与每个比较标签都不同。令 $q = C + 1$。

归纳假设与前缀熵。假设前一层求值器的输出为 w比特，并且在某个固定的近满熵率下，针对至少 t_0 次比较是左源强的，

在 p比特行上的误差为 $2^{-\Omega(p)}$。当前层每一侧有 N 个输入比特，并使用 $p = N/(8q)$ 个前缀比特。熵至少为 $(1-\tau)N$ 的源有一个 p比特的前缀，其最小熵至少为$p - \tau N$，因为丢弃 $N - p$ 比特最多会使最小熵减少 $N - p$。我们将 $\tau > 0$固定得足够小，使得 $\tau N \leqslant \frac{1}{2}\tau_{\mathrm{prev}} p$ 且 $\tau \leqslant 1/16$，其中 $\tau_{\mathrm{prev}} > 0$是前一层允许的亏损率。因此，前一层求值器适用于具有均匀线性熵裕量的前缀分布。我们只使用有限多个层级，因此这些固定正数中最小的一个仍为正数，且不依赖于 J。

将至多 $2t_0$ 次比较划分为 T 和 S，每组的大小至多为 t_0。在当前层计算

$$\mathsf{R}_i = \mathcal{E}_{\mathrm{prev}}(x_i^{\mathrm{pre}}, y_i^{\mathrm{pre}}), \quad P_i = \mathrm{prefix}_{64v}(\mathsf{R}_i), \quad \mathsf{L}_i = H_{8v}(x_i; P_i), \qquad i = 0, 1, \ldots, t.$$

诚实的 R_0 在给定 R_i 以及给定整个诚实左前缀的条件下，接近均匀分布。为了说明也可以保留篡改的 篡改的 左侧前缀，请使用引理 3.43末尾的补全表论证：在给定采样得到的左侧补全表后，每个篡改前缀都是诚实前缀的确定性函数。因此，前述源强度陈述也可以保留它们的完整向量。对补全表取平均后，对于 $Q = T$ 和 $Q = S$，我们都有

$$(\mathsf{R}_0, \mathsf{R}_Q, Z) \approx_{2^{-\Omega(p)}} (U_w, \mathsf{R}_Q, Z), \tag{3.64}$$

其中 Z 是所有左侧前缀（包括诚实和篡改前缀）组成的向量。其总长度至多为 $qp = N/8$。

双侧转录及其局部性。 一旦公开 Z，诚实和篡改的 $\mathbf{R}_i$ 就只依赖于原始右侧源。接下来公开 $\mathbf{P} = (P_0, \ldots\ldots, P_t)$，这是一个至多包含 $64qv = w/4$ 比特的元组。固定 Z 后，该元组具有右侧局部性。现在，在给定 $(Z,\ \mathbf{P})$ 的条件下，$\mathbf{L}_i$ 只依赖于原始左侧源；$\mathbf{R}_i$ 则保持右侧局部性。固定转录值后，所有这些量都是相应原始源的确定性函数。因此，每个被保留的正概率转录叶节点都是定义在两个原始侧上的乘积分布。这正是按此顺序公开信息的原因。

我们将反复使用一条标准的最小熵尾界规则。若随机变量 W 的最小熵至少为 k，且消息 M 至多有 2^b 个取值，那么除消息质量至多为2^{-s} 的情形外，W 的条件最小熵至少为 $k - b - s$：事实上，概率至少为 2^{-b-s} 的消息不可能使任何原子的概率升至 2^{-k+b+s}以上。在任何已固定的乘积转录之后，同样适用这条规则；下文的所有误差估计都是对叶节点取平均，而非声称对每个概率极低的叶节点都成立。

新左局部字 L_0。 固定 Z 和 P_S。根据（3.64），当 $Q = S$，P_0 与独立均匀分布的 $64v$ 比特种子在指数意义上接近，即使给定 $P_S,\ Z$也是如此。在这些值给定的条件下，左侧源与剩余的右局部种子元组相互独立。揭示L_S 至多需要$8qv$ 额外的左侧比特，并且不会限制右侧源，因为在指定 L_S 之后，$P_S,\ Z$ 它是左局部的。

诚实的左侧源初始熵为 $(1-\tau)N$。对于良好转录质量，Z、$\mathbf{L}_S$以及两个 v比特尾部余量的总成本至多为 $qp + 8qv + 2v$。因此，其剩余最小熵至少为

$$(1-\tau)N - qp - 8qv - 2v > 24v \tag{3.65}$$

对于规定的固定比率和足够大的长度。其种子 H_{8v} 为 P_0，恰好包含$64v$ 比特，而该定理之前的普通强哈希计算给出的误差为 $O(2^{-v})$ 在$24v$比特熵阈值处。将 P_0 替换为独立均匀种子，在每个良好的左侧条件分布上应用该强哈希界，然后使用（3.64）恢复真实的种子分布。

重要的是，我们还可以在这次比较中保留<style id='1'>整个向量$\mathbf{P}$。给定（Z，P_S，L_S），对于任何进一步的右局部种子值，条件左侧分布都相同；强哈希误差是关于诚实种子 P_0 取平均的，披露有关该种子的辅助信息不会增加该误差。可以通过以下方式将其形式化：首先将整个右局部元组附加到理想均匀种子分布上，然后在每种情况下附加相同的左侧条件分布。我们得到

$$(\mathsf{L}_0, \mathsf{L}_S, Z, \mathbf{P}) \approx_{2^{-\Omega(v)}} (U_{8v}, \mathsf{L}_S, Z, \mathbf{P}). \tag{3.66}$$

这是比较次数加倍的前半部分。

中的熵足够用于 R_0 的第二次普通哈希。使用 （3.64）和 $Q = T$，将联合分布 （$\mathsf{R}_0, \mathsf{R}_T$，$Z$）替换为具有均匀分布、独立的 R_0的分布。使用给定 （$\mathsf{R}_0, \mathsf{R}_T$，$Z$）的原始条件分布，附加右输入的其余部分，然后给定 Z附加独立的左侧部分。这至多使完整实验改变其先前的提取误差，并保留条件乘积结构。这种平滑视角避免了对 R_T的取值进行条件化，因为在单个叶子中，该取值不一定具有很高的概率。

在平滑实验中，即使给定 $\mathbf{R}_T$，Z，$\mathbf{R}_0$ 一开始也具有 w 比特的熵。公开整个 $\mathbf{P}$ 最多会损失 $w/4$ 个比特。分别在两侧固定 $\mathbf{R}_T$ 和 $\mathbf{L}_S$ 后，T 中的比较输出是左侧局部的（其 $\mathbf{R}_i$ 已被固定），而 S 中的输出是右侧局部的（其 $\mathbf{L}_i$ 已被固定）。后者至多进一步泄露关于 $\mathbf{R}_0$的 qv 个比特。使用 $2v$ 比特的尾部松弛，在良好乘积叶子上，其剩余最小熵满足

$$H_\infty(\mathsf{R}_0 \mid \text{current right-side messages}) \geqslant w - w/4 - qv - 2v > 3v. \tag{3.67}$$

该不等式在平滑后的良好叶意义下成立；异常质量和平滑代价都关于 v 呈指数级小。

根据 (3.66)，诚实种子 L_0 相对于 L_S、Z、$\mathbf{P}$是新鲜的。给定固定的乘积转录，它是左局部的，而剩余部分 R_0 是右局部的。应用强哈希 $H_v(\mathsf{R}_0;\mathsf{L}_0)$，并使用熵下界 (3.67)。保留 T-比较输出；在固定的 R_T 叶上，它们是左局部的；而 S-比较输出在固定 L_S 后是右局部的。种子特定的平均论证还允许我们保留整个原始左侧输入。将平滑后的分布替换为实际分布，并恢复实际的比较边缘分布，只会带来前述指数级小的误差。因此

$$\big(H_v(\mathsf{R}_0;\mathsf{L}_0),\ (H_v(\mathsf{R}_i;\mathsf{L}_i))_{i=1}^t,\ X\big) \approx_{2^{-\Omega(v)}} \big(U_v,\ (H_v(\mathsf{R}_i;\mathsf{L}_i))_{i=1}^t,\ X\big). \tag{3.68}$$

这就完成了对每个中间层的归纳，并使允许的比较次数翻倍。有限层级即可为最终层级所需的所有组提供一个误差呈指数级小的左强统计评估器。

最终层级：关注支撑集，而非统计提取。在最后一层，对均匀分布在任意边长为 $2^{(1-\tau_C)N}$的矩形上运行相同的转录和新鲜性论证。现在我们提取支撑集大小信息。由（3.64）以及 $Q = T$可知，诚实的R_0 在给定（R_T，Z）的条件下接近均匀分布。**消息**$\mathbf{P}$至多取 $2^{w/4}$ 个值。对于固定的（R_T，Z），对$\mathbf{P}$ 的所有取值对应的R_0 中少于 $2^{2w/3}$ 个点的条件支撑集取并集，其大小至多为

$$2^{w/4}2^{2w/3} = 2^{11w/12}.$$

一个均匀分布的 w 比特字落在此并集中的概率至多为 $2^{-w/12}$；统计近似还会增加一个指数级小的误差。因此，除去这部分联合质量很小的情况，

$$|\operatorname{supp}(\mathsf{R}_0 \mid \mathsf{R}_T, Z, \mathbf{P})| \geqslant 2^{2w/3}. \tag{3.69}$$

同理，(3.66) 表明，除去转录质量呈指数级小的情况，诚实方 $8v$ 比特 L_0 在给定 (L_S、 Z、$\mathbf{P}$) 时的支撑集至少包含 2^{8v-1} 个点。事实上，在支撑集较小的叶上，其条件统计距离与均匀分布至少相差 $1/2$；而 (3.66) 中的平均距离呈指数级小。因此

$$|\operatorname{supp}(\mathsf{L}_0 \mid \mathsf{L}_S, Z, \mathbf{P})| \geqslant 2^{8v-1} \tag{3.70}$$

除指数级小的质量外。

选择一个同时良性转录，然后固定比较词 R_T 在右侧以及 L_S在左侧。它们是乘积叶两侧相对的限制，因此不会缩小另一侧的条件支撑集。由此得到一个非空乘积矩形，在其上 (3.69) 和 (3.70) 均成立。 T 中的最终比较项仅取决于左侧变量，而 S 中的最终比较项仅取决于右侧变量。因此，在二元特征矩阵中，它们的总贡献的秩至多为二，无论任一组中发生多少次比较。

诚实几何级数哈希的秩。写作 $\mathsf{L}_0 = (u, b) \in K^2$，其中 $K = \mathbb{F}_{2^{4v}}$，以及 $\mathsf{R}_0 = (r, t) \in L^2$，其中 $L = \mathbb{F}_{2^{w/2}}$。由于 J 是 2 的幂，多项式恒等式

$$(1+z)^{J-1} = \sum_{i=0}^{J-1} z^i$$

在特征为二的情形下成立。因此，对于任意非零特征系数 $\lambda \in K^\times$ ，

$$\operatorname{Tr}_{K/\mathbb{F}_2}\left(\lambda \operatorname{Tr}_{L/K}(bt(1+ur)^{J-1})\right) = \operatorname{Tr}_{L/\mathbb{F}_2}\left(\lambda \sum_{i=0}^{J-1} (bu^i)(tr^i)\right).$$

这是缩放求值向量之间的非退化绝对迹配对

$$\Phi_J(u,b) = b(1, u, \ldots, u^{J-1}) \in K^J, \quad \Phi_J(r,t) = t(1, r, \ldots, r^{J-1}) \in L^J.$$

K^J 与 L^J 的配对在 K^J 侧是非退化的：对于非零的 $z \in K^J$，与 L^J 进行迹配对所得的结果并非恒为零。

L_0 支撑集的密度至少为 $1/2$，在 K^2 中由（3.70）。引理 3.42（亏损为 $d \leqslant 1$）给出仿射张成余维数至多为二，在 K^J 中因此线性张成维数至少为

$$4Jv - 2.$$

同样，R_0 支撑集在 L^2中的密度至少为 $2^{-w/3}$。在 L 上应用同一引理，可得 L^J中的仿射张成余维数至多为 $w/3+1$。当相容长度充分大时，这两次应用都满足该引理的根界假设，因为一旦选定所需的裕量， J 就是一个固定的常数。

将非退化配对限制到维数至少为 $4Jv-2$ 的短边张成空间和余维数至多为 $w/3+1$ 的长边张成空间上，其秩最多会减少后一个余维数。因此，诚实特征矩阵的秩至少为 $4Jv - w/3 - 3$。减去左局部和右局部比较贡献，最多还需损失两个秩，得到

$$\operatorname{rank}(\text{honest-bearing mixed character}) \geqslant 4Jv - \frac{w}{3} - 5 > K_*N, \tag{3.71}$$

其中 J 是在公开设置中选取的。该矩形是在任意一个满足所断言大小的初始矩形内选取的；因此，该初始矩形的秩也大于 K_*N。这证明了可调的受限秩断言。

全局秩、实际标签和域兼容性。对于每个固定的 C 和 J，哈希仅使用固定次数的多项式求值、幂运算、加法、乘法，以及在扩域比率固定的域中的相对迹。因此，一个固定标签的二元特征可以明确表示为有限个乘积之和，其中每个乘积由一个左侧域值函数和一个右侧域值函数组成。在外层构造中，标签并非固定的，而是具有以下形式

$$a(x,y) = \iota(P_x(t), P_y(t)) = a_L(x) + a_R(y),$$

其中，最后一个等式利用了 ι 的线性性。将其代入 $S_a(z) = V_0(z) + aV_1(z)$，可得到分离项之和，尽管函数 V_0，V_1 本身不必关于其源比特是线性的或低次数的。随后进行固定次幂运算和乘积运算，会将分离域项的数量 增加一个取决于 C，J、但与输入长度无关的常数因子。使用迹对偶基展开每个这样的域乘积，代价至多为公共基域的二进制度，即 $O(N)$。因此，每个实际特征都具有全局二元秩 $O_{C,\ \kappa_*}(N)$ 以及一种显式的分离编码。

所需的扩域和迹映射之所以有效，是因为本定理之前规定了长度约定。将所有比特块都表示在公共子域上的固定基中，可确保前缀运算和对子域的限制保持分离表示。有限递归允许将 τ_C 缩小一次，从而使所有熵裕量、平滑误差和域大小假设同时成立。 τ_C 和原生输出长度 $4v = N/A_C$ 都不取决于最终宽度 J。

□

3.3.4 常速率不可篡改性与分离源强

定理 3.45 （固定-C 不可篡改性，具有指数误差）。固定 $C \geqslant 1$ 和 $\delta > 0$，使得$\delta \leqslant 1$。存在一个显式多项式时间映射 $G\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$，满足 $m = \Omega_{\delta,C}(n)$，且 $\epsilon = 2^{-\Omega_{\delta,C}(n)}$，使得对于最小熵为 $k \geqslant \delta n$ 的独立源以及任意可容许的 C 独立篡改，记 $V = (G(f_i(X),\ g_i(Y)))_{i=1}^{C}$，对于同一个映射，以下条件同时成立

$$\Delta((G(X,Y), V, H), (U_m, V, H)) \leqslant \epsilon \qquad \text{for each } H \in \{X, Y\}.$$

此处 U_m 独立于（V，H）。去掉 H 即可得到普通不可篡改性。每个输出特征的二元秩为 $O_{\delta,\ C}(n)$，并具有该长度的显式分离编码。比较像不受熵条件限制。

构造 3.8 常数率固定-C 不可篡改提取器

前文中的构件。定理 2.25中的强浓缩器、定理 3.44中的行求值器，以及引理 3.7中的强制恒等式。

公开参数和固定设置。令 $q = C + 1$。将凝聚器实例化为率为 $\delta/(16q)$、亏损为 $\tau_C/64$ 且平均行误差至多为 $1/(1000q)$。记其行长度为 $M \geqslant c_D n$，二的幂次行数为 r（如有必要，可通过忽略多余的种子比特来重复行）。选取固定整数 $H \geqslant 4$，使得 $4 \mid H$ 且 $q/(A_C H) \leqslant \tau_C/64$，并令 H 为行构造中所有固定兼容性除数的一个充分大的倍数。具体而言，对于每个充分大的整数 d，要求$N = A_C H d$ 为兼容的行长度，行求值器中的每个工作域次数以及前缀或块长度都是 d的整数倍，并且最内层标签域的二进制度至少为 $2d$。特别地， $v = Hd/4$ 是一个整数。这些条件只涉及行构造的固定长度比率，而这些比率取决于 C，与可调列宽无关。因此H 可以在确定这些列宽以及 j之前固定。对于下文选取的固定 j，使用

$$\ell = \lfloor M/(A_C H j) \rfloor, \quad d = j\ell, \quad N = A_C H d, \qquad F_0 = \mathbb{F}_{2^\ell} \subset K_0 = \mathbb{F}_{2^d} \subset F = \mathbb{F}_{2^{Hd}}.$$

行本原域为 F；最内层标签域包含 K_0，并且有足够空间容纳固定的 K_0-线性单射ι： $K_0^2 \to$ 标签域。按照行证明中的做法，使用完整子域块基。对于足够大的 n， $M - N < A_C H j$ 和$d \geqslant M/$（$2A_C H$），因此 $b = \lceil n/d \rceil \leqslant b_* := \lceil 2A_C H/c_D \rceil$。选取一个 2 的幂 $R \geqslant$，使其不小于$\max\{4, r, 8\binom{q}{2}(b_* - 1)\}$，令 $D = R/2$，并为 $1 \leqslant a \leqslant R$取互不相同的 $t_a \in K_0$。使用 i（a） $= 1 +$ （（$a - 1$） $\bmod$ r）。按递增顺序设定 $K_a = 3 + q\sum_{b' < a} C_{b'}$，其中 $C_{b'} N$是每个列b'实际建议字符的秩的上界。根据定理 3.44 选择其宽度。这些界可以在选择 j之前确定：将实际表达式展开为 K_0-坐标。应用特征后，每个由一个左侧域元素和一个右侧域元素组成的乘积至多需要 d二元秩。域次数与 d 的比率是固定的，而投影到 F_0只会将特征拉回到 F。因此，这种有限的符号展开给出了一个有效的 $C_{b'}$。按照证明的步骤 7，根据局部偏差指数选择 j。固定 j 后，界定完整合并输出的秩，并按照步骤 8 选择全局偏差指数；然后才选择最终投影长度，如步骤 9 所述。

输入。 两个长度为 n 的原始字 x, y。

求值步骤。

1. 分别计算每个源上的浓缩器行，并保留其前 N 个比特。将每个原始输入填充至 b 个有限域 $\mathbf{K}$ 中的系数，并将其系数多项式记为 P_x 或 P_y。 **2.** 对于每个 $1 \leqslant a < R$，构造数值标签 $\alpha_a = \iota(P_x(t_a)$ $P_y(t_a))$。在索引为 $i(a)$、标签为 α_a 的两行上运行列-a 求值器。沿迹从 F 到 F_0 追踪其输出，以得到 W_a。 **3.** 返回将 $\sigma_D(W_1, \ldots, W_{2D-1})$ 投影到 m 个比特的固定满射二元投影。

输出。 一个提取器，同时满足两项独立的源强保证。

证明。 固定满足陈述条件的独立源 X, Y，以及一个可容许的独立篡改元组 $(f_i, g_i)_{i=1}^{C}$。记 $S_X = \mathrm{supp}(X)$ 和 $S_Y = \mathrm{supp}(Y)$。由于 $S_X \times S_Y$ 中的每个点都有正概率，可容许性意味着

$$(f_i(x), g_i(y)) \neq (x, y) \qquad \text{for every } (x, y) \in S_X \times S_Y \text{ and every } i \in [C].$$

当任一支撑集受到限制时，此性质仍然成立。

我们首先证明一个弥散性陈述：在指定的诚实域输出上具有非零系数的每个二元混合特征，在包含于$S_X \times S_Y$ 且边长至少为 $2^{\delta n/2}$的每个乘积矩形上都不是常数。接下来，我们界定这些实际的输入依赖标签特征的全局二元秩，并应用相对于支撑集的近似对偶秩到偏差

引理。最后，我们利用熵余量保留完整的原始源中的任一个。下面的仅用于证明的限制和代表性表格并不是由 G执行的求值。

步骤 *1*：移除含有重原子的比较像。固定一个乘积矩形 $A_X \times A_Y \subseteq S_X \times S_Y$，使其满足 $|A_X|$, $|A_Y| \geqslant 2^{\delta n/2}$。对于步骤 1–7，从 A_X 和 A_Y 上的独立均匀分布开始，并应用下文所述的连续源局部限制。假设诚实合并器输出与 C 比较域值合并器输出的某个混合特征（其诚实系数非零）在该矩形上为常数。记 $q = C + 1$，并令

$$h_{\mathrm{raw}} = \frac{\delta n}{4q}.$$

如果某个活跃比较项的左边缘分布或右边缘分布中，有一个取值的概率大于$2^{-h_{\mathrm{raw}}}$，则将对应的原始侧限制为该取值的原像。此限制造成的最小熵损失小于h_{raw} 比特。经过此固定后，所选比较项的输出仅是另一原始侧的函数。在该另一侧上固定其单比特特征贡献的值，额外代价至多为一比特，并从该特征中移除此比较项。剩余特征在一个正权重乘积矩形上仍为常数，且其诚实系数未变。

重新计算保留的比较项边缘分布并重复此过程。每次迭代都会移除一个活跃比较项，因此最多进行 C 次迭代。任一原始侧的总熵损失至多为$Ch_{\mathrm{raw}} + C$。由于 $C/q < 1$，诚实源的最小熵仍至少为

$$\frac{\delta n}{2} - \frac{C\delta n}{4q} - C \geqslant \frac{\delta n}{4} - O_C(1).$$

现在，两侧所有保留的比较项边缘分布的最小熵均至少为 h_{raw}。这一约简至关重要：不可篡改性从不假设篡改映像保留熵。

步骤 *2*：获得固定的不相同执行对。 分别在每个原始侧上，针对诚实输入与剩余比较输入之间的每一对，选择相等／不等指示量的一个最可能模式，并以该模式为条件。每侧仅有 $O(C^2)$个比特，因此每个选定模式的条件概率至少为 $2^{-O(C^2)}$；这只损失 $O(C^2)$个熵比特，并保留乘积结构。现在，任意两个执行要么在受限支撑集上的该原始侧处处相同，要么处处不同。如果两个完整执行对相同，则将其系数合并到混合特征中。删除任何合并后系数变为零的代表项。诚实代表项不会与篡改代表项合并，因为在原始乘积支撑集上，每种可行篡改作为执行对都没有不动点。因此，其系数仍然非零。我们得到 $r_0 \leqslant q$ 个保留代表项，它们的完整输入对两两不同；对于任意两个不同的代表项，在当前乘积分布中，它们在至少一个原始侧上处处不同。

步骤 *3*：选择共同的高熵浓缩器行。 强浓缩器的实例化熵阈值低于每个剩余诚实边缘分布和比较边缘分布的熵，因此，对于每个这样的边缘分布，其带种子输出表在 r 浓缩器索引上的平均分布，接近于一个被截顶在预设的近完整行熵 h处的分布。下面给出截顶论证的一种有用的基本形式。对于 M 比特上的行分布 P，以条件概率将值 z 标记为接受

$$a_P(z) = \min\{1, 2^{-h}/P(z)\}$$

（并以概率 1 在 $P(z) = 0$ 处接受）。接受后的子分布为每个 z 分配质量 $\min\{P(z)\ 2^{-h}\}$。其拒绝质量为 $\sum_z (P(z) - 2^{-h})_+$，至多为 P 的距离

与最小熵至少为 h的分布类的距离。因此，凝聚器误差界定了平均拒绝概率，其中包括随机选择行索引的情况。

对每一行和每个存活执行，引入这些接受/拒绝指示变量，并使用相应原始侧本地的随机硬币。浓缩器参数确保，在任一侧，具有任何拒绝的索引所占的期望比例至多为 $1/1000$ （在对至多 q个执行求和之后）。马尔可夫不等式表明，以至少 $1-16/1000$ 的概率，该侧至少有 $15r/16$ 个索引在所有执行中均被接受。固定一个概率至少为 $(1-16/1000)2^{-qr}$ 且具有此性质的指示变量模式。由于 $q,\ r$ 是常数，其概率下界为某个正常数。在给定该模式的条件下，每个接受行边缘分布的最大原子概率至多为

$$\frac{2^{-h}}{\Pr[\text{chosen pattern}]},$$

因为其归一化前的点质量由 2^{-h} 限制。因此，它的最小熵仅损失 $O_{\delta,\ C}(1)$ 比特。在两个原始侧上独立地执行此操作；所有限制仍然是乘积限制。接受行在每一侧的所有存活执行中均为公共行。

步骤 **4**：固定指纹相等模式，而不是固定它们的所有值。每个原始字都由其固定多项式 $P_x(T)$ 表示，该多项式定义在 K_0 上，次数小于 $b \leqslant b_* = O_{\delta,\ C}(1)$。对于每一对存活执行、每个原始侧以及 R 个位置，选择一个最可能的联合模式；该模式由布尔比特指示两个求值在 t_a 处是否相等，并以该模式为条件。每侧至多有 $R\binom{q}{2}$ 个这样的比特，因此所选模式的概率至少为 $2^{-R\binom{q}{2}}$，且只需付出 $O_{\delta,\ C}(1)$ 的熵代价，同时保持独立性。

对于两个不同的执行对，至少有一个原始侧在受限支撑集上处处具有不同的完整字。令 $Z_{ij} \subseteq [R]$ 表示该侧上它们的指纹相等的位置集合。由于相等示性函数已被固定， Z_{ij} 对于条件化支撑集中的每个输入都相同，尽管差多项式本身可能随该输入而变化。任选该支撑集中的一个输入，并将其两个不同的完整字记为 u 和 v。它们的系数多项式满足

$$Q_*(T) := P_u(T) - P_v(T) \not\equiv 0, \qquad \deg Q_* \leqslant b-1, \qquad Z_{ij} = \{a \in [R] : Q_*(t_a) = 0\}.$$

固定的求值点 t_a 互不相同，因此由根界可得 $|Z_{ij}| \leqslant b-1$。由于ι具有单射性，完整标签发生碰撞要求两侧的指纹都一致，因此碰撞只能发生在 Z_{ij}。于是，这一对至多有 $b-1$ 个可能的标签碰撞位置。对至多$\binom{q}{2}$ 对取并集后，发生任意成对标签碰撞的列位置不超过

$$\binom{q}{2}(b_*-1) \leqslant R/8$$

列位置遭受任意成对标签碰撞。

平衡行调度将每个 r 浓缩器索引恰好重复 R/r 次。由于每一侧至少有 $15r/16$ 个索引被共同接受，坏数据索引在每一侧至多排除 $R/16$个位置，合计至多 $R/8$ 个位置。排除这些位置以及标签碰撞位置后， R 个位置中至少还剩下$3R/4$ 个。我们不使用位置 R，因此在实际的 $R-1=2D-1$个已求值列中至少有

$$\frac{3R}{4} - 1 \geqslant D = R/2 \tag{3.72}$$

个位置对所有存活执行同时均有效，且具有两两不同的标签。按递增顺序固定其中任意 D 个：

$$I = \{a_1 < \cdots < a_D\} \subseteq [2D-1].$$

这个集合是在一个实际乘积子源上一次性选定的；它不得依赖于执行、目标域输出或后续的傅里叶特征。

将每个被接受的 M 比特行截断为其公开兼容的 N 比特前缀，最多会损失 $M-N = O_{\delta,\ C}(1)$ 比特。加上前面提到的常数损失，对于足够大的 n，每个选定行的边缘分布的最小熵都至少为 $(1-\tau_C/8)N$。

步骤 *5*：仅固定当前列中的数值标签。考虑一个选定索引 $a \in I$。公开所有 r_0 个存活执行对应的 K_0 值指纹 $P_{x_i}(t_a)$ 和 $P_{y_i}(t_a)$。左侧公开信息仅取决于左侧原始输入；右侧公开信息仅取决于右侧输入。因此，每一侧最多泄露 qd 比特。一个初等的最小熵尾界表明，除消息质量至多为 2^{-t} 的情况外，以至多有 2^b 个值的消息为条件，最小熵最多降低 $b+t$。令 $t = \tau_C N/16$。由于公开设置保证

$$\frac{qd}{N} = \frac{q}{A_C H} \leqslant \frac{\tau_C}{64},$$

我们得到，除去总质量呈指数级小的部分后，行熵下界为

$$H_\infty(\text{当前指定数据行} \mid \text{当前指纹消息}) \geqslant (1-\tau_C/8)N - qd - \tau_C N/16 \geqslant (1-\tau_C/2)N.$$

我们只固定这些当前指纹值。所有当前标签都变为固定的数值标签，其中诚实指定标签与每个比较标签均不相同。我们绝不对前面各列的指纹值或求值输出进行条件化；这些仍是非恒定函数，需要通过其分离秩界加以控制。

步骤 *6*：最后活动列的秩支配完整前缀。对于每一列 b，实际指定求值器的公开符号展开给出一个全局二进制特征秩界$C_b N$，该界对源和篡改均一致成立。各列的宽度是逐步选定的，因此对于列 a，

$$K_a = 3 + q\sum_{b<a} C_b \tag{3.73}$$

并且行定理给出该列的受限混合秩大于 $K_a N$。这些选择是在参数固定的情况下作出的，且不依赖于本证明中选取的子源。

取 $r_0 D$ 所选域输出的完整向量的一个非零二元特征。令$a \in I$ 为其具有非零系数的最后一列，并选择一个在该列上的系数非零的执行。将该执行视为秩论证中指定的行。该执行的行分布的熵至少为$(1-\tau_C/2)N$，且其数值标签与位置 a上的所有其他标签均不同。因此，定理3.44 给出了相对于其余至多 $> K_a N$ 个执行的受限秩 C，即使这些执行的比较任意依赖于指定输入也是如此。

我们详细说明如何从原始源变量得到指定压缩行。对于每个行值 u，根据当前左侧子源在给定其指定行 u 条件下的条件分布，抽取一个原始左侧字，并将这些样本存入表 $T_L(u)$。以类似方式抽取一个独立的右侧表 $T_R(v)$。这些表的抽取过程独立于指定行值。

先抽取行值，再查询这些表，所得分布与原始乘积分布完全一致。对于每一对固定的表，每个篡改行和每个前序列输出都是指定的左、右行值的独立确定性函数。分别进行前复合不会增加矩阵秩，因此， a之前的所有列合计至多贡献

$$qN\sum_{b<a}C_b$$

二元特征矩阵。当前列在规定阈值下的每个行矩形上秩为 $>K_aN$。根据秩的次可加性，完整特征的受限秩严格大于

$$K_aN-qN\sum_{b<a}C_b=3N.$$

由于行熵有固定的余量 $\tau_CN/2$，同样的推理适用于指定行支撑集的任意子矩形，只要其边长至少为 $2^{(1-\tau_C)N}$。

在指定的乘积行分布上应用推论 2.20的局部线性秩近似对偶提升，然后对代表性表格以及当前指纹叶节点取平均。稀有消息质量的例外部分为 $2^{-\Omega(N)}$。选定列只有 $D=O_{\delta,C}(1)$ 个，每列均有$N=\Omega_{\delta,C}(n)$，因此我们得到一个常数 $\eta=\eta(\delta,C)>0$，它根据各列的特征界选取，且独立于 j；于是联合选定列输出的每个非零特征 χ 均满足

$$\left|\mathbb{E}(-1)^{\chi(\mathbf{W}_I)}\right|\leqslant 2^{-\eta n},\qquad \mathbf{W}_I\in F_0^{r_0D}. \tag{3.74}$$

这是相对于选定列的完整前缀 的新鲜性，而不仅仅是相对于单独选定的较早输出或固定的较早转录。

步骤 7：傅里叶反演与多数强制。设 $L_0=r_0D\ell$ 为联合选定向量的二进制长度。对于任意给定的 $w\in F_0^{r_0D}$，傅里叶反演和（3.74）给出

$$\Pr[\mathbf{W}_I=w]\geqslant 2^{-L_0}\left(1-(2^{L_0}-1)2^{-\eta n}\right). \tag{3.75}$$

在公共域设置中选择常数因子 j，使得 $L_0\leqslant qD\ell\leqslant \eta n/2$ 对于充分大的 n成立。于是，对于每个 w，（3.75）的右侧都严格为正。因此，可以在保持当前乘积子源不变的同时，独立指定所有 D 次执行中的所有 r_0 选定列的值。

任意选择 β_0、……、$\beta_{r_0-1}\in F_0$，并为每个 $a\in I$ 和每次执行i指定 $W_{i,\,a}=\beta_i$。由于 $D=R/2$ 是 2 的幂， T^D 在

$$(1+\beta_iT)^D\prod_{a\notin I}(1+W_{i,a}T)$$

中的系数恰为 β_i^D：在特征为二的域中，$(1+\beta_iT)^D=1+\beta_i^DT^D$，且其余的 $D-1$ 因子的次数均小于 D。因此

$$\sigma_D(W_{i,1},\dots,W_{i,2D-1})=\beta_i^D.$$

弗罗贝尼乌斯映射 $\beta\mapsto\beta^D$ 是 F_0的一个置换。随着 β_i 独立变化，合并域输出组成的元组遍历 $F_0^{r_0}$的每个点。因此，任何非平凡的含诚实系数混合域特征都不可能在起始矩形上为常数。这就是所需的弥散结论。

步骤 *8*：全局秩与指数误差提升。求值标签取决于实际输入这一点仍然很重要。在每一列，

$$\alpha_a(x,y)=\iota(P_x(t_a),P_y(t_a))=\alpha_{a,L}(x)+\alpha_{a,R}(y).$$

标签仿射种子为 $V_0(x^{\mathrm{pre}})+\alpha_a(x,\ y)V_1(x^{\mathrm{pre}})$。尽管 $V_0,\ V_1$ 可能是非线性的，这个表达式仍具有有界数量的项，这些项分解为左局部和右局部域因子。现在我们固定 j （如步骤 7 所述），并将完整求值器按 F_0上的坐标展开。特别是，

$$[K_0:F_0]=j,\qquad [F:F_0]=Hj.$$

所有工作域相对于 F_0 的扩域次数均有界，其上界仅取决于 $\delta,\ C,\ j$。相容域基使坐标投影和相对迹 F_0-线性，而域乘法具有有界的坐标展开。因此，固定数量的递归域运算以及固定次数的多项式 σ_D 使每个合并域输出都具有如下表示

$$\sum_{\nu=1}^{T_{\delta,C,j}} a_\nu(x)b_\nu(y),\qquad a_\nu(x),b_\nu(y)\in F_0,$$

其中 $T_{\delta,\ C,\ j}$ 与 n无关。经过任意独立的左右篡改组合后，同一分离界仍然成立。对于二元特征，每一项都变为 $\mathrm{Tr}^{F_0/\mathbb{F}_2}(\lambda a_\nu(x)b_\nu(y))$，并且在 F_0的迹对偶 $\mathbb{F}_2$基下，其二元矩阵秩至多为 ℓ。对至多 $q=C+1$次执行求和后，对于完整合并输出元组的每个实际含诚实成分的混合特征，

$$\mathrm{rank}_2[\chi(x,y)]_{x,y}\leqslant qT_{\delta,C,j}\ell\leqslant A_{\delta,C,j}n$$

对于一个与 n 无关的常数 $A_{\delta,\ C,\ j}$。

单列常数 $C_{b'}$ 和局部指数 η 在 j之前就已选定。对于完整合并器 而言，常数$A_{\delta,\ C,\ j}$ 只有在 j 之后才选定：将缩短列输出相乘可能会引入取决于 $[K_0$ 的分离宽度：$F_0]=j$。由于步骤 7 将 j 确定为$\delta,\ C$的函数，因此得到的全局秩界仍为 $O_{\delta,\ C}(n)$。

对于近似对偶性应用，取允许的支撑集为所有非空乘积$A\times B\subseteq S_X\times S_Y$。该类在分别限制下封闭，并且每个此类限制都保持可容许性。步骤 1–7 表明，每个含诚实成分的混合特征在每个允许矩形上都不是常数，其中 $|A|,\ |B|\geqslant 2^{\delta n/2}$。结合全局秩界 $A_{\delta,\ C,\ j}n$，这验证了引理2.19 的假设，其中

$$n_L=n_R=n,\qquad a_0=b_0=\delta/2,\qquad a_1=b_1=3\delta/4.$$

每一侧的熵间隙为 $\delta n/4>0$。因此，对于每个这样的混合特征，以及每个支撑在 S_X 和 S_Y 上的独立源对（其两侧最小熵至少为 $3\delta n/4$），该引理一致地给出：

$$|\operatorname{bias}(\chi)|\leqslant 2^{-\zeta n}\tag{3.76}$$

对于一个常数 $\zeta=\zeta(\delta,\ C)>0$，它在可容许的支撑集、篡改映射和混合特征上均匀分布。只有在 j 和上述完整合并器的秩常数选定后，才确定这个常数。最终投影长度随后在步骤 9 中选定。经过满射二元投影 $\pi:F_0\to\mathbb{F}_2^m$ 后，非零特征会拉回为非零域特征，因此同一界限适用于投影输出。

步骤 *9*：保留任一原始源并完成。固定这样一个混合特征 χ，并首先保留 X。令

$$b(x) = \mathbb{E}_Y (-1)^{\chi(x,Y)}.$$

将 X空间划分为 $b(x)$的两个符号类。概率至少为 2^{-d_s}的类，其中$d_s = \delta n/4$，其条件最小熵至少为 $\delta n - d_s = 3\delta n/4$；对该类进行条件化只会影响左侧源，并保持其与 Y的独立性。其乘积支撑集仍包含于 $S_X \times S_Y$，因此篡改仍然是允许的，且受限分布属于步骤 8 中的允许类。对每个这样的类应用（3.76）。质量低于 2^{-d_s}的两个类贡献至多 2^{1-d_s}。由此得到

$$\mathbb{E}_X |b(X)| \leqslant 2^{-\zeta n} + 2^{1-\delta n/4}. \tag{3.77}$$

如果改为保留 Y，论证是对称的。对诚实输出和实际比较输出进行傅里叶反演，并以保留的原始源为条件，可得

$$\Delta\big((G(X,Y),V,H),(U_m,V,H)\big) \leqslant 2^{qm-1}\big(2^{-\zeta n} + 2^{1-\delta n/4}\big), \qquad H \in \{X,Y\}.$$

我们只使用在诚实输出上具有非零系数的特征；仅涉及 V 的特征会与实际比较边缘分布相消，而后者保持不变。按设置选择公开满射投影的长度，

$$m = \min\left\{\ell, \left\lfloor \frac{n \min\{\zeta, \delta/4\}}{2q} \right\rfloor\right\}.$$

由于 $\ell = \Theta_{\delta,\ C}(n)$ 且 q为常数，$m = \Omega_{\delta,\ C}(n)$，最后的统计距离界为 $2^{-\Omega_{\delta,\ C}(n)}$。对于固定的 δ，C，该构造本身可在多项式时间内完成：所有公开枚举的行调度和域运算计数都是常数，其余运算均为明确的多项式时间域运算。对于有限多个较短长度，规定的一比特常数补全具有统计误差 $1/2$ 且特征秩为零；减小全局正常数即可覆盖这一有限范围，而不改变渐近结论。

□

推论 3.46（分离源强与线性输出）。 构造 3.8保留两个原始源中的任一个及完整的比较元组，输出 $\Omega_{\delta,\ C}(k)$，误差为 $2^{-\Omega_{\delta,\ C}(n)}$。

证明。 使用 H 定理3.45中的两种选择。 □

4 熵高效输出在常数率下

本节中的构造框区分内部生成的种子与每个带种子提取器的公开种子接口。它们明确记录所有种子长度和有效载荷长度；接下来的碰撞论证属于分析，而不是求值步骤。

本节首先证明具有加性熵损失的普通双源输出，然后证明源强C-篡改输出、其 $C = 0$ 普通源强推论，最后证明具有加性损失的和集输出。普通拼接封装和和集封装共享单有效载荷引理，并且可以共享定理3.33的种子。中心化不可篡改封装使用第C 节中的固定3.3定理。每个非空输出陈述都明确给出了其非平凡的指数级小误差范围。

4.1 具有加性熵损失的双源输出

4.1.1 前置条件：单载荷碰撞引理的证明

下面的加权论证保留了实际的另一个源。它直接适用于任意最小熵分布；无需进行平坦源归约，也无需对单个被篡改图像作出假设。

引理 4.1（单侧源强碰撞回收）。设 X, Y 是相互独立的、最小熵至少为 k 的 n 比特源，令 $K = 2^k$，设 G 是一个 d 比特种子，并令 $D = 2^d$。假设

(i) G 在分量阈值 $k-4$ 下针对每个无不动点左映射都是单侧不可篡改的，误差为 γ，并且在该阈值下是误差为 γ 的普通提取器；

(ii) 在当前乘积分布上，$\Delta((X, G(X,Y)), (X, U_d)) \leqslant \eta$；

(iii) $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^p$ 是熵为 k、误差为 ζ 的强种子提取器。

记 $M = 2^p$，并假设 $K \geqslant 16M$。对于 $0 < \theta < 1$，令 $\beta = 3(1+\theta)\zeta/\theta + 2\eta$。那么

$$\Delta((Y, G(X,Y), E(X, G(X,Y))), (Y, U_{d+p})) \leqslant \beta + \frac{1}{2}\sqrt{\theta + 2\beta + 2(1+\theta)D^2\gamma + DM/K}. \tag{4.1}$$

保留 X 时的对称结论可通过交换两个源，并使用 (i) 的右侧篡改版本得到。

证明。记 $p_x = \Pr[X = x]$ 和 $q_y = \Pr[Y = y]$，于是 $p_x,\ q_y \leqslant 1/K$。对于数值种子 $v \in \{0,1\}^d$，将 X 的支撑集划分为哈希纤维 $A_{v,a} = \{x\colon p_x > 0,\ E(x, v) = a\}$，其质量为 $w_{v,a} = \sum_{x \in A_{v,a}} p_x$。保留质量至多为 $(1+\theta)/M$ 的纤维，并将它们的并集记为 A_v。记 $r_v = \Pr[X \in A_v]$。若 $r_v < 1/2$，则舍弃整个数值种子 v。

对于 M 原子上的任意概率分布，大于 $(1+\theta)/M$ 的原子的质量至多为其与均匀分布之间统计距离的 $(1+\theta)/\theta$ 倍。因此，对数值均匀种子取平均后，强种子提取器保证表明

$$D^{-1}\sum_{v}(1 - r_v) \leqslant (1+\theta)\zeta/\theta.$$

马尔可夫不等式表明，被舍弃的数值种子的均匀质量至多为 $2(1+\theta)\zeta/\theta$。这两个删除事件分别可由 (X, v) 和 v 测得。将假设 (ii) 应用于当前乘积分布本身，可将每个事件从均匀分布 v 转移到 $v = G(X,Y)$，代价至多为 η。因此，分布 $P\mathrm{of}\,(Y, G(X,Y)\ E(X, G(X,Y)))$ 存在一个质量为 $t_0 \geqslant 1 - \beta$ 的存活子分布 $\mu \leqslant P$。

我们对其条件 碰撞能量进行界定，并按原始的 Y-边际分布加权 l:

$$\mathcal{C}_Y(\mu) = \sum_{y\colon q_y > 0} \frac{1}{q_y}\sum_{v,a}\mu(y,v,a)^2.$$

对于一个保留种子 v 和一个保留纤维 $A_{v,a}$，令

$$e_{v,a}(y) = \sum_{x \in A_{v,a}} p_x \mathbf{1}_{\{G(x,y)=v\}},$$

并将 $e_{v,a}(y) = 0$ 设为丢弃的种子或纤维。独立性给出 $\mu(y,v,a) = q_y e_{v,a}(y)$，因此

$$\mathcal{C}_Y(\mu) = \sum_{y} q_y \sum_{v,a} e_{v,a}(y)^2. \tag{4.2}$$

对角项 $x = x'$ 至多贡献 $\sum_x p_x^2 \leqslant 1/K$，因为每个原始对 (x, y) 都有唯一的真实种子。

对于一个保留的数值种子 v，定义保留的哈希纤维中有序不同元素对的总权重为

$$N_v = \sum_{a:\ A_{v,a} \text{ retained}} \sum_{\substack{x,x' \in A_{v,a} \\ x \neq x'}} p_x p_{x'}.$$

由于 $r_v \geqslant 1/2$，且哈希值至多有 M 个，

$$N_v \geqslant \frac{r_v^2}{M} - \sum_x p_x^2 \geqslant \frac{1}{4M} - \frac{1}{K} \geqslant \frac{1}{8M}, \qquad N_v \leqslant \sum_{a:\ A_{v,a} \text{ retained}} w_{v,a}^2 \leqslant \frac{1+\theta}{M}.$$

令 $(X_v,\ X_v')$ 为这样一个有序对，其抽样概率为 $p_x p_{x'}/N_v$。对于保留纤维$A_{v,\ a}$中的 x，其第一边缘分布的点质量为

$$\Pr[X_v = x] = \frac{p_x(w_{v,a} - p_x)}{N_v} \leqslant \frac{8(1+\theta)}{K} \leqslant \frac{16}{K}.$$

由对称性，X_v' 也满足相同的上界；两个边缘分布的最小熵均至少为 $k - 4$。该有序对的抽样独立于 Y。

用随机无不动点左映射表示条件分布 $X_v' \mid X_v$，该映射独立于 X_v 和 Y 进行抽样：对于 X_v 的支撑集中的每个点，按照此条件分布选择其像；对于其他点，则任选一个不同的像。对于每个固定映射 f，假设 (i) 蕴含

$$\Pr[G(X_v, Y) = v, G(f(X_v), Y) = v] \leqslant D^{-1} \Pr[G(f(X_v), Y) = v] + \gamma.$$

对这些映射取平均。由于 X_v' 的平均后边缘分布的熵至少为 $k - 4$，因此假设 (i) 中的普通提取给出

$$\Pr[G(X_v, Y) = G(X_v', Y) = v] \leqslant D^{-2} + 2\gamma. \tag{4.3}$$

第二边缘分布不必对每个固定映射都具有高熵；只有在取平均之后才应用普通提取。始终保持与 Y 的独立性。

展开（4.2）中的平方项。对于每个保留的 v，其非对角贡献为 N_v 乘以（4.3）中的概率。令 $N_v = 0$ 表示丢弃的种子，并对所有 D 数值种子求和：

$$\begin{aligned} \mathcal{C}_Y(\mu) &\leqslant \frac{1}{K} + \sum_v N_v(D^{-2} + 2\gamma) \\ &\leqslant \frac{1}{K} + \frac{1+\theta}{DM} + \frac{2(1+\theta)D\gamma}{M}. \end{aligned}$$

相对于概率测度 $P_Y \times U_{d+p}$应用加权柯西–施瓦茨不等式（其质量为 $q_y/(DM)$），得到

$$\Delta(\mu, t_0(P_Y \times U_{d+p})) \leqslant \frac{1}{2}\sqrt{DM\mathcal{C}_Y(\mu) - t_0^2}.$$

从 P 到 μ 的距离，以及从完整理想分布到 t_0 倍该理想分布的距离，总计为 $1 - t_0 \leqslant \beta$。最后，

$$t_0^2 = 1 - 2(1 - t_0) + (1 - t_0)^2 \geqslant 1 - 2\beta.$$

将此不等式与碰撞界和 $DM/K = 2^{d+p-k}$ 结合即可证明（4.1）。该论证可直接应用于给定的源分布，并完整保留原始 Y-边缘分布。 □

引理 4.2（两个分别源强的有效载荷的异或）。设 $S = S(X,Y)$ 是任意 d-比特函数，且 $A = E_L(X,S)$、$B = E_R(Y,S)$ 各自具有 p 比特。若 (Y,S,A) 与 (Y,U_d,U_p) 的距离为 ξ_L，且 (X,S,B) 与 (X,U_d,U_p) 的距离为 ξ_R，则 $F = (S, A \oplus B)$ 分别是源强的：在揭示 Y 时误差为ξ_L，在揭示 X 时误差为 ξ_R。

证明。对于固定的数值 (y,s)，与 $E_R(y,s)$ 做异或相当于对 $\{0,1\}^p$ 进行一个置换。对真实分布和理想分布应用这一确定性置换会保持它们之间的距离，从而证明第一个断言。第二个断言的证明相同，只需保留 X，并固定 $E_L(X,S)$。 □

4.1.2 定理

定理 4.3（来自两个独立源的近乎全部输出）。存在绝对常数$C_0, C_1 > 0$，使得对于每个固定的 $0 < \delta \leqslant 1$，都存在 $b_\delta > 0$ 和 n_δ，满足以下性质。对于 $n \geqslant n_\delta$, $\delta n \leqslant k \leqslant n$ 为整数，且

$$2^{-b_\delta n} \leqslant \varepsilon \leqslant \tfrac{1}{4},$$

存在一个可在多项式时间内确定性计算的函数

$$\mathrm{Ext}_{n,k,\delta,\varepsilon} : \{0,1\}^n \times \{0,1\}^n \longrightarrow \{0,1\}^m$$

使得对于满足 $H_\infty(X)$ $H_\infty(Y) \geqslant k$ 的独立 X, Y，

$$\Delta(\mathrm{Ext}_{n,k,\delta,\varepsilon}(X,Y), U_m) \leqslant \varepsilon, \qquad m \geqslant 2k - C_0 \log(1/\varepsilon) - C_1.$$

其输出具有如下形式

$$\big(S, E_L(X,S), E_R(Y,S)\big), \qquad S = G(X,Y).$$

两个带种子输出的长度不相等，并且种子 S 会被保留。对于每个固定的 δ，构造和求值的复杂度关于 n +$\log(1/\varepsilon)$是多项式的。$2k$尺度的输出在该长度下不是源强的，如引理2.4所示；推论4.9 则另外给出 $k - O(\log(1/\varepsilon) + 1)$源强输出。

4.1.3 确定性构造与种子参数详述

Co构造 4.1 具有加性熵损失的普通双源输出 s

来自前文各部分的组成要素。定理3.33的单侧和集种子（构造3.5）；引理2.23中的强种子提取器；熵余量强化，引理2.5。引理4.1 是已经证明的碰撞接口，用于分析第一个载荷。

公开参数和固定设置。 固定 $\delta > 0$、整数 $\delta n \leqslant k \leqslant n$以及定理 4.3的误差 ε 的取值范围。令 $\sigma = \delta/4$、 $s = \log(1/\varepsilon)$，并令 $\zeta = \varepsilon^4/2^{20}$。以率 σ实例化定理 $G_{\Sigma,\,\sigma}$ 中的构造，至少输出 $c_\sigma n$个比特。使用下文校准的绝对常数，令

$$d = \left\lceil C_1 \left[\log^2(n+2) + (s+1)\log\left(2 + \frac{n}{s+1}\right)\right]\right\rceil,$$
$$p = \lfloor k - d - 2s - 6 \rfloor, \qquad q = \lfloor k - A(s+1) \rfloor.$$

以 E_L、 E_R 实例化引理 2.23 ，其熵阈值为 k ，误差为 ζ，在忽略填充后使用共同的 d比特种子，并将输出分别截断为 p 和 q 比特。下文的参数验证确保 p、 $q \geqslant 1$ 以及 $d \leqslant c_\sigma n$。此处 q 表示载荷长度，而非篡改次数。

输入。 有序对（x、 y） $\in \{0、\ 1\}^n \times \{0、1\}^n$。

求值步骤。

1. 计算内部种子 $S =$ 前缀$_d\, G_{\Sigma,\ \sigma}(x + y)$。**2.** 计算 p 比特的左载荷 $A_L = E_L(x,\ S)$。**3.** 使用相同的种子，计算 q 比特的右载荷 $A_R = E_R(y,\ S)$。**4.** 返回拼接结果（S， A_L， A_R） ，包括全部 d 个种子比特。

输出。 恰好为 $m = d + p + q$ 比特，其中 $m \geqslant 2k - (A+2)s - A - 8$ ，且误差至多为 ε。只有种子是和兼容的；最终求值器分别读取 x， y ，并未声称在该输出长度下源强。

不参与求值。 哈希纤维修剪、碰撞伙伴，以及用于附加第二个载荷的比较实验。求值期间不会采样额外的随机种子。

第 3.1 节中的独立普通常数率构造不会因该种子选择而改变。缩短后的完整双侧单篡改种子可以满足相同的接口，但并非必需。单侧和集种子提供不可篡改性，其诚实边缘分布提供普通提取，而熵余量提供所需的源强。

4.1.4 追加第二个有效载荷

推论 4.4 （单载荷再利用中的普通拼接）。 设引理 4.1 适用于长度为 p的左载荷，其距离界为 ξ。若 E_R 是针对 Y的强种子提取器，具有 q 个输出比特，误差为 ζ_R，则

$$\Delta((S, E_L(X,S), E_R(Y,S)), U_{d+p+q}) \leqslant \xi + \zeta_R. \tag{4.4}$$

证明。 将 $(y, s, a) \mapsto (s, a, E_R(y, s))$ 应用于比较 $(Y, S, E_L(X, S)) \approx_\xi (Y, U_d, U_p)$，得到

$$(S, E_L(X,S), E_R(Y,S)) \approx_\xi (U_d, U_p, E_R(Y, U_d)).$$

右侧的独立种子允许从 Y 中进行强提取；附加独立的 U_p，并使用三角不等式。

□

推论 4.5（通用误差校准）。 对于 $0 < \varepsilon \leqslant 1/4$，单载荷误差至多为 $\varepsilon/4$，如果

$$\theta = \varepsilon^2/64, \quad \zeta \leqslant \varepsilon^4/2^{20}, \quad \eta \leqslant \varepsilon^2/4096, \quad \gamma \leqslant \varepsilon^2/(512D^2), \quad DM/K \leqslant \varepsilon^2/64.$$

若还满足 $\zeta_R \leqslant \varepsilon^4/2^{20}$，则拼接输出的误差小于 ε。关于 DM/K 的界意味着 $K \geqslant 16M$。

证明。 由于 $1 + \theta < 2$，单有效载荷引理具有 $\beta \leqslant \varepsilon^2/1024$。其被开方数至多为

$$\varepsilon^2(1/64 + 1/512 + 1/128 + 1/64) < \varepsilon^2/16.$$

因此，其距离至多为 $\varepsilon^2/1024 + \varepsilon/8 < \varepsilon/4$。应用推论 4.4 进行拼接。 □

4.1.5 最终参数代入与熵核算

推论 4.6（短输出强化）。 若一个普通的 d比特双源提取器在分量阈值 h 下的误差为 η_0，那么在分量熵 $k > h$ 下，其对任一原始源均为强提取器，误差至多为 $2^d(\eta_0 + 2^{h-k})$。

证明。 这是引理 2.5在零比较情形下的结论。 □

固定任意的 $0 < \delta \leqslant 1$ 和整数 $\delta n \leqslant k \leqslant n$，并取定理 3.33 中固定率 $\sigma = \delta/4$ 下的 $G_{\Sigma,\sigma}$。记 $a = e_\sigma$, $c = c_\sigma$ 和 $s = \lceil \log(1/\varepsilon) \rceil$ 分别为这个和兼容种子的指数误差常数和输出速率常数。设 $G(x, y) =$ 前缀$_d\, G_{\Sigma,\sigma}(x + y)$。种子定理给出所需的单侧保证和普通提取，且 $\gamma \leqslant 2^{-an}$，没有额外损失。对于所有充分大的 n, $k - 4 \geqslant \sigma n$，这正是碰撞伙伴源所需的条件。推论 4.6 给出

$$\eta \leqslant 2^d(2^{-an} + 2^{\sigma n - k}) \leqslant 2^{d+1-a_* n}, \qquad a_* := \min\{a, \delta/2\}. \tag{4.5}$$

使用引理 2.23 误差为 $\zeta = \varepsilon^4/2^{20}$。调整绝对常数后，其种子长度至多为 $B\left\lceil \log^2(n+2) + (s+1)\log\left(2 + \frac{n}{s+1}\right)\right\rceil$，且可用输出至少为 $k - A(s+1)$。取

$$d = \left\lceil C_1 \left[\log^2(n+2) + (s+1)\log\left(2 + \frac{n}{s+1}\right)\right]\right\rceil$$

取足够大的绝对 C_1。这超过了实际带种子提取器对种子的要求，因此我们对未使用的种子坐标进行填充。进一步增大 C_1，以确保下面的左侧输出是一个允许的截断。令

$$p = \lfloor k - d - 2s - 6 \rfloor, \qquad q = \lfloor k - A(s+1) \rfloor. \tag{4.6}$$

在现在选定的范围内，这两个长度都将为正。特别地，$2^p \leqslant 2^k\varepsilon^2/(64 \cdot 2^d)$，正如推论 4.5 所要求的。

只需确保

$$d \leqslant cn, \qquad 2d + 2s + 14 \leqslant a_* n, \qquad p, q \geqslant 1. \tag{4.7}$$

这里，a_*，c 是和兼容种子的已固定正常数，不是 s 的函数。条件 $d \leqslant cn$ 确保 $G_{\Sigma,\ \sigma}$ 的输出足够长，满足指定前缀的要求。对于$s \leqslant bn$，由单调性可得

$$\frac{\left\lceil \log^2(n+2) + (s+1)\log\left(2 + \frac{n}{s+1}\right)\right\rceil}{n} \leqslant \frac{\log^2(n+2)}{n} + (b + 1/n)\log\left(2 + \frac{1}{b + 1/n}\right).$$

右侧趋于 $b\log(2+1/b)$，当 $n\to\infty$时；而该表达式在 $b\downarrow 0$时趋于零。选取一个足够小的固定值 $b_\delta>0$，使其满足 (4.7) 中的所有线性预算，包括两个输出长度均为正。于是对于足够大的 n，所有预算都成立，且对 $2\leqslant s\leqslant b_\delta n$一致。

公式（4.5）和（4.7）蕴含分别强性要求 $\eta\leqslant\varepsilon^2/4096$ 以及单侧误差要求 $\gamma\leqslant\varepsilon^2/$（512$\cdot\ 2^{2d}$）。应用碰撞封装及其校准推论。输出误差至多为 ε。

最后，保留种子抵消了其自身长度的开销：

$$d+p+q\geqslant 2k-(A+2)s-A-8. \tag{4.8}$$

常数 $A+2$ 和 $A+8$ 是绝对的带种子提取器常数。这证明了定理4.3。输出损失中不再有对数n 项。内部种子长度并未声称是最优的。单侧和集定理提供了种子。特别地，定理4.3 仍然断言普通提取，而不是其完整长输出的不可篡改性。

4.1.6 全误差解释与求值

对于固定的 δ，选择 C_δ 使其大于（4.8）中的常数并大于 $2/b_\delta+2$。当 $s>b_\delta n$时，表达式 $2k-C_\delta$（$s+1$）为非正。此时返回空字符串恰好是均匀分布的，并满足截断全误差保证 $m\geqslant$最大值$\{0,\ \lfloor 2k-C_\delta$（$s+1$）$\rfloor\}$。再次增大 C_δ，以覆盖有限范围 $n<n_\delta$。对于 $n\geqslant n_\delta$ 和 $s\leqslant b_\delta n$，必要时将构造的输出截断为所声称的非负长度。统计距离不会因截断而增加。

对于任意实数精度要求，只需构造满足 $2^{-\lceil\log(1/\varepsilon)\rceil}$的算法；新增的熵损失为一个绝对常数。因此，算法可以将精度指数取为整数。其种子求值器由固定凝聚器、常数个域幂和迹，以及一个固定的初等对称多项式组成。其第二阶段使用引理 2.23。当率固定时，所有计算的复杂度均为 $n+\log(1/\varepsilon)$ 的多项式。着色、代表表、源剪枝和子源选择仅在证明中出现。

最终的双源包装器不声称保留种子的 $O(n)$特征秩或特殊内积形式。该性质用于种子构造内部，其作用已得到明确证明。

4.2 具有加性熵损失的源强不可篡改输出

固定 $C\geqslant 0$。我们在较低的固定熵率下使用定理 $(2C+1)$的篡改提取器 3.45，使用引理 2.5缩短它，并使用引理2.23中的强种子提取器，其绝对常数为 A，B。情形 $C=0$ 表示比较元组为空；其短种子仍具有单比较不可篡改性。下面的中心化回收引理是由这些种子和带种子提取器接口直接推出的独立结论。

4.2.1 前置条件及证明：实际泄漏下的中心化回收

我们首先证明一个稍微更一般的结论，该结论允许由比较种子选择任意的源局部载荷泄漏。下面的大型后验分布和碰撞图是证明对象；求值器从不计算它们。

引理 4.7（具有实际泄漏的中心化回收）。设 X,Y 是集合 $\mathcal{A}$,$\mathcal{B}$ 上的独立平坦源，且集合大小为 $K=2^k$。令

$$D=2^d,\qquad M=2^p,\qquad L=DM=2^m,\qquad m=d+p.$$

假设 $k \geqslant h$ 和 $G:\{0,1\}^n \times \{0,1\}^n \to [D]$ 在最多 $R = 2C+1$ 种独立篡改下是不可篡改的，误差为 η，并且分别对任一源都是强的，只要两个源的最小熵均至少为 h。固定可行的 $(f_i, g_i)_{i=1}^C$，并记

$$S_0 = G(X,Y), \qquad \bar{S} = (G(f_i(X), g_i(Y)))_{i=1}^C.$$

对于每个数值 $\bar{s} \in [D]^C$ 令 $a_{\bar{s}}: \mathcal{A} \to [M]^C$ 为任意值，并令

$$T = (\bar{S}, a_{\bar{S}}(X))$$

为实际泄漏元组。令 $0 < \alpha, \beta < 1$。假设 $E:\{0,1\}^n \times [D] \to [M]$ 是一个误差为 ζ、熵为

$$\kappa = k - Cm - \log\frac{1}{\alpha\beta}. \tag{4.9}$$

Then

$$\begin{aligned} &\Delta((Y,T,S_0,E(X,S_0)),(Y,T,U_d,U_p)) \\ &\leqslant 2\beta + 3\alpha + 3\zeta + \frac{\eta}{2} + \frac{1}{2}\sqrt{\Delta}, \end{aligned} \tag{4.10}$$

其中

$$\boxed{\Delta = (2C+1)\frac{L^{C+1}}{K} + \frac{2D^{C+2}L^C\,2^h}{\alpha^2 K} + \frac{4\eta D^{2C+2}}{\alpha^2}.} \tag{4.11}$$

理想端的所有变量 Y, T 均保持其原始联合分布。

后验分布与独立种子的参考对象 固定一个数值尾部标签 $t = (\bar{s}, \bar{a})$。定义

$$A_t = \{x \in \mathcal{A} : a_{\bar{s}}(x) = \bar{a}\}, \qquad a_t = |A_t|.$$

对于每个固定的 $\bar{s}$，这些集合构成$\mathcal{A}$的一个划分，因此

$$\sum_t a_t = D^C K. \tag{4.12}$$

对于 $x \in A_t$ put

$$w_t(x) = \Pr_Y[\bar{S}(x,Y) = \bar{s}], \qquad \lambda_t = \Pr[T = t], \qquad N_t(y) = |\{x \in A_t : \bar{S}(x,y) = \bar{s}\}|.$$

在 A_t之外使用 $w_t(x) = 0$。基本恒等式为

$$\sum_x w_t(x) = K\lambda_t, \qquad \Pr[X = x \mid T = t] = \frac{w_t(x)}{K\lambda_t}, \qquad \Pr[Y = y, T = t] = \frac{N_t(y)}{K^2}. \tag{4.13}$$

We 省略零概率标签。用 P_t 表示后验分布 $X \mid T = t$.

设 P为 $(Y,\ T,\ S_0,\ E(X,\ S_0))$ 的实际分布。通过对原始 $(X,\ Y,\ T)$ 进行采样，然后采样一个独立均匀分布的 $S \in [D]$，并输出 $(Y,\ T,\ S,\ E(X,\ S))$，构造参考分布$\mathcal{R}$。最后，令

$$Q(y,t,s,a) = \frac{\Pr[Y = y, T = t]}{DM} = \frac{N_t(y)}{K^2 L}.$$

因此$\mathcal{R}$两者均 d Q 保留实际的（$Y,\ T$）边缘分布。任务是首先将 P 与 $\mathcal{R}$ 进行比较 .

如果称 t 为良好

$$\lambda_t \geqslant \beta L^{-C}. \tag{4.14}$$

坏标签的总质量至多为 β。对于良好标签，

$$\max_x P_t(x) \leqslant (K\lambda_t)^{-1} \leqslant 2^{-(k-Cm-\log(1/\beta))}.$$

特别地， P_t 的最小熵至少为 κ。对于良好 t，定义

$$\psi_t(y) = \mathbf{1}[N_t(y) \geqslant \alpha K\lambda_t]. \tag{4.15}$$

满足 $N_t(y) < \alpha K\lambda_t$ 的点对的实际质量至多为 α:

$$\sum_t \sum_{y:N_t(y)<\alpha K\lambda_t} \frac{N_t(y)}{K^2} \leqslant \sum_t \alpha\lambda_t = \alpha.$$

对于每个存活的 (y, t)，条件分布 $X \mid (Y = y, T = t)$ 在大小为

$$N_t(y) \geqslant \alpha\beta K L^{-C} = 2^{\kappa}.$$

因此，强种子提取给出

$$\Delta(\mathcal{R}, Q) \leqslant \zeta + \alpha + \beta. \tag{4.16}$$

这一步使用独立种子和实际的尾部纤维；它不适用于经过条件化的双源分布 G 。

后验分布 **不使用带种子误差乘数的修剪** 丢弃一个原始对 (X, T) if

$$w_T(X) < \alpha D^{-C}. \tag{4.17}$$

这至多耗费 α。事实上，对于每个固定的 x ，每个D^C 数值种子元组 $\bar{s}$都只对应一个可能的 $\bar{a} = a_{\bar{s}}(x)$。因此

$$\frac{1}{K}\sum_x \sum_{t:w_t(x)<\alpha D^{-C}} w_t(x) \leqslant \alpha.$$

该估计针对实际的 (X, T) 分布；在评估诚实种子之前， P和$\mathcal{R}$ 共享此分布。

对于良好的 t 和数值 s ，定义哈希纤维质量

$$r_{t,s}(a) = \Pr_{P_t}[E(X, s) = a].$$

丢弃满足 $r_{t,\ s}(E(x,\ s)) > 2/M$ 的点。对于 $[M]$ 上的任意分布 r，其质量超过 $2/M$ 的原子总质量至多为 $2\Delta(r, U_M)$：对于这些原子， $r(a) - 1/M > r(a)/2$。由于 P_t 的熵至少为 κ，对均匀分布的 s 取平均可知，求和遍历良好的 t 后，此删除在$\mathcal{R}$下发生的概率至多为 2ζ 。

在 P 下进行相同的删除，代价至多为 $2\zeta + \eta$。要看出这一点，使用保留 X 的源强种子比较:

$$(X, \bar{S}, S_0) \approx_\eta (X, \bar{S}, U_d).$$

$\bar{T} = (S,\ a_{\bar{S}}(X))$ 和重纤维删除测试都是保留变量与诚实种子的确定性函数。这是修剪所需的唯一转移。特别是，没有分解D^C 乘以 ζ。

对于良好的 t集合

$$\phi_{t,s}(x) = \mathbf{1}[x \in A_t]\,\mathbf{1}[w_t(x) \geqslant \alpha D^{-C}]\,\mathbf{1}[r_{t,s}(E(x,s)) \leqslant 2/M]. \tag{4.18}$$

令 $\mu \leqslant P$ 和 $\nu \leqslant \mathcal{R}$ 为通过保留良好的 t、 $\psi_t(y) = 1$以及$\phi_{t,\ s}(x) = 1$而得到的次分布，其中对 μ 使用实际的诚实种子，对 ν使用独立种子。质量损失满足

$$b_P := 1 - \|\mu\|_1 \leqslant \beta + 2\alpha + 2\zeta + \eta, \tag{4.19}$$

$$b_R := 1 - \|\nu\|_1 \leqslant \beta + 2\alpha + 2\zeta. \tag{4.20}$$

定义$\phi_{t,\ s}$ 的截断函数在 $t,\ s$ 为数值量后，是源局部的。它们使用对 Y取平均的后验分布 P_t，而不是在给定特定 $Y = y$条件下的后验分布。

中心化能量恒等式 令

$$h_s(x,y) = \mathbf{1}[G(x,y) = s] - D^{-1}.$$

对于良好的 $t = (\bar{s},\ \bar{a})$和 $a \in [M]$，定义

$$B_{t,s,a}(y) = \sum_{\substack{x \in A_t \\ E(x,s) = a}} \phi_{t,s}(x)\,\mathbf{1}[\bar{S}(x,y) = \bar{s}]\,h_s(x,y).$$

于是，逐点地，

$$(\mu - \nu)(y,t,s,a) = K^{-2}\psi_t(y)B_{t,s,a}(y).$$

因此，

$$\mathcal{E} := \sum_{y,t,s,a} \frac{(\mu-\nu)(y,t,s,a)^2}{Q(y,t,s,a)} = \frac{L}{K^2} \sum_{t\ \text{good}} \sum_{y : N_t(y) > 0} \frac{\psi_t(y)}{N_t(y)} \sum_{s,a} B_{t,s,a}(y)^2. \tag{4.21}$$

省略零质量原子。由加权柯西–施瓦茨不等式可得

$$\frac{1}{2}\|\mu - \nu\|_1 \leqslant \frac{1}{2}\sqrt{\mathcal{E}}. \tag{4.22}$$

参考文献在此阶段并非均匀分布。通过 D^{-1} 进行中心化，才能使大图项在非篡改种子替换下消失。

对角与交叉输入例外 将（4.21）中的平方项展开为有序对 （$x,\ x'$）。若一对有序对满足

$$x = x', \qquad \text{or} \qquad x' = f_i(x), \qquad \text{or} \qquad x = f_i(x') \quad \text{for some } i. \tag{4.23}$$

对于每个 x, x', y,

$$\sum_s |h_s(x,y)h_s(x',y)| \leqslant \left(\sum_s h_s(x,y)^2\right)^{1/2} \left(\sum_s h_s(x',y)^2\right)^{1/2} = 1 - D^{-1} \leqslant 1. \tag{4.24}$$

哈希相等性和 ϕ 分解只会使这个绝对值界进一步减小。

对于每个固定的 $(y,\ t)$，对角线至多贡献 $N_t(y)$ 个有序对。通过计算 x，关系 $x' = f_i(x)$ 同样至多贡献 $N_t(y)$ 个有序对；通过计算 x'，关系 $x = f_i(x')$ 至多贡献 $N_t(y)$ 个有序对。对于非单射的 f_i，这仍然成立。至多有 KL^C 个受支撑的 $(y,\ t)$。因此，按绝对值界定的例外贡献至多为

$$\mathcal{E}_{\mathrm{exc}} \leqslant (2C+1)\frac{L^{C+1}}{K}. \tag{4.25}$$

跨输入例外确保所有后续的复合比较映射都是可容许的。

源局部碰撞图 对于每个良好的 t 和数值型 s，构造由有序对 $x,\ x' \in A_t$ 组成的对称有向图 $\Gamma_{t,\ s}$，满足

$$\phi_{t,s}(x) = \phi_{t,s}(x') = 1, \qquad E(x,s) = E(x',s),$$

移除（4.23）中的每个例外对后。令 $J_{t,\ s} = |\Gamma_{t,\ s}|$。

每个保留的哈希纤维的基数至多为

$$B_t := \frac{2D^C K\lambda_t}{\alpha M}. \tag{4.26}$$

事实上，每个幸存的 x 都有 $P_t(x)$ 个 $\geqslant \alpha/(D^C K\lambda_t)$，而其保留纤维的总 P_t 质量至多为 $2/M$。因此，$\Gamma_{t,\ s}$ 的每个顶点的次数至多为 B_t，并且

$$J_{t,s} \leqslant a_t B_t. \tag{4.27}$$

若$J_{t,\ s} < 2^h B_t$，则称图为小图。在保留的$(y,\ t)$上，$N_t(y)^{-1} \leqslant (\alpha K\lambda_t)^{-1}$。利用 $|h_s h'_s| \leqslant 1$，一个小图对（4.21）的绝对贡献至多为

$$\frac{LJ_{t,s}}{\alpha K^2\lambda_t} \leqslant \frac{2D^{C+1}2^h}{\alpha^2 K}.$$

图标签至多有 DL^C 个。因此

$$\mathcal{E}_{\mathrm{small}} \leqslant \frac{2D^{C+2}L^C\,2^h}{\alpha^2 K}. \tag{4.28}$$

这些小图不应用种子定理。

大图与可容许的额外比较 对于一个大图，从$\Gamma_{t,\ s}$中均匀分布地抽取$(V,\ V\prime)$，且与原始的 Y独立。其第一边缘分布的最大原子概率至多为$B_t/J_{t,\ s} \leqslant 2^{-h}$。第二边缘分布也有相同的上界，不过下文只需用到第一边缘分布。

将条件分布 $V' \mid V$表示为随机确定性表 $v \mapsto \rho(v)$，并独立于 $V,\ Y$进行抽样。每个表在 V的支撑集上只选择 v 的邻居。对于每个固定的表，诚实对$(V,\ Y)$具有以下可容许的额外比较：

$$(\rho(V), Y), \qquad (f_i(V), g_i(Y)), \qquad (f_i(\rho(V)), g_i(Y)), \quad 1 \leqslant i \leqslant C. \tag{4.29}$$

第一次比较之所以不同，是因为 $V\prime \neq V$。中间的比较满足原有的可容许性条件。最后的比较根据（$f_i(V') \neq V$，其结果为 4.23）。所有映射彼此独立，且其选取独立于 Y。

应用强不可篡改性并保留 Y，然后对各个表取平均。它将 $G(V,Y)$ 替换为一个与其距离至多为 η 的独立均匀种子，同时保留

$$Y,\quad G(V',Y),\quad (G(f_i(V),g_i(Y)))_i,\quad (G(f_i(V'),g_i(Y)))_i.$$

除 $h_s(V,\ Y)$ 之外，以下所有分解都是这些保留变量的函数：

$$h_s(V',Y)\,\mathbf{1}[\bar{S}(V,Y)=\bar{S}(V',Y)=\bar{s}]\,\frac{\psi_t(Y)}{N_t(Y)}.$$

在 $N_t(Y)\ =\ 0$ 的地方将商设为零。其绝对值至多为 $(\alpha K\lambda_t)^{-1}$。在独立替换下，$h_s(V,\ Y)$ 的期望为零。因此，统计距离的标准有界测试推论给出

$$\left|\mathbb{E}\left[h_s(V,Y)h_s(V',Y)\mathbf{1}[\bar{S}(V,Y)=\bar{S}(V',Y)=\bar{s}]\,\frac{\psi_t(Y)}{N_t(Y)}\right]\right|\leqslant\frac{2\eta}{\alpha K\lambda_t}. \tag{4.30}$$

前一个主项难点正是在这里消失的。图和随机表都不是利用 Y 选取的。

将 (4.30) 乘以 $KJ_{t,\ s}$ 即可恢复对 y 和有序对的求和。因此

$$\mathcal{E}_{\text{large}}\leqslant\frac{2\eta L}{\alpha K^2}\sum_{t,s}\frac{J_{t,s}}{\lambda_t}.$$

由 (4.12)、(4.26) 和 (4.27)，

$$\sum_{t,s}\frac{J_{t,s}}{\lambda_t}\leqslant\sum_{t,s}\frac{a_tB_t}{\lambda_t}\leqslant\frac{2D^{2C+1}K^2}{\alpha M}.$$

由此可得

$$\mathcal{E}_{\text{large}}\leqslant\frac{4\eta D^{2C+2}}{\alpha^2}. \tag{4.31}$$

只有指数精确的种子误差会产生这一 D次幂。

中心化引理的完成 方程 (4.25)、(4.28) 和 (4.31) 给出 $\mathcal{E}\ \leqslant\Delta$。对次分布也使用半-$\ell_1$ 距离，三角不等式给出

$$\begin{aligned}\Delta(P,Q)&\leqslant\frac{b_P}{2}+\frac{1}{2}\|\mu-\nu\|_1+\frac{b_R}{2}+\Delta(\mathcal{R},Q)\\&\leqslant 2\beta+3\alpha+3\zeta+\frac{\eta}{2}+\frac{1}{2}\sqrt{\Delta}.\end{aligned}$$

这证明了引理 4.7。最终的理想分布保留了原始的比较边缘分布。中间次分布中使用的每次删除都包含在所显示的误差中。

4.2.2 主定理与完整的保留种子异或求值器

定理 4.8（源强端点）。对于每个固定的 $\delta>0$ 和整数 $C\geqslant 0$，存在 $b_{\delta,\ C}>0$，使得对于充分大的 n、整数 $\delta n\leqslant k\leqslant n$，以及

$$2^{-b_{\delta,C}n}\leqslant\varepsilon\leqslant 1/4,$$

存在一个显式多项式时间映射 $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$，满足

$$m \geqslant \frac{k}{C+1} - O(\log(1/\varepsilon)+1)$$

使得，对于每个独立的最小熵为k 的对以及每个可容许的 C-元组，

$$(F(X,Y),(F(f_i(X),g_i(Y)))_{i=1}^{C},Y) \approx_{\varepsilon} (U_m,(F(f_i(X),g_i(Y)))_{i=1}^{C},Y), \tag{4.32}$$

$$(F(X,Y),(F(f_i(X),g_i(Y)))_{i=1}^{C},X) \approx_{\varepsilon} (U_m,(F(f_i(X),g_i(Y)))_{i=1}^{C},X). \tag{4.33}$$

均匀分布的变量与所显示的其他所有变量相互独立。同一个函数满足这两个比较条件。加性输出损失常数可以取为绝对常数；误差范围常数和运行时间常数可以依赖于 δ、C。

构造 4.2 源强 C-篡改保留种子异或

前文部分中的组成要素。定理3.45 （构造 3.8）中的固定比较提取器；引理 2.5中的强种子提取器；以及引理 2.23中的中心化回收接口 4.7。

公开参数和固定设置。 固定 $\delta > 0$、 $C \geqslant 0$、整数 $\delta n \leqslant k \leqslant n$，以及定理 ε 4.8的范围内的目标 ε 。首先令 $q = C+1$、 $R = 2C+1$以及 $\sigma = \delta/(8q)$ 。根据定理 $G_0$3.45 构造一个基础提取器 R ，具有 率为 σ的比较。令

$$h_0 = \left\lceil \frac{\delta n}{8q} \right\rceil, \quad h = \left\lceil \frac{\delta n}{4q} \right\rceil, \quad u = \lceil \log(1/\varepsilon) \rceil + 10, \quad r_C = \lceil \log(2C+1) \rceil.$$

使用引理 A， B 中的常数2.23，取 A 为正整数，并选择

$$\begin{aligned} d &= \left\lceil B \left[\log^2(n+2) + (u+1)\log\left(2 + \frac{n}{u+1}\right) \right] \right\rceil, \\ m &= \left\lfloor \frac{k - (A+4)(u+1) - r_C}{q} \right\rfloor, \\ p &= m - d, \qquad \kappa = k - Cm - 2u. \end{aligned}$$

定义 $G =$ 前缀$_d G_0$。在熵阈值 E 和误差 κ 下实例化 2^{-u}，将其种子接口填充至 d 比特，并将其输出截断至 p 比特。界限（4.36）–（4.38）证实这些调用符合要求。从阈值 h_0 到 h 的强化只会改变保证，不会改变对 G 的求值。

输入。$(x,\ y) \in \{0,1\}^n \times \{0,1\}^n$。

求值步骤。

1. 计算 $S = G(x,y)$。**2.** 计算 $A_L = E(x,S)$ 和 $A_R = E(y,S)$，
二者长度均为 p。**3.** 返回保留种子与按位异或的串接：

$$S = G(x,y), \qquad F(x,y) = \big(S, E(x,S) \oplus E(y,S)\big). \tag{4.34}$$

输出。 恰好 $m = d+p$ 比特。定理 4.8 给出的误差至多为 ε， $m \geqslant k/(C+1) - O(\log(1/\varepsilon)+1)$，且对于同一函数，分别针对任一源都具有强度。

未评估。 $2C+1$ 辅助篡改、实际泄漏后验分布、碰撞图以及依赖于源的双射。求值器进行一次种子调用和两次种子提取器调用，而不是篡改输入调用。

对于非整数 $k \geqslant \delta n$，令 $k' = \lfloor k \rfloor$，并使用速率参数 $\delta/2$调用定理，而不是使用 δ：对于 $n \geqslant 2/\delta$，有 $k' \geqslant (\delta/2)n$。使用其误差范围常数 $b_{\delta/2,\ C}$ 以及长度截断值。输出界为 $k'/(C+1) - O(\log(1/\varepsilon) + 1)$，因此舍入至多会额外产生 $1/(C+1)$ 个比特，并被计入所述损失中。

4.2.3 回到对称求值器

令 F 为（4.34）。假设保留了 $Y = y$。对于每个输出，无论是诚实的还是经过篡改的，其种子都包含在输出中。因此，变换

$$(s, z) \longmapsto (s, z \oplus E(y, s))$$

对于诚实输出，以及

$$(s_i, z_i) \longmapsto (s_i, z_i \oplus E(g_i(y), s_i))$$

对于比较元组 i，都是由保留的 y 和显示的种子确定的双射。它们将诚实输出变换为 $(S_0,\ E(X,\ S_0))$，并将比较元组变换为

$$T = (\bar{S}, (E(f_i(X), S_i))_{i=1}^{C}).$$

这正是引理 4.7中的泄漏，其中 $a_{\bar{s}}(x) = (E(f_i(x)\ s_i))_i$。诚实均匀分布在同一双射下保持不变。因此，该引理以其显示的误差证明了（4.32）。

要证明（4.33），交换两个源及其篡改映射。对 $G^{\mathrm{op}}(x,\ y) = G(y,\ x)$应用同一引理，并使用 G 的另一个源强保证。求值器仍然恰好是（ 4.34 ）；只是分析过程进行了交换。

对于最小熵至少为整数 k的一般独立源，将每个源分解为定义在 2^k 个点上的平坦源的凸组合。对每个乘积分量使用相同的求值器和映射应用该界，然后取平均。理想分布保留每个分量的实际源和比较元组，因此取平均后得到它们原有的联合分布。可容许性在每个权重为正的分量上均得以保持。

4.2.4 参数计算

获得短源强种子 使用引理中的逐原子熵松弛强强化 2.5。

固定 $q = C + 1,\quad R = 2C + 1$，并选择速率为 $\sigma = \delta/(8q)$ 的定理 3.45 。将其输出和误差常数记为 $c_0,\ a_0 > 0$，并令

$$h_0 = \left\lceil \frac{\delta n}{8q} \right\rceil, \qquad h = \left\lceil \frac{\delta n}{4q} \right\rceil, \qquad a = \min\left\{ a_0, \frac{\delta}{8q} \right\}.$$

由于 $R + 1 = 2q$，引理 2.5 给出，对于 $d \leqslant c_0 n$，

$$\eta \leqslant D^{2q}(2^{-a_0 n} + 2^{h_0 - h}) \leqslant 2^{\,2+2qd-an}. \tag{4.35}$$

同一个 G 在阈值 h 处适用于中心化引理中所需的所有比较。

载荷长度和全部三个误差项 设置

$$s=\lceil\log(1/\varepsilon)\rceil,\qquad u=s+10,\qquad \alpha=\beta=\zeta=2^{-u}.$$

将 A 增大为正整数，并令

$$r_C=\lceil\log(2C+1)\rceil,\qquad m=\left\lfloor\frac{k-(A+4)(u+1)-r_C}{q}\right\rfloor,\qquad p=m-d.$$

取 d 为式（2.9）中界的上取整，并令 $s=u$。未使用的种子提取器种子比特可以忽略。取 E 处于阈值

$$\kappa=k-Cm-2u.$$

其输出之所以可用，是因为

$$\begin{aligned}\kappa-p&=k-qm+d-2u\\&\geqslant(A+4)(u+1)+r_C+d-2u\geqslant A(u+1).\end{aligned}$$

将正的误差范围常数 $b_{\delta,\,C}$ 取充分小，再将 n 取充分大，以确保

$$d\leqslant c_0n,\qquad p\geqslant 1, \tag{4.36}$$

$$4qd+4u+4\leqslant an, \tag{4.37}$$

$$k-Cm-(C+2)d-h\geqslant 4u+1. \tag{4.38}$$

这些选择是同时进行的。事实上，对于 $u\leqslant b_{\delta,\,C}n+O(1)$，种子长度界除以 n至多为

$$B\frac{\log^2(n+2)}{n}+B\left(b_{\delta,C}+O(1/n)\right)\log\left(2+\frac{1}{b_{\delta,C}+O(1/n)}\right)+O(1/n),$$

而这可以任意小。此外，$k-Cm\geqslant k/q\geqslant\delta n/q$ 与 $h\leqslant\delta n/(4q)+1$，从而在（4.38）中留有固定的正线性余量。最后，当$b_{\delta,\,C}$ 充分小时，m/n 仍与零保持正距离。

(4.11) 中的三项均至多为 2^{-2u}。对于第一项，

$$(2C+1)2^{qm-k}\leqslant 2^{-(A+4)(u+1)}\leqslant 2^{-2u}.$$

对于第二项，由 (4.38) 可得

$$2^{1+2u+(C+2)d+Cm+h-k}\leqslant 2^{-2u}.$$

对于第三项，由 (4.35) 和 (4.37) 可推出

$$\eta\leqslant 2^{-4u-2qd-2},\qquad 4\eta\,2^{2u+2qd}\leqslant 2^{-2u}.$$

因此，引理4.7 的误差至多为

$$8\cdot 2^{-u}+\frac{\eta}{2}+\frac{\sqrt{3}}{2}\,2^{-u}<10\cdot 2^{-u}<\varepsilon.$$

对称论证可得相同的界，从而证明定理 4.8。

由于 $r_C\leqslant C+1=q$，输出公式给出

$$m\geqslant\frac{k}{C+1}-\frac{(A+4)(u+1)}{C+1}-2=\frac{k}{C+1}-O(\log(1/\varepsilon)+1),$$

具有绝对加性损失常数。求值调用 G 一次，调用 E 两次，因此是多项式时间的。

4.3 推论：具有加性熵损失的普通源强输出

推论 4.9 （ $C = 0$ 特化情形）。 对每个固定的 $\delta > 0$，存在 $b_{\delta,\ 0} > 0$ 使得，对于充分大的 n、整数 $\delta n \leqslant k \leqslant n$以及 $2^{-b_{\delta,\ 0} n} \leqslant \varepsilon \leqslant 1/4$，存在一个显式确定性多源提取器，其输出为

$$m \geqslant k - O(\log(1/\varepsilon) + 1)$$

并且分别对任一源都是强的：

$$(F(X,Y), X) \approx_\varepsilon (U_m, X), \qquad (F(X,Y), Y) \approx_\varepsilon (U_m, Y).$$

加性　　损失常数是绝对常数，同一个函数满足这两种比较　　s.

构造 4.3 普通源强加性损失特化

输入 **前面部分中的推导**。 构造 4.2 和定理 4.8，专用于 $C = 0$。

公开参数和固定设置。 使用相同的率和误差输入，但设置为 $C = 0$、 $q = 1$、$R = 1$和 $r_C = 0$。因此 $u = \lceil \log(1/\varepsilon) \rceil + 10$、 $d = \lceil B \left[\log^2(n+2) + (u+1) \log\left(2 + \frac{n}{u+1}\right) \right] \rceil$、$m = \lfloor k - (A+4)(u+1) \rfloor$、 $p = m - d$，以及 $\kappa = k - 2u$。内部种子仍采用一次比较不可篡改性。

输入。$(x,\ y) \in \{0,\ 1\}^n \times \{0,\ 1\}^n$。

求值步骤。

1. 计算 d比特种子 S，使用构造 4.2 以及这些专用公开参数。**2.** 返回 $(S,\ E(x,\ S) \oplus E(y,\ S))$，使用其 p比特有效载荷。

输出。 $m = d + p \geqslant k - O(\log(1/\varepsilon) + 1)$ 比特，分别对任一源都是强的，如推论 4.9所示。这并不是一个新的放大过程；最终的比较元组为空，但仍然需要内部的不可篡改种子。

证明。 设 $C = 0$ 在定理 4.8中。于是 $q = 1$， $R = 1$，并且 $r_C = \lceil \log(2C+1) \rceil = 0$。比较元组 T 为空；在中心化引理中，它只有一个取值，其中 $\lambda_t = w_t(x) = 1$ 和 $N_t(y) = K$。例外对恰为对角线。其余步骤和参数不等式均仍然成立，能量界特化为

$$\Delta = \frac{L}{K} + \frac{2D^2\, 2^h}{\alpha^2 K} + \frac{4\eta D^2}{\alpha^2}.$$

同一个求值器保留 S并将其两个带种子的有效载荷异或，满足 $m = \lfloor k - (A+4)(u+1) \rfloor$以及 $u = \lceil \log(1/\varepsilon) \rceil + 10$。这证明了两个强比较，且损失如所述。 □

4.4 带加性熵损失的和集输出

4.4.1 前置条件：线性无损凝聚器

定理 4.10 （强线性无损凝聚器）。 参考文献。 [12, 定义 4.6 和引理 4.7]；另见 [36, 19]。对于任意固定的 $\alpha > 0$、 $k \leqslant n$ 和 $0 < \zeta < 1/2$，存在一个显式的强线性无损凝聚器 $C\colon \mathbb{F}_2^n \times \mathbb{F}_2^{d_C} \to \mathbb{F}_2^v$，其误差为 ζ，并且

$$d_C \leqslant (1 + 1/\alpha) \log(nk/\zeta) + O_\alpha(1), \qquad v \leqslant (1+\alpha)k + d_C + O_\alpha(1).$$

因此，在给定独立均匀种子的条件下，输出平均而言与一个最小熵至少为 k 的源相差 ζ。对于固定种子，该映射是二元线性的。对于固定的 α，求值可在多项式时间内完成，时间复杂度为 $n + \log(1/\zeta)$。

这是 [12, 定义 4.6 和引理 4。7], 其中记录了 GUV 凝聚器及其线性实现。

4.4.2 前置条件：和相容回收及其证明

令 $K = 2^k$、$D = 2^d$ 和 $M = 2^p$ 表示字母表大小。

引理 4.11（熵余量赋予加数强度）。如果一个普通的 d-比特和集提取器 G 具有分量阈值 h 和误差 γ，那么对于独立的 X，Y，其中 $H_\infty(X) \geqslant k > h$，$H_\infty(Y) \geqslant h$，

$$\Delta((X, G(X+Y)), (X, U_d)) \leqslant 2^d(\gamma + 2^{h-k}).$$

证明。对于每个输出值 u，令 $B_u = \{x : \Pr_Y[G(x+Y) = u] > 2^{-d} + \gamma\}$。如果 $\Pr[X \in B_u] \geqslant 2^{h-k}$，则在给定 B_u 的条件下，左侧熵 h 以及 Y的独立性保持不变，这与单个原子 u上的提取相矛盾。因此，每个 B_u 的质量都低于 2^{h-k}。在 B_u 之外，正原子超额至多为 γ，而在其中至多为一个。对 2^d 输出原子求和。这就是支撑引理2.5的单侧熵松弛论证。 □

推论 4.12（保留种子的和集回收）。对于每个 s，设 $H_s : \mathbb{F}_2^n \to \mathbb{F}_2^p$ 为线性映射，且为强种子 (k, ζ) 提取器。设 G 在熵为 k 时对第一个被加数是强的，误差为 η；并且在阈值为 $k-4$ 时是左侧单边不可篡改和集种子，误差为 γ，且具有相同的普通提取保证。若 $K \geqslant 16M$、$0 < \theta < 1$ 和 $\beta = 3(1+\theta)\zeta/\theta + 2\eta < 1$，则对于$S = G(X+Y)$，

$$\Delta((Y, S, H_S(X+Y)), (Y, U_{d+p})) \leqslant \beta + \tfrac{1}{2}\sqrt{\theta + 2\beta + 2(1+\theta)D^2\gamma + DM/K}. \tag{4.39}$$

舍弃 Y 即可得到普通和集结论。

~证明。将引理 4.1 应用于 $G(x, y) = G(x+y)$ 以及 $E(x, s) = H_s(x)$。它以所示误差比较 $(Y, S, H_S(X))$ 与 (Y, U_d, U_p)。应用 $(y, s, a) \mapsto (y, s, a + H_s(y))$。由线性，真实有效载荷变为$H_s(X+Y)$；固定平移在给定每个 (y, s) 的条件下，保持 U_p 的均匀分布。无需对 S 进行条件化。 □

推论 4.13（和集校准）。对于 $0 < \varepsilon \leqslant 1/4$，在推论 4.5 的假设下，式（4.39）中的界至多为 ε，尤其是在 $DM/K \leqslant \varepsilon^2/256$ 且其余参数保持不变时。

证明。应用共同校准（较小的对角项只会有帮助）。 □

4.4.3 前置条件：多项式剪切与适配种子长度的线性哈希

在推论 4.12中，不能用一般的非线性强种子提取器替代：其证明使用$H_u(x+y) = H_u x + H_u y$。尽管全输入通用哈希是线性的，但它未必能适配可用的和集种子。下面这个初等族在无损压缩器之后给出了我们所需的参数范围。

引理 4.14（多项式剪切哈希）。设 $E = \mathbb{F}_{2^t}$ 和 $b \geqslant 1$。对于 $u \in E$，定义

$$\mathsf{Sh}_u(z_0, \dots, z_b) = (z_1 + uz_0, z_2 + u^2 z_0, \dots, z_b + u^b z_0).$$

对于每个分布 熵至少为 k 的 $Z \in E^{b+1}$，以及独立均匀分布的 $U \in E$，

$$\Delta((U, \mathsf{Sh}_U(Z)), U_{(b+1)t}) \leqslant \tfrac{1}{2}\sqrt{b\, 2^{bt-k}}. \tag{4.40}$$

每个固 定种子映射都是二元线性的且为满射。种子长度为 t，而不是 $(b+1)t$。

证明。利用非退化的迹配对为 E^{b+1}的特征编索引。非零输出特征$\lambda = (\lambda_1, \dots, \lambda_b)$的拉回为

$$\left(\textstyle\sum_{i=1}^{b} \lambda_i u^i, \lambda_1, \dots, \lambda_b\right).$$

固定 $\lambda \neq 0$ 和固定的首个系数 η 后，方程 $\sum_i \lambda_i u^i = \eta$ 至多有 b 个根，因为其左侧减去 η 后是一个次数至多为 b 的非恒定多项式。若 $\widehat{\mu}(\xi) = \mathbb{E}\chi_\xi(Z)$，则

$$\frac{1}{|E|}\sum_u \sum_{\lambda \neq 0} \left|\widehat{\mu}\left(\sum_i \lambda_i u^i, \lambda\right)\right|^2 \leqslant \frac{b}{|E|}\sum_\xi |\widehat{\mu}(\xi)|^2 \leqslant b\, 2^{bt-k}.$$

最后一步是帕塞瓦尔恒等式，$\sum_\xi |\widehat{\mu}(\xi)|^2 = 2^{(b+1)t}\mathsf{cp}(Z)$，以及 $\mathsf{cp}(Z) \leqslant 2^{-k}$。傅里叶柯西–施瓦茨不等式将每个固定种子下的距离界定为其非零特征和的平方根的一半。对种子取平均并使用詹森不等式。线性性是显然的；通过令 $z_0 = 0$ 可得满射性。 □

引理 4.15（种子拟合线性提取）。固定 $0 < \delta \leqslant 1$ 和 $0 < a < \delta/8$。存在 $b_{\delta,\, a} > 0$使得，对于所有充分大的 n、所有整数 $\delta n \leqslant k \leqslant n$，以及所有整数 $2 \leqslant s \leqslant b_{\delta,\, a} n$，令

$$d = \lfloor an \rfloor, \qquad p = k - d - 2s - 8, \qquad \zeta = 2^{-4s-20}.$$

存在一个显式的强线性（k，ζ）带种子提取器，其种子比特数恰为 d，输出比特数为 p。常数和运行时间可能依赖于 δ，a，但 $d + p = k - 2s - 8$ 是精确的。

构造 4.4 种子拟合二元线性载荷提取器

取自前文的材料。定理 4.10 中的强线性无损凝聚器和多项式剪切哈希，引理 4.14。

公开参数和固定设置。固定 $\delta > 0$、$0 < a < \delta/8$、整数 $\delta n \leqslant k \leqslant n$，以及处于引理 4.15 范围内的整数 s。设定 $d = \lfloor an \rfloor$、$p = k - d - 2s - 8$ 和 $\zeta = 2^{-4s-20}$。选择一个固定整数$b \geqslant \max\{2, 16/a\}$以及 $\alpha = 1/(8b)$。在熵 k、参数 α 和误差 $\zeta/2$ 下实例化线性凝聚器 C；令其种子长度和输出长度分别为 d_C、v。置

$$w = 4s + 21, \quad \Lambda = 2w + \lceil \log b \rceil + 2, \quad t = \left\lfloor \frac{k - \Lambda}{b} \right\rfloor, \quad E = \mathbb{F}_{2^t}.$$

The 下面的证明验证 $v \leqslant (b+1)t$、$d_C + t \leqslant d$ 和 $p \leqslant bt$。固定该域的二元基。

输入。 $z \in \mathbb{F}_2^n$ 以及一个 d 比特的数值种子 S。

求值步骤。

1. 将前 d_C 个种子比特解析为 S_C，将接下来的 t 个比特解析为 $U \in E$，并忽略剩余的 $d - d_C - t$ 个比特。**2.** 计算 $\mathcal{C}(z,\ S_C)$，补零至 $(b+1)t$ 个比特，并将结果解析为 $(z_0,\ z_1,\ \ldots,\ z_b) \in E^{b+1}$。**3.** 计算 $(z_1 + Uz_0,\ z_2 + U^2 z_0,\ \ldots,\ z_b + U^b z_0)$，将结果写成 bt 个比特，并返回其前 p 个比特。

输出。 $H_S(z) \in \mathbb{F}_2^p$ 使用恰好 d 个给定的种子比特。对于每个固定的 S，z 中的映射是二元线性的。引理 4.15 给出了独立均匀种子下的误差 ζ。在下面的和集封装中，S 由内部生成，而推论 4.12（而非独立性假设）证明了这种用法的合理性。

证明。选择一个固定整数 $b \geqslant \max\{2,\ 16/a\}$ 并令 $\alpha = 1/(8b)$。应用定理 4.10，误差为 $\zeta/2$。将其种子长度和输出长度分别写为 $d_C,\ v$。存在一个常数 A_b 使得

$$d_C \leqslant A_b(\log(n+2) + s + 1), \qquad v \leqslant (1+\alpha)k + d_C + A_b.$$

定义这些整数

$$w = \log(2/\zeta) = 4s + 21, \qquad \Lambda = 2w + \lceil \log b \rceil + 2, \qquad t = \left\lfloor \frac{k - \Lambda}{b} \right\rfloor.$$

选择正数 $b_{\delta,\ a}$ 使其足够小。由于 $\log n = o(n)$，在指定的$k,\ s$ 以及所有足够大的 n 上均匀成立，我们可以确保

$$d_C \leqslant an/4, \tag{4.41}$$

$$d_C + 2\Lambda + A_b + b + 1 \leqslant k/(2b), \tag{4.42}$$

$$d + 2s + 8 \geqslant \Lambda + b, \qquad p \geqslant 1. \tag{4.43}$$

为了将其明确为一种选择顺序，首先固定 $b,\ \alpha,\ A_b$，然后选择 $b_{\delta,\ a}$，使得这三个不等式中s/n的系数都满足严格的裕量，最后提高长度阈值，以吸收对数项和常数项。所有所需的右侧裕量都是n的正倍数；没有任何参数取决于输入源。

原生哈希输入长度满足要求，因为

$$\begin{aligned}(b+1)t - v &\geqslant (1/b - \alpha)k - (1 + 1/b)\Lambda - (b+1) - d_C - A_b \\ &\geqslant 7k/(8b) - k/(2b) > 0.\end{aligned}$$

将浓缩器输出填充至 $(b+1)t$ 比特（用零填充），并应用引理 4.14，使用独立的t比特种子。给定浓缩器种子，其比较源的最小熵至少为 k。由于 $bt \leqslant k-\Lambda$，剪切误差至多为

$$\tfrac{1}{2}\sqrt{b}\,2^{-\Lambda/2} \leqslant \zeta/8.$$

总误差小于 ζ。对于任意固定的种子对，组合映射都是线性的。

其实际种子长度满足

$$d_C + t \leqslant an/4 + n/b \leqslant 5an/16 \leqslant \lfloor an \rfloor$$

当 n足够大时。用不参与计算的种子比特进行填充，使其恰好达到 d。其可用输出满足 $bt \geqslant k-\Lambda-b \geqslant p$，根据（4.43）；截取至 p 比特。所有映射都是显式的；对于固定的 b，α，其域运算和浓缩器计算均为多项式时间。 □

注 4.16（线性长度种子在此处的优势）。 对于常数误差，该构造可以保留 $d = \Theta_\delta(n)$ 个种子比特，并且只产生 $k - d - O(1)$ 个哈希比特。这并非损失：种子本身也包含在输出中。这里并未声称存在最优的 $O(\log n + \log(1/\varepsilon))$ 线性长度种子。将 d固定为较小的线性比例，也能为浓缩器和哈希留出足够空间，使其适用于直至某个足够小的线性比例的所有误差指数。

4.4.4 定理

定理 4.17（近最优和集输出）。 对于每个固定的 $0 < \delta \leqslant 1$，都存在 $b_\delta > 0$ 和 n_δ，使得只要 $n \geqslant n_\delta$，$\delta n \leqslant k \leqslant n$ 是一个整数，并且

$$2^{-b_\delta n} \leqslant \varepsilon \leqslant \tfrac{1}{4},$$

就存在一个显式确定性多项式时间映射

$$\mathrm{SumExt}_{n,k,\varepsilon} : \mathbb{F}_2^n \longrightarrow \mathbb{F}_2^m, \qquad m = k - 2\lceil \log(1/\varepsilon) \rceil - 8,$$

其在 $X+Y$ 上的输出对所有满足 $H_\infty(X)$ $H_\infty(Y) \geqslant k$ 的独立 X，Y 都是 ε-接近均匀分布的。特别是，

$$m \geqslant k - 2\log(1/\varepsilon) - 10.$$

运行时间和非平凡误差范围中的常数可能依赖于 δ，但所示的输出损失常数则不依赖于它。

4.4.5 完整的仅和保留种子构造

构造 4.5 带加性熵损失的和集输出

前文中的构件。 定理 3.33 (构造3.5）中的单侧不可篡改和集种子；引理 4.15 （构造 4.4）；以及推论 4.12 中的保留种子和集回收接口。

公开参数和固定设置。 固定 $\delta > 0$、整数 $\delta n \leqslant k \leqslant n$以及误差 ε 在定理 4.17的范围内。根据定理 G_0 3.33 以率 $\delta/8$实例化，输出和误差常数为 c，$e > 0$。固定$0 < a <$最小值$\{c/4,\ e/16,\ \delta/64,\ 1/64\}$ 并令

$$s = \lceil \log(1/\varepsilon) \rceil, \qquad d = \lfloor an \rfloor, \qquad p = k - d - 2s - 8.$$

设 $G =$前缀$_d G_0$，并使用这些 n、k、δ、a、s实例化构造 4.4。其 d比特种子接口在独立均匀种子下提供 p 有效载荷比特和误差 2^{-4s-20} 。

输入。 仅有 $z \in \mathbb{F}_2^n$。

求值步骤。

1. 计算 d比特种子 $S = G(z)$。**2.** 通过构造 $H_S(z)$ 计算 4.4，包括其种子拆分、凝聚器、多项式剪切和最终截断。**3.** 返回

$$\mathrm{SumExt}(z) = (S, H_S(z)), \qquad S = G(z). \tag{4.44}$$

输出。 恰好 $d + p = k - 2\lceil \log(1/\varepsilon) \rceil - 8$ 比特，误差至多为 ε 如定理 4.17所述。求值器从不接收任一加数。H_S 的线性是相对于 z 而言，并且对每个 固定的 种子成立；这并不是说整个依赖种子的求值器是线性的。

未求值。 加数、由 $H_S(y)$ 引起的平移、碰撞伙伴，以及使用独立种子的比较实验。

4.4.6 最终参数及主定理的证明

定理 *4.17*的证明。 固定 $\delta > 0$ 并以率 $\sigma = \delta/8$应用定理 3.33。令其输出和误差常数为 c，$e > 0$。选取一个正有理数

$$a < \min\{c/4, e/16, \delta/64, 1/64\}, \qquad d = \lfloor an \rfloor.$$

令 G 为前 d 个种子比特。截断保留单侧保证及其普通边缘分布，其中 $\gamma \leqslant 2^{-en}$。对于所有充分大的 n， $k - 4 \geqslant \sigma n$。引理 4.11 给出了第一个加数中的强度，其为

$$\eta \leqslant 2^d\big(2^{-en} + 2^{\sigma n - k}\big) \leqslant 2^{-(e-a)n} + 2^{-(7\delta/8 - a)n}.$$

令 $s = \lceil \log(1/\varepsilon) \rceil$，并令 $\varepsilon = 2^{-s} \leqslant \varepsilon$。选择 $b_\delta > 0$ 充分小，使得 $s \leqslant 2b_\delta n$ 在引理4.15对较大的 n所给出的范围内，并且在整个该范围内，

$$2^{-en} \leqslant 2^{-2s-9-2d}, \qquad 2^{-(e-a)n} + 2^{-(7\delta/8-a)n} \leqslant 2^{-2s-12}.$$

根据对 a的选择，这些不等式具有固定的正线性余量；减小 b_δ 即可使其均匀成立。由 s 的向上取整所产生的额外一项，则通过取 n 充分大来吸收。

使用引理 4.15 ，误差为 $\zeta = \varepsilon^4/2^{20}$ 且有效载荷长度为 $p = k - d - 2s - 8$。当 $K = 2^k$，$D = 2^d$，$M = 2^p$ 时，恰有

$$DM/K = 2^{-2s-8} = \varepsilon^2/256.$$

推论 4.13 的所有假设均成立。因此，仅和求值器（4.44）的误差至多为$\varepsilon \leqslant \varepsilon$。其长度恰为

$$d + p = k - 2s - 8 \geqslant k - 2\log(1/\varepsilon) - 10.$$

这就证明了定理，并且对所述区间内所有 k 都成立。 □

非平凡误差范围与全误差约定 上述有意义的构造适用于 $\varepsilon \geqslant 2^{-b_\delta n}$。这并不意味着对于每个更小的指定误差，仍能以系数 2 保证相同的非零输出界。若需要带有依赖于速率的常数的全误差表达式，可取 $C_\delta \geqslant \max\{12, 1/b_\delta + 2\}$。对于 $\varepsilon < 2^{-b_\delta n}$，界 $k - C_\delta(\log(1/\varepsilon) + 1)$ 非正，因为 $k \leqslant n$。此时返回空字符串，其分布恰好均匀。在非平凡范围内截断主要输出，并对有限多个较短长度作调整，即可得到关于所有误差的惯常表述，

$$m \geqslant \max\{0, \lfloor k - C_\delta(\log(1/\varepsilon) + 1)\rfloor\}$$

但不声称在线性指数范围已证明的下限以下仍能得到非空输出。

5 递减熵率

经评估的流程为：全对表（构造5.1）、头部/建议/破坏器短提取器（构造 5.3）以及保留种子的异或输出扩展（构造 5.4）。正则化仅用于证明该流程的性质。

固定 $0 < \vartheta < 1/2$ 以及一个非负整数篡改次数 t，且二者均独立于 n。本节分别证明在熵率为 $C/\log n$ 时的源强多重篡改提取，其输出为 $k/(t+1) - O(\log(1/\varepsilon) + 1)$。普通的近-$k$ 结论是其唯一推论。我们先证明增长行构件，然后构造并分析短不可篡改种子，最后给出中心化回收的校准过程和完整的长输出求值器。

5.1 增长行构件及其证明

以下引理和引用的定量定理是逆对数构造的前置条件。这里先给出完整的参数界，再介绍求值算法。内部引理将在下文证明；引入的原语则通过特定版本的定理引用加以标明。

5.1.1 引用的原语及其接口

我们使用第 2 节中的定量近似对偶、线性浓缩器和仿射某处随机输入。对于本节的局部记号，令 a_1 为仿射某处随机输入中一个充分大的绝对常数，并记 $H(T) = (2T)^{a_1 \log(2T)}$。我们使用以下相关性破坏器接口来生成短种子。为了获得最终的熵高效输出，我们使用引理 2.23 中的强种子提取器以及第4节中的中心化回收引理。

定理 5.1（定量仿射相关性破坏器）。 参考文献。 [52, 引理 3.16 和第 3.6.2 节，式（3.61）]；另见 [49, 定义 3.4 和定理 3.5]。设 $M = h + 1 \geqslant 2$、$v = \lceil \log(2M) \rceil$，并令 $u \geqslant m \geqslant 1$ 为整数。对于数据长度 N 和辅助信息长度 a，令 $z =$

$u+\lceil\log(N+2)\rceil$。存在一个显式均匀的函数 $\mathsf{ACB}\colon \{0,1\}^N\times\{0,1\}^d\times\{0,1\}^a\to\{0,1\}^m$，其参数选择为

$$d\leqslant C\big(Ma+Mv^3z\big),$$
$$K\leqslant C\big(Ma+(M^2+Mv^3)z\big),$$

具有以下性质。设 $b\geqslant 0$ 为种子缺陷容许量。假设 A_j，B_j，$Q_j\in\mathbb{F}_2^N$ 和 $R_j\in\mathbb{F}_2^d$ 是随机变量，其中对于 $0\leqslant j\leqslant h$，有 $Q_j=A_j+B_j$，且

$$(A_j)_{j=0}^h\ \perp\ ((B_j)_{j=0}^h,(R_j)_{j=0}^h),\qquad H_\infty(A_0)\geqslant K,\quad H_\infty(R_0)\geqslant d-b.$$

对于固定的 α_0，……，$\alpha_h\in\mathbb{F}_2^a$，若对每个 $j>0$ 都有 $\alpha_0\neq\alpha_j$，令 $Z_j=\mathsf{ACB}(Q_j,\ R_j,\ \alpha_j)$。那么

$$(Z_0,(Z_j)_{j=1}^h)\approx_{256M^3 2^{b-u}}(U_m,(Z_j)_{j=1}^h).$$

比较元组具有其实际的联合边缘分布，并且 U_m 与该元组独立。结论允许每个所示元组内部存在任意相关性，且比较输入不含熵。这里的参数范围满足 $d,\ K\leqslant N$。

这是 [52, 引理 3.16 和第 3.6.2 节，式 (3.61)] 的定量形式。关于种子亏损的损失 2^b可由此得到：将种子保留在条件元组中，并将其边缘分布改为均匀分布。破坏器所依据的变换也在 [49, 定义 3.4 和定理 3。5] 中给出。我们将此定理用作引用的定量工具。特别是，其独立性要求针对残余 A 部分与由种子决定的 B 部分，而非完整的数据元组和种子元组。

为清晰起见，缺陷转移使用所引用的定理，并保留其中的条件元组。令 $\mathcal{C}=((B_j)_{j=0}^h,\ (R_j)_{j=0}^h)$。将 R_0 的边缘分布替换为 U_d，保留给定 R_0 时 $\mathcal{C}$ 其余部分的条件分布，并保持独立的 A 元组不变。对于先前不在支撑集中的种子取值，任意定义条件分布。原分布是将新分布按 $w(r)=2^d\Pr[R_0=r]\leqslant 2^b$ 重新加权得到的。均匀种子定理中保留的元组包含R_0，因此对其真实分布和理想分布同时重新加权即可恢复原始的保留边缘分布，并使它们的半-ℓ_1 距离至多增加 2^b。这就证明了所展示的弱种子误差，而无需假设 R_0 与其他种子或 B 元组独立。

5.1.2 初等条件化与傅里叶事实

引理 5.2（平坦源分解）。对于一个整数 k，任何原子质量至多为2^{-k} 的分布都可以表示为若干个 2^k元素集合上的均匀分布的凸组合。

证明。 考虑多面体 $0\leqslant p_i\leqslant 2^{-k}$, $\sum_i p_i=1$。顶点不能有两个坐标严格处于各自的界限之间，因为进行幅度很小且方向相反的扰动后，约束仍然成立。它也不能恰好有一个这样的坐标，因为其他所有坐标之和是 2^{-k} 的整数倍。因此，每个顶点都恰好在 2^k 个点上服从均匀分布。

□

分别分解独立源。由此得到的分量对仍然相互独立；统计距离根据凸性取平均。篡改映射在所有分量上仍是原来的固定映射。

引理 5.3（条件化与截断）。 概率为 p 的事件至多损失 $\log(1/p)$ 比特的最小熵。从比特串中删除 b 个坐标至多损失 b 比特。将独立平坦源条件化于乘积矩形上，会保留其独立性和平坦性。

证明。 第一个界由 $\Pr[X = x \mid E] \leqslant \Pr[X = x]/p$ 得出。对于第二个界，每个保留的比特串至多有 2^b 个原像。对于第三个界，条件点概率可分解为两个均匀分布。

□

引理 5.4（输出包上的傅里叶逼近）。 设 P 支撑于仿射子空间$V \subseteq \mathbb{F}_2^D$上。若每个在 V上非常数的环境特征的绝对期望至多为 2^{-v}，则

$$\Delta(P, U_V) \leqslant 2^{\dim(V)/2 - v - 1} \leqslant 2^{D/2 - v - 1}.$$

证明。 在 V上选取仿射坐标。这些坐标中的每个非平凡特征都可扩展为一个环境特征，至多相差一个固定符号。对于 $d = \dim V$，由柯西–施瓦茨不等式和帕塞瓦尔恒等式可得

$$\Delta(P, U_V) \leqslant \frac{1}{2} \left(\sum_{\lambda \neq 0} |\widehat{P}(\lambda)|^2 \right)^{1/2} \leqslant \frac{1}{2} \sqrt{2^d - 1}\, 2^{-v}.$$

零维情形的距离为零。 □

5.1.3 通过分别特征进行联合转换

共同正则化定理适用于任意分别计算的特征，包括原始输入和篡改输入的元组。我们在此记录其归一化形式，以便后续的参数代换保留其原始符号。

设 $\phi\colon \mathbb{F}_2^n \to \mathbb{F}_2^{p_L}$和 $\psi\colon \mathbb{F}_2^n \to \mathbb{F}_2^{p_R}$为固定映射，令 $P = p_L + p_R$，并令 $F(x,\ y) = \mathcal{B}(\phi(x)\ \psi(y)) \in \mathbb{F}_2^D$，其中$\mathcal{B}$为固定的双线性映射。可以包含常数特征坐标。这些映射可以是非线性的、非单射的，并与任意独立篡改复合。

定理 5.5（特征正则化）。 存在绝对常数 K_0，使得以下结论成立。对于独立平坦源 $X,\ Y$、整数 $R,\ u \geqslant 1$和 $v \geqslant 0,\ z \geqslant 1$，存在一个有限的确定性乘积矩形划分，满足以下性质。

(i) 对每个叶节点 τ，输出分布与其输出仿射包上的均匀分布相差至多 $2^{D/2 - v - 1}$。

(ii) 除总质量至多为 2^{1-z}的叶节点外，每个输入的最小熵损失至多为

$$\mathcal{L} = K_0 \left[\frac{P}{u} + \frac{P}{R} \left(2^u (v+1) + (u+1) \log(n + P + 2) \right) + D(R+2) + z \right] \tag{5.1}$$

比特。

该划分仅用于证明，不由提取器计算。

证明。 如果 $P = 0$，输出为常数，结论显然成立。先假设 $1 \leqslant R \leqslant P$。将两个特征空间填充至相同维数 P，并应用定理 2.30，取 $r_* = R$、 $H = v$以及尾部参数 $t = z$。加权对偶界见引理 2.29。成功的限制计入特征包维数之和，精确消息则计入输出仿射包的维数。两个自信息估计给出所示损失，其中绝对常数 K_0增大。每个叶节点都是一个乘积矩形，且其精确输出仿射包上的傅里叶误差至多为 $2^{D/2 - v - 1}$。

为完整起见，若 $R > P$，则用该定理所采用的低秩协议计算每个 D 输出比特。每个双线性部分的秩至多为 P，因此总共只需至多 $D(P + 2)$ 个声明比特。该表在每个所得矩形上都是常数，因此在其

单点仿射包上均匀分布。至多含 $B_0 = D(P+2)$ 个声明比特的转录至多有 2^{B_0} 个叶节点。概率低于 2^{-B_0-z} 的叶节点，其总质量至多为 2^{-z}；对于所有其他叶节点，联合惊奇度以及每一方的熵损失均至多为 $B_0 + z \leqslant D(R+2) + z$。这也满足 (5.1)。因此，所有允许的截断函数都涵盖在内，包括非线性和非单射特征映射。

□

5.1.4 浓缩器表与联合仿射正则化

构造 5.1 全对有限域内积浓缩器表

来自前文的要素。定理 2.27 中的均匀线性 somewhere condenser，以及第 2.2 节中的固定有限域约定。

公开参数与固定设置。给定输入长度 n 和一个可接受的工作阈值k_0，在该阈值下实例化浓缩器，得到 $\mathcal{C}_1, \dots, \mathcal{C}_L \colon \mathbb{F}_2^n \to \mathbb{F}_2^{B_0}$。选取整数$1 \leqslant s \leqslant B_0$ 并令 $B = s\lfloor B_0/s \rfloor$, $T = L^2$，以及 $F = \mathbb{F}_{2^s}$。固定 F 的一个二元基，并将有序对$(i,j) \in [L]^2$ 按字典序排列。在命题 5.8适用的范围内，使用 $k_0 = k/4$ 和 $s = \lfloor n^{1/2-\vartheta/2} \rfloor$。

输入。 $(x, y) \in \{0,1\}^n \times \{0,1\}^n$。

求值步骤。

1. 对每个输入求每一行浓缩器行的值，并保留其前 B 比特。将每一行解析为 B/s个有限域 K 的域元素。

2. 对于每个 对 $(i,\ j) \in [L]^2$，计算

$$W_{ij}(x,y) = \sum_{e=1}^{B/s} \mathcal{C}_i(x)_e \, \mathcal{C}_j(y)_e \in F.$$

3. 按固定的对顺序返回这些 T 个域元素，每个元素由 s 比特表示。将表展平后得到 Ts 比特。

输出。 该表为 $W(x,\ y) \in (\mathbb{F}_2^s)^T$；它是一个中间对象，并非最终提取器本身。不同于多优行常率构造，该表保留全部 L^2 个对。

未求值。 联合仿射正则化、仿射包、优行的选择以及篡改表元组。它们用于证明该表的分布，但不是从 $(x,\ y)$ 计算得到的。

该表在 x、y 上是二元双线性的，因此其分别特征的总维数为 $P = 2n$，且与 T 无关。对于下面的证明，将凝聚器的优行率记为 $1/2 + \beta$，将其误差记为 $\varepsilon_c = 2^{-cB_0}$，其中常数取自定理 2.27。

引理 5.6（域内积特征界）。 若独立源 U、V 的长度为 B 比特，且熵分别为 h_U、h_V，则它们的域内积的每个非平凡二元特征的绝对偏差至多为

$$2^{(B-h_U-h_V)/2}.$$

证明。 每个此类特征都是 $\mathbb{F}_2^B \times \mathbb{F}_2^B$ 上的非退化二元双线性形式。相应的符号矩阵具有正交行，且每行的平方长度为 2^B，因此其算子范数为 $2^{B/2}$。两个源的概率向量的欧几里得范数分别至多为 $2^{-h_U/2}$ 和 $2^{-h_V/2}$。应用算子范数不等式即可。

□

引理 5.7（每个熵充分高的矩形都有一个完整行）。 假设 $s = o(B_0)$ 且 $B_0 \to \infty$。对于所有充分大的 n，如果平坦矩形的两边的熵都至少为 k_0，那么某一行

$W(A,\ B)$ 的仿射包投影是整个 $\mathbb{F}_2^s$。

证明。截断所损失的熵比特少于 s。因此，对于充分大的 n，理想凝聚器混合分布中的每个指定的良好行的熵至少为 $(1/2+\beta/2)B$。两个理想凝聚器混合分布可以独立采样，并且它们的乘积分布与实际凝聚器乘积分布相差不超过 $2\varepsilon_c$，其中 $\varepsilon_c = 2^{-cB_0}$。

假设每个输出行投影都有一个真仿射包。对于每一对行 $(i,\ j)$，在其实际输出支撑集上选择一个非平凡的二元特征常数，设其等于 b_{ij}。令 $\mathcal{E}$ 表示所有这些选定方程都成立的事件。其实际概率为 1。

在理想凝聚器混合分布的每个乘积分量中，令 $(i,\ j)$ 为指定的良好行对。根据引理 5.6，其选定方程成立的概率至多为 $(1+2^{-\beta B/2})/2$，其中常数已适当放宽。因此，无论其他行之间存在何种依赖关系，每个分量中 $\mathcal{E}$ 的概率都至多为该数量。取平均并恢复原始分布可得

$$1 \leqslant \frac{1+2^{-\beta B/2}}{2} + 2\varepsilon_c,$$

当 n 足够大时产生矛盾。 □

可复用的联合转换保证

命题 5.8（正则化参数）。固定 t，并令 k 为工作熵阈值。在阈值 $k/4$ 下构造凝聚器，并令 $T = L^2$ 为域乘积行的数量，每行长度为 s。对于 $k = C_{\vartheta,\ t}n/\log\ n$，取充分大的常数，并令 $s = \lfloor n^{1/2-\vartheta/2} \rfloor$ 和 $T = (\log\ n)^{O(1)}$。于是，对于边熵至少为 $k/2$ 的每个平坦源的乘积，以及任意 t 个独立篡改，联合表分布与由联合仿射元组 $(Q^0,\ \ldots,\ Q^t)$ 构成的混合分布的距离为 2^{-s}，其中 $Q^j \in (\mathbb{F}_2^s)^T$，且 Q^0 中的某一行恰好服从均匀分布。该优行可能因分量而异。

证明。对于 $t+1$ 次执行，使用分别特征 $(x, f_1(x) \ldots, f_t(x))$ 和 $(y, g_1(y) \ldots, g_t(y))$。它们的总维数为 $P = 2(t+1)n$，输出维数为 $D = (t+1)Ts$。应用定理 5.5，并取

$$v = D + 2s + 2, \qquad z = s + 2.$$

总近似误差与异常叶误差至多为 2^{-s}。只需使其熵损失为 $\mathcal{L} \leqslant k/4$。那么每个良好矩形的边熵至少为 $k/4$，而引理 5.7 给出了诚实仿射投影中的均匀行。在取平均之前，将异常叶替换为任意有效的仿射元组。

取 $u = \lfloor (\vartheta/8)\log\ n \rfloor$ 以及 $R = \lceil n^{1/2+\vartheta/4} \rceil$。在（5.1）中，除依赖于 t、ϑ 的常数因子外，各项均有如下上界：

$$\begin{aligned} P/u &= O_{\vartheta,t}(n/\log n), \\ (P/R)2^u(v+1) &\leqslant n^{1-5\vartheta/8+o(1)}, \\ (P/R)(u+1)\log(n+P+2) &\leqslant n^{1/2-\vartheta/4+o(1)}, \\ D(R+2) &\leqslant n^{1-\vartheta/4+o(1)}, \\ z &\leqslant n^{1/2-\vartheta/2+o(1)}. \end{aligned}$$

除第一个以外，其余都是 $o(n/\log\ n)$。因此，增加 $C_{\vartheta,\ t}$ 会得到 $\mathcal{L} \leqslant k/4$。这里有 $s = o(B_0)$，这正是满行引理所要求的。对于诚实表本身，同样的结论既可以直接得出，也可以通过舍弃比较坐标得出。

□

5.1.5 强普通种子与辅助信息

我们首先分别针对两个输入中的任一个，使普通种子变得强；并未声称它与两个输入整体具有独立性。

引理 5.9（两个独立的强保证）。设 $E_0\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^q$ 是一个误差为 ε_0、熵阈值为 K 的普通双源提取器。若 X, Y 相互独立，且各自的最小熵至少为 $K + b$，则 $S = E_0(X, Y)$ 满足

$$\Delta((S, X), (U_q, X)) \leqslant 2^q(\varepsilon_0 + 2^{-b}),$$
$$\Delta((S, Y), (U_q, Y)) \leqslant 2^q(\varepsilon_0 + 2^{-b}).$$

证明。固定一个输出值 s，并令 $p_s(x) = \Pr[E_0(x, Y) = s]$。设 $A_s = \{x\colon p_s(x) > 2^{-q} + \varepsilon_0\}$。若 $\Pr[X \in A_s] > 2^{-b}$，则条件源 $X|A_s$ 的熵至少为 K，且与 Y 独立。在这个矩形上输出 s 的概率严格大于 $2^{-q} + \varepsilon_0$，这与提取性质矛盾。因此 $\Pr[X \in A_s] \leqslant 2^{-b}$。于是

$$\mathbb{E}_X \big(p_s(X) - 2^{-q}\big)_+ \leqslant \varepsilon_0 + 2^{-b}.$$

对所有 2^q 输出值求和，得到第一个不等式。交换 X、Y 的角色，即得第二个不等式。任一源的均匀性以及对被篡改源的任何熵条件都不是必需的。

□

对于一个输出 q 比特且具有 $\varepsilon_0 \leqslant 2^{-4q}$ 的普通提取器，取熵余量 $b = 4q$，可得

$$\max\{\Delta((S, X), (U_q, X)), \Delta((S, Y), (U_q, Y))\} \leqslant 2^{1-3q}. \tag{5.2}$$

我们将验证第 5.2 节中所用种子满足所需的熵余量。

多项式指纹给出分离辅助信息 固定 $F = \mathbb{F}_{2^q}$ 的一个表示，并令$h = \lceil n/q \rceil$。将一个 n 比特串 x 用零填充至 hq 比特，将其连续的块解释为 x_0、 ..... 、 $x_{h-1} \in F$，并定义

$$P_x(T) = \sum_{j=0}^{h-1} x_j T^j \in F[T].$$

映射 $x \mapsto P_x$ 是单射。因此，只要 $x \neq x'$， $P_x - P_{x'}$ 就是一个次数至多为 $h - 1$ 的非零多项式。它在 F 中至多有 $h - 1$ 个根。对于 $h = 1$，它没有根。

定理 5.10（从双侧强种子生成辅助信息）。设 $S = S(X, Y) \in F$ 是两个独立源的确定性函数，并假设

$$\Delta((S, X), (U_q, X)) \leqslant \varepsilon_s, \qquad \Delta((S, Y), (U_q, Y)) \leqslant \varepsilon_s.$$

定义一个确定性辅助信息函数

$$\mathrm{Adv}(x, y) = \big(S(x, y), P_x(S(x, y)), P_y(S(x, y))\big). \tag{5.3}$$

它恰有 $a = 3q$ 比特。对于任意满足 $\Pr[(f_j(X), g_j(Y)) = (X, Y)] = 0$ 的 t 分割状态篡改（f_j, g_j），其辅助信息满足

$$\Pr\big[\exists j : \mathrm{Adv}(X, Y) = \mathrm{Adv}(f_j(X), g_j(Y))\big] \leqslant t\big(\varepsilon_s + (h - 1)2^{-q}\big). \tag{5.4}$$

这适用于任意篡改函数，对其输出没有熵要求。如果 S 是多项式时间的，则求值可在多项式时间内完成。

证明。 固定一个篡改（f, g），并将 $S' = S(f(X)\ g(Y))$ 简记为 $S' = S$。 X, Y 的独立性给出

$$\Pr[f(X) = X,\ g(Y) = Y] = \Pr[f(X) = X] \Pr[g(Y) = Y].$$

由于该乘积为零，因此其中一个因子为零。先设 $\Pr[f(X) = X] = 0$。辅助字符串相等意味着 $S = S'$，于是

$$P_X(S) = P_{f(X)}(S).$$

该蕴含关系的使用不以事件 $S = S'$ 为条件。对于 X 支撑集中的每个 x，集合

$$B_x = \{s \in F : P_x(s) = P_{f(x)}(s)\}$$

的大小至多为 $h - 1$。因此，强种子保证给出

$$\begin{aligned}\Pr[\mathrm{Adv}(X, Y) = \mathrm{Adv}(f(X), g(Y))] &\leqslant \Pr[S \in B_X] \\ &\leqslant \Pr[U_q \in B_X] + \varepsilon_s \\ &\leqslant (h - 1)2^{-q} + \varepsilon_s.\end{aligned}$$

当 $\Pr[g(Y) = Y] = 0$ 时，使用第三个辅助分量，并相对于 Y 应用强保证。对 $j \in [t]$ 求和这些界。域多项式使用霍纳法则求值；域运算和 $\mathbb{F}_{2^q}$ 的表示可以像域乘积构造中那样，在多项式时间内均匀实现。

□

这是采样码辅助信息所依据的标准指纹原理。它在这里的优势在于能够得到特别透明的精确界。证明不以种子相等为条件，不假设被篡改的种子服从均匀分布，也不假设它与任一诚实源相互独立。

5.1.6 固定前缀相关辅助信息后的联合仿射转换

这是允许使用仿射相关性破坏器的条件化步骤。它既避免了对非线性辅助信息进行特征展开，也避免了对所得执行元组作独立性假设。

引理 5.11（带分离侧信息的分支转换）。设 W 是命题 5.8 中的表，并将 $Q = (Q^0, \ldots, Q^t)$ 记为其联合执行。设 $\pi(Q) \in \{0,1\}^b$ 是一个固定的线性投影。对于每个给定值 $r \in \{0,1\}^b$，令 $L_r(x)$ 和 $R_r(y)$ 为任意的分离函数，每个函数至多输出 c 个比特。它们可以依赖于固定的篡改函数。定义实际侧信息为

$$J = (L_{\pi(Q)}(X), R_{\pi(Q)}(Y)).$$

假设 X, Y 是熵至少为 k、 $c + 2s \leqslant k/2$ 和 $b \leqslant s/4$ 的独立平坦源。那么（Q, J）在以下距离内

$$\delta = 2^{b+2-s} \tag{5.5}$$

具有以下性质的概率混合：每个分量上的 J 和 π（Q）均固定， Q 联合仿射，并且某个诚实行的熵至少为 $s - b$。更具体地说，该行是一个均匀分布的 s 比特行经过至多 b 个仿射约束后的投影。

证明。若 $\delta = 2^{b+2-s} \geqslant 1$，取一个均匀全表元组，固定 $\pi(Q) = 0$，并任意固定 J。所得仿射分布包含一个熵至少为 $s-b$ 的诚实行，且它与实际分布的距离至多为 $1 \leqslant \delta$。因此，不妨假设 $\delta < 1$。

暂时固定 r，不对 $\pi(Q) = r$进行条件化。分别对 $L_r(X)$ 和 $R_r(Y)$ 的取值进行条件化。这些都是乘积矩形。在左侧，概率小于 2^{-c-2s} 的标签总质量至多为 2^{-2s}，因为标签最多有 2^c 个。除这些标签外，条件熵至少为 $k - c - 2s \geqslant k/2$。右侧同理。

对每个良好标签矩形，将命题 5.8 应用于整个元组 Q。在不良标签矩形上，使用任意一个具有均匀诚实行的仿射元组。无论哪种情况，都保留固定的标签取值。我们构造了一个概率分布 P_r 于 $(Q,\ J_r)$上，其中 $J_r = (L_r(X)\ R_r(Y))$，满足

$$\Delta(\mathcal{L}(Q, J_r), \widehat{P}_r) \leqslant 2^{-s} + 2^{1-2s} \leqslant 2^{1-s}.$$

每个 P_r 的分量都具有固定标签和均匀的诚实行。

现在将两个分布都限制在事件 $\pi(Q) = r$ 上，不对这些限制后的测度进行归一化，然后对所有 r 求和。真实测度之和等于 $(Q,\ J)$ 的实际概率分布，因为前缀事件划分了样本空间。将近似测度之和记为 ν，其质量记为 w。实际分布与 ν 之间的半-ℓ_1 距离至多为 2^{b+1-s}。为单独界定 的质量，记 $P_r = \mathcal{L}(Q,\ J_r)$ 和 $E_r = \{\pi(Q) = r\}$。由于限制之前 P_r 和 P_r 都是概率分布，

$$|\widehat{P}_r(E_r) - P_r(E_r)| \leqslant 2^{1-s}.$$

真实事件的概率之和为 1，因此 $|w-1| \leqslant \sum_r |\widehat{P}_r(E_r) - P_r(E_r)| \leqslant 2^{b+1-s}$。因此，质量估计使用限制前概率分布的事件概率界；这与受限测度的半-ℓ_1 估计是分开的。由于 $\delta < 1$，我们有 $2^{b+1-s} = \delta/2 < 1/2$ 和 $w > 0$。将 ν 按其质量归一化。额外的半-ℓ_1 开销为 $|w-1|/2$，总计至多为 $\frac{3}{2}2^{b+1-s} \leqslant 2^{b+2-s}$。

将均匀仿射分布限制到 $\pi(Q) = r$ 后，会得到非空纤维上的均匀仿射分布。施加的独立仿射约束至多有 b 个，因此每个投影的维数至多减少 b。特别地，原先均匀分布的诚实行仍保有至少 $s-b$ 的熵。分量权重会重新归一化，但其仿射支撑集和固定标签保持不变。

□

对 2^b choices 的求和仅用于分析。实际计算出的提取器只计算其实际前缀和实际辅助信息。因此不会引入指数级运行时间。

5.1.7 仿射去相关与行合并

引理 5.12（相对于所有种子行的分解）。假设 $(V_0, \ldots, V_h, R_0, \ldots, R_h)$ 是联合仿射的。存在一个表示 $V_j = A_j + B_j$，使得

$$(A_0, \ldots, A_h) \perp (B_0, \ldots, B_h, R_0, \ldots, R_h),$$

and

$$H_\infty(A_0) = H(V_0 \mid R_0, \ldots, R_h) \geqslant H(V_0) - \sum_{j=0}^{h} |R_j|.$$

此处 H 表示以比特为单位的香农熵。对于均匀仿射分布，其香农熵和最小熵均等于其仿射维数。种子投影的所有非空纤维都具有

相同的方向空间，因此所示的条件熵就是它们共同的仿射维数；$|R_j|$ 表示种子 R_j 的比特长度。

证明。 将支撑集写成向量空间 V 的一个仿射平移。种子投影是作用于 V 上的线性映射。选取其核的一个补子空间： $V = V_{\ker} \oplus V_{\sec}$。从这两个子空间中独立均匀地选取元素，生成 V 上的均匀分布。将核中的选择投影到所有数据坐标，得到 $(A_j)_j$。将截面中的选择（包括仿射偏移量）投影到数据上，得到 $(B_j)_j$；截面中的选择也决定了所有种子。这两个集合相互独立。核在诚实坐标上的像，恰好是给定种子元组时 V_0 的公共方向。熵下界由投影到该元组的维数不等式得出。

□

引理 5.13 （带建议的仿射某处随机合并器）。 设 $T \geqslant 1$ 和 $t \geqslant 0$ 为满足 $(t+1)T \geqslant 2$ 的整数。设 $Q^0, \dots, Q^t \in (\mathbb{F}_2^s)^T$ 联合仿射，并假设 Q^0 的某一行的熵至少为 $s - b$，其中 $b \geqslant 0$。设 $\alpha_0, \dots, \alpha_t$ 为固定的 a_0 比特串，且对每个 $j > 0$ 都有 $\alpha_0 \neq \alpha_j$。令

$$N = Ts, \quad M = (t+1)T, \quad v = \lceil \log(2M) \rceil, \quad a = a_0 + \lceil \log T \rceil.$$

对于所需的输出长度 $q \geqslant 1$，应用定理 5.1 ，使用 $M - 1$ 次比较和原语指数

$$u = q + b + \lceil \log(1024 M^3) \rceil.$$

令 d， K 分别为其种子长度和熵阈值。若 $d \leqslant s$ 且

$$s - b - Md \geqslant K, \tag{5.6}$$

则下述公共函数在该元组上的 t-不可篡改误差至多为 $2^{-q}/4$ 。对于$i \in [T]$，定义 $\mathsf{bin}_T(i)$ 为 $i-1$ 的 $\lceil \log_2 T \rceil$ 比特二进制编码，并在前面填充零（当 $T = 1$ 时为空字符串）。使用 $\|$ 表示字符串拼接，令

$$\mathrm{Merge}(Q, \alpha) = \bigoplus_{i=1}^{T} \mathrm{ACB}\big(Q, \mathrm{prefix}_d(Q_i), \alpha \,\|\, \mathrm{bin}_T(i)\big). \tag{5.7}$$

每个表都被展平为 N 比特。该公式中的每个建议字符串都具有预先规定的公共长度 a，并且不同的行索引具有不同的编码。

构造 5.2 带建议的仿射某处随机合并器

前面部分的要素。 定理 5.1 中的定量仿射相关性破坏器。引理 5.12 仅在分析中提供其独立性分解。

公开参数和固定设置。 给定 T， s， t， a_0， b 以及目标输出长度 q，令

$$N = Ts, \quad M = (t+1)T, \quad a = a_0 + \lceil \log T \rceil, \quad u = q + b + \lceil \log(1024M^3) \rceil.$$

针对数据长度 N、建议长度 a、 $M-1$ 次比较、输出长度 q和指数 u，实例化该破坏器。令 d， K 分别为其种子长度和熵阈值。仅使用满足 $d \leqslant s$和 $s - b - Md \geqslant K$的参数。

输入。 一张表 $Q = (Q_1, \ldots, Q_T) \in (\mathbb{F}_2^s)^T$以及一个 a_0 比特的建议字符串 α。

求值步骤。

1. 将整张表 Q 展平为一个 N 比特的数据字。**2.** 对每个 $i \in [T]$，取 $R_i = \text{前缀}_d(Q_i)$，并将 $i-1$ 的 $\lceil \log_2 T \rceil$ 比特编码追加到 α。**3.** 对每个 $i \in [T]$计算 $Z_i = \mathsf{ACB}(Q,\ R_i,\ \alpha \,\|\, \mathrm{bin}_T(i))$，并返回 $\bigoplus_{i=1}^{T} Z_i$。

输出。 q 比特。引理 5.13 阐述了联合仿射比较表以及固定且不同的诚实/比较辅助信息所满足的不可篡改性保证。该算法求值 T 次，尽管其分析统计了所有执行中的 $M = (t+1)T$ 次调用。

未求值。优行与核–截面分解。求值器不对各行之间的独立性作任何假设。

证明。 在分析中选择一行优行 i_*。共有 M 次破坏器调用，按 (j, i) 编号，其中$0 \leqslant j \leqslant t$, $1 \leqslant i \leqslant T$。所选诚实调用为 $(0, i_*)$；其余所有 $M-1$ 次调用都是它的比较调用，包括诚实表中的其他行。其辅助信息与每个比较辅助信息均不同：行索引不等即可满足条件，而对于索引相同的情况，给定的 α 字符串彼此不同。

数据与种子元组是联合仿射的。诚实 d比特前缀的熵至少为 $d-b$，因为删除 $s-d$ 个坐标至多损失这么多熵比特。将引理 5.12 应用于全部 M 次调用。所选数据坐标 Q^0 的熵至少为 $s-b$，因此其独立残余部分的熵至少为 $s - b - Md \geqslant K$。仿射相关性破坏器的应用无需假设原始数据元组与种子元组相互独立。其误差至多为 $256M^3 2^{b-u} \leqslant 2^{-q}/4$。

因此，在给定其他每个破坏者输出的条件下，所选输出近似均匀分布。将其与其他诚实行的行输出进行异或不会破坏均匀性，而比较表异或是已保留输出的函数。这就证明了该结论。

□

如果固定建议字符串在这类分量的混合中有时相同，则丢弃这些分量。它们的总概率计入误差。当固定建议字符串已经不同时，无需满足关于 Q^0、 Q^j 的仿射无碰撞条件。

5.2 短输出不可篡改性：构造与证明

本小节中的构造适用于任意固定的正比较次数 t。在最终的 t 篡改定理中，将采用比较次数 $2t+1$ 和较小的源阈值。这里的证明独立于后面的长输出定理。

Cons构造 5.3 熵率递减时的短非篡改提取器 **e**

前面部分的要素。构造 5.1 中的表和命题 5.8；定理 2.28中的仿射某处随机提取器；定理 5.10中的多项式指纹；以及引理 5.13 (构造5.2)。

公开参数和固定设置。固定 $t \geqslant 1$， $0 < \vartheta < 1/2$，以及工作熵阈值 k 处于命题5.8的范围内。使用凝聚器阈值 $k/4$ 和 $s = \lfloor n^{1/2-\vartheta/2} \rfloor$构造该表。令$T = L^2$为其行数。设 $N = Ts$， $M = (t+1)T$，并设 $v = \lceil \log(2M) \rceil$。取下文使用的固定充分大的绝对常数 C_*，令

$$\ell = \left\lfloor \frac{s}{C_* M^3 v^3} \right\rfloor, \qquad q = \left\lfloor \frac{\ell}{8H(T)} \right\rfloor, \qquad b = M\ell, \tag{5.8}$$

其中 $H(T) = (2T)^{a_1} \log(2T)$ 来自定理 2.28。将该提取器应用于 T 行、每行 ℓ 比特，并将其输出截断至 q 比特。使用 $a_0 = 3q$、亏损容许值 b 和输出长度 q 实例化构造 5.2。特别地，其破坏器使用 $M-1$ 次比较和 $u = q + b + \lceil \log(1024M^3) \rceil$；将其种子长度记为 d。以下估计保证 $q \geqslant 4\log(n+2)$、$\ell/H(T) \geqslant (2T)^4, (t+1)q + 2s \leqslant k/2$、$b \leqslant s/4$ 和 $s - b - Md \geqslant K$。

输入。$(x, y) \in \{0,1\}^n \times \{0,1\}^n$。

求值步骤。

1. 计算该表。 根据构造 $W(x, y)$ 构造 5.1，保留其固定的行序 er.

2. 从头部计算一个短的普通种子。 对于每一行，取其前 ℓ 比特以形成 $\mathrm{Head}(W)$。应用所选的仿射某处随机提取器，并取其 q比特前缀，记为$S = S(W(x, y))$。

3. 根据完整原始输入计算辅助信息。 对 $\mathbb{F}_{2^q}$，将 x， y填充至 $\lceil n/q \rceil$ 个系数，并按照定理 5.10构造它们的多项式 P_x， P_y 。将 S 解释为有限域 K 中的域元素，并设置$\alpha = (S, P_x(S)\ P_y(S)) \in \{0,1\}^{3q}$。

4. 执行所有破坏器调用并进行异或。 对于每一行 i，使用完整展平的 Ts比特表作为数据，使用其行前缀 $\mathrm{prefix}_d(W_i)$ 作为种子，并使用 $\alpha \,\|\, \mathrm{bin}_T(i)$ 作为辅助信息。返回

$$E_{\mathrm{short}}(x, y) = \mathrm{Merge}(W(x, y), \alpha(x, y)). \tag{5.9}$$

输出。 $q = n^{1/2-\vartheta/2-o(1)}$ 反对数级区间中的比特，误差至多为 $2^{-q/2}$ 根据定理 5.15。头部种子 S 和逐行破坏器种子是不同的对象。此处仅返回破坏器输出的异或。

未求值。 假设头部的枚举、提示条件化、正则化矩形和仿射分量。求值器仅计算其实际表、实际头部种子和实际辅助信息。

普通种子提取与提示分离

引理 5.14（种子与辅助信息的保证）。 当源熵为 $k/2$ 时， $S(x, y)$ 是误差至多为 2^{-4q} 的普通提取器。当源熵为 k时，对于任一原始源，它都是强提取器，误差至多为 2^{1-3q}。对于每个可容许篡改列表，提示碰撞概率至多为

$$\zeta = t\left(2^{1-3q} + (\lceil n/q \rceil - 1)2^{-q}\right). \tag{5.10}$$

证明。 根据命题 5.8，熵为 $k/2$的源上的诚实表 2^{-s}接近于由仿射表构成、且具有一行均匀分布行的混合。它们的头部有一行均匀分布的 ℓ比特行。仿射某处随机

提取器因此给出的误差至多为 $2^{-\ell/H(T)}+2^{-s}\leqslant 2^{-4q}$。由凸性，第一个论证可从平坦源推广。由于 $k/2\geqslant 4q$ 当 n足够大时，引理 5.9 加上松弛量 $4q$ 给出了两个强保证。定理 5.10现在给出（5.10）。

□

辅助信息在最终联合转换之前固定 对于固定的篡改，设 $f_0=g_0=\mathrm{id}$ 和$Q^j=W(f_j(X)\,g_j(Y))$。所有 M行头部的拼接是完整表元组的 b 比特线性投影；将其记为 $r=\pi(Q)$。对于给定值 r，所有 q 比特种子 $s_j=S(Q^j)$ 都已确定。定义以下独立函数

$$L_r(x)=(P_{f_j(x)}(s_j))_{j=0}^{t},\qquad R_r(y)=(P_{g_j(y)}(s_j))_{j=0}^{t}.$$

每个函数都有 $c=(t+1)q$ 输出比特。完整的建议元组由 r、$L_r(x)$ $R_r(y)$ 确定。

应用引理 5.11。它给出了一个近似分布，距离至多为 $\delta=2^{b+2-s}$，其对应于所有表和所有辅助信息的联合分布；在该分布中，每个分量都具有固定的辅助信息、联合仿射表，并且拥有一个熵至少为 $s-b$ 的诚实行。请注意，指纹不必是表的仿射函数：在给定的头部固定后，它们是原始输入的独立函数，因此其固定发生在矩形正规化之前。

检验破坏者的熵不等式 令 $a=3q+\lceil\log T\rceil$ and $u=q+b+\lceil\log(1024M^3)\rceil$。上述参数条件意味着

$$a=O(\ell),\qquad u+\log(N+2)=O(M\ell).$$

定理5.1 因此给出

$$\begin{aligned}d&=O(M^2v^3\ell),\\K&=O\big((M^3+M^2v^3)\ell\big).\end{aligned}$$

因此

$$b+Md+K\leqslant C_1M^3v^3\ell\leqslant s$$

对于一个绝对常数 C_1，通过选择 $C_*\geqslant 2C_1$。这也给出 $d\leqslant s$ 并验证了 (5.6)。如有必要，进一步增大 C_*即可吸收所有整数取整误差。

定理 5.15 (短输出不可篡改性)。 函数 (5.9) 有 q 个输出比特，且其 t-不可篡改性误差至多为

$$\zeta+2^{-q}/4+3\cdot 2^{b+2-s}. \tag{5.11}$$

当 n 足够大时，此值至多为 $2^{-q/2}$。

证明。 首先考虑平坦源，并使用刚构造的联合近似。理想混合分布中的建议碰撞概率至多为 $\zeta+\delta$，因为该事件由保留的建议元组决定。对于其他每个分量，应用引理 5.13。逐分量误差至多为$2^{-q}/4$。将诚实/篡改输出元组还原为其实际分布的代价为 δ，而将理想尾部边缘分布还原为其实际分布还需付出 δ。这就证明了（5.11）。分别将任意独立源分解为平坦源并取平均；所有映射仍是原先固定的映射。

由于 $q\geqslant 4\log(n+2)$，我们有 $\zeta\leqslant 2t\,2^{-3q/4}$。此外 $b\leqslant s/4$ 和 $q\leqslant s/8$ 意味着 $2^{b+2-s}\leqslant 2^{2-6q}$。对于固定的 t 和充分大的 q，式 (5.11) 中它们的和至多为 $2^{-q/2}$。 □

逆对数参数计算以 $C_{\vartheta\text{ 的率},\ t}/\log\ n$，　$T=(\log\ n)^{O(1)}$ 和

$$M^3v^3H(T)=2^{O_{\vartheta,t}((\log\log n)^2)}=n^{o(1)}.$$

由此，　$s=\lfloor n^{1/2-\vartheta/2}\rfloor$ 得到

$$q=n^{1/2-\vartheta/2-o(1)}\geqslant n^{1/2-\vartheta}$$

对所有充分大的 n 均成立。特别地，　$q\gg\log\ n$, $\ell/H(T)\ \geqslant(2T)^4$, 以及 $(t+1)q+2s=o(k)$。这验证了上文使用的每个规模条件。

求值算法计算一个表、一个普通种子、两个指纹以及 T次相关性破除器调用。行数、以比特计的域大小和子程序输入长度均有多项式界。定量破除器关于其参数是均匀多项式时间的。算法既不枚举正则化划分，也不枚举 2^b 规定的头部。

5.3 具有加性熵损失的源强输出

上面的短输出构造仅用于生成种子。现在，我们应用第 4 节中的中心化回收引理。所得求值器保留其种子，并将两个源局部有效载荷进行异或。特别是，最终的源强性断言保留了所有实际的长比较输出，而不仅仅是它们的前缀。

5.3.1 前置条件：有限预算回收陈述

引理 5.16（带显式预算的中心化回收）。设 $t\geqslant 0$ 为整数，令 $q=t+1$且 $\rho_t=\lceil\log(2t+1)\rceil$，并设 A, $B\geqslant 1$ 为引理 2.23中的绝对常数，其中 A 为整数。对于整数 $n,k\geqslant 1$ 和 $u\geqslant 2$，若 $k\leqslant n$，令

$$d=\left\lceil B\left\lceil\log^2(n+2)+(u+1)\log\!\left(2+\frac{n}{u+1}\right)\right\rceil\right\rceil+A(u+1),\tag{5.12}$$

$$m=\left\lfloor\frac{k-4u-\rho_t}{q}\right\rfloor,\qquad p=m-d,\qquad \kappa=k-tm-2u.\tag{5.13}$$

假设 $1\leqslant h\leqslant k$、 $p\geqslant 1$和一个 d 比特种子 G 在分量熵阈值 h 下，针对至多 $2t+1$ 次独立篡改分别具有源强性，且误差 η 满足

$$\eta\leqslant 2^{-2qd-4u-4},\tag{5.14}$$

$$k-tm-(t+2)d-h\geqslant 4u+2.\tag{5.15}$$

存在一个显式强种子提取器 $E\colon\{0,1\}^n\times\{0,1\}^d\to\{0,1\}^p$，且映射

$$F(x,y)=\big(G(x,y),E(x,G(x,y))\oplus E(y,G(x,y))\big)\tag{5.16}$$

在所有独立的最小熵为k 的输入上分别是源强 t-不可篡改的，误差小于 $10\cdot\ 2^{-u}$。同一个求值器同时满足两种源强度比较，并保留完整的实际 t元组 m比特的篡改输出。

证明。Firs　检查带种子提取器接口。整数 κ 满足 $0<\kappa\leqslant k\leqslant n$，并且

$$\kappa-p=k-qm+d-2u\geqslant d+2u+\rho_t\geqslant A(u+1).$$

因此，引理 2.23 在阈值 κ 和误差 $\zeta = 2^{-u}$ 可以提供 p 输出比特，最多使用$\lceil B\left[\log^2(n+2) + (u+1)\log\left(2+\binom{n}{u+1}\right)\right]\rceil \leqslant d$ 个种子比特。忽略未使用的种子坐标，并根据需要截断输出。添加的 A $(u+1)$ 个种子坐标（见5.12）只是无害的填充；它们确保有效载荷能够容纳其中，同时不会将带种子提取器的熵损失常数计入最终输出亏损。

应用引理 4.7，其篡改次数为 $C = t$，$\alpha = \beta = \zeta = 2^{-u}$，且其阈值等于$h$。其三个能量项分别为

$$\begin{aligned}
\Delta_1 &= (2t+1)2^{qm-k} \leqslant 2^{-4u},\\
\Delta_2 &= 2^{1+2u+(t+2)d+tm+h-k} \leqslant 2^{-2u-1},\\
\Delta_3 &= 4\eta\, 2^{2qd+2u} \leqslant 2^{-2u-2}.
\end{aligned}$$

对于第一个不等式，使用 $2t+1 \leqslant 2^{\rho_t}$ 和 $qm \leqslant k - 4u - \rho_t$；另外两个恰好是（5.15）和（5.14）。特别地，$\Delta_1 + \Delta_2 + \Delta_3 \leqslant 3\cdot 2^{-2u}$。中心化引理的距离至多为

$$8\cdot 2^{-u} + \frac{\eta}{2} + \frac{\sqrt{3}}{2}2^{-u} < 10\cdot 2^{-u}.$$

为完整起见，当保留 $Y = y$ 时，从诚实有效载荷中减去 $E(y,\ s)$，并从比较 i 中减去 $E(g_i(y)\ s_i)$。这些是由 y 和所显示的种子值确定的双射。它们将诚实输出变为（$S_0,\ E(X,\ S_0)$），并将比较元组变为

$$T = (\bar{S}, (E(f_i(X), S_i))_{i=1}^{t}),$$

这正是引理 4.7 所允许的源局部泄漏。撤销这些双射会保留诚实均匀分布，并恢复原始的比较边缘分布。交换源，并对 $G^{\mathrm{op}}(x,\ y) = G(y,\ x)$ 应用相同论证可证明比较保留X；所评估的映射（5.16）不变。通过独立平坦源混合，可将证明扩展到平坦源，同时保留每个分量实际的比较/源边缘分布。当 $t = 0$比较元组为空时，种子仍具有单比较不可篡改性。

□

5.3.2 前置条件：校准低熵短种子

引理 5.17 (反对数型源强种子校准**)。** 固定 $t \geqslant 0$ 和 $0 < \vartheta < 1/2$。存在常数 $c_0,\ C_{t,\ \vartheta} > 0$ 以及一个整数 n_0，使得对所有 $n \geqslant n_0$、$C_{t,\ \vartheta} n/\log n \leqslant k \leqslant n$，以及

$$2^{-n^{1/2-\vartheta}} \leqslant \varepsilon \leqslant 1/4.$$

令 $q = t+1$、$R = 2t+1$,

$$h_0 = \lceil c_0 n/\log n\rceil, \qquad h = 2h_0, \qquad u = \lceil \log(1/\varepsilon)\rceil + 12, \tag{5.17}$$

并根据（5.12）–（5.13）选取 $d,\ m,\ p,\ \kappa$。在阈值 h_0和比较次数 R下实例化短输出构造。如果其输出长度为 q_0，则

$$G(x,y) = \mathrm{prefix}_d\, E_{\mathrm{short}}(x,y)$$

定义良好，并满足引理5.16的所有种子和长度假设，包括在阈值 h 处对源强度的两项比较。

证明。使用定理5.15 及其逆对数参数计算，其中比较次数为$R = 2t + 1$，指数参数为 ϑ。选择 c_0 足够大，使该构造在阈值h_0处成立。其长度和误差满足

$$q_0 = n^{1/2-\vartheta/2-o(1)}, \qquad \gamma_0 \leqslant 2^{-q_0/2}. \tag{5.18}$$

所有常数都可以依赖于固定的 t, ϑ，但不能依赖于 k 或 ε。这里调用的是短输出定理，而不是下文证明的长输出定理。

在所述误差范围内一致成立，

$$\begin{aligned} &u \leqslant n^{1/2-\vartheta} + 13, \\ &d = O\big(n^{1/2-\vartheta} \log n + \log^2(n+2)\big) = o(q_0). \end{aligned} \tag{5.19}$$

此外，$q_0 = o(h_0)$，因此 $d = o(h_0)$，尤其是 $d \leqslant q_0$。$1/2 - \vartheta$ 与$1/2 - \vartheta/2$ 之间的严格指数间隙可以吸收 d 中的因子 $\log n$ 以及所有固定的比较次数因子。因此，对于所有充分大的 n，我们有

$$\min\{q_0/2, h_0\} \geqslant 4qd + 4u + 5. \tag{5.20}$$

将引理2.5 应用于 d 位 R-篡改种子的前缀，将其阈值从 h_0 提高到$h = 2h_0$。由于 $R + 1 = 2q$，两个强误差都至多为

$$\begin{aligned} \eta &\leqslant 2^{2qd}\big(2^{-q_0/2} + 2^{-h_0}\big) \\ &\leqslant 2^{1+2qd-\min\{q_0/2,h_0\}} \leqslant 2^{-2qd-4u-4}. \end{aligned} \tag{5.21}$$

该估计对所有可容许的 R 比较以及最小熵至少为 h的所有独立源都一致成立。因此，它也适用于中心化回收引理中的辅助图边缘分布和复合比较；其篡改图像不需要具有任何熵。

取 $C_{t,\,\vartheta} \geqslant 16qc_0$ 并增大 n_0 以吸收上取整项。于是 $h \leqslant k/(4q)$。此外，$d + u = o(k)$ 关于 k 和 ε均匀分布，且 $m = k/q - O_t(u+1)$。因此 $p = m - d \geqslant 1$。使用 $m \leqslant k/q$可得

$$\begin{aligned} k - tm - (t+2)d - h &\geqslant \frac{k}{q} - (t+2)d - h \\ &\geqslant \frac{3k}{4q} - (t+2)d \geqslant 4u + 2 \end{aligned}$$

对所有充分大的 n。这些是引理 5.16的其余假设。关于 n 的所有下界都可以在特定的 k、ε之前选定，因为上述估计使用了它们所允许的最不利值。

□

5.3.3 主要定理与完整的保留种子构造

定理 5.18（速率为 $C/\log n$ 的源强熵高效输出）。固定一个整数 $t \geqslant 0$和 $0 < \vartheta < 1/2$。存在常数 $C_{t,\,\vartheta} > 0$ 和 n_0，使得对于每个 $n \geqslant n_0$、整数$C_{t,\,\vartheta} n/\log n \leqslant k \leqslant n$，以及

$$2^{-n^{1/2-\vartheta}} \leqslant \varepsilon \leqslant 1/4,$$

存在一个显式的多项式时间 t 不可篡改双源提取器 $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$，满足

$$m \geqslant \frac{k}{t+1} - O\big(\log(1/\varepsilon) + 1\big).$$

隐含的加性损失常数是绝对常数。对于每一对独立的最小熵为 k 的源，以及每一个可容许的 t 元组独立篡改，令 $\mathbf{Z}=(F(f_i(X)\ g_i(Y)))_{i=1}^{t}$。同一个函数满足

$$(F(X,Y),\mathbf{Z},Y)\approx_{\varepsilon}(U_m,\mathbf{Z},Y),\tag{5.22}$$

$$(F(X,Y),\mathbf{Z},X)\approx_{\varepsilon}(U_m,\mathbf{Z},X).\tag{5.23}$$

均匀字与所显示的完整边缘分布独立，而$\mathbf{Z}$ 由实际的 m比特比较输出组成。对于 $t=0$，这分别是关于任一原始源的普通提取，并且具有源强性质。对于不断增长的篡改次数，或对于固定常数下精确的平方根误差指数，不作任何声称。

Cons构造 5.4 具有加性熵损失的源强低速率输出 **s**

前面部分所用的材料。 定理 5.15 （构造 5.3）中的短提取器；引理 5.17中的种子校准；引理 2.23中的强种子提取器；以及引理 5.16中的显式预算回收接口。

公开参数和固定设置。 固定 $t\geqslant 0$、 $0<\vartheta<1/2$、整数$C_{t,\vartheta}n/\log n\leqslant k\leqslant n$，以及 $2^{-n^{1/2-\vartheta}}\leqslant\varepsilon\leqslant 1/4$。固定绝对常数 A、 B 和校准常数 c_0。首先设

$$q=t+1,\qquad R=2t+1,\quad \rho_t=\lceil\log(2t+1)\rceil,$$
$$h_0=\lceil c_0 n/\log n\rceil,\quad h=2h_0,\qquad u=\lceil\log(1/\varepsilon)\rceil+12.$$

选择

$$d=\left\lceil B\left\lceil\log^2(n+2)+(u+1)\log\left(2+\frac{n}{u+1}\right)\right\rceil\right\rceil+A(u+1),$$
$$m=\left\lfloor\frac{k-4u-\rho_t}{t+1}\right\rfloor,\qquad p=m-d,\qquad \kappa=k-tm-2u.\tag{5.24}$$

实例化构造 5.3 在阈值 h_0 和比较次数 R，而非 t；将其输出长度记为 q_0 以区别于 $q=t+1$。定义 $G=\text{前缀}_d E_{\text{short}}$。引理 5.17 给出 $d\leqslant q_0$ 以及在阈值 h下所需的源强保证。实例化 E 时，熵为 κ，误差为 2^{-u}， d 比特（忽略填充后），以及 p输出比特。

输入。 $(x,\ y)\ \in\{0,\ 1\}^n\times\{0,\ 1\}^n$。

求值步骤。

1. 在（$x,\ y$）上求值一次该较低阈值、 R次比较的短提取器，并保留其前 d比特作为 S。

2. 计算 p比特字 E（$x,\ S$）和 E（$y,\ S$）。

3. 返回

$$S=G(x,y),\qquad F(x,y)=\big(S,E(x,S)\oplus E(y,S)\big).\tag{5.25}$$

输出。 恰好 $m=d+p$ 比特，并满足定理 5.18 中的两项源强度保证。输出包含生成的种子，且不保留中间表或辅助信息。误差至多为 $10\cdot 2^{-u}<\varepsilon$（在所述参数范围内）。

未评估。 强化、正则化、后验分布、碰撞图以及辅助比较执行。这些用于认证三个已评估的调用，并非额外计算。

证明。 引理5.17 满足引理5.16中的所有种子、长度和熵假设。后者证明了两项比较，误差为 $10\cdot 2^{-u}\leqslant(10/4096)\varepsilon<\varepsilon$。由于 $\rho_t\leqslant t+1$，显式长度

满足

$$m \geqslant \frac{k}{t+1} - \frac{4(\lceil \log(1/\varepsilon) \rceil + 12)}{t+1} - 2 = \frac{k}{t+1} - O(\log(1/\varepsilon) + 1),$$

其中绝对常数。校准引理中，对足够大的n ，所有选取的长度均为正。短构造和两次带种子调用在固定的 t, ϑ 下均可在多项式时间内确定性完成。对于这一最终的非线性复合，无需秩条件。 □

在 $\varepsilon = 2^{-n^{1/2-\vartheta}}$ 处，定理给出 $m \geqslant k/(t+1) - O(n^{1/2-\vartheta}) = (1 - o(1))k/(t+1)$，并对完整输出提供两项源强度保证。这取代了旧的非对称增强及其短输出源强伴随构造。保留种子抵消了自身的长度开销：每个有效载荷中仅提取$p = m - d$ 比特，但会返回全部 d 个种子比特。

5.3.4 普通源强推论

推论 5.19 （逆对数速率下近乎-k 源强的普通输出）。 对于每个固定的$0 < \vartheta < 1/2$，存在 $C_\vartheta > 0$，使得对于每个充分大的 n、每个整数 $C_\vartheta n/\log n \leqslant k \leqslant n$，以及每个

$$2^{-n^{1/2-\vartheta}} \leqslant \varepsilon \leqslant 1/4,$$

存在一个显式的普通双源提取器，分别对任一原始源都具有强性质，且满足

$$m \geqslant k - 4\lceil \log(1/\varepsilon) \rceil - 56, \qquad \text{error at most } \varepsilon.$$

同一个求值器满足针对两个源的源强度比较。

构造 5.5 普通源强低速率特化

来自前文的组成部分。构造5.4 和定理5.18，以及 $t = 0$。

公开参数和固定设置。令 $q = 1$, $R = 1$, $\rho_0 = 0$，以及 $u = \lceil \log(1/\varepsilon) \rceil + 12$。保留构造5.4中的种子长度 d 和较低的种子阈值 h_0，并设定 $m = k - 4u$, $p = m - d$, $\kappa = k - 2u$。

输入。$(x, y) \in \{0,1\}^n \times \{0,1\}^n$ 属于推论5.19的参数范围。

求值步骤。

1. 在阈值 h处，利用单比较短种子计算 S $_0$.
2. 使用这些特定长度返回 $(S, E(x,S) \oplus E(y,S))$。

输出。 在任何可选的固定截断之前，恰好有 $k - 4\lceil \log(1/\varepsilon) \rceil - 48$ 比特。这意味着推论中所述的下界 $k - 4\lceil \log(1/\varepsilon) \rceil - 56$。同一个求值器对任一单独的源都是强的；内部种子仍处理一次比较。

证明。 令 $t = 0$ 为定理 5.18中的对象。其种子使用一次比较，其泄漏元组为空，并且$\rho_0 = 0$。(5.24) 中的显式输出长度为 $k - 4\lceil \log(1/\varepsilon) \rceil - 48$，这至少达到所述界；额外的固定截断不会造成影响。两个强比较均可直接得到。该推论中不存在单独的输出放大或碰撞论证。 □

参考文献

[1] Divesh Aggarwal、Yevgeniy Dodis 和 Shachar Lovett。基于加性组合学的不可篡改码。载于第*46*届 *ACM* 计算理论年会论文集，2014年。[2] Omar Alrabiah、Jesse Goodman、Jonathan Mosheiff 和 João Ribeiro。低次多项式是良好的提取器，2025年。arXiv:2405.10297v2；定理9。[3] Srinivasan Arunachalam、Davi Castro-Silva、Arkopal Dutt 和 Tom Gur。算法化多项式 Freiman–Ruzsa 定理。技术报告 TR25-130，计算复杂性电子期刊，2025年。引理 3.3（受限同态测试）。[4] Boaz Barak、R. Impagliazzo 和 Avi Wigderson。使用少量独立源提取随机性。载于第*45*届 *IEEE* 计算机科学基础年会论文集，第384–393页，2004年。[5] Boaz Barak、Guy Kindler、Ronen Shaltiel、Benny Sudakov 和 Avi Wigderson。模拟独立性：凝聚器、Ramsey图、分散器和提取器的新构造。载于第*37*届 *ACM* 计算理论年会论文集，第1–10页，2005年。[6] Boaz Barak、Anup Rao、Ronen Shaltiel 和 Avi Wigderson。超越 Frankl-Wilson构造的 $n^{o(1)}$熵双源分散器与 Ramsey 图。载于第*38*届 *ACM* 计算理论年会论文集，2006年。[7] Avraham Ben-Aroya、Dean Doron 和 Amnon Ta-Shma。从双源提取器到不可篡改提取器的高效归约：达到近对数最小熵。*SIAM*计算杂志，51(2):STOC17–31–STOC17–49，2022年。[8] Eli Ben-Sasson、Shachar Lovett 和 Noga Ron-Zewi。联系秩与通信复杂度的加性组合学方法。*J. ACM*，61(4):22:1–22:18，2014年。[9] Eli Ben-Sasson 和 Noga Ron-Zewi。从仿射提取器到双源提取器：通过近似对偶性。*SIAM*计算杂志，44(6):1670–1697，2015年。[10] Abhishek Bhowmick、Zeev Dvir 和 Shachar Lovett。匹配向量族的新界，2013年。arXiv:1204.1367v2；引理7.9及其证明。[11] Jean Bourgain。关于素域中的和积现象及其应用的进一步研究。国际数论杂志，1:1–32，2005年。[12] Eshan Chattopadhyay、Jesse Goodman 和 Jyun-Jie Liao。适用于近对数熵的仿射提取器。载于第*62*届 *IEEE* 计算机科学基础年会，*FOCS 2021*，美国科罗拉多州丹佛，*2022*年*2*月*7*日至*10*日，第622–633页。IEEE，2021年。[13] Eshan Chattopadhyay、Vipul Goyal 和 Xin Li。不可篡改提取器与码，以及其众多篡改扩展。载于第*48*届 *ACM* 计算理论年会论文集，2016年。[14] Eshan Chattopadhyay 和 Xin Li。显式不可篡改提取器、多源提取器与近乎最优的隐私放大协议。载于第*57*届 *IEEE* 计算机科学基础年会论文集，2016年。

[15] Eshan Chattopadhyay 和 Xin Li。和集源提取器。载于 Daniel Wichs 和 Yishay Mansour 编，第*48*届*ACM SIGACT*计算理论年会论文集，*STOC*，剑桥，美国马萨诸塞州，*2016*年*6*月*18*日至*21*日，第299–311页。ACM，2016年。[16] Eshan Chattopadhyay 和 Jyun-Jie Liao。双源和提取器。载于 Stefano Leonardi 和 Anupam Gupta 编，*STOC '22*：第*54*届*ACM SIGACT*计算理论年会，意大利罗马，*2022*年*6*月*20*日至*24*日，第1584–1597页。ACM，2022年。[17] Eshan Chattopadhyay 和 David Zuckerman。显式双源提取器与抗扰函数。《数学年刊》，189：653–705，2019年。[18] Mahdi Cheraghchi 和 Venkatesan Guruswami。针对逐比特篡改和分割状态篡改的不可篡改编码。载于*TCC*，第440–464页，2014年。[19] Mahdi Cheraghchi 和 Piotr Indyk。稀疏 Walsh–Hadamard 变换的近最优确定性算法。载于第*27*届*ACM-SIAM*离散算法年会（*SODA 2016*）论文集，第1036–1045页。工业与应用数学学会，2016年。[20] Benny Chor 和 Oded Goldreich。如何从弱随机性源中获得无偏比特及概率通信复杂性。*SIAM*计算杂志，17(2)：230–261，1988年。[21] Eldon Chung、Maciej Obremski 和 Divesh Aggarwal。提取器：低熵要求与不可篡改性的冲突。技术报告，arXiv，2021年。[22] Gil Cohen。局部相关性破坏器及其在三源提取器和合并器中的应用。载于第*56*届*IEEE* 计算机科学基础年会论文集，2015年。[23] Gil Cohen。充分利用建议：新相关性破坏器及其应用。载于第*57*届 *IEEE* 计算机科学基础年会论文集，2016年。[24] Gil Cohen。针对多对数熵的双源分散器及改进的 Ramsey 图。载于 Daniel Wichs 和 Yishay Mansour 编，第*48*届*ACM SIGACT*计算理论年会论文集，*STOC 2016*，剑桥，美国马萨诸塞州，*2016*年*6*月*18*日至*21*日，第278–284页。ACM，2016年。[25] Gil Cohen。面向最优双源提取器和 Ramsey 图。载于 Hamed Hatami、Pierre McKenzie 和 Valerie King 编，第*49*届*ACM SIGACT*计算理论年会论文集，*STOC 2017*，蒙特利尔，加拿大魁北克省，*2017*年*6*月*19*日至*23*日，第1157–1170页。ACM，2017年。[27] Gil Cohen 和 Leonard Schulman。近对数最小熵提取器。载于第*57*届*IEEE* 计算机科学基础年会论文集，2016年。[28] Yevgeniy Dodis、Ariel Elbaz、Roberto Oliveira 和 Ran Raz。从两个独立源中改进随机性提取。载于*RANDOM 2004*，第334–344页，2004年。

[29] Yevgeniy Dodis、Xin Li、Trevor D. Wooley 和 David Zuckerman。通过特征和实现隐私放大和不可篡改提取器。*SIAM*计算杂志，43(2)：800–830，2014年。[30] Yevgeniy Dodis 和 Daniel Wichs。基于弱秘密的不可篡改提取器与对称密钥密码学。载于第*41*届*ACM*计算理论年会，第601–610页，2009年。[31] Dean Doron 和 Jesse Goodman。低误差提取器。正在准备中的手稿，2026年。[32] Zeev Dvir、Swastik Kopparty、Shubhangi Saraf 和 Madhu Sudan。重数方法的扩展及其在 Kakeya 集和合并器中的应用。载于第*50*届*IEEE*计算机科学基础年会论文集，2009年。[33] Zeev Dvir 和 Avi Wigderson。Kakeya 集、新合并器和旧提取器。*SIAM*计算杂志，40(3)：778–792，2011年。[34] Stefan Dziembowski、Krzysztof Pietrzak 和 Daniel Wichs。不可篡改码。载于*ICS*，第434–452页，2010年。[35] W. T. Gowers、Ben Green、Freddie Manners 和 Terence Tao。关于 Marton 猜想。《数学年刊》，201：515–549，2025年。[36] Venkatesan Guruswami、Christopher Umans 和 Salil Vadhan。基于 Parvaresh–Vardy 码的非平衡扩展图和随机性提取器。*ACM*期刊，56(4)：文章20，2009年。定理5.12和5.14；注5.15。[37] A. A. Karatsuba。关于某个算术和。苏联数学文献，12：1172–1174，1971年。苏联科学院院报，199(4)：770–772，1971年的英文译本。[38] A. A. Karatsuba。狄利克雷特征在加性序列上的取值分布。苏联科学院院报，319(3)：543–545，1991年。英文译本：数学文献，44(1)：145–148，1992年。[39] Mark Lewko。一个具有近 $4/9$最小熵率的显式双源提取器。*Mathematika*，65(4)：950–957，2019年。[40] Xin Li。三源提取器的改进构造。载于第*26*届*IEEE*计算复杂性年会论文集，第126–136页，2011年。[41] Xin Li。一种构造仿射提取器和分散器的新方法。载于第*26*届*IEEE*计算复杂性年会论文集，第137–147页，2011年。[42] Xin Li。不可篡改提取器、双源提取器和隐私放大。载于第*53*届 *IEEE* 计算机科学基础年会论文集，第688–697页，2012年。[43] Xin Li。针对具有多对数最小熵的常数个独立源的提取器。载于第*54*届 *IEEE* 计算机科学基础年会论文集，第100–109页，2013年。[44] Xin Li。具有指数级改进的新型独立源提取器。载于第*45*届*ACM*计算理论年会，第783–792页，2013年。

[45] Xin Li。多对数最小熵的三源提取器。载于第*56*届 *IEEE* 计算机科学基础年会论文集，2015年。[46] Xin Li。改进的双源提取器，以及多对数熵的仿射提取器。载于第*57*届 *IEEE* 计算机科学基础年会论文集，2016年。[47] Xin Li。改进的不可篡改提取器、不可篡改码和独立源提取器。载于第*49*届 *ACM* 计算理论年会论文集，2017年。[48] Xin Li。不可篡改提取器与不可篡改码：部分最优构造。载于 Amir Shpilka 编，第*34*届计算复杂性会议，*CCC 2019*，*2019*年*7*月*18*日至*20*日，美国新泽西州新不伦瑞克，LIPIcs 第137卷，页码 28:1–28:49。Schloss Dagstuhl - Leibniz-Zentrum für Informatik，2019年。[49] Xin Li。具有渐近最优熵的双源提取器，以及（许多）其他结果。载于第*64*届计算机科学基础年会，*FOCS 2023*，美国加利福尼亚州圣克鲁斯，*2023*年*11*月*6*日至*9*日，页码 1271–1281。IEEE，2023年。[50] Xin Li 和 Yan Zhong。显式定向仿射提取器及线性分支程序的改进困难性。载于第*39*届计算复杂性会议（*CCC 2024*），LIPIcs 第300卷，页码 10:1–10:14，2024年。另见 arXiv:2304.11495v2 [cs.CC]。[51] Xin Li 和 Yan Zhong。小熵的双源与仿射不可篡改提取器。载于第*51*届自动机、语言与编程国际研讨会（*ICALP 24*），LIPIcs 第297卷，页码 108:1–108:18，2024年。另见 arXiv:2404.17013 [cs.CC]。[52] Xin Li 和 Yan Zhong。具有小误差和复杂性下界的不可篡改仿射提取器。技术报告，arXiv，2026年。arXiv:2610.02407v1。[53] Mehdi Makhul 和 Arne Winterhof。有限域上沿多项式值的 Thue–Morse 函数的正规性。数论研究，8:38，2022年。引理1。[54] Anup Rao。常数个多项式级小最小熵独立源的提取器。载于第*38*届 *ACM* 计算理论年会论文集，2006年。[55] Anup Rao。低权仿射源的提取器。载于第*24*届 *CCC* 论文集，2009年。[56] Ran Raz。使用弱随机种子的提取器。载于第*37*届 *ACM* 计算理论年会论文集，页码 11–20，2005年。[57] Victor Shoup。寻找有限域上不可约多项式的新算法。计算数学，54:435–447，1990年。[58] David Zuckerman。线性次数提取器与最大团及色数的不可近似性。计算理论，3:103–128，2007年。定理 4.3，归于 Raz。