

Jet and Anti- k_t Algorithm

Zhonghua Zhang | Department of Modern Physics | USTC

- **What's Jet: Origin and Evolution**
- **Anti-kt clustering algorithm**
- **Summary**

The anti- k_t jet clustering algorithm

Matteo Cacciari and Gavin P. Salam

LPTHE, UPMC Université Paris 6,

Université Paris Diderot — Paris 7,

CNRS UMR 7589, Paris, France

E-mail: cacciari@lpthe.jussieu.fr, salam@lpthe.jussieu.fr

Gregory Soyez

Brookhaven National Laboratory,

Upton, NY, U.S.A.

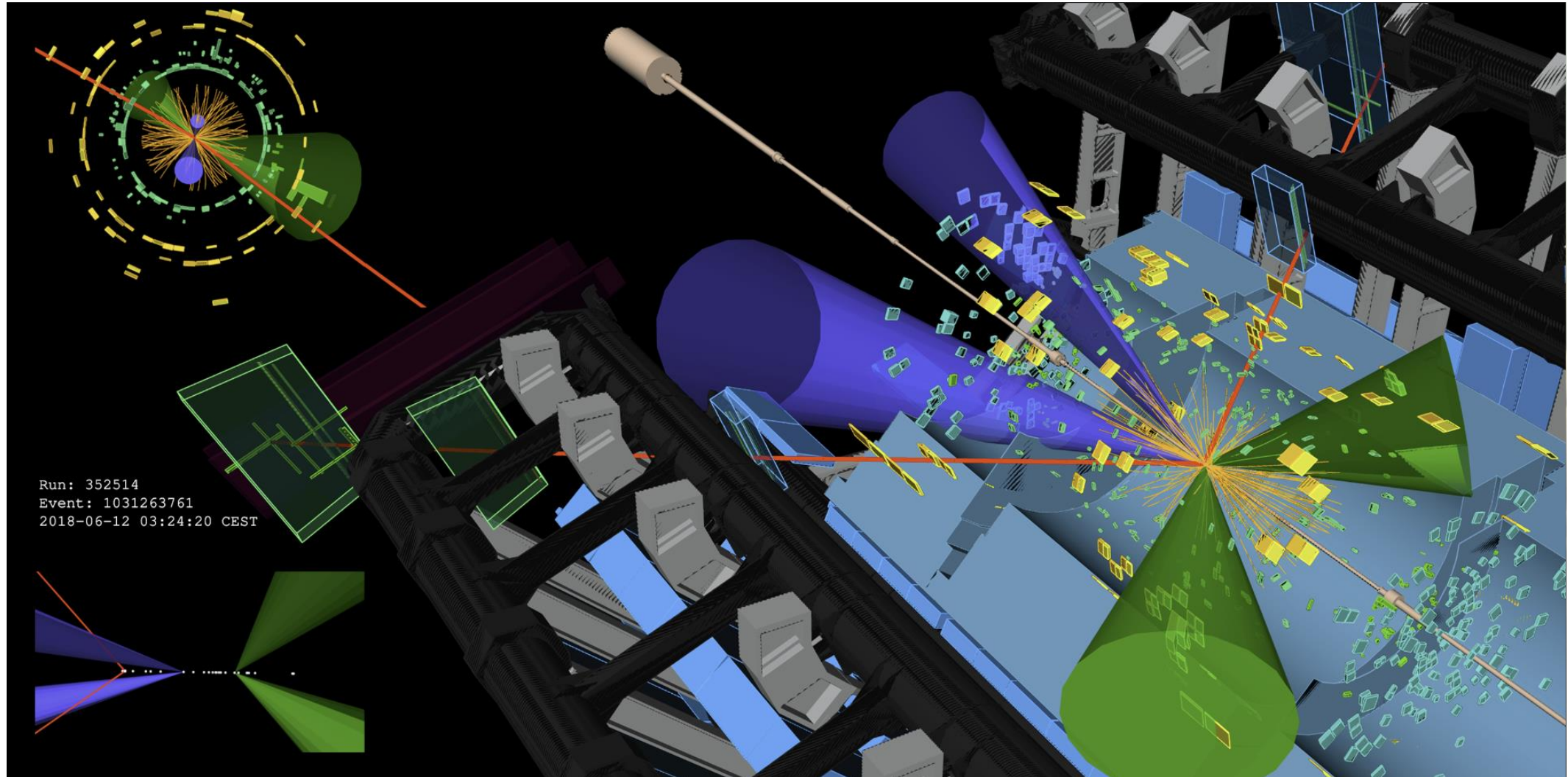
E-mail: gsoyez@quark.phy.bnl.gov

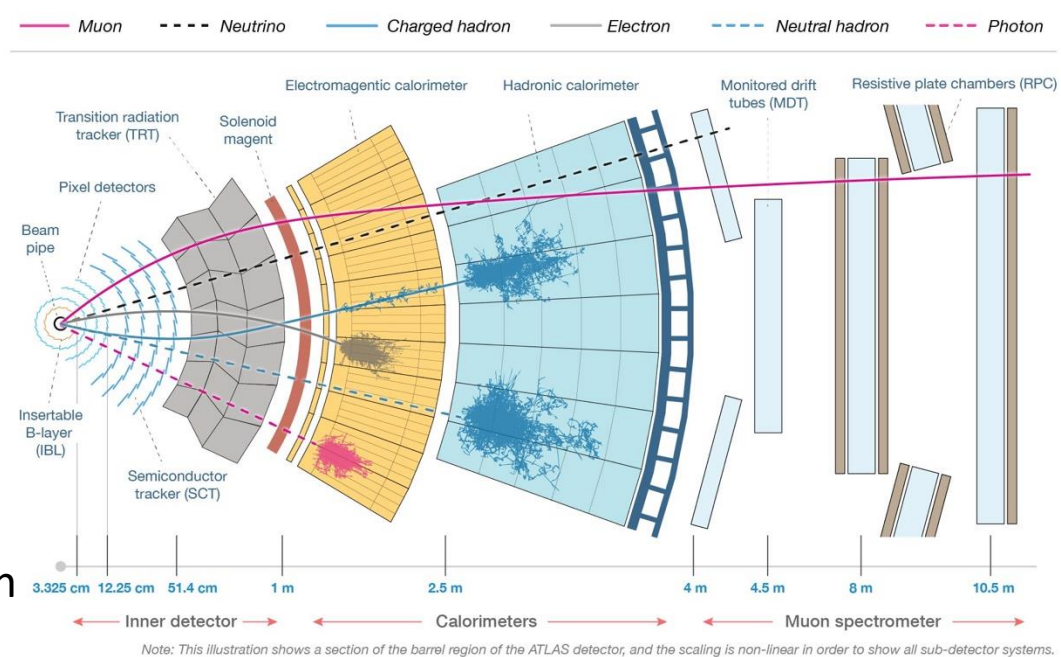
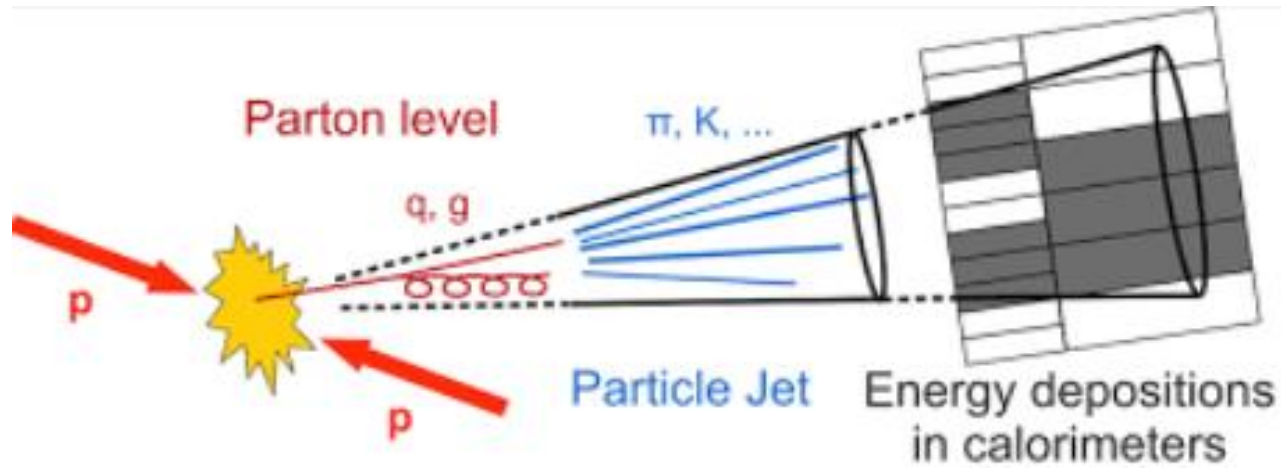


Part I

Jet Origin and Evolution

Blue and green cone: recombined by jet clustering algorithm



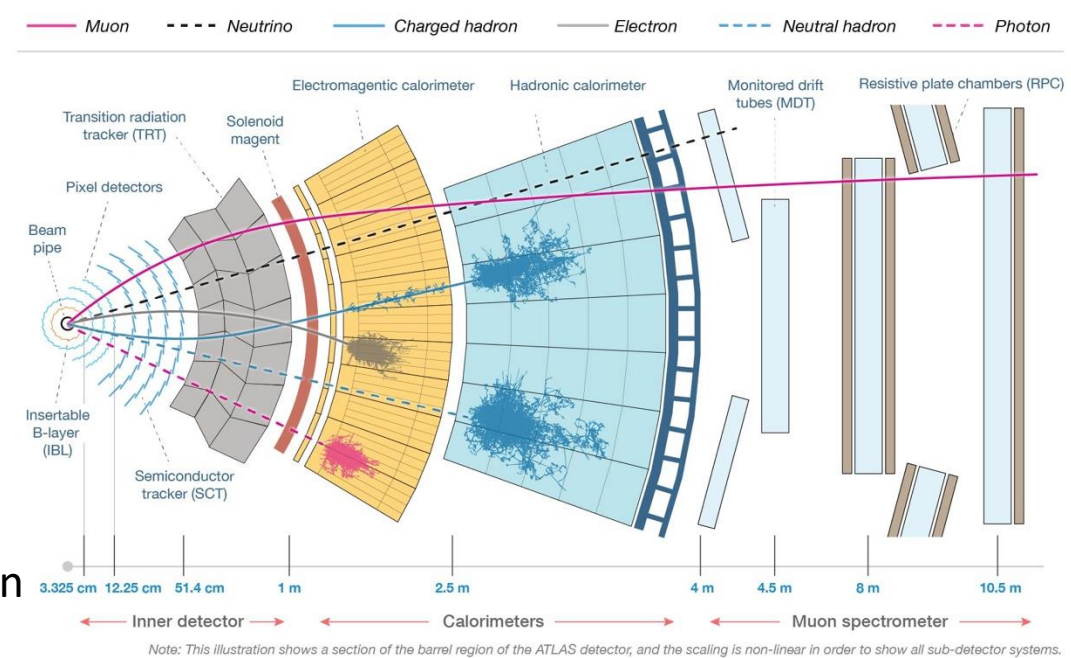
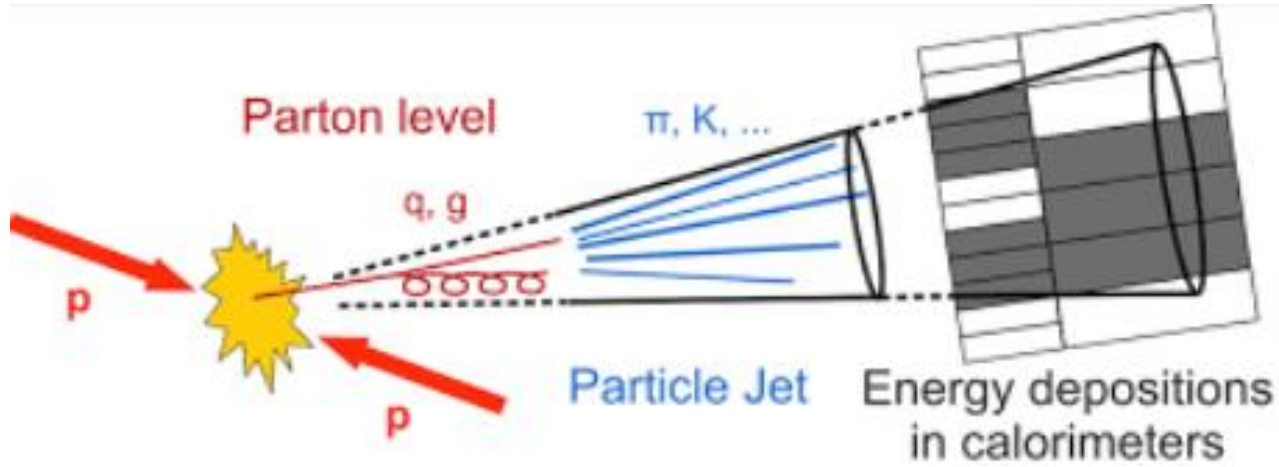


Hard Scattering → Parton Shower → Hadronization → Detection

阶段 (Stage)	物理本质 (Physics)	空间尺度 (Spatial Scale)	时间尺度 (Timescale)
1. Hard Scattering	点对点硬散射	$10^{-18} \sim 10^{-16}$ m	$10^{-26} \sim 10^{-24}$ s
2. Parton Shower	微扰 QCD 辐射	$\sim 10^{-15}$ m	$\sim 10^{-23}$ s
3. Hadronization	非微扰夸克禁闭	$10^{-15} \sim 10^{-13}$ m	$10^{-23} \sim 10^{-22}$ s
(Unstable Decay)	(重强子弱衰变)	(毫米级 $\sim 10^{-3}$ m)	(皮秒级 $\sim 10^{-12}$ s)
4. Detection	硬件宏观响应	1 mm \sim 11 m	30 ns \sim 几 μ s

Before inner detector

→ Inner. Cal. MS.



Hard Scattering → Parton Shower → Hadronization → Detection

$$q \rightarrow qg$$

$$g \rightarrow gg, g \rightarrow q\bar{q}$$



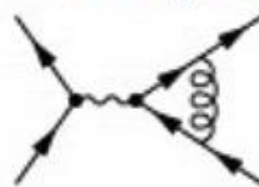
Infrared and Collinear Divergences

$$d\sigma_{q \rightarrow qg} \approx \sigma_0 \cdot \frac{\alpha_s}{2\pi} \cdot C_F \cdot P_{qq}(z) \cdot dz \cdot \frac{d\theta^2}{\theta^2}$$

$$P_{qq}(z) = \frac{1+z^2}{1-z}$$

1-z: radiative quark moment fraction

Virtual Diagrams



Σ Real Emissions



$$\theta \rightarrow 0$$

Collinear Divergence: $\sigma \rightarrow \infty$

$$z \rightarrow 1$$

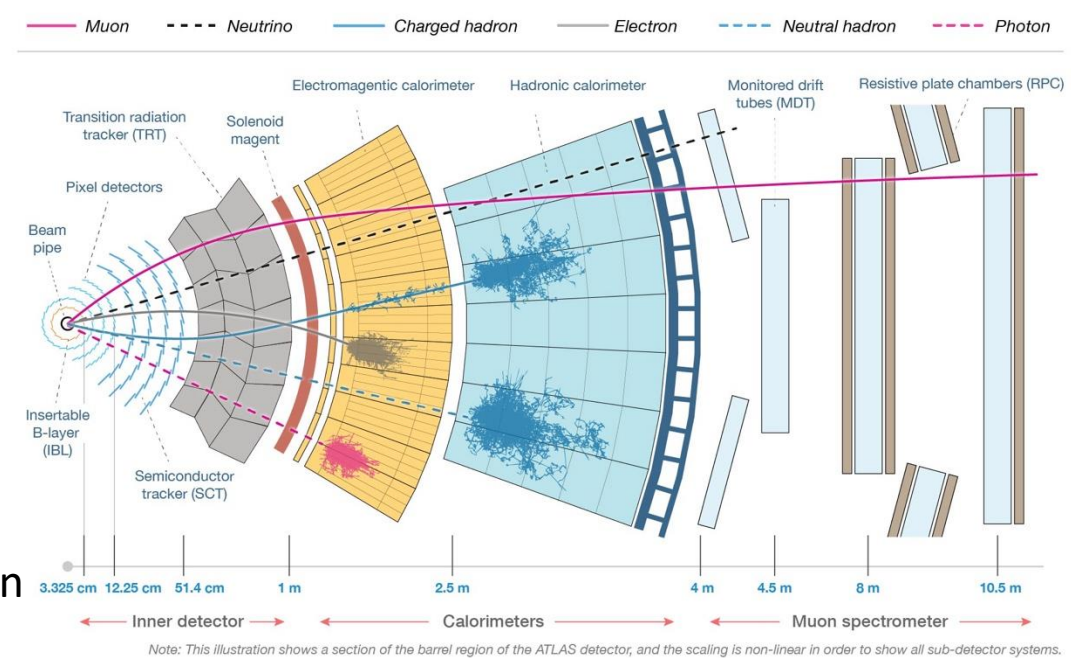
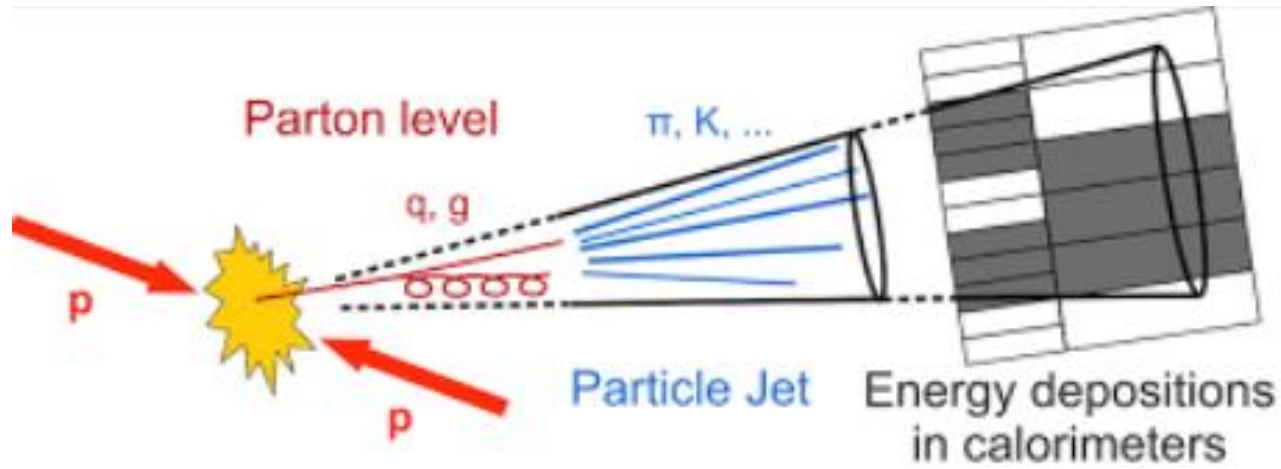
Infrared Divergence: $\sigma \rightarrow \infty$

Theory: safe

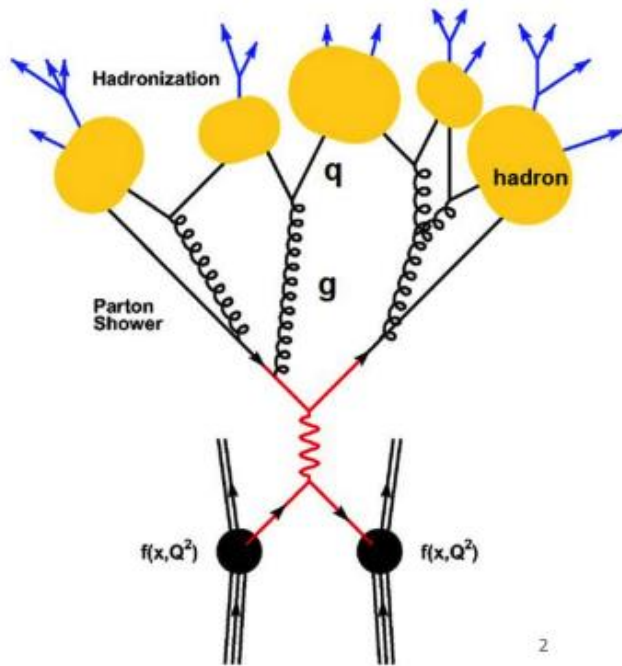
KLN guarantees the sum of real and virtual process in NLO is valid and finite cross section

Experiment: trouble

Need special process



Hard Scattering → Parton Shower → Hadronization → Detection



- As color-charged partons recede from each other, the color field is squeezed into a high-tension "**color string**"
- potential energy increases linearly with distance.
- When this string becomes sufficiently stretched, it snaps by spontaneously creating **new quark-antiquark pairs** from the **vacuum**
- transforming the remnant energy into a collimated spray of **stable, colorless hadrons**.

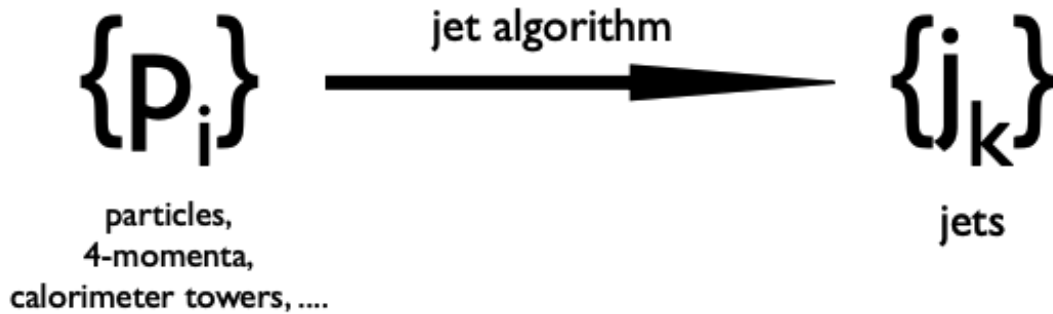
Part II

Anti-kt clustering algorithm

Inputs to jet algorithm

Small jet $R=0.4$:

A **jet algorithm** maps the momenta of the final state particles into the momenta of a certain number of jets:



Most algorithms contain a resolution parameter, R , which controls the extension of the jet

"Jet [definitions] are legal contracts between theorists and experimentalists"
– M.J. Tannenbaum

中：喷注的定义，是理论物理学家与实验物理学家之间签下的法律合同

Topoclusters

Clusters of topologically connected calorimeter signals

Particle flow (PFlow) objects

Combines tracks and topoclusters

Trackers:

- Better resolution for low p_T particles
- Better angular resolution
- Traces particles to hard-scatter interaction or pile-up

Calorimeters:

- Better resolution for high p_T
- Captures neutral particles

Combining information

- Improves energy and angular resolution
- Reduces pile-up contributions

Jet clustering algorithm: Clustering algorithm development

算法家族 (Family)	具体代表算法	红外安全 (Infrared)	共线安全 (Collinear)	洛伦兹推进不变性 (Boost Invariant)	阶次独立性 (Order Indep.)	实验重构几何外观 & 核心缺陷
传统锥算法 (Seeded Cone)	JetClu, MidPoint	✗	✗	✗ / ⚠	✗	完美的正圆形。但存在“幽灵搭桥”和“化整为零”的致命理论缺陷，LO以上计算直接发散。
无种子锥算法	SISCone (2007)	✓	✓	✓	✓	接近圆形（存在分裂合并变形）。实现了理论安全，但计算复杂度高达 $O(N^2 \ln N)$ ，在LHC高堆积下直接卡死。
传统顺序重组	kt 算法 (1993)	✓	✓	✓	✓	不规则“橡皮泥”状。软粒子先凝聚，导致边缘极度容易吃进 Pile-up 噪声，实验上根本无法进行能量校准。
空间顺序重组	C/A 算法 (1997)	✓	✓	✓	✓	较不规则状。纯空间几何聚类，无视动量。虽然无法直接校准，但如今是**喷注子结构 (Substructure)** 的王牌工具。
新一代	Anti-kt (2008)	✓	✓	✓	✓	完美的正圆形。硬粒子作为种子向外无情抽吸，完美融合了理论安全与实验易校准性，计算复杂度仅为 $O(N \ln N)$ 。

Anti-Kt: definition and procedures

① 粒子与粒子两两之间的“代数间距” (d_{ij})

$$d_{ij} = \min \left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2} \right) \frac{\Delta R_{ij}^2}{R^2}$$

- 其中 $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, 是粒子 i 和 j 在快度-方位角画布上的几何距离平方。
- R 是你预设的分辨率参数 (ATLAS 标准通常为 0.4 或 1.0)。
- 最关键的灵魂项: 前面的动量指数 $p = -1$, 导致 p_T 被扔到了分母上, 变成了 $\frac{1}{p_T^2}$ 。

② 粒子与束流管 (Background) 之间的“截断距离” (d_{iB})

$$d_{iB} = \frac{1}{p_{Ti}^2}$$

- 这个距离完全不看任何几何位置, 纯粹由粒子自己的横动量 p_T 决定。

第一步: 全场海选, 寻找绝对极小值

算法扫描整个矩阵中所有的 d_{ij} 和 d_{iB} , 找出当前全场数值最小的那个距离 (The global minimum)。

第二步: 分支判决 (融合 vs 独立)

- **分支 A:** 如果这个全场最小值是一个两两间距 d_{ij} :
算法认为粒子 i 和粒子 j 靠得最近、动量最契合。于是强行将它们**融合 (Combine)**, 将它们的四矢量严格相加得到一个新的次级团簇 ($P_{\text{new}} = P_i + P_j$)。随后, 在列表中删去 i 和 j , 加入新团簇。
- **分支 B:** 如果这个全场最小值是一个截断距离 d_{iB} :
算法认为粒子 i 已经长得足够成熟, 周围半径 R 范围内再也没有任何人能吸引它了。于是算法对它**宣告独立**——将粒子 i (或它已经滚雪球积攒成的团簇) 从待聚类列表里移出, 打包标记为一个**最终的物理 Jet**。

第三步: 矩阵刷新

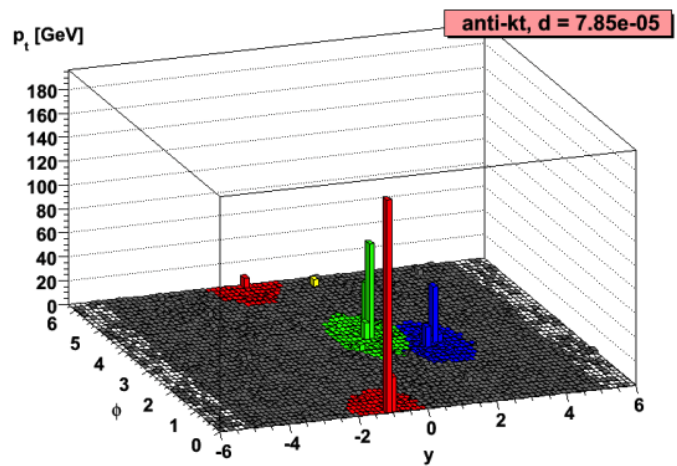
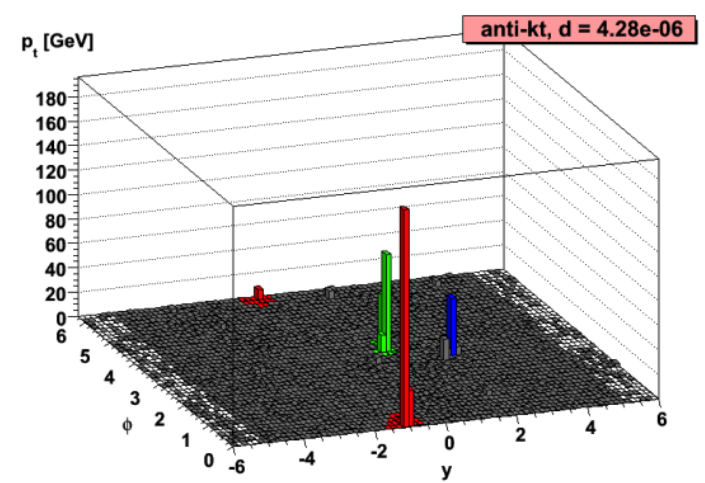
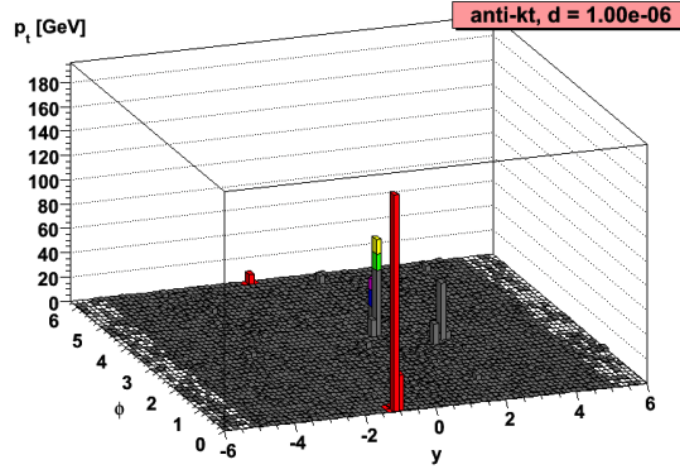
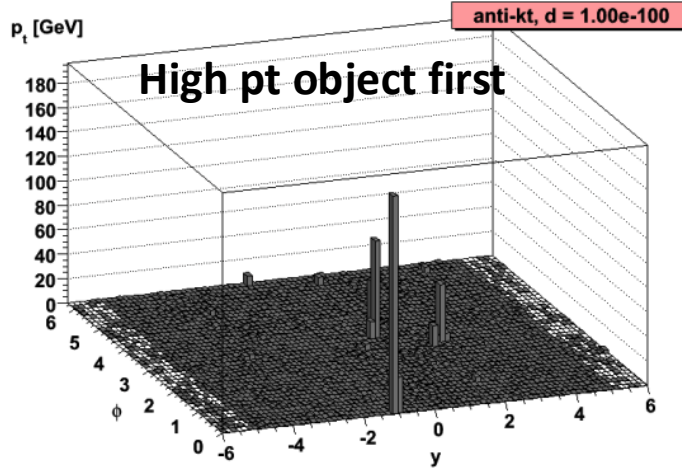
因为全场少了一到两个粒子, 或者诞生了一个新团簇, 算法会重新计算受影响行的 d_{ij} 和 d_{iB} 。

第四步: 无限迭代

回到第一步, 重复上述过程, 直到列表里的粒子数变为 0。

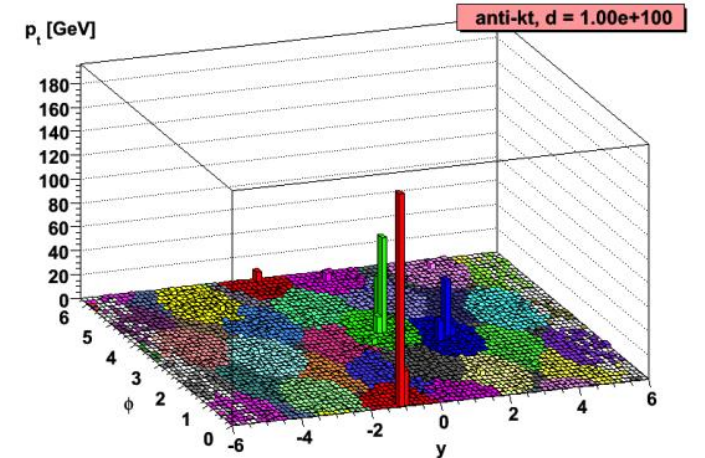
Anti-kt in action

$$\min \left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2} \right) \frac{\Delta R_{ij}^2}{R^2}$$



$$d_{iB} = \frac{1}{p_{Ti}^2}$$

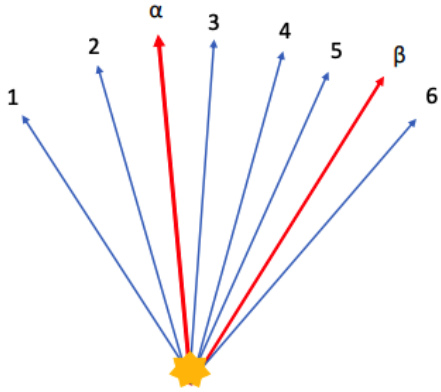
Until d_{iB} becomes the least
 Making $d_{ij} = 1e100$
 And do pt cut



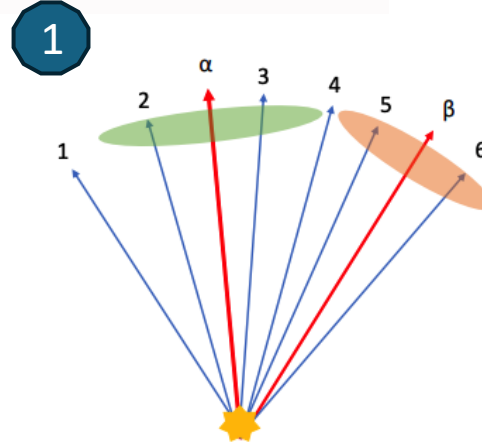
General behavior for boundary

$$\min \left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2} \right) \frac{\Delta R_{ij}^2}{R^2}$$

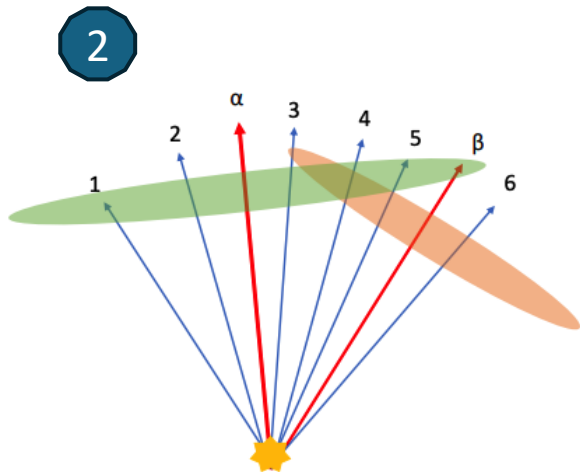
$$d_{iB} = \frac{1}{p_{Ti}^2}$$



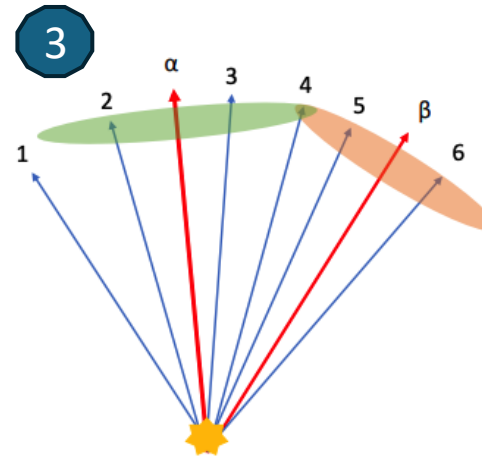
- Take a section of an event with two **hard particles** (α, β) and several **soft particles** (1-6)
- Assume $k_{t\alpha} > k_{t\beta} \gg k_{t1-6}$
- Therefore $\frac{1}{k_{t\alpha}^2} < \frac{1}{k_{t\beta}^2} \ll \frac{1}{k_{t1-6}^2}$
- The distances $d_{\alpha 1-6}$ are dependant on the transverse momentum of the hard particle, not the soft particles (soft resilience)



- If the distance $\Delta_{\alpha\beta} > 2R$, each hard particle will have a conical jet with a regular boundary
- Drawing the boundaries where $d_{\alpha 1-6} < d_{\alpha B}$ and $d_{\beta 1-6} < d_{\beta B}$ we get the green and orange circles (respectively)
- In this case, neither particle 1 or 4 are included in the jets

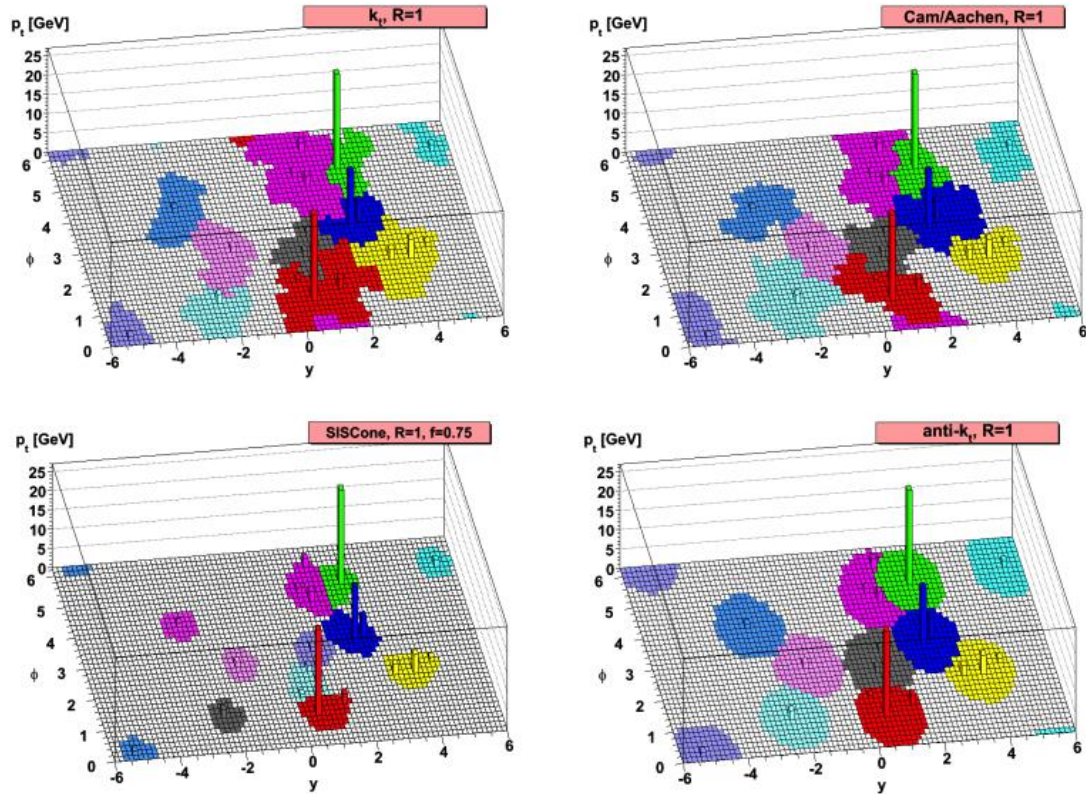


- If the distance $\Delta_{\alpha\beta} < R$, then both hard particles will be clustered into the same jet
- The boundary will then depend on their relative momenta:
 - If $k_{t\alpha} \gg k_{t\beta}$ then the resulting jet will be conical centered on α
 - If $k_{t\alpha} \approx k_{t\beta}$ then the resulting jet will be centered on the midpoint between the two pseudo jets that each particle creates



- If the distance is in between: $R < \Delta_{\alpha\beta} < 2R$, then both hard particles will produce a jet, but they cannot both be conical:
 - If $k_{t\alpha} \gg k_{t\beta}$ then the α jet will be conical and the β jet will have a crescent shaped boundary where the two jets overlap
 - If $k_{t\alpha} \approx k_{t\beta}$ then the resulting boundary will be more complicated, depending on the distances of the particles within the overlap
 - If $k_{t\alpha} = k_{t\beta}$ then the resulting boundary will simply be a straight line

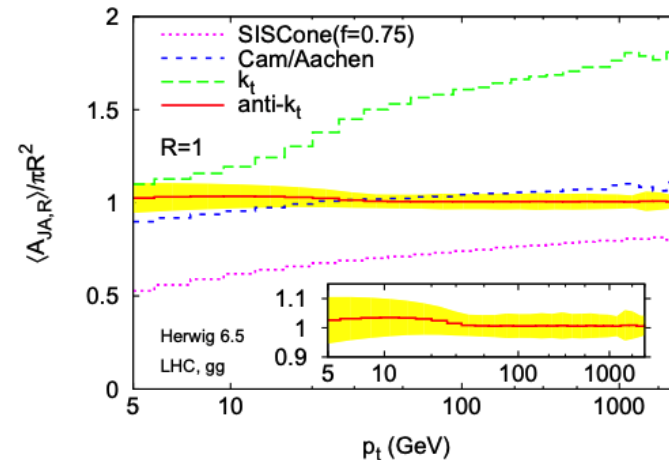
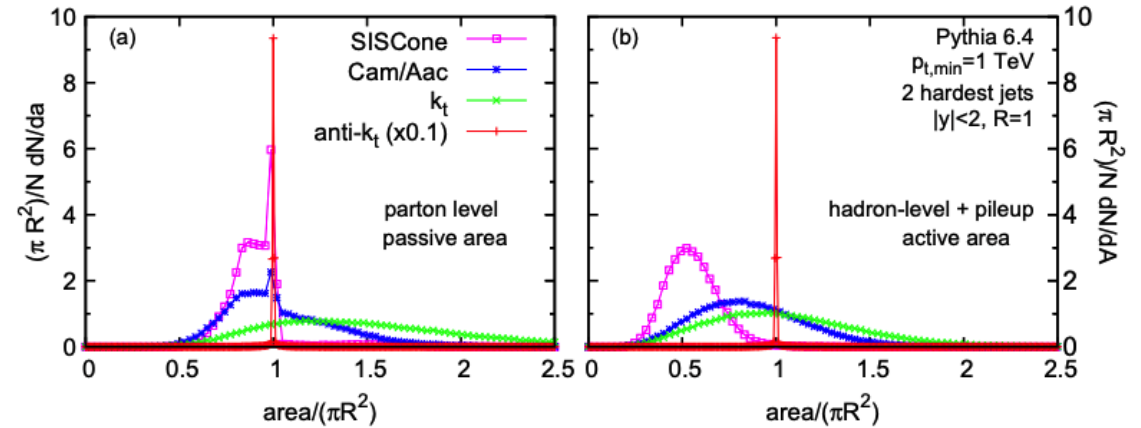
Area-related properties I



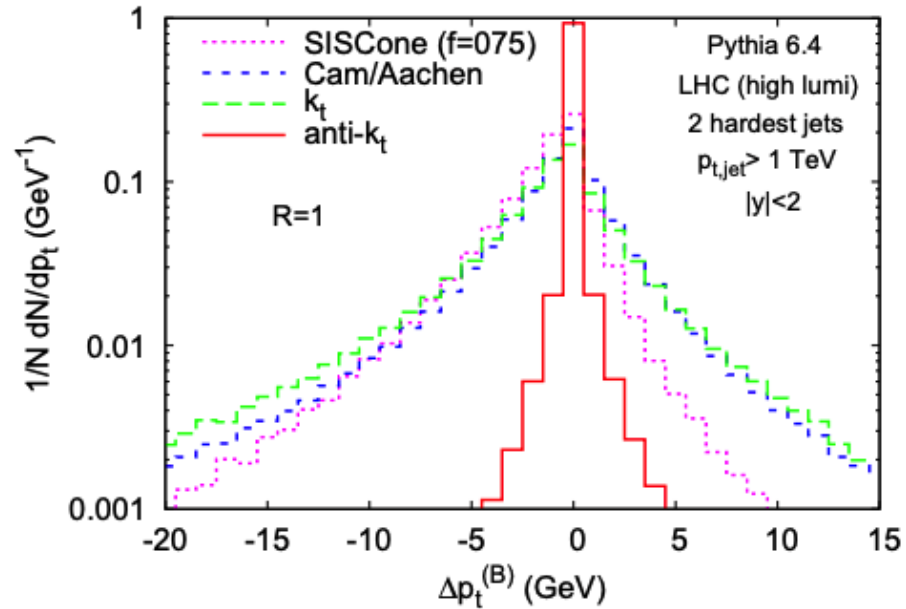
Passive Area (a): Measures the jet's sensitivity to isolated point radiation (such as a single perturbation soft gluon).

Active Area (A): Measures the jet's susceptibility to diffuse background radiation in the detector (such as low-event UEs, pile-up noise).

$$A_{\text{anti-}k_t} = a_{\text{anti-}k_t} = \pi R^2$$



Area-related properties II



Other algorithms(SIScone, kt, C/A)::

- Susceptible to back-reaction
- The misassignment probability scale is directly governed by the pile-up background density

Anti-kt:

- Strongly suppresses back-reaction.
- Fluctuations are actively suppressed by the jet's own hard scale p_{tJ} instead of background fluctuations.

- **Back-Reaction:** jet includes additional noise and impact the clustering order of PFOs
- **Anti-kt:** high pt PFOs first

Other properties

Anti-kt: Non-Global Logarithms

- **Theoretical Bottleneck:** Traditional soft-adaptable boundaries dynamically distort under soft emissions, causing highly complex, non-linear coupling in all-order single-logarithmic resummations
- **Anti-kt Solution:** Because its boundaries are completely unaffected by soft radiation, the jet behavior mathematically reduces to that of an **ideal, rigid geometric cone**.
- **Theoretical Edge:** This rigid boundary immensely decouples the single-logarithmic non-global terms, **significantly simplifying higher-order perturbative QCD calculations**.

Anti-kt: Milan Factor

- **Physical Essence:** The Milan factor (M) is a universal correction coefficient used in non-perturbative hadronization calculations to account for soft gluon branchings at large angles.
- **Traditional Bottleneck:** In most algorithms, the random clustering of soft particles near the jet boundaries destroys the strict linear dependence on soft momenta, completely **breaking the universality of M**
- **Anti-kt Solution:** Thanks to its soft-resilient rigid boundary, the linear momentum relation is perfectly preserved
- **Phenomenological Edge:** It miraculously restores the **universal value of the Milan factor ($M=1.49$)**, making non-perturbative theoretical predictions exceptionally robust and precise

Example application: top reconstruction

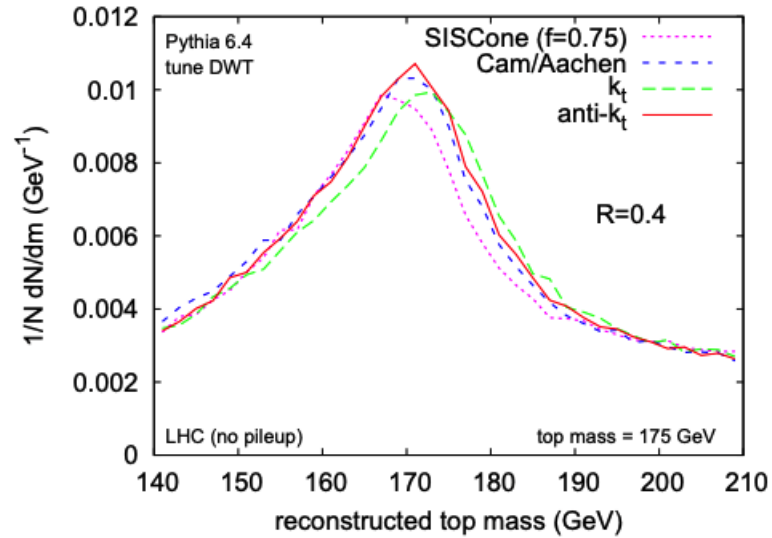


Figure 5: Top mass reconstruction in Pythia-simulated LHC $t\bar{t}$ events. Both the t and the \bar{t} decay hadronically, $t \rightarrow bW^+ \rightarrow bq\bar{q}$ and $\bar{t} \rightarrow \bar{b}W^- \rightarrow \bar{b}q\bar{q}$. All jet algorithms have been used with $R = 0.4$.

📄 Physics Context & Strategy

- **Motivation:** Evaluate if the soft-resilient (rigid) boundary of anti- k_t compromises mass resolution in complex multi-jet environments. [PDF + 3](#)
- **Channel:** Fully hadronic top-quark pair decay ($t\bar{t} \rightarrow b\bar{b}q\bar{q}q\bar{q}$), generating at least 6 hard jets ($p_t > 10 \text{ GeV}$, $|y| < 5$). [PDF](#)
- **Analysis Setup:** Standard small radius $R = 0.4$; b -jets are tagged, and light jets are paired to minimize the W and top mass differences. [PDF + 1](#)

📄 Key Performance Findings

- **Mass Spectrum Performance:** All four IRC-safe algorithms behave remarkably similarly in terms of the reconstructed top-mass peak. [PDF](#)
- **The Winner:** Cambridge/Aachen and anti- k_t perform **marginally better** than k_t and SIS Cone. [PDF](#)
- **Physics Outlook:** While differences are small in this no-pileup threshold scenario, anti- k_t is expected to excel significantly in **multi-scale processes with large pileup**, as its rigid boundary effectively cuts out background energy fluctuations. [PDF + 3](#)

Part III
Summary

Summary

- Overview of jet origin and evolution
- Anti-kt successfully bridges the gap between theoretical safety and experimental feasibility
 - delivering the regular conical shape of ideal cone algorithms
 - maintaining strict infrared and collinear (IRC) safety
 - possessing a soft-resilient, rigid boundary that is completely unaffected by soft radiation
 - achieving an active jet area exactly equal to its passive area (πR^2)
 - Excellence performance in test, rigid boundary does not sacrifice mass resolution.

Mathematically safe, Fast($\mathcal{O}(N \ln N)$), Robust

Thanks

Rough description of references

PDF from Matteo Cacciari FeynRules/MadGraph School November 2018, Hefei

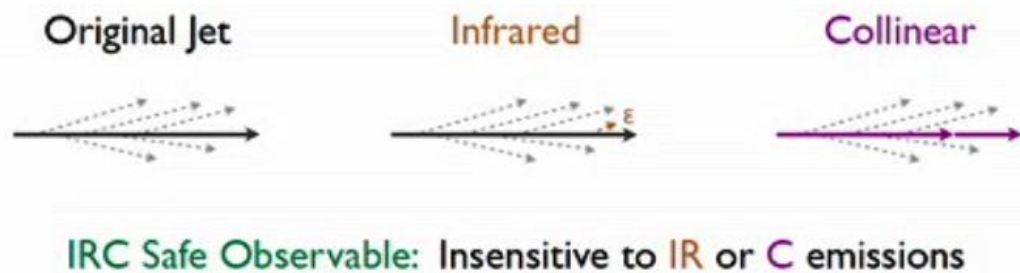
Material from Gavin Salam and Grégory Soyez

Illustration of the detection of particles in the ATLAS experiment from Lakmin Wickremasinghe



Part V
Backup

KLN → Experiment



在微扰 QCD 计算中，我们要算一个夸克出射的截面。在次领头阶 (NLO) 我们需要计算两个部分：

$$\sigma_{\text{NLO}} = \int d\sigma_{\text{real}} + \int d\sigma_{\text{virtual}}$$

- 1 **实辐射图 (Real)**：一个夸克飞出来，由于红外/共线发散，它辐射出了一个能量为 $E \rightarrow 0$ 的软胶子。这个图的积分结果是 $+\infty$ 。
- 2 **虚回路图 (Virtual)**：没有发射真实粒子，但是在费曼图的内部环路里，真空涨落交换了一个软胶子（回路积分）。这个图的积分结果是 $-\infty$ 。

在理论上，只要你定义的可观测物理量对“有没有这个软胶子”完全不敏感（即包容性 Inclusive），那么在相空间积分完后，这两个无穷大相加：

$$(+\infty) + (-\infty) = \text{一个有限的物理测量值 (例如 } 0.5 \text{ pb)}$$

KLN 定理在数学上宣告胜利。

2. 实验算法是怎么强行“破坏”这种抵消的？（以红外不安全为例）

现在我们把数据拿到探测器里，用实验算法去重构 Jet。

假设对撞机在某个局部产生了两个硬粒子（硬核 A 和硬核 B），它们挨得比较近。

- **情况一（对应虚图）**：真空中只有这两个硬核，没有辐射实胶子。你的算法把它们重构成了 2 个独立的 Jet。
- **情况二（对应实图）**：由于红外发散，这两个硬核中间真的辐射出了一个能量极低 ($E \rightarrow 0$) 的实软胶子。

如果你的算法是红外不安全的（比如传统的 Seeded Cone 锥算法）：

这个能量无限趋近于 0 的软胶子，会极其倒霉地在中间起到“桥梁”作用，导致算法把硬核 A、硬核 B 和这个软胶子全部打包在一起，最终重构出了 1 个巨大的 Jet。

灾难发生了：

- 在虚图里，实验算法数出了 2 个 Jet；
- 在实图里，实验算法数出了 1 个 Jet。

当你（理论物理学家或实验分析员）想要去测量“2-Jet 的产生截面”时：

- 虚图由于贡献了 2 个 Jet，其截面积分 ($-\infty$) 被计入了你的分析。
- 实图由于变成了 1 个 Jet，它在你的算法眼里不属于 2-Jet 事件，它的贡献 ($+\infty$) 被排除在你的 2-Jet 分析之外！

原本在理论上应该手拉手一起抵消的 ($+\infty$) 和 ($-\infty$)，被你的实验重构算法强行拆散到了不同的物理分类里（一个被分到了 1-Jet 桶，一个被分到了 2-Jet 桶）。

结果就是，在 2-Jet 的测量通道里，只剩下了虚图的 $-\infty$ ，发散无法抵消，理论计算彻底崩溃（算出了负无穷大的截面）。