A sharp bilinear cone restriction estimate

Thomas Wolff
Department of Mathematics
253-37 Caltech
Pasadena, Ca 91125 USA
wolff@cco.caltech.edu

The purpose of this paper is to prove an essentially sharp L^2 Fourier restriction estimate for light cones, of the type which is called bilinear in the recent literature.

Fix $d \geq 3$, denote variables in $\mathbb{R}^d$ by $(\overline{x}, x_d)$ with $\overline{x} \in \mathbb{R}^{d-1}$, and let $\Gamma = \{x : x_d = |\overline{x}| \text{ and } 1 \leq x_d \leq 2\}$. Let Γ_1 and Γ_2 be disjoint conical subsets, i.e.

$$\Gamma_i = \{x \in \Gamma : \frac{\overline{x}}{x_d} \in \Omega_i\}$$

where Ω_i are disjoint closed subsets of the sphere S^{d-2}. Let f and g be two functions on Γ whose supports are contained in Γ_1 and Γ_2 respectively. We will prove the following estimate, where σ is surface measure on Γ, and $\widehat{fd\sigma}$ is the $\mathbb{R}^d$ Fourier transform:

Theorem 1 If $p > 1 + \frac{2}{d}$ then

$$\|\widehat{fd\sigma}\widehat{gd\sigma}\|_p \leq C_{p,\Gamma_1,\Gamma_2}\|f\|_2\|g\|_2 \qquad (1)$$

Bilinear estimates of this general type have been used by several authors; see in particular [9]. The estimate (1) was formulated by Bourgain in [3], and it was proved in [3] when $d = 3$ and $p > 2 - \epsilon$ for some $\epsilon > 0$, the case $p = 2$ being easier and implicit in [1]. Tao and Vargas [14] recently obtained the explicit range $p > 2 - \frac{8}{121}$ when $d = 3$, and noted that one can also obtain a range $p > 2 - \epsilon_0$ in the four dimensional case. The range of p in Theorem 1 is known to be best possible when $d = 3$ except for the question of the endpoint - see [14], where the conjecture that (1) should hold for $d = 3$ and $p \geq \frac{5}{3}$ is attributed to Machedon and Klainerman - and is similarly best possible in higher dimensions; see the remarks in section 5 of the present paper.

Although Theorem 1 is sharp of its type in any dimension, it is more satisfactory in low dimensions, since when d is large the L^2 norms on the right hand side of (1) are quite weak in comparison with other relevant norms and the exponent $1+\frac{2}{d}$ is only a small improvement on the exponent $1+\frac{2}{d-2}$ which follows from the Strichartz inequality. When $d=4$, Theorem 1 implies (via a rescaling argument as in [15]) a statement analogous to a result of Barcelo [1] for the three dimensional case:

Corollary When $d=4$ the restriction of the Fourier transform to Γ defines a bounded operator from L^p to $L^p(\Gamma)$ for any $p<\frac{3}{2}$.

The range of p here is again sharp. It should be pointed out that the geometric information needed for our results is simpler than what is likely to be needed either to solve the restriction problem for S^2, or to solve some of the other outstanding problems concerning the cone such as the multiplier problem and local smoothing, even in the $2+1$-dimensional case. On the other hand, there are very few hypersurfaces for which a sharp restriction theorem is known, and the approach below may be useful in connection with the sphere as well, insofar as it is possible to consider the sphere without first resolving the Kakeya problem.

As might be expected the proof of Theorem 1 uses Kakeya techniques related to Bourgain's paper [2] and the now classical work of C. Fefferman and Cordoba. The necessary geometric information while not particularly deep is different from what has been used previously, and we prove what we need in section 1 below. In section 2 we discuss a lemma from [10], in section 3 we prove our main lemma (Lemma 3.5) and in section 4 we prove Theorem 1. In section 5 we prove the corollary and make some further related remarks. Finally, in an appendix we discuss the related question of mixed norm estimates for the restriction of the X-ray transform to the light rays. We prove an optimal local result (except for endpoint questions) in three and four dimensions and a partial result in higher dimensions. This is stated below as Theorem A.1.

We will use several ideas and lemmas from the previous work on the cone problem, e.g. from [3], [10] and [14]. Some aspects of the argument and also the fact that Theorem 1 should be an accessible result were suggested by the author's recent paper [19].

List of notation

$Q(N)$: the cube in $\mathbb{R}^d$ centered at the origin with side length N.

$|E|$: measure or cardinality of the set E depending on the context.

χ_E: indicator function of E.

1. A property of light rays

In this section we fix a suitable large constant B depending on the dimension d.

A light ray will mean a line in $\mathbb{R}^d$ making a 45 degree angle with the plane $x_d = 0$. We fix two disjoint conical sets Γ_1 and Γ_2 as described in the introduction and will say that a light ray is white (resp. black) if its direction belongs to Γ_1 (resp. Γ_2). Thus any white and black rays are transverse. We fix a small positive number ϵ.

Let $\delta > 0$, and let $\mathcal{W}$ and $\mathcal{B}$ be sets respectively of white and black light rays with respective cardinalities m and n. For each white line W (or black line B) we associate to W (or B) the infinite cylinder whose axis is W (or B) and whose cross section radius is δ. We will denote these tubes by w and b. For each tube w (similarly b) we define

$$\phi_w(x) = \min(1, \frac{\delta}{\mathrm{dist}(x, w)})^M \tag{2}$$

where M is a large constant depending on ϵ. We assume that $\mathcal{W}$ (similarly $\mathcal{B}$) is δ-*separated*; by this we mean the following: if D is a disc in projective space with radius δ, then the tubes w whose axes belong to D have bounded overlap, i.e. no point belongs to more than B of them. We note this implies that the cardinality of lines in $\mathcal{W}$ which intersect a given compact set is bounded by a (negative) power of δ.

A *μ-fold point* is a point which belongs to at least μ white tubes, and a *smooth μ-fold point* is a point where the quantity

$$\Phi_{\mathcal{W}} \stackrel{def}{=} \sum_{w \in \mathcal{W}} \phi_w$$

is at least equal to μ.

We fix a partition of $Q(1)$ into pairwise disjoint δ^ϵ-cubes; in this section we reserve the letter Q for these cubes (except for the standing notation $Q(N)$ for cubes centered at the origin). In what follows we will be working with a relation $\sim$ between white or black tubes and the cubes Q. For any such relation we denote

$$n_{\mathcal{W}}(Q) = |\{w \in \mathcal{W} : w \sim Q\}|$$

$$n_{\mathcal{B}}(Q) = |\{b \in \mathcal{B} : b \sim Q\}|$$

If x is a point or E is a set contained in a cube Q then we will use the notation

$$w \sim x \quad \text{(resp. } w \sim E)$$

to mean that $w \sim Q$, where Q is the δ^ϵ-cube containing x (resp. E), and we define

$$\tilde{\Phi}_{\mathcal{W}}(x) = \sum_{\substack{w \in \mathcal{W} \\ w \not\sim x}} \phi_w(x)$$

$$\tilde{\Phi}_{\mathcal{B}}(x) = \sum_{\substack{b \in \mathcal{B} \\ b \not\sim x}} \phi_b(x)$$

We also define (cf. [2]) a *bush* to be a set of tubes which are all the same color and which all pass through a common point p, and more generally an η-*bush* is a set of tubes which are all the same color and are all at distance $< \eta$ from a common point p. We call any such point p a *base point* for the bush.

The purpose of this section is to prove the following lemma.

Lemma 1.1 Assume $\mathcal{W}$ and $\mathcal{B}$ are δ-separated. Then there is a relation $\sim$ between white or black tubes and δ^ϵ-cubes Q so that the following hold, where C depends on d only; the implicit constants also depend on ϵ:

1. $\sum_Q n_{\mathcal{W}}(Q) \lesssim m (\log \frac{1}{\delta})^5$.
2. $\sum_Q n_{\mathcal{B}}(Q) \lesssim n (\log \frac{1}{\delta})^5$.
3. The δ-entropy of the set $\{x \in Q(1) : \tilde{\Phi}_{\mathcal{W}}(x) \geq \mu \text{ and } \Phi_{\mathcal{B}}(x) \geq \nu\}$ is $\lesssim \delta^{-C\epsilon} \frac{mn}{\mu^2 \nu}$.
4. The δ-entropy of the set $\{x \in Q(1) : \Phi_{\mathcal{W}}(x) \geq \mu \text{ and } \tilde{\Phi}_{\mathcal{B}}(x) \geq \nu\}$ is $\lesssim \delta^{-C\epsilon} \frac{mn}{\mu \nu^2}$.

Remarks 1. It is easy to see that the δ-entropy of the points which belong to μ white and ν black tubes can be as large as $\frac{mn}{\mu\nu}$ - just take $\mathcal{W}$ and $\mathcal{B}$ to be bushes with a common basepoint and set $\mu = m$, $\nu = n$. Thus 3. gains a factor of μ over the "trivial" bound valid with $\tilde{\Phi}_{\mathcal{W}}$ replaced by $\Phi_{\mathcal{W}}$. In the proof of Theorem 1, this factor will compensate for the factor appearing in Mockenhaupt's estimate for the relevant square function, i.e. in Lemma 2.1 below. It is also important that the dependence on δ in 1. and 2. is only logarithmic, or more precisely that it does not involve the specific power $\delta^{-\epsilon}$. On the other hand the distinction between μ-fold points and smooth μ-fold points is purely technical - the functions ϕ_b are needed later on in order to estimate Schwartz tails.

2. It is natural to state Lemma 1.1 in the above manner, since only properties 1-4 of the relation $\sim$ will be used in the subsequent sections and not its exact definition. However, the relation will be constructed in an explicit and fairly simple way: roughly, arrange the white or black tubes into bushes, and define $w \sim Q$ if w belongs to a bush whose basepoint is in Q. This procedure together with the induction argument in section 4 below is a variant on the "two ends" argument in [17], [18].

Lemma 1.2 below is true because ϕ_w is essentially a rapidly decreasing sum of constants times characteristic functions of dilates of w; we leave the details to the reader. Lemma 1.3 is a geometrical fact; similar facts are used in various places in the literature, e.g. in [3] and [14].

Lemma 1.2 If $x \in Q(1)$ is a smooth μ-fold point for the white tubes with $\mu \geq \delta^B$ then x is a basepoint for an η-bush (of white tubes) with cardinality $\gtrsim (\log \frac{1}{\delta})^{-1}\mu(\frac{\eta}{\delta})^M$ for some $\eta \leq \delta^{1-\epsilon}$. Conversely if C is a large fixed constant and $x \in Q(1)$ is a basepoint for an η-bush with cardinality $\geq C\mu(\frac{\eta}{\delta})^M$, then x is a smooth μ-fold point.

Lemma 1.3 Let $\mathcal{C} \subset \mathcal{W}$ be an η-bush with (say) $\eta \leq \sqrt{\delta}$, and let p be a basepoint for $\mathcal{C}$. Define a set Ω by deleting from $Q(1)$ the double of the δ^ϵ-square Q containing p. Let b be any black tube. Then

$$\int_\Omega \phi_b \Phi_{\mathcal{C}} \lesssim \delta^{-\epsilon(d-2)}\delta^d(\frac{\eta}{\delta})^{2d-3} \tag{3}$$

Proof First let b and w be a black and a white tube. For any $\lambda \leq 1$ the set

$$\{x \in Q(1) : \phi_b(x) \geq \lambda\}$$

is contained in a tube with the same axis as b and with width about $\delta\lambda^{-\frac{1}{M}}$, and similarly with w. Since w and b are transverse we have the bound

$$|\{x \in Q(1) : \min(\phi_b(x), \phi_w(x)) \geq \lambda\}| \lesssim (\delta\lambda^{-\frac{1}{M}})^d \tag{4}$$

Let $\Delta(b, w)$ be the quantity $\inf_{x\in\Omega}(\mathrm{dist}(x, b) + \mathrm{dist}(x, w)) + \delta$. If λ is large compared with $(\frac{\delta}{\Delta(b,w)})^M$ then the set in (4) does not intersect Ω. It follows therefore that

$$\int_\Omega \phi_b\phi_w \;\; \lesssim \;\; \int_\Omega \min(\phi_b, \phi_w)$$

$$
\begin{aligned}
&\lesssim \int_0^{(\frac{\delta}{\Delta(b,w)})^M} (\delta\lambda^{-\frac{1}{M}})^d d\lambda \\
&\lesssim \delta^d (\frac{\delta}{\Delta(b,w)})^{M-d} \qquad (5)
\end{aligned}
$$

Now we prove the estimate (3) when $\eta = \delta$. It is clear from (5) that the contribution to the left side from tubes $w \in \mathcal{C}$ such that $\Delta(b,w) \gtrsim \delta^{\epsilon}$ is small. On the other hand let ρ be small compared with δ^{ϵ}, and consider how many tubes $w \in \mathcal{C}$ there can be with $\Delta(b,w) \leq \rho$. The bush $\mathcal{C}$ is clearly contained in a $C\delta$-neighborhood of the portion of the light cone with origin at p which corresponds to the conical subset Γ_1. If b contains a certain point y which lies within ρ of Γ_1 and is further than δ^{ϵ} from p, then by transversality b must intersect Γ at a point within $C\rho$ of y. Thus the number of tubes w with $\Delta(b,w) \leq \rho$ is bounded by the δ-entropy of the set of lines in Γ which intersect a fixed $C\rho$-disc lying at distance further than δ^{ϵ} from the vertex; equivalently, by the δ-entropy of a $\delta^{-\epsilon}\rho$-disc on S^{d-2}, which is $(\frac{\delta^{-\epsilon}\rho}{\delta})^{d-2}$. We conclude using (5) that there is a bound

$$
\sum (\frac{\delta^{-\epsilon}\rho}{\delta})^{d-2} \delta^d (\frac{\delta}{\rho})^{M-d}
$$

with the sum being over dyadic $\rho \geq \delta$. Thus we get the bound $\delta^{d-\epsilon(d-2)}$ as claimed.

We now remove the restriction $\eta = \delta$. If $\mathcal{C}$ is an η-bush then, for parameters ρ such that $\rho \geq \eta$ but ρ is small compared with η^{ϵ}, the maximum number of η-separated lines in $\mathcal{C}$ with $\Delta(b,w) \leq \rho$ is bounded by $(\frac{\delta^{-\epsilon}\rho}{\eta})^{d-2}$; for this just apply the above argument replacing δ by η. The space of light rays is $2d-3$-dimensional, so any fixed light ray can be within η of at most $(\frac{\eta}{\delta})^{2d-3}$ δ-separated ones. It follows that for any $\rho << \eta^{\epsilon}$ there are $\lesssim (\frac{\eta}{\delta})^{2d-3}(\frac{\delta^{-\epsilon}(\rho+\eta)}{\eta})^{d-2}$ tubes w with $\Delta(b,w) \leq \rho$. We now apply (5) as above to bound the left side of (3) by

$$
\sum (\frac{\eta}{\delta})^{2d-3} (\frac{\delta^{-\epsilon}(\rho+\eta)}{\eta})^{d-2} \delta^d (\frac{\delta}{\rho})^{M-d}
$$

plus a negligible error, with the sum being over dyadic $\rho \geq \delta$. Estimate (3) follows from this. $\square$

The following lemma is the main step in the argument. Essentially, it corresponds to Lemma 1.1 except that here we ignore the tails (they will be taken care of in the next lemma) and work with a fixed value of μ (hence the induction argument in the last part of the proof of Lemma 1.1 below).

Lemma 1.4 Given a value of μ_0 we can partition $\mathcal{W}$ as

$$\mathcal{W} = \mathcal{W}_g \cup \mathcal{W}_b$$

where

1. $\mathcal{W}_g$ has no μ_0-fold points in $Q(1)$, and
2. $\mathcal{W}_b = \cup_{i=1}^R \mathcal{C}_i$ where each $\mathcal{C}_i$ is a bush with basepoint in $Q(2)$ and $R \lesssim \frac{m}{\mu_0}(\log \frac{1}{\delta})^2$.

Proof We fix a large enough constant $C = C_d$ and then another large constant A. We will use a recursive argument. Accordingly, if $\mathcal{W}^i \subset \mathcal{W}$, then we let $\kappa(\mathcal{W}^i)$ be the maximum possible cardinality for a set of δ-separated μ_0- fold points for $\mathcal{W}^i$. We have $\kappa(\mathcal{W}) \lesssim \delta^{-d}$ since all the tubes in $\mathcal{W}$ are contained in a fixed compact set.

Assume now that $\kappa(\mathcal{W}^i) = k$. We will prove: $\mathcal{W}^i = \mathcal{W}^{i+1} \cup \mathcal{W}_b^i$ where $\kappa(\mathcal{W}^{i+1}) \leq \frac{k}{2}$, and $\mathcal{W}_b^i$ is the union of $\lesssim A\frac{m}{\mu_0} \log \frac{1}{\delta}$ δ-bushes.

Namely, let $\mathcal{R}_i$ be a set of δ-separated μ_0-fold points for $\mathcal{W}_i$ with maximum possible cardinality k. There are two cases.

(i) If $k \leq A\frac{m}{\mu_0} \log \frac{1}{\delta}$ then we let $\mathcal{W}_b^i$ be all tubes $w \in \mathcal{W}^i$ such that $\text{dist}(x, w) < \delta$ for some $x \in \mathcal{R}_i$ and $\mathcal{W}^{i+1} = \mathcal{W}^i \backslash \mathcal{W}_b^i$. Evidently $\mathcal{W}_b^i$ is the union of $\lesssim A\frac{m}{\mu_0} \log \frac{1}{\delta}$ δ-bushes; and $\kappa(\mathcal{W}^{i+1}) = 0$ since any μ-fold point for $\mathcal{W}^i$ must lie within δ of some point of $\mathcal{R}_i$.

(ii) If $k > A\frac{m}{\mu_0} \log \frac{1}{\delta}$ we choose $A\frac{m}{\mu_0} \log \frac{1}{\delta}$ points from $\mathcal{R}_i$ at random. We let $\mathcal{W}_b^i$ be the tubes $w \in \mathcal{W}^i$ such that $\text{dist}(x, w) < \delta$ for some x in the random sample, and $\mathcal{W}^{i+1} = \mathcal{W}^i \backslash \mathcal{W}_b^i$. Evidently $\mathcal{W}_b^i$ is the union of $\lesssim A\frac{m}{\mu_0} \log \frac{1}{\delta}$ δ-bushes. We will show that with high probability $\kappa(\mathcal{W}^{i+1}) \leq \frac{k}{2}$.

For this, define for each $w \in \mathcal{W}^i$

$$P(w) = k^{-1} |\{x \in \mathcal{R}_i : \text{dist}(w, x) < \delta\}|$$

Thus the probability that w is in $\mathcal{W}^{i+1}$ is at most

$$(1 - P(w))^{A\frac{m}{\mu_0} \log \frac{1}{\delta}}$$

If $P(w) \geq C^{-1}\frac{\mu_0}{m}$ it follows that the probability that w is in $\mathcal{W}^{i+1}$ is at most $\delta^{\frac{A}{C}}$. If A is large enough then since the cardinality of the set of lines in $\mathcal{W}$ which intersect $Q(2)$ is bounded by δ^{-B} it follows that with high probability no tubes with $P(w) \geq C^{-1}\frac{\mu_0}{m}$ belong to $\mathcal{W}^{i+1}$.

Now let $\mathcal{R}_{i+1}$ be a maximal set of δ-separated μ_0-fold points for $\mathcal{W}^{i+1}$, and let $\mathcal{R}$ be a maximal 2δ-separated subset of $\mathcal{R}_{i+1}$. Consider the quantity

$$\sum_{w \in \mathcal{W}^{i+1}} kP(w) \tag{6}$$

We have seen that with high probability (6) is less than $\frac{k}{C}\frac{\mu_0}{m}|\mathcal{W}^{i+1}| \leq \frac{k}{C}\mu_0$. On the other hand, we have

$$\begin{aligned} (6) &= \sum_{x \in \mathcal{R}_i} |\{w \in \mathcal{W}^{i+1} : \mathrm{dist}(w,x) < \delta\}| \\ &\geq \mu_0 |\mathcal{R}| \end{aligned}$$

The first line followed from the definition (6) by reversing the order of summation, and the second line then followed because every point in $\mathcal{R}$ is within δ of a point of $\mathcal{R}_i$ and no two points of $\mathcal{R}$ can be within δ of the same point of $\mathcal{R}_i$. We conclude that with high probability $|\mathcal{R}| \leq C^{-1}k$. Since $|\mathcal{R}|$ and $|\mathcal{R}_i|$ are comparable it then follows that $|\mathcal{R}_{i+1}| \leq \frac{k}{2}$, as was to be shown.

We now proceed recursively. Let $\mathcal{W}^0 = \mathcal{W}$ and apply the preceding to express $\mathcal{W}^0 = \mathcal{W}_b^0 \cup \mathcal{W}^1$. Then apply the preceding to express $\mathcal{W}^1 = \mathcal{W}_b^1 \cup \mathcal{W}^2$ and continue in this manner, stopping when we reach a situation where we are in case (i) above. Suppose we stop after T stages. Since $\kappa(\mathcal{W}^i)$ is initially $\lesssim \delta^{-d}$ and decreases each time at least by a factor of 2, we then have $T \lesssim \log\frac{1}{\delta}$. We now define $\mathcal{W}_g$ to be the set $\mathcal{W}^{i+1}$ defined at the last iteration. It satisfies $\kappa(\mathcal{W}_g) = 0$ as required. On the other hand we define $\mathcal{W}_b = \cup_i \mathcal{W}_b^i$. This set is the union of the $\lesssim \log\frac{1}{\delta}$ sets $\mathcal{W}_b^i$, each of which is the union of $\lesssim A\frac{m}{\mu_0}\log\frac{1}{\delta}$ bushes. The lemma follows. $\square$

The next lemma is a version of the preceding one incorporating Schwartz tails.

Lemma 1.5 Fix $\mu_0 \geq \delta^B$. Then $\mathcal{W} = \mathcal{W}_g \cup \mathcal{W}_b$ where

1. $\Phi_{\mathcal{W}_g} \leq \mu_0$ everywhere.

2. $\mathcal{W}_b = \cup_{k:2^k<\delta^{-\epsilon}} \mathcal{W}_b^k$, and for each k, $\mathcal{W}_b^k = \cup_{i=1}^{R_k} \mathcal{C}_i$, where $\mathcal{C}_i$ is a $2^k\delta$-bush with basepoint in $Q(2)$ and $R_k \lesssim \frac{m}{2^{Mk}\mu_0}(\log\frac{1}{\delta})^3$.

Proof Let w^η be the η-tube with the same axis as w. Notice that Lemma 1.3 is applicable also to the w^η's (provided $\log\eta$ is comparable to $\log\delta$, which will be the case below), since we used δ-separation in the proof only to conclude that the cardinality of the white lines which intersect $Q(2)$ was bounded by a negative power of δ.

We now define recursively a family of subsets $\mathcal{W}_b^j$. Let $\mathcal{W} = \mathcal{W}_g^1 \cup \mathcal{W}_b^1$ be the decomposition from Lemma 1.3 for the given μ_0. If $k \geq 2$ and if $\mathcal{W}_b^j$ have been defined for $j < k$ then we let $\mathcal{W}_g^{k-1} = \mathcal{W}\backslash(\cup_{j=1}^{k-1}\mathcal{W}_b^j)$. The following inductive hypothesis will hold:

($*$) If $j \leq k-1$, then the family of tubes $\{w^{2^j\delta} : w \in \mathcal{W}_g^{k-1}\}$ has no $2^{Mj}\mu_0$-fold points.

Let $\eta = 2^k\delta$ and apply Lemma 1.3 to the tubes $\{w^\eta : w \in \mathcal{W}_g^{k-1}\}$ replacing μ_0 by $2^{Mk}\mu_0$. This decomposes $\mathcal{W}_g^{k-1} = \mathcal{W}_b^k \cup \mathcal{W}_g^k$ where the tubes $\{w^{2^k\delta} : w \in \mathcal{W}_g^k\}$ have no $2^{Mk}\mu_0$-fold points and $\mathcal{W}_b^k$ is the union of at most $\frac{m}{2^{Mk}\mu_0}(\log\frac{1}{\delta})^2$ $2^k\delta$-bushes. The inductive hypothesis ($*$) is then satisfied for $j \leq k$. We continue in this manner, stopping when $2^{Mk}\mu_0$ becomes greater than m. This will occur at a stage k with $2^k < \delta^{-\epsilon}$, since we have assumed $\mu_0 \geq \delta^B$. We define W_g to be the last $\mathcal{W}_g^k$.

If $\Phi_{\mathcal{W}_g}(x) \geq (C\log\frac{1}{\delta})\mu_0$ with C a large fixed constant then by Lemma 1.2 x must be a $2^{Mk}\mu_0$-fold point for the tubes $\{w^{2^k\delta} : w \in \mathcal{W}_g\}$ for some k, hence also a $2^{Mk}\mu_0$-fold point for the larger family $\{w^{2^k\delta} : w \in \mathcal{W}_g^k\}$, which is impossible by construction. The lemma now follows by replacing μ_0 with $(C\log\frac{1}{\delta})^{-1}\mu_0$. $\square$

To prove Lemma 1.1 it suffices by symmetry to construct a relation between white tubes and δ^ϵ-squares so that 1. and 3. hold. This will again be done recursively. A remark on terminology: in this argument, when we say that "$\mathcal{C}$ is a $2^k\delta$-bush" we mean that $\mathcal{C}$ is a $2^k\delta$-bush but not a $2^{k-1}\delta$-bush.

We apply Lemma 1.5 to $\mathcal{W}$ with $\mu_0 = \frac{m}{2}$, obtaining a set $\mathcal{W}_g^1$ with $\Phi_{\mathcal{W}_g^1} \leq \frac{m}{2}$ and a collection of stage 1 η-bushes $\mathcal{C}_i^1$ (thus each $\mathcal{C}_i^1$ is a $2^k\delta$-bush for some k with $2^k\delta \leq \delta^{1-\epsilon}$). Then we apply Lemma 1.5 to $\mathcal{W}_g^1$ with $\mu_0 = \frac{m}{4}$ obtaining $\mathcal{W}_g^2$ with $\Phi_{\mathcal{W}_g^2} \leq \frac{m}{4}$ and stage two η-bushes $\mathcal{C}_i^2$ and continue in this manner, taking $\mu_0 = \frac{m}{2^j}$ at the jth stage. We stop the induction at stage R, where R by definition is the smallest integer such that $\frac{m}{2^R} < \delta^B$.

Clearly $R \lesssim \log\frac{1}{\delta}$. For each $j_0 \le R$ we now have a decomposition

$$\mathcal{W} = \mathcal{W}_g^{j_0} \cup (\cup_{j \le j_0} \cup_i \mathcal{C}_i^j) \tag{7}$$

where $\Phi_{\mathcal{W}_g^{j_0}} \le \frac{m}{2^{j_0}}$, and (by part 2. of Lemma 1.5) we have the following:

For each j and k there are $\lesssim 2^j 2^{-Mk}(\log\frac{1}{\delta})^3$ values of i such that $\mathcal{C}_i^j$ is a $2^k\delta$-bush.

For each $\mathcal{C}_i^j$ we fix a basepoint p_i^j. We now define the relation $\sim$:

<u>Definition</u> A tube w and δ^ϵ-square Q are related, $w \sim Q$, if w belongs to an η-bush $\mathcal{C}_i^j$ such that p_i^j is in Q or one of its neighbors.

We show first that 1. holds. Suppose that $\mathcal{C}_i^j \subset \mathcal{W}_g^{j-1}$ is a $2^k\delta$-bush. Then, using Lemma 1.2 and the fact that $\Phi_{\mathcal{W}_g^{j-1}} \le \frac{m}{2^{j-1}}$, we get the following bound for the cardinality of $\mathcal{C}_i^j$:

$$|\mathcal{C}_i^j| \lesssim 2^{Mk}\Phi_{\mathcal{C}_i^j}(p_i^j) \lesssim 2^{Mk}\frac{m}{2^j}$$

By the preceding bound for the number of $2^k\delta$-bushes, we then have

$$\sum_i |\mathcal{C}_i^j| \lesssim \sum_k 2^j 2^{-Mk}(\log\tfrac{1}{\delta})^3 \cdot 2^{Mk}\frac{m}{2^j} \lesssim m(\log\tfrac{1}{\delta})^4$$

Summing over j we get $\sum_{i,j} |\mathcal{C}_i^j| \lesssim m(\log\frac{1}{\delta})^5$. Thus, there are at most $m(\log\frac{1}{\delta})^5$ pairs $(w, \mathcal{C})$ where w is a white tube and $\mathcal{C} = \mathcal{C}_i^j$ is an η-bush containing w. This obviously implies 1. It remains to prove 3.

Fix μ. If $\mu \lesssim \delta^B$ (and if $B = B_d$ was chosen large enough) then 3. will clearly hold, since the right hand side will be greater than δ^{-d}. On the other hand, if μ is large compared with δ^B then we can choose j_0 so that $\frac{m}{2^{j_0}}$ is less than $\frac{\mu}{2}$ but greater than $\frac{\mu}{8}$. We consider the decomposition (7) with this value of j_0. Thus $\Phi_{\mathcal{W}_g^{j_0}} \le \frac{\mu}{2}$, and, for each k we have

$$|\{(i,j) : j \le j_0 \text{ and } \mathcal{C}_i^j \text{ is a } 2^k\delta-\text{bush}\}| \lesssim 2^{-Mk}(\log\tfrac{1}{\delta})^3\frac{m}{\mu} \tag{8}$$

Fix a black tube b, and fix also a choice of $\mathcal{C}_i^j$ with $j \le j_0$. Define Ω_{ij} by deleting from $Q(1)$ the δ^ϵ-square containing p_i^j and its neighbors. Lemma 1.3 implies that if $\mathcal{C}_i^j$ is a $2^k\delta$-bush then

$$\int_{\Omega_{ij}} \phi_b \Phi_{\mathcal{C}_i^j} \lesssim \delta^{-(d-2)\epsilon} 2^{Ck}\delta^d$$

where C depends on d.

Now sum over b, i and $j \le j_0$ obtaining (provided M has been chosen large enough)

$$\sum_{ij} \int_{\Omega_{ij}} \Phi_{\mathcal{B}} \Phi_{\mathcal{C}_i^j} \lesssim n \sum_k 2^{-Mk} (\log \tfrac{1}{\delta})^3 \frac{m}{\mu} \cdot \delta^{-(d-2)\epsilon} 2^{Ck} \delta^d \lesssim \delta^{-(d-2)\epsilon} \frac{nm}{\mu} \delta^d (\log \tfrac{1}{\delta})^3$$

where the first inequality followed from (8).

Suppose now that x is a point such that $\tilde{\Phi}_{\mathcal{W}}(x) \ge \mu$. By the definition of the relation $\sim$ we have

$$\tilde{\Phi}_{\mathcal{W}}(x) \le \Phi_{\mathcal{W}_g^{j_0}}(x) + \sum_{\substack{j \le j_0 \\ x \in \Omega_{ij}}} \Phi_{\mathcal{C}_i^j}(x)$$

The first term on the right side is $\le \frac{\mu}{2}$, so

$$\tilde{\Phi}_{\mathcal{W}}(x) \le 2 \sum_{\substack{j \le j_0 \\ x \in \Omega_{ij}}} \Phi_{\mathcal{C}_i^j}(x)$$

whence

$$\int \Phi_{\mathcal{B}} \tilde{\Phi} \le 2 \sum_{ij} \int_{\Omega_{ij}} \Phi_{\mathcal{B}} \Phi_{\mathcal{C}_i^j} \lesssim \delta^{-(d-2)\epsilon} \frac{nm}{\mu} \delta^d (\log \tfrac{1}{\delta})^3$$

It follows that the measure of the set where $\Phi_{\mathcal{B}} \ge \nu$ and $\tilde{\Phi} \ge \mu$ is

$$\lesssim \delta^{-(d-2)\epsilon} \frac{nm}{\nu \mu^2} \delta^d (\log \tfrac{1}{\delta})^3$$

Using that the functions ϕ_w are roughly constant on δ-discs it then follows that the δ-entropy is

$$\lesssim \delta^{-(d-2)\epsilon} \frac{nm}{\nu \mu^2} (\log \tfrac{1}{\delta})^3$$

as claimed. □

What we actually use below is a slight variant on Lemma 1.1 where the infinite cylinders are replaced by finite ones. We introduce the following notation which will also be used in section 3.

Definition 1. Suppose that g is a radial function in $\mathbb{R}^d$ and R is a centered compact convex set. Then we use the notation g_R to mean $g \circ A$, where A is an affine function mapping (the John ellipsoid for) R onto the unit ball.

2. ϕ will denote the function $\phi(x) = \min(1, |x|^{-M})$, where M is a sufficiently large constant.

Suppose now that we have collections $\mathcal{B}$ and $\mathcal{W}$ of cylinders of length 1 and cross section radius δ, which are δ-separated in the same sense as before; i.e. the ones whose direction belongs to a given δ-disc in projective space have bounded overlap, and furthermore the axis directions belong to Γ_1 and Γ_2 respectively. Let $m = |\mathcal{W}|$, $n = |\mathcal{B}|$. Fix (in addition to ϵ) another small positive η; the choice of M and the implicit constants below may now also depend on η. The quantities $\Phi_{\mathcal{W}}$ and $\tilde{\Phi}_{\mathcal{W}}$ are defined in the same way as before, except of course that we use the modified definiton of ϕ_w via the definition above.

<u>Lemma 1.1′</u> With the above assumptions there is a relation $\sim$ between white or black tubes $w \in \mathcal{W}$ or $b \in \mathcal{B}$ and δ^{ϵ}-cubes $Q \subset Q(1)$ so that the following hold, where $n_{\mathcal{W}}(Q) = |\{w : w \sim Q\}|$:

1. $\sum_Q n_{\mathcal{W}}(Q) \lesssim m\delta^{-\eta}$.
2. $\sum_Q n_{\mathcal{B}}(Q) \lesssim n\delta^{-\eta}$.
3. The δ-entropy of the set $\{x \in Q(1) : \tilde{\Phi}_{\mathcal{W}}(x) \geq \mu \text{ and } \Phi_{\mathcal{B}}(x) \geq \nu\}$ is $\lesssim \delta^{-C\epsilon} \frac{mn}{\mu^2\nu}$.
4. The δ-entropy of the set $\{x \in Q(1) : \Phi_{\mathcal{W}}(x) \geq \mu \text{ and } \tilde{\Phi}_{\mathcal{B}}(x) \geq \nu\}$ is $\lesssim \delta^{-C\epsilon} \frac{mn}{\mu\nu^2}$.

To prove this we define $w \sim Q$ if the infinite cylinder[1] with the same axis as w is related to Q in the sense of Lemma 1.1 and if in addition the distance from w to the origin is less than $\delta^{-\frac{\eta}{2}}$. Then 1. and 2. follow immediately from 1. and 2. of Lemma 1.1, and 3. and 4. follow from 3. and 4. of Lemma 1.1 using that the contribution to Φ from tubes further than $\delta^{-\frac{\eta}{2}}$ from the origin is negligibly small if M is large. $\square$

2. A lemma of Mockenhaupt

We cover the unit sphere S^{d-2} with a family of spherical caps c of radius $N^{-\frac{1}{2}}$ with bounded overlap; this gives also a covering of Γ by a family of "sectors" $\rho = \rho_c$, where $\rho_c = \{x \in \Gamma : \frac{\overline{x}}{x_d} \in c\}$.

We will be using a variant on the square function estimate in [10]. To state it, let $\{\rho_j\}$ be the sectors ρ which intersect Γ_1 and let $\{\tilde{\rho}_k\}$ be the sectors which intersect Γ_2. Let

[1] We allow the possibility that an infinite cylinder may contain several w's. It is therefore easy to reduce to the case where the infinite cylinders are δ-separated.

f and g be two functions on $\mathbb{R}^d$ and assume that $f = \sum_{j=1}^{\mu} f_j$ and $g = \sum_{k=1}^{\nu} g_k$, where $\mathrm{supp} f_j$ is contained in the $N^{-\frac{1}{2}}$-neighborhood of the sector $\rho = \rho_j$, and likewise $\mathrm{supp} g_k$ is contained in the $N^{-\frac{1}{2}}$-neighborhood of $\tilde{\rho}_k$. Let $F = \hat{f}$, $G = \hat{g}$, and $SF = (\sum_j |\hat{f}_j|^2)^{\frac{1}{2}}$ $SG = (\sum_k |\hat{g}_k|^2)^{\frac{1}{2}}$.

Lemma 2.1 $\|FG\|_2^2 \lesssim \min(\mu, \nu) \|(SF)(SG)\|_2^2$.

Proof[10] We claim that for a given point $z \in \mathbb{R}^d$ there are $\lesssim \min(\mu, \nu)$ pairs (j, k) such that $z \in \mathrm{supp} f_j + \mathrm{supp} g_k$.

We will use the following geometrically obvious fact (a consequence of the strict convexity of the sphere): let ϵ_0 be a fixed positive constant and let $\zeta, \omega_1, \omega_2$ be points of S^{d-2} with $|\omega_i - \zeta| \geq \epsilon_0$ for $i = 1, 2$. Let ℓ be a line in $\mathbb{R}^{d-1}$ which passes through the point ζ and assume that both ω_1 and ω_2 are at distance at most δ from ℓ. Then $|\omega_1 - \omega_2| \leq C\delta$, where C depends on ϵ_0.

In order to prove the claim it suffices to show that for fixed j the set of k such that $z \in \mathrm{supp} f_j + \mathrm{supp} g_k$ has bounded cardinality. To this end we fix ζ with $(\zeta, 1) \in \rho_j$, and ω_1 and ω_2 such that $(\omega_i, 1) \in \tilde{\rho}_k$ and $z \in \mathrm{supp} f_j + \mathrm{supp} g_{k_i}$ for $i = 1, 2$. If we let $z = (w, t)$ then for suitable $a, b \in [\frac{1}{2}, 2]$ we have

$$a + b = t + \mathcal{O}(N^{-\frac{1}{2}})$$

$$a\omega_1 + b\zeta = w + \mathcal{O}(N^{-\frac{1}{2}})$$

and therefore

$$a\omega_1 + (t - a)\zeta = w + \mathcal{O}(N^{-\frac{1}{2}})$$

so that

$$\omega_1 - \zeta = a^{-1}(w - t\zeta) + \mathcal{O}(N^{-\frac{1}{2}}) \tag{9}$$

Estimate (9) says that the distance from ω_1 to the line through ζ spanned by $w - t\zeta$ is $\lesssim N^{-\frac{1}{2}}$. Likewise the distance from ω_2 to this line is $\lesssim N^{-\frac{1}{2}}$. The disjoint conical support assumption implies that $|\omega_i - \zeta|$ is bounded below for each i so we conclude that $|\omega_1 - \omega_2| \leq CN^{-\frac{1}{2}}$. This means that there are at most a bounded number of possible values for k, proving the claim.

The claim implies the lemma by a well-known calculation with the Plancherel theorem, which we omit. $\square$

3. Main lemma

It will be convenient to change the setup described in the introduction slightly in this section. We fix a scale N, let $Q(N)$ be the square centered at the origin with side N, and let $\Gamma^{(N)}$ be the $\frac{1}{N}$-neighborhood of Γ_1; similarly $\Gamma_1^{(N)}$ is the $\frac{1}{N}$-neighborhood of Γ_1, etc. Corresponding to the covering of Γ by sectors described in section 2 is a covering of $\Gamma^{(N)}$ by $\frac{1}{N}$-neighborhoods of sectors, and in this section we use ρ to denote one of the latter. Thus ρ is essentially a $1\times\overbrace{N^{-\frac{1}{2}}\times\ldots\times N^{-\frac{1}{2}}}^{d-2\,times}\times N^{-1}$-rectangle. We fix disjoint sets $E_\rho\subset\rho$ with $\cup_\rho E_\rho=\Gamma^{(N)}$ and let $\zeta_\rho=\chi_{E_\rho}$.

Let f be a function supported on $\Gamma_1^{(N)}$ with L^2 norm 1, $F=\hat{f}$, $F_\rho=\widehat{\zeta_\rho f}$, and

$$SF(x)=(\sum_\rho|F_\rho(x)|^2)^{\frac{1}{2}}$$

Further let b be a fixed radial Schwartz function nonzero on $Q(1)$ whose Fourier transform has compact support and whose $\mathbb{Z}^d$ translations form a partition of unity. For each ρ we fix a tiling $\mathcal{F}^\rho$ of $\mathbb{R}^d$ by rectangles σ with dimensions $N\times\overbrace{N^{\frac{1}{2}}\times\ldots\times N^{\frac{1}{2}}}^{d-1\,times}$, the long direction being orthogonal to the light cone Γ at points of (the center line of) ρ, and we let $\mathcal{F}=\cup_\rho\mathcal{F}^\rho$. We also let $\mathcal{P}^\rho$ be a tiling by $N\times\overbrace{N^{\frac{1}{2}}\times\ldots\times N^{\frac{1}{2}}}^{d-2\,times}\times 1$ rectangles dual to the sector ρ. For each ρ and each $\sigma\in\mathcal{F}^\rho$ we define $F_\rho^\sigma=b_\sigma F_\rho$, where b_σ (and also b_π, ϕ_σ, etc. in the subsequent argument) are as in the definition at the end of section 1; thus $\sum_\sigma F_\rho^\sigma=F_\rho$. For each (ρ,σ) we also further decompose F_ρ^σ as $\sum_{\pi\in\mathcal{P}^\rho}F_\rho^{\sigma,\pi}$, where $F_\rho^{\sigma,\pi}=b_\pi F_\rho^\sigma$. The following fact (trivial to prove, since b has compact support) will be very important below:

<u>Lemma 3.1</u> The inverse Fourier transforms of the functions F_ρ^σ and $F_\rho^{\sigma,\pi}$ are supported in a fixed dilate $\overline{\rho}$ of ρ, and in particular are supported in the $\frac{C}{N}$-neighborhood of Γ.

The following fact is also clear from the Schwartz inequality since $\sum_{\sigma\in\mathcal{F}^\rho}\phi_\sigma^2$ and $\sum_{\pi\in\mathcal{P}^\rho}\phi_\pi^2$ are bounded for fixed ρ. Suppose that for each ρ a subset $\mathcal{A}^\rho\subset\mathcal{F}^\rho$ is given. Then

$$\sum_\rho|\sum_{\sigma\in\mathcal{A}^\rho}F_\rho^\sigma|^2\lesssim\sum_\rho\sum_{\sigma\in\mathcal{A}^\rho}|F_\rho^\sigma|^2\phi_\sigma^{-2}\lesssim\sum_{\substack{\rho,\sigma,\pi\\ \sigma\in\mathcal{A}^\rho\\ \pi\in\mathcal{P}^\rho}}|F_\rho^{\sigma,\pi}|^2\phi_\pi^{-2}\phi_\sigma^{-2} \qquad (10)$$

The next two lemmas keep track of some relationships among the various decompositions of F which follow from orthogonality considerations and the uncertainty principle. We note the following: let π_0 be a rectangle containing the origin, and let π be a translate of π_0. Then, the operator with kernel

$$K(x,y) = \phi_\pi(x)^{-2}\phi_{\pi_0}(x-y)^{100}\phi_\pi(y)^4$$

maps L^2 to L^∞ with norm $\lesssim |\pi|^{\frac{1}{2}}$, since one can easily show that $\int |K(x,y)|^2 dy \lesssim |\pi|$ for fixed x.

Lemma 3.2 For fixed ρ and $\sigma \in \mathcal{F}^\rho$ we have $\sum_\pi \|\phi_\pi^{-2}\phi_\sigma^{-3}F_\rho^{\sigma,\pi}\|_\infty^2 \lesssim N^{-\frac{d}{2}}\|\phi_\sigma^{-4}F_\rho^\sigma\|_2^2$.

Proof Fix a Schwartz function κ whose Fourier transform is 1 on the unit ball and let $\overline{\kappa}$ be the corresponding function whose Fourier transform is 1 on the set $\overline{\rho}$ in Lemma 3.1, obtained from κ by composition with a linear map followed by multiplication by a character and by a scalar with magnitude about $|\rho|$. Then $F_\rho^{\sigma,\pi} = \overline{\kappa} * F_\rho^{\sigma,\pi}$. Let σ_0 and π_0 be the rectangles in the tilings $\mathcal{F}^\rho$ and $\mathcal{P}^\rho$ which contain the origin. Then $|\overline{\kappa}(z)| \lesssim |\rho|\phi_{\pi_0}(z)^{200} \lesssim |\rho|\phi_{\pi_0}(z)^{100}\phi_{\sigma_0}(z)^{100}$. We conclude that

$$|\phi_\pi^{-2}(x)\phi_\sigma^{-3}(x)F_\rho^{\sigma,\pi}(x)| \lesssim \int K(x,y)|\phi_\pi^{-4}(y)\phi_\sigma^{-4}(y)F_\rho^{\sigma,\pi}(y)|dy$$

where

$$\begin{aligned} K(x,y) &= |\rho|\phi_\pi(x)^{-2}\phi_\sigma(x)^{-3}\phi_{\pi_0}(x-y)^{100}\phi_{\sigma_0}(x-y)^{100}\phi_\pi(y)^4\phi_\sigma(y)^4 \\ &\lesssim |\rho|\phi_\pi(x)^{-2}\phi_{\pi_0}(x-y)^{100}\phi_\pi(y)^4 \end{aligned}$$

We've seen that the norm of this kernel from L^2 to L^∞ is $\lesssim |\pi|^{\frac{1}{2}}|\rho| \approx |\rho|^{\frac{1}{2}} \approx N^{-\frac{d}{4}}$. Accordingly

$$\begin{aligned} \sum_\pi \|\phi_\pi^{-2}\phi_\sigma^{-3}F_\rho^{\sigma,\pi}\|_\infty^2 &\lesssim N^{-\frac{d}{2}}\sum_\pi \|\phi_\pi^{-4}\phi_\sigma^{-4}F_\rho^{\sigma,\pi}\|_2^2 \\ &= N^{-\frac{d}{2}}\sum_\pi \|(\phi_\pi^{-4}b_\pi)\phi_\sigma^{-4}F_\rho^\sigma\|_2^2 \end{aligned}$$

and now we use that $\sum_\pi |\phi_\pi^{-4}b_\pi|^2 \lesssim 1$ pointwise, obtaining the lemma. $\square$

For each ρ and each $\sigma \in \mathcal{F}^\rho$ we define a parameter

$$h(\sigma) = (N^{-\frac{d+1}{2}} \|\phi_\sigma^{-4} F_\rho^\sigma\|_2^2)^{\frac{1}{2}}$$

We think of $h(\sigma)$ as being essentially the L^2 average of F_ρ on σ. We group the σ's into families corresponding to the different possible dyadic values for $h(\sigma)$; thus

$$\mathcal{F}(h) = \{\sigma \in \mathcal{F} : h(\sigma) \in [\frac{h}{2}, h]\}$$

and we define

$$F_h = \sum_\rho \sum_{\sigma \in \mathcal{F}(h) \cap \mathcal{F}^\rho} F_\rho^\sigma$$

Lemma 3.3 $h^2 |\mathcal{F}(h)| \lesssim N^{-\frac{d+1}{2}}$.

Proof Clearly

$$h^2 |\mathcal{F}(h)| \lesssim N^{-\frac{d+1}{2}} \sum_\rho \sum_{\sigma \in \mathcal{F}^\rho} \|\phi_\sigma^{-4} F_\rho^\sigma\|_2^2$$

For fixed ρ we have $\sum_\sigma |b_\sigma \phi_\sigma^{-4}|^2 \lesssim 1$ pointwise. So for fixed ρ we have $\sum_{\sigma \in \mathcal{F}^\rho} \|\phi_\sigma^{-4} F_\rho^\sigma\|_2^2 \lesssim \|F_\rho\|_2^2$. If we sum over ρ and use orthogonality of the F_ρ's the lemma follows. □

If g is a function supported on $\Gamma_2^{(N)}$ with L^2 norm 1 we will likewise denote $\widehat{g d\sigma}$ by G, etc. Thus we obtain also functions G_ρ, G_ρ^σ, $G_\rho^{\sigma,\pi}$, G_h, and families of tubes $\mathcal{G}$, $\mathcal{G}^\rho$, $\mathcal{G}(h)$. The next lemma is a "local" estimate; it will then be combined with Lemma 1.1 to give the following Lemma 3.5 which is the main result of this section.

Lemma 3.4 Fix a square Q with side $\sqrt{N}$. Let $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{G}}$ be subsets of $\mathcal{F}(h_1)$ and $\mathcal{G}(h_2)$ respectively and let μ and ν be the maximum values on the square Q of the functions $\Phi_{\tilde{\mathcal{F}}}$ and $\Phi_{\tilde{\mathcal{G}}}$. Then

$$\int_Q |(\sum_\rho \sum_{\sigma \in \tilde{\mathcal{F}} \cap \mathcal{F}^\rho} F_\rho^\sigma)(\sum_{\rho_2} \sum_{\sigma_2 \in \tilde{\mathcal{G}} \cap \mathcal{G}^{\rho_2}} G_{\rho_2}^{\sigma_2})|^2 \lesssim h_1^2 h_2^2 \mu\nu \min(\mu, \nu) N^{\frac{d}{2}} \tag{11}$$

Proof We subdivide $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{G}}$ according to the possible dyadic values for ϕ_σ on Q. Thus we define

$$\tilde{\mathcal{F}}(k) = \{\sigma \in \tilde{\mathcal{F}} : \min_Q \phi_\sigma \in [2^{-(k+1)}, 2^{-k}]\}$$

$$\tilde{\mathcal{G}}(\ell) = \{\sigma \in \tilde{\mathcal{G}} : \min_Q \phi_\sigma \in [2^{-(\ell+1)}, 2^{-\ell}]\}$$

We note that if $\sigma \in \tilde{\mathcal{F}}(k)$ then

$$\|\phi_\sigma \phi_Q\|_\infty \lesssim 2^{-k} \tag{12}$$

This follows from the rapid decay of ϕ and the fact that σ contains a translate of Q. Hence also $\|\phi_\sigma b_Q\|_\infty \lesssim 2^{-k}$. Furthermore, from the definition of μ and ν, we have

$$|\tilde{\mathcal{F}}(k)| \lesssim 2^k \mu \text{ and } \tilde{\mathcal{G}}(\ell) \lesssim 2^\ell \nu \tag{13}$$

The left side of (11) is $\lesssim \sum_{k=0}^\infty \sum_{\ell=0}^\infty 2^{k+\ell} A(k,\ell)$, where

$$A(k,\ell) = \int |(b_Q^3 \sum_\rho \sum_{\sigma \in \tilde{\mathcal{F}}(k) \cap \mathcal{F}^\rho} F_\rho^\sigma)(b_Q^3 \sum_{\rho_2} \sum_{\sigma_2 \in \tilde{\mathcal{G}}(\ell) \cap \mathcal{G}^{\rho_2}} G_{\rho_2}^{\sigma_2})|^2 \tag{14}$$

Using Lemma 3.1 and that $\hat{b}$ has compact support, one sees that the Fourier transform of the function $b_Q^3 \sum_{\sigma \in \tilde{\mathcal{F}}(k) \cap \mathcal{F}^\rho} F_\rho^\sigma$ is supported in the $CN^{-\frac{1}{2}}$-neighborhood of the sector ρ; and similarly with the second factor in (14). Lemma 2.1 is therefore applicable and implies that

$$A(k,\ell) \lesssim \min(|\tilde{\mathcal{F}}(k)|, |\tilde{\mathcal{G}}(\ell)|) \int \sum_\rho |b_Q^3 \sum_{\sigma \in \tilde{\mathcal{F}}(k) \cap \mathcal{F}^\rho} F_\rho^\sigma|^2 \sum_{\rho_2} |b_Q^3 \sum_{\sigma_2 \in \tilde{\mathcal{G}}(\ell) \cap \mathcal{G}^{\rho_2}} G_{\rho_2}^{\sigma_2}|^2$$

It follows by (10) that $A(k,\ell)$ is

$$\lesssim \min(|\tilde{\mathcal{F}}(k)|, |\tilde{\mathcal{G}}(\ell)|) \int b_Q^{12} \sum_{\substack{\rho,\sigma,\pi \\ \sigma \in \tilde{\mathcal{F}}(k) \cap \mathcal{F}^\rho}} \sum_{\substack{\rho_2,\sigma_2,\pi_2 \\ \sigma_2 \in \mathcal{G}(\ell) \cap \mathcal{G}^{\rho_2}}} |F_\rho^{\sigma,\pi}|^2 |G_{\rho_2}^{\sigma_2,\pi_2}|^2 \phi_\pi^{-2} \phi_\sigma^{-2} \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-2} \tag{15}$$

We claim next that for each pair (π, π_2) we have

$$\int b_Q^4 \phi_\pi^2 \phi_{\pi_2}^2 \lesssim N^{\frac{d-2}{2}} \tag{16}$$

Namely, π and π_2 each have one "short" direction in which the width is 1, and these directions lie in Γ_1 and Γ_2 respectively, and are therefore transverse. It follows that $\pi \cap \pi_2$ is contained within a bounded distance of a $d-2$-plane, hence that

$$\int_{\pi \cap \pi_2} b_Q^4 \lesssim N^{\frac{d-2}{2}} \tag{17}$$

Estimate (16) is just a version of (17) incorporating Schwartz tails, and is proved by estimating ϕ_π^2 by an appropriate sum of constants times characteristic functions of translates of π (and similarly with $\phi_{\pi_2}^2$) and then applying (17) to the terms in the resulting series.

We now consider the terms in the sum (15). For each pair (ρ, σ, π) and $(\rho_2, \sigma_2, \pi_2)$ we have

$$\begin{aligned} &\int b_Q^{12} |F_\rho^{\sigma,\pi}|^2 |G_{\rho_2}^{\sigma_2,\pi_2}|^2 \phi_\pi^{-2} \phi_\sigma^{-2} \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-2} \\ \lesssim\ & \|b_Q^2 \phi_\pi^{-2} \phi_\sigma^{-1} F_\rho^{\sigma,\pi}\|_\infty^2 \|b_Q^2 \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-1} G_{\rho_2}^{\sigma_2,\pi_2}\|_\infty^2 \int b_Q^4 \phi_\pi^2 \phi_{\pi_2}^2 \\ \lesssim\ & N^{\frac{d-2}{2}} \|b_Q^2 \phi_\pi^{-2} \phi_\sigma^{-1} F_\rho^{\sigma,\pi}\|_\infty^2 \|b_Q^2 \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-1} G_{\rho_2}^{\sigma_2,\pi_2}\|_\infty^2 \end{aligned} \tag{18}$$

by (16). It then follows that

$$\begin{aligned} \int b_Q^{12} |F_\rho^{\sigma,\pi}|^2 |G_{\rho_2}^{\sigma_2,\pi_2}|^2 &\lesssim N^{\frac{d-2}{2}} \|\phi_\sigma^2 b_Q^2\|_\infty^2 \|\phi_{\sigma_2}^2 b_Q^2\|_\infty^2 \|\phi_\sigma^{-3} \phi_\pi^{-2} F_\rho^{\sigma,\pi}\|_\infty^2 \|\phi_{\sigma_2}^{-3} \phi_{\pi_2}^{-2} G_{\rho_2}^{\sigma_2,\pi_2}\|_\infty^2 \\ &\lesssim 2^{-4k-4\ell} N^{\frac{d-2}{2}} \|\phi_\sigma^{-3} \phi_\pi^{-2} F_\rho^{\sigma,\pi}\|_\infty^2 \|\phi_{\sigma_2}^{-3} \phi_{\pi_2}^{-2} G_{\rho_2}^{\sigma_2,\pi_2}\|_\infty^2 \end{aligned} \tag{19}$$

The first inequality followed from (18) by rearranging some factors, and the second inequality followed from (12).

Using (19) and Lemma 3.2 we may now bound (15) by

$$\min(|\tilde{\mathcal{F}}(k)|, |\tilde{\mathcal{G}}(\ell)|) 2^{-4k-4\ell} N^{-\frac{d+2}{2}} \sum_{\rho,\rho_2} \sum_{\substack{\sigma \in \tilde{\mathcal{F}}(k) \cap \mathcal{F}^\rho \\ \sigma_2 \in \tilde{\mathcal{G}}(\ell) \cap \mathcal{G}^{\rho_2}}} \|\phi_\sigma^{-4} F_\rho^\sigma\|_2^2 \|\phi_{\sigma_2}^{-4} G_{\rho_2}^{\sigma_2}\|_2^2$$

which by definition of h_1 and h_2 is

$$\lesssim \min(|\tilde{\mathcal{F}}(k)|, |\tilde{\mathcal{G}}(\ell)|) 2^{-4k-4\ell} N^{-\frac{d+2}{2}} \cdot N^{d+1} |\tilde{\mathcal{F}}(k)|\, |\tilde{\mathcal{G}}(\ell)| h_1^2 h_2^2$$

We now use (13), and obtain a bound on (15) by

$$2^{-3k-3\ell} \mu\nu \min(2^k \mu, 2^\ell \nu) N^{\frac{d}{2}} h_1^2 h_2^2$$

Summing over k and ℓ gives the lemma. □

Fix $\epsilon > 0$ and then $\eta > 0$ and partition $Q(N)$ in nonoverlapping $N^{1-\epsilon}$-squares; the letter R below will always denote one of these squares. We recall that f and g have L^2 norm 1 and are supported on $\Gamma_1^{(N)}$ and $\Gamma_2^{(N)}$ respectively.

Lemma 3.5 On $Q(N)$, for any h_1 and h_2 there are decompositions

$$F_{h_1} = F_g + F_b \text{ and } F_b = \sum_R F_b^R$$

$$G_{h_2} = G_g + G_b \text{ and } G_b = \sum_R \mathcal{G}_b^R$$

where supp$F_b^R \subset R$, supp$G_b^R \subset R$, and the following estimates hold.

1. $\int_{Q(N)} |F_g G_g|^2 + |F_b G_g|^2 + |F_g G_b|^2 \lesssim N^{-\frac{d+2}{2}+C\epsilon}$

2. For each R we have $F_b^R = \alpha_R \widehat{f_R}$ and $G_b^R = \beta_R \widehat{g_R}$, where α_R and β_R are supported on R and have L^∞ norm ≤ 1, and f_R and g_R are supported on the $N^{-(1-\epsilon)}$-neighborhoods of Γ_1 and Γ_2 respectively, and

$$\sum_R \|f_R\|_2^2 + \|g_R\|_2^2 \leq C_\eta N^{-\epsilon+\eta} \tag{20}$$

Proof Let $\mathcal{W} = \mathcal{F}(h_1)$, $\mathcal{B} = \mathcal{G}(h_2)$. We can assume that both h_1 and h_2 are greater than N^{-B_1} where B_1 is a large dimension-dependent constant, since otherwise it is easy to check that the lemma is valid with F_b and G_b equal to zero. It follows that the cardinalities of $\mathcal{W}$ and $\mathcal{B}$ are bounded by N^{B_2}.

We apply Lemma 1.1$'$ after rescaling by N; thus δ in Lemma 1.1 is $N^{-\frac{1}{2}}$; and we also set ϵ in Lemma 1.1$'$ equal to twice the present ϵ.

For each $N^{1-\epsilon}$-square R we then define

$$F_b^R = \begin{cases} \sum_{\substack{\sigma \in \mathcal{W} \\ \sigma \sim R}} F_\rho^\sigma & \text{on } R \\ 0 & \text{elsewhere} \end{cases}$$

$$G_b^R = \begin{cases} \sum_{\substack{\sigma \in \mathcal{B} \\ \sigma \sim R}} G_\rho^\sigma & \text{on } R \\ 0 & \text{elsewhere} \end{cases}$$

Define F_b to be equal to F_b^R on R for each R and similarly with G_b, and define $F_g = F - F_b$, $G_g = G - G_b$.

We will now show that

$$\int_{Q(N)} |F_g G|^2 \lesssim N^{-\frac{d+2}{2}+C\epsilon}$$

Namely, fix a $\sqrt{N}$-square Q. Define μ to be the maximum on Q of $\Phi_{\tilde{\mathcal{F}}}$, where $\tilde{\mathcal{F}}$ is the tubes $w \in \mathcal{F}(h_1)$ such that $w \not\sim Q$, and define ν to be the maximum on Q of $\Phi_{\mathcal{G}(h_2)}$. By Lemma 3.4 we have

$$\int_Q |F_g G|^2 \lesssim h_1^2 h_2^2 \mu^2 \nu N^{\frac{d}{2}}$$

We now sum over Q and use 3. of Lemma 1.1. This gives

$$\int_{Q(N)} |F_g G|^2 \lesssim N^{C\epsilon} h_1^2 h_2^2 N^{\frac{d}{2}} |\mathcal{F}(h_1)|\, |\mathcal{G}(h_2)|$$

which is $\lesssim N^{C\epsilon} \cdot N^{-\frac{d+2}{2}}$ by Lemma 3.3. We can clearly estimate $\int_{Q(N)} |F_g G|^2$ and $\int_{Q(N)} |F_g G_g|^2$ in the same way, and it follows that 1. holds.

We have the following almost orthogonality estimate:

$$\| \sum_{\sigma \sim R} F_\rho^\sigma \|_2^2 \lesssim h_1^2 N^{\frac{d+1}{2}} |\{\sigma : \sigma \sim R\}| \tag{21}$$

Namely, for fixed ρ we have

$$\| \sum_{\substack{\sigma \sim R \\ \sigma \in \mathcal{F}^\rho}} F_\rho^\sigma \|_2^2 \lesssim \sum_{\substack{\sigma \sim R \\ \sigma \in \mathcal{F}^\rho}} \| \phi_\sigma^{-4} F_\rho^\sigma \|_2^2 \lesssim h_1^2 N^{\frac{d+1}{2}} |\{\sigma \in \mathcal{F}^\rho : \sigma \sim R\}|$$

where the first inequality follows from the Schwartz inequality since $\sum_\sigma \phi_\sigma^8 \lesssim 1$ pointwise and the second follows from the definition of h_1. Lemma 3.1 implies that the functions $\sum_{\substack{\sigma \sim R \\ \sigma \in \mathcal{F}^\rho}} F_\rho^\sigma$ are essentially orthogonal for different ρ and (21) follows.

Using Lemma 3.1 again we see that, on each fixed $N^{1-\epsilon}$ square R, $F_b = \sum_{\sigma \sim R} F_\rho^\sigma$ agrees with the Fourier transform of a function f_R^0 supported on the $\frac{C}{N}$-neighborhood of Γ_1. We have

$$\begin{aligned} \sum_R \|f_R^0\|_2^2 = \sum_R \| \sum_{\sigma \sim R} F_\rho^\sigma \|_2^2 &\lesssim h_1^2 N^{\frac{d+1}{2}} \sum_R |\{\sigma : \sigma \sim R\}| \\ &\lesssim h_1^2 N^{\frac{d+1}{2}} |\mathcal{F}(h_1)| N^\eta \\ &\lesssim N^\eta \end{aligned}$$

The first inequality followed from (21), the second inequality followed from 1. of Lemma 1.1$'$ and the last inequality followed from Lemma 3.3. Now fix R and take a suitable

Schwartz function κ supported in $D(0, \frac{1}{2})$ and whose Fourier transform is ≥ 1 on a large disc centered at the origin. Let $\kappa_R(x) = e^{ik\cdot x}N^{d(1-\epsilon)}\kappa(N^{1-\epsilon}x)$ for an appropriate k; if k is chosen correctly then $\widehat{\kappa_R} \geq 1$ on R. Define $\alpha_R = \frac{1}{\widehat{\kappa_R}}$ on R and zero otherwise, and $f_R = \kappa_R * f_R^0$. Then $F_b^R = \alpha_R \widehat{f_R}$. To make the estimate (20) we will use the following fact, which follows from Schur's test:

If s is a function supported in $D(0,r)$ with $\|s\|_\infty \leq |D(0,r)|^{-1}$ and if supp$f$ intersects every disc of radius r in measure $\leq \gamma r$, then $\|s * f\|_2 \lesssim \gamma^{\frac{1}{2}}\|f\|_2$.

We apply this with $s = \kappa_R$, $f = f_R^0$, $r \approx N^{1-\epsilon}$, $\gamma \approx N^{-\epsilon}$, which is justified since f_R^0 is supported on the $\frac{C}{N}$-neighborhood of Γ. It follows that $\|f_R\|_2^2 \lesssim N^{-\epsilon}\|f_R^0\|_2^2$, so we have the part of (20) which relates to f. We can of course treat g the same way, so the proof is complete. □

We note also that the L^2 norms of F_g, F_b, G_g and G_b on $Q(N)$ are all bounded by a constant; it suffices to prove this for F_b and G_b, and for them it follows from (20).

4. Proof of Theorem 1

We will use a lemma from the previous work:

Lemma 4.1 [3],[14] In order to prove Theorem 1 it suffices to prove that

$$\int_{Q(N)} |\widehat{fd\sigma}\widehat{gd\sigma}|^p \lesssim N^\gamma \|f\|_2^p \|g\|_2^p$$

for fixed $p > 1 + \frac{2}{d}$ and $\gamma > 0$.

This lemma originates in section 4 of [3], and the version stated above is a special case of Lemma 2.4 in part I of [14]. We also make a further reduction which follows by the uncertainty principle in the usual way: it suffices to prove that if f and g are functions with L^2 norm 1 which are supported on the $\frac{1}{N}$-neighborhoods of Γ_1 and Γ_2 respectively, then

$$\int_{Q(N)} |\hat{f}\hat{g}|^p \lesssim N^{-p+\gamma} \tag{22}$$

if $p > 1 + \frac{2}{d}$ and $\gamma > 0$.

The rest of this section is the proof of (22).

Fix $p > 1 + \frac{2}{d}$ and let $\phi(N)$ be the supremum of the quantity

$$N^p \int_{Q(N)} |\hat{f}\hat{g}|^p \tag{23}$$

over functions f and g with L^2 norm 1 which are supported in the $\frac{1}{N}$-neighborhoods of Γ_1 and Γ_2 respectively. Fix a sufficiently small ϵ and then a much smaller η; we will show that

$$\phi(N) \le C(1 + N^\eta \phi(N^{1-\epsilon})) \tag{24}$$

for a suitable constant C.

Namely, choose f and g with L^2 norm 1 so that the quantity (23) is essentially maximized. Then choose h_1 and h_2 using the pigeonhole principle so that

$$\int_{Q(N)} |F_{h_1} G_{h_2}|^p \gtrsim (\log N)^{-2p} \phi(N)$$

where F_{h_1} and G_{h_2} were defined in section 3. This is possible since it is easy to see that parameter values h which are less than a high negative power of N make a negligible contribution. Now apply Lemma 3.5 with this choice of h_1 and h_2. With notation as in Lemma 3.5 we have (by the triangle inequality)

$$\phi(N) \lesssim (\log N)^{2p} \int_{Q(N)} (|F_b G_g|^p + |F_g G_b|^p + |F_g G_g|^p) \ + \ (\log N)^{2p} \sum_R \int_R |\widehat{f_R}\widehat{g_R}|^p$$

In the first term, we estimate the L^p norm by the L^1 and L^2 norms using Holder's inequality, and use that the L^1 norms of $F_g G_b$, $F_b G_g$ and $F_b G_b$ are bounded by a constant by the remark at the end of section 3. In the second term, by definition of $\phi(N^{1-\epsilon})$, we can estimate the integral over a fixed R by

$$N^{-(1-\epsilon)p} \phi(N^{1-\epsilon}) \|f_R\|_2^p \|g_R\|_2^p$$

Making these estimates we conclude that

$$\begin{aligned} N^{-p}\phi(N) \ \lesssim \ & (\log N)^{2p} \left(\int_{Q(N)} (|F G_g|^2 + |F_g G|^2 + |F_g G_g|^2) \right)^{p-1} + \\ & + (\log N)^{2p} N^{-(1-\epsilon)p} \phi(N^{1-\epsilon}) \sum_R \|f_R\|_2^p \|g_R\|_2^p \end{aligned}$$

We now use Holder's inequality on the sum over R and then insert the estimates in Lemma 3.5; this gives

$$\begin{aligned} N^{-p}\phi(N) &\lesssim (\log N)^{2p}\left(\int_{Q(N)}(|FG_g|^2+|F_gG|^2+|F_gG_g|^2)\right)^{p-1} \\ &\quad + (\log N)^{2p}N^{-(1-\epsilon)p}\phi(N^{1-\epsilon})(\sum_R \|f_R\|_2^2)^{\frac{p}{2}}(\sum_R \|g_R\|_2^2)^{\frac{p}{2}} \\ &\lesssim (\log N)^{2p}N^{(p-1)(C\epsilon-\frac{d+2}{2})} + (\log N)^{2p}N^{-(1-\epsilon)p}\phi(N^{1-\epsilon})\cdot N^{p(-\epsilon+\eta)} \end{aligned}$$

The assumption $p > 1+\frac{2}{d}$ implies that the exponent $p-\frac{d+2}{2}(p-1)$ is negative. We therefore obtain (24), since we can replace η by $\frac{\eta}{p+1}$, say.

If γ is given and if we take η sufficiently small then estimate (24) implies by an obvious induction that $\phi(N) \lesssim N^{\gamma}$; thus we have proved (22) and therefore Theorem 1. $\square$

5. Further remarks

We will now prove the corollary which was stated in the introduction. We first rephrase it in a somewhat sharper form and in general dimensions. We will use mixed norms on Γ splitting the S^{d-2} and radial variables:

$$\|f\|_{L^p(L^q)} \stackrel{def}{=} (\int_{S^{d-2}}(\int_1^2 |f(t\omega)|^q dt)^{\frac{p}{q}} d\omega)^{\frac{1}{p}}$$

In the statement below, note that when $d=4$ the condition on p reduces to $p>3$; by duality we obtain a bound $\|\hat{f}\|_{L^p(L^2)} \lesssim \|f\|_p$ for any $p<\frac{3}{2}$, which clearly includes the result that was stated in the introduction. When $d \geq 5$ the requirement that p be larger than $2+\frac{4}{d}$ becomes significant so the statement becomes weaker.

Corollary 1 Assume that $p > \max(2+\frac{4}{d}, 2+\frac{2}{d-2})$. Let f be a function on Γ. Then $\|\widehat{fd\sigma}\|_p \leq C_p\|f\|_{L^p(L^2)}$.

Proof This is the same as the proof of Theorem 2.2 in [15]; see also [14], where the rescaling maps for the cone employed below are used.

Fix a large number N and a spherical cap $c \subset S^{d-2}$ centered at a point $e \in S^{d-2}$ with radius N^{-1}, i.e. $c = \{\omega \in S^{d-2} : |\omega - e| < N^{-1}\}$. Let $\Gamma_c = \{x \in \Gamma : \frac{\overline{x}}{x_d} \in c\}$. Define T_c to be the linear map such that $T(e,1) = (e,1)$, $T(e,-1) = N^2(e,-1)$ and $Ty = Ny$ if

$y \in \mathbb{R}^d$ is orthogonal to $(e, 1)$ and $(e, -1)$. T_c maps light rays to light rays and has the following metric properties:

$$\det T_N = N^d \tag{25}$$

and T_c expands the distance between any two light rays contained in c by a factor of roughly N, and roughly preserves distances on each individual such light ray.

Let c_1 and c_2 be two caps contained in c separated by an amount comparable to N^{-1} and let f and g be functions on Γ with $L^p(L^2)$ norm 1 which are supported on Γ_{c_1} and Γ_{c_2} respectively. Define $\tilde{f} d\sigma$ and $\tilde{g} d\sigma$ to be the measures obtained by pushing forward $f d\sigma$ and $g d\sigma$ by the map T_c. Then $\tilde{f}$ and $\tilde{g}$ are functions on Γ whose conical supports are at least a constant distance apart, and their $L^p(L^2)$ norms are comparable to $N^{-\frac{d-2}{p'}}$; hence their L^2 norms are at most $N^{-\frac{d-2}{p'}}$. Furthermore we have the formulae

$$\widehat{f d\sigma} = \widehat{\tilde{f} d\sigma} \circ T_c^{-1}$$

$$\widehat{g d\sigma} = \widehat{\tilde{g} d\sigma} \circ T_c^{-1}$$

and therefore, by (25) and Theorem 1,

$$\begin{aligned} \int |\widehat{f d\sigma}\widehat{g d\sigma}|^{\frac{p}{2}} &= N^d \int |\widehat{\tilde{f} d\sigma}\widehat{\tilde{g} d\sigma}|^{\frac{p}{2}} \\ &\lesssim N^{d-p\frac{d-2}{p'}} = N^{d-(p-1)(d-2)} \end{aligned}$$

for any $p > 2 + \frac{4}{d}$. We now cover S^{d-2} with caps c_j of "width" N^{-1} as above and let f_j be functions on Γ with $\mathrm{supp} f_j \subset \Gamma_{c_j}$. By applying the preceding estimate and summing over j we obtain

$$\sum_{(j,k):\mathrm{dist}(c_j,c_k)\approx N^{-1}} \int |\widehat{\chi_j f d\sigma}\widehat{\chi_k f d\sigma}|^{\frac{p}{2}} \lesssim N^{d-(p-1)(d-2)} \sum_j \|f_j\|_p^p$$

The exponent of N is negative if $p > 2 + \frac{2}{d-2}$. The result now follows exactly as in [15], since the supports of the Fourier transforms of the functions $\widehat{\chi_j f d\sigma}\widehat{\chi_k f d\sigma}$ have finite overlap if N is fixed and $\mathrm{dist}(c_j, c_k) \approx N^{-1}$. $\square$

We now consider the Mockenhaupt square function

$$SF(x) = (\sum_\rho |F_\rho|^2)^{\frac{1}{2}}$$

where $F = \widehat{fd\sigma}$ with f supported on Γ, $f = \sum_\rho f_\rho$ with f_ρ supported in the sector ρ of width about $N^{-\frac{1}{2}}$ and $F_\rho = \widehat{f_\rho d\sigma}$. The following simple result appears natural in higher dimensions where the expected critical exponent is $2 + \frac{2}{d-2}$; we do not consider the question of $L^4(\mathbb{R}^3)$ estimates except to note that Theorem 1 can of course be substituted into the numerology in [14].

In addition, we consider a "weaker" square function $\tilde{S}$ defined as follows: let $F = \widehat{fd\sigma}$ be as above, let Δ run through a covering of Γ by discs of radius $N^{-\frac{1}{2}}$, suppose that f_Δ is supported in Δ and $f = \sum_\Delta f_\Delta, F_\Delta = \widehat{f_\Delta d\sigma}$ and

$$\tilde{S}F = (\sum_\Delta |F_\Delta|^2)^{\frac{1}{2}}$$

Corollary 2 If $2 \leq p \leq 2 + \frac{4}{d}$ then there is an estimate

$$\|F\|_p \lesssim N^{(\frac{1}{2}-\frac{1}{p})\frac{d-2}{4}+\epsilon}\|SF\|_p \tag{26}$$

for any $\epsilon > 0$.

Proof We consider first the "bilinear" version; in this version, one can prove a stronger result where $\tilde{S}F$ replaces SF. Thus we let f and g as in Theorem 1 and $F = \widehat{fd\sigma}$, $G = \widehat{gd\sigma}$, and will show that

$$\|FG\|_{\frac{p}{2}} \lesssim N^{(\frac{1}{2}-\frac{1}{p})\frac{d-2}{2}+\epsilon}(\|\tilde{S}F\|_p^2 + \|\tilde{S}G\|_p^2) \tag{27}$$

Namely, we have

$$\|FG\|_{L^{\frac{p}{2}}(Q)} \lesssim N^{-\frac{1}{2}+\epsilon}\|b_Q F\|_2 \|b_Q G\|_2$$

when $p > 2+\frac{4}{d}$. This follows by applying (22) (with N replaced by $N^{\frac{1}{2}}$ and $Q(N)$ replaced by Q) to the functions $b_Q F$ and $b_Q G$. By interpolation with L^2 there is also an estimate

$$\|FG\|_{L^{\frac{p}{2}}(Q)} \lesssim N^{-\frac{d+2}{2}(\frac{1}{2}-\frac{1}{p})+\epsilon}\|b_Q F\|_2 \|b_Q G\|_2$$

when $2 \leq p \leq 2 + \frac{4}{d}$. The $b_Q F_\Delta$'s are essentially orthogonal (their Fourier supports are essentially disjoint) so we can estimate $\|b_Q F\|_2$ by $\|b_Q \tilde{S}F\|_2$; using this and then Holder's inequality we obtain

$$\begin{aligned}
\|FG\|_{L^{\frac{p}{2}}(Q)} &\lesssim N^{-\frac{d+2}{2}(\frac{1}{2}-\frac{1}{p})+\epsilon}\|b_Q \tilde{S}F\|_2 \|b_Q \tilde{S}G\|_2 \\
&\lesssim N^{-\frac{d+2}{2}(\frac{1}{2}-\frac{1}{p})+\epsilon} \cdot N^{\frac{d}{2}(1-\frac{2}{p}+\epsilon)}\|b_Q \tilde{S}F\|_p \|b_Q \tilde{S}G\|_p \\
&\leq N^{(\frac{1}{2}-\frac{1}{p})\frac{d-2}{2}+\epsilon}(\|b_Q \tilde{S}F\|_p^2 + \|b_Q \tilde{S}G\|_p^2)
\end{aligned}$$

Now take an $\ell^{\frac{p}{2}}$ sum over Q. Using the rapid decay of b we obtain (27).

The same argument clearly applies to S, so we also have

$$\|FG\|_{\frac{p}{2}} \lesssim N^{(\frac{1}{2}-\frac{1}{p})\frac{d-2}{2}+\epsilon}(\|SF\|_p^2 + \|SG\|_p^2) \tag{28}$$

In the case of S, since the maps T_c essentially take sectors contained in Γ_c to sectors one can pass from the estimate (28) to the "linear" one (i.e. (26)) by rescaling, just as in [14] or in the proof of Corollary 1. □

Further remarks 1. It will be clear to the experts that one could also obtain a partial result on the (higher dimensional) cone multiplier/local smoothing problem using the estimate (27) together with the usual technology as discussed for example in [10] and an estimate for a Nikodym type light ray maximal function, followed by another rescaling argument to pass from the bilinear to the linear estimate. We do not present this here because the estimate we have at present for the maximal function is rather crude.

2. Let $p_d = 2 + \frac{2}{d-2}$. It is natural to ask the following question: is there an estimate

$$\|\hat{f}\|_{L^{p'}(L^p)} \lesssim \|f\|_{p'} \tag{29}$$

provided $p > p_d$. One could also weaken this by asking instead for the estimate ($p \geq p_d$)

$$\forall \epsilon \exists C_\epsilon : \|\hat{f}\|_{L^{p'}(L^p)} \leq C_\epsilon \lambda^\epsilon \|f\|_{p'} \tag{30}$$

if supp$f \subset D(0,\lambda)$, $\lambda \geq 1$.

This statement would easily imply the restriction conjecture for the sphere S^{d-2}. Namely, suppose that $f \in L^{p'}(\mathbb{R}^{d-1})$ with p' as above and that f is supported in $Q(N)$, and apply (29) to the function $f(\overline{x})e^{2\pi i x_d}\phi(\frac{x_d}{N})$ where ϕ is a suitable bump function. (if one assumes instead (30) then this argument still works using Tao's ϵ-removal lemma, see [14] for example.) Of course (29) would also solve the cone restriction problem, so it appears to be a natural common generalization.

The statements (29) or (30) are also related to several other conjectures in the literature. For example, (30) may be seen to be weaker than the "Radon transform" conjecture in [13], and is therefore also weaker than the so-called local smoothing conjecture [11]. We sketch the argument as follows: let Rf be the Radon transform of f restricted to the planes orthogonal to light rays as discussed in [13]; we will use the notation of that paper. Observe that the partial Fourier transform of Rf in the s variable can be identified with the restriction of $\hat{f}$ to the cone. Because of this, a rescaling argument followed by an application of the Hausdorff-Young theorem in the s variable shows, assuming [13], formula (33) (and that $p' \leq 2!$), that if supp$f \subset D(0,\lambda) \subset \mathbb{R}^d$ then

$$\|\hat{f}\|_{L^{p'}(L^p)} \lesssim \lambda^{\frac{d-1}{p}+\alpha}\|f\|_{p'}$$

Thus if [13], (33) were true for all $\alpha > -\frac{d-1}{p}$ as is conjectured in [13] then it would follow that (30) holds.

In the four dimensional case, estimate (29) is superficially similar to Corollary 1, the difference being that the radial dependence is now L^p instead of L^2, but since it would imply the restriction conjecture for S^2 it should not be accessible using only "soft" Kakeya information like our Lemma 1.1.

3. We indicate why the exponent $1+\frac{2}{d}$ in Theorem 1 is sharp. When $d=3$ this is pointed out in [14] and the idea is the same in general dimensions - namely, to use a surface which has two families of rulings, with each ruling being a light ray. It suffices to take the doubly ruled surface to be a plane Π.

Let ρ_i, $i=1,2$ be sectors of Γ of width $N^{-\frac{1}{2}}$ as above. Let ℓ_i be the center line of ρ_i and ν_i the normal to Γ along ℓ_i. Then ν_1 and ν_2 span a certain 2-plane Π through the origin. Let R be the intersection of the cube $Q(N)$ with the $\sqrt{N}$-neighborhood of Π. One can think of R as the union of rectangles dual to the $\frac{1}{N}$-neighborhood of ρ_i for either value $i=1$ or 2. For either i one can therefore construct in a standard way a function f_i supported in ρ_i with L^2 norm 1, such that $|\widehat{f_i d\sigma}|^2 \gtrsim N \cdot |R|^{-1}$ on, say, the dilation of R by a small factor ϵ_0. Here $|R| \approx N^{\frac{d+2}{2}}$. It follows that for any p one has

$$\int |\widehat{f_1 d\sigma}\widehat{f_2 d\sigma}|^p \gtrsim N^{-\frac{d}{2}p} N^{\frac{d+2}{2}}$$

This approaches ∞ as $N \to \infty$ if $p < 1+\frac{2}{d}$, which shows that (1) cannot hold when $p < 1+\frac{2}{d}$.

Appendix: Estimates for the restricted X-ray transform

The motivation for this appendix was to clarify the relationship between Lemma 1.1 and other approaches that have been taken to the restriction of the X-ray transform to the light rays - see for example [3], [6], [7], [8], [14] and [16]. This leads to a family of mixed norm estimates which we formulate as Theorem A.1 below.

Let $\mathcal{L}$ be the space of light rays with the integral defined by

$$\int_{\mathcal{L}} f(\ell)d\ell = \int_{S^{d-2}} \int_{Y(\omega)} f(\ell(y,\omega))dyd\omega$$

Here $\ell(y,\omega)$ is the line through y with direction $(\omega, 1)$, and $Y(\omega)$ is the hyperplane perpendicular to $\ell(0,\omega)$. We define mixed norms on G by

$$\|f\|_{L^q(L^r)} = (\int_{S^{d-2}} (\int_{Y(\omega)} |f(\ell(y,\omega))|^r dy)^{\frac{q}{r}} d\omega)^{\frac{1}{q}}$$

We define the X-ray transform as an operator from functions on $\mathbb{R}^d$ to functions on $\mathcal{L}$ via

$$Xf(\ell) = \int_{\ell} f$$

and will be interested in estimates for X from L^p to $L^q(L^r)$.

We first discuss necessary conditions in order to formulate a plausible conjecture; we omit details here. Suppose that X is bounded from L^p to $L^q(L^r)$. Then dilations give the condition

$$\frac{d}{p} - \frac{d-1}{r} = 1 \tag{31}$$

See e.g. [4] and [7]. Furthermore, the maps T_c used in section 5 give the condition

$$\frac{d-2}{q} \geq \frac{d}{p} - \frac{d}{r} \tag{32}$$

Again see [7]. Another condition can be obtained by considering the example $f = \chi_E$ where E is the δ-neighborhood of the cone segment Γ. This takes the form

$$\frac{1}{p} \leq \frac{d}{2r} \tag{33}$$

It is natural to expect that (31), (32), (33) are essentially also sufficient for boundedness. We will not consider endpoint questions and will therefore work locally. Index juggling leads to the following

Plausible conjecture: Let $p = q = \frac{d^2-2d+2}{d}$ and $r = \frac{d^2-2d+2}{2}$. Then X is bounded from the Sobolev space $W^{p,\epsilon}(Q(1))$ to $L^q(L^r)$ for any $\epsilon > 0$.

By $W^{p,\epsilon}(Q(1))$ we mean functions supported in $Q(1)$ with $\|f\|_{p,\epsilon} \stackrel{def}{=} \|(1-\Delta)^{\frac{\epsilon}{4}} f\|_p < \infty$.

There is an obvious bound on L^1, namely, by Fubini's theorem

$$\|Xf\|_{L^\infty(L^1)} = \|f\|_1 \tag{34}$$

Interpolating (34) with the preceding conjecture we obtain the following conjectural bound on L^p.

Plausible conjecture$_p$: Assume that $1 \leq p \leq \frac{d^2-2d+2}{d}$. Define r via $\frac{d}{p} - \frac{d-1}{r} = 1$ and q via $\frac{d-2}{q} = \frac{d}{p} - \frac{d}{r}$. Then X is bounded from $W^{p,\epsilon}(Q(1))$ to $L^q(L^r)$ for any $\epsilon > 0$.

This would imply all local $W^{p,\epsilon} \to L^q(L^r)$ estimates with the given p which are not ruled out by (31) (in the local form where $\leq$ replaces $=$), (32) and (33).

We will prove the following:

Theorem A.1 If $d = 3$ or $d = 4$ then the above conjectures are true. If $d \geq 5$ then the second conjecture is true on L^p provided $p \leq \frac{d+1}{2}$.

Remarks 1. We note that q and r coincide when $p = \frac{d}{2}$, $q = r = d-1$, and that this case is covered by our result. This is new except when $d = 3$ (see below); it is analogous to the result of Drury [5] (see also [12] and [4]) for the full X-ray transform.

2. Consider the case $d = 3$. In this case, the angular parameter ω runs over a one dimensional space and the restricted X-ray transform as defined here is a special case of the restricted X-ray transform associated to a "rigid line complex" [6], [7]. If $d = 3$ and $q = r$, then the estimate in Theorem A.1 is an estimate from $W^{\frac{3}{2},\epsilon}$ to L^2. The latter estimate is known, actually in the sharper form where $\epsilon = 0$ - cf. [16] (I thank Allan Greenleaf for this reference) and [6] - and a dual formulation of this same estimate is used in [14]. However, Theorem A.1 is new also in the three dimensional case if $p > \frac{3}{2}$.

3. It may be possible to obtain a scale invariant result (i.e. $\epsilon = 0$) by modifying the argument below, at least if one assumes strict inequality in (32) and (33) and ignores the three dimensional case, but we do not attempt that here because the formulation of Lemma 1.1 in the body of the paper is unsuitable for that purpose. We note though that our estimate on $W^{p\epsilon}$ can immediately be "upgraded" to a (local, of course) estimate on L^p provided one assumes strict inequality in (31), (32), (33). This is because one can

interpolate with the known fact that X is bounded from a negative order L^2 Sobolev space to L^2. We leave details to the reader.

4. A proof of the above conjectures for the full range of p in general dimensions has to be hard, since this would include a version of the Kakeya conjecture. Namely, if the first conjecture is true in $\mathbb{R}^d$, then a Kakeya set in $\mathbb{R}^{d-1}$ must have Minkowski dimension at least $d + \frac{4}{d} - 3$, as may be seen by applying the restricted X-ray bound to the indicator function of a cylinder over the δ-neighborhood of the Kakeya set. From this and known arguments (namely the subadditivity of the minimal possible Minkowski dimension for a Kakeya set in $\mathbb{R}^n$ as a function of n) follows that the first conjecture if true in all dimensions would imply that Kakeya sets have full Minkowski dimension.

We will need the following numerical inequalities (trivial in principle, but we give proofs for the reader's convenience). Here $\theta \in [\frac{1}{2}, 1]$ (we emphasize that $\theta \geq \frac{1}{2}$) and the variables x, y, a, b, a_j, b_k are nonnegative real numbers.

$$\min(ax, by)^\theta \max(ax, by)^{1-\theta} \leq \min(x, y)^\theta \max(x, y)^{1-\theta} \max(a, b)^\theta \min(a, b)^{1-\theta} \tag{35}$$

$$\min(\sum_j a_j, \sum_k b_k)^\theta \max(\sum_j a_j, \sum_k b_k)^{1-\theta} \leq \sum_{j,k} \min(a_j, b_k)^\theta \max(a_j, b_k)^{1-\theta} \tag{36}$$

Proofs For (35) we may assume that $x \leq y$. If also $ax \leq by$, then

$$\min(ax, by)^\theta \max(ax, by)^{1-\theta} = a^\theta b^{1-\theta} \min(x, y)^\theta \max(x, y)^{1-\theta}$$

and (35) follows. If $ax \geq by$ then

$$\begin{aligned} \min(ax, by)^\theta \max(ax, by)^{1-\theta} &= (\frac{by}{ax})^{2\theta-1} a^\theta x^\theta b^{1-\theta} y^{1-\theta} \\ &\leq a^\theta x^\theta b^{1-\theta} y^{1-\theta} \\ &= \min(x, y)^\theta \max(x, y)^{1-\theta} \max(a, b)^\theta \min(a, b)^{1-\theta} \end{aligned}$$

since $a \geq b$.

For (36) we can assume $\sum_j a_j \leq \sum_k b_k = 1$. In fact, we can assume in addition that $\sum_j a_j = 1$. This follows from (35): let $t = \sum_j a_j$ and consider the effect of replacing a_j by $t^{-1} a_j$. The left side of (36) increases by a factor of $t^{-\theta}$, and (35) implies the right side increases by at most this much.

The right side of (36) is smallest if $\theta = 1$ so we are reduced to proving that $\sum_j a_j = \sum_k b_k = 1$ implies $\sum_{j,k} \min(a_j, b_k) \geq 1$. But

$$\sum_j \sum_k \min(a_j, b_k) \geq \sum_j \min(a_j, \sum_k b_k) \geq \min(\sum_j a_j, \sum_k b_k)$$

so we are done. □

We start the proof of Theorem A.1 by giving a convenient restatement of Lemma 1.1; this differs from Lemma 1.1 only in that the Schwartz tails have been discarded and entropy replaced by measure, and is therefore an immediate corollary of Lemma 1.1.

Let $\mathcal{W}$ and $\mathcal{B}$ be δ-separated sets of white and black δ-tubes (thus they satisfy the transversality assumptions); assume each tube intersects the unit square. We let $\sim$ be the relation in Lemma 1.1 and will use the notation $w \sim x$ and $n_{\mathcal{W}}(Q)$ defined there. Let

$$\Phi_{\mathcal{W}}(x) = \sum_{w \in \mathcal{W}} \chi_w(x),\ \Phi_{\mathcal{B}}(x) = \sum_{b \in \mathcal{B}} \chi_b(x)$$

$$\Phi^*_{\mathcal{W}}(x) = \sum_{\substack{w \in \mathcal{W} \\ w \sim x}} \chi_w(x),\ \Phi^*_{\mathcal{B}}(x) = \sum_{\substack{b \in \mathcal{B} \\ b \sim x}} \chi_b(x)$$

$$\tilde{\Phi}_{\mathcal{W}} = \Phi_{\mathcal{W}} - \Phi^*_{\mathcal{W}},\ \tilde{\Phi}_{\mathcal{B}}(x) = \Phi_{\mathcal{B}} - \Phi^*_{\mathcal{B}}$$

Lemma A.1 The following hold, where C depends on d only; the implicit constants also depend on ϵ, and Q runs over a partition of $Q(1)$ into δ^{ϵ}-squares:

1. $\sum_Q n_{\mathcal{W}}(Q) \lesssim |\mathcal{W}|(\log \frac{1}{\delta})^5$.
2. $\sum_Q n_{\mathcal{B}}(Q) \lesssim |\mathcal{B}|(\log \frac{1}{\delta})^5$.
3. $|\{x \in Q(1) : \tilde{\Phi}_{\mathcal{W}}(x) \geq \mu \text{ and } \Phi_{\mathcal{B}}(x) \geq \nu\}| \lesssim \delta^{-C\epsilon} \frac{|\mathcal{W}||\mathcal{B}|}{\mu^2 \nu} \delta^d$.
4. $|\{x \in Q(1) : \tilde{\Phi}_{\mathcal{B}}(x) \geq \nu \text{ and } \Phi_{\mathcal{W}}(x) \geq \mu\}| \lesssim \delta^{-C\epsilon} \frac{|\mathcal{W}||\mathcal{B}|}{\mu \nu^2} \delta^d$. □

The rough idea now is to regard 3. and 4. of Lemma A.1 as a "virtual" $L^{\frac{3}{2}}$ to L^3 estimate and to interpolate between this and an L^1 to L^1 estimate, namely the following:

Lemma A.2 $|\{x \in Q(1) : \Phi_{\mathcal{W}}(x) \geq \mu \text{ and } \Phi_{\mathcal{B}}(x) \geq \nu\}| \lesssim \delta^{d-1} \min(\frac{|\mathcal{W}|}{\mu}, \frac{|\mathcal{B}|}{\nu})$.

Proof It is clear that $\| \sum_{w \in \mathcal{W}} \chi_w \|_{L^1(Q(1))} \lesssim |\mathcal{W}| \delta^{d-1}$, hence the measure of the μ-fold points is $\lesssim \frac{|\mathcal{W}|}{\mu} \delta^{\delta-1}$, which implies the lemma. □

Fix $\theta \in [\frac{1}{2}, 1]$ and define

$$\Psi_\theta = \min(\Phi_{\mathcal{B}}, \Phi_{\mathcal{W}})^\theta \max(\Phi_{\mathcal{B}}, \Phi_{\mathcal{W}})^{1-\theta}$$

$$S_\theta = \min(\Phi^*_{\mathcal{B}}, \Phi^*_{\mathcal{W}})^\theta \max(\Phi^*_{\mathcal{B}}, \Phi^*_{\mathcal{W}})^{1-\theta}$$

$$T_\theta = (\tilde{\Phi}_{\mathcal{B}})^\theta \Phi_{\mathcal{W}}^{1-\theta} + (\tilde{\Phi}_{\mathcal{W}})^\theta \Phi_{\mathcal{B}}^{1-\theta}$$

We will use below that

$$\Psi_\theta \lesssim S_\theta + T_\theta \tag{37}$$

This is a consequence of the numerical inequality

$$\min(a+b, c+d)^\theta \max(a+b, c+d)^{1-\theta} \lesssim a^\theta (c+d)^{1-\theta} + c^\theta (a+b)^{1-\theta} + \min(b,d)^\theta \max(b,d)^{1-\theta}$$

which follows for example from (36).

We now estimate T_θ for appropriate θ by interpolation between Lemmas A.1 and A.2.

Lemma A.3 Let p and q satisfy $1 \le q \le 3$ and $\frac{1}{q} \ge \frac{2}{p} - 1$. Let $\theta = \frac{1}{4}(3 - \frac{1}{q})$. Then

$$\|\delta^{d-2} T_\theta\|^q_{L^q(Q(1))} \lesssim \delta^{-C\epsilon} (\delta^{2d-3} |\mathcal{B}| \cdot \delta^{2d-3} |\mathcal{W}|)^{\frac{q}{2p}} \tag{38}$$

Proof It suffices to consider the case where $\frac{1}{q} = \frac{2}{p} - 1$ since the δ-separation implies that the quantity $(\delta^{2d-3}|\mathcal{B}| \cdot \delta^{2d-3}|\mathcal{W}|)$ is $\lesssim 1$.

Define $Y(\mu, \nu)$ to be the set where $\tilde{\Phi}_{\mathcal{W}} \ge \mu$ and $\Phi_{\mathcal{B}} \ge \nu$. Lemmas A.1 and A.2 give

$$|Y(\mu,\nu)| \lesssim \delta^{-C\epsilon} \min\left(\frac{|\mathcal{W}|\,|\mathcal{B}|}{\mu^2 \nu} \delta^d, \frac{|\mathcal{W}|}{\mu} \delta^{d-1}, \frac{|\mathcal{B}|}{\nu} \delta^{d-1}\right)$$

and therefore also

$$\begin{aligned} |Y(\mu,\nu)| &\lesssim \delta^{-C\epsilon} \left(\frac{|\mathcal{W}|\,|\mathcal{B}|}{\mu^2 \nu} \delta^d\right)^{\frac{q}{p}-1} \left(\frac{|\mathcal{W}|}{\mu} \delta^{d-1}\right)^{1-\frac{q}{2p}} \left(\frac{|\mathcal{B}|}{\nu} \delta^{d-1}\right)^{1-\frac{q}{2p}} \\ &= \delta^{-C\epsilon} \frac{(|\mathcal{B}|\,|\mathcal{W}|)^{\frac{q}{2p}} \delta^{d-2+\frac{q}{p}}}{\mu^{\theta q} \nu^{(1-\theta) q}} \end{aligned}$$

where we used the value of θ to obtain the last line. Summing over dyadic levels for μ and ν between 1 and a negative power of δ gives

$$\|\Phi_{\mathcal{B}}^{1-\theta}(\tilde{\Phi}_{\mathcal{W}})^{\theta}\|_q^q \lesssim \delta^{-C\epsilon}(|\mathcal{B}|\,|\mathcal{W}|)^{\frac{q}{2p}}\delta^{d-2+\frac{q}{p}}$$

This and the analogous estimate with the roles of $\mathcal{B}$ and $\mathcal{W}$ reversed imply

$$\|T_\theta\|_q^q \lesssim \delta^{-C\epsilon}(|\mathcal{B}|\,|\mathcal{W}|)^{\frac{q}{2p}}\delta^{d-2+\frac{q}{p}}$$

which is equivalent to (38) when $\frac{1}{q} = \frac{2}{p} - 1$. □

We will now pass to a similar estimate for Ψ_θ. We will use a rescaling argument and induction on δ like the final argument in [17] or [18]. The rescaling argument requires another relation between the exponents, which is essentially the dual relation to (31). We remark at this point that the quantity which we need to estimate in order to prove Theorem A.1 is $\min(\Phi_{\mathcal{B}}, \Phi_{\mathcal{W}})$ and not the slightly larger Ψ_θ. It is possible that the slightly stronger result obtained by considering Ψ_θ could prove useful, but the main reason we use Ψ_θ is that the rescaling argument in the proof is difficult to carry out with $\min(\Phi_{\mathcal{B}}, \Phi_{\mathcal{W}})$.

Lemma A.4 Assume that $q \leq 3$, $\frac{1}{q} \geq \frac{2}{p} - 1$, and $1 \leq \frac{q}{p} \leq \frac{d}{d-1}$. Then for any $\epsilon > 0$ there is a constant A_ϵ making the following estimate valid; here $\theta = \frac{1}{4}(3 - \frac{1}{q})$:

$$\|\delta^{d-2}\Psi_\theta\|_{L^q(Q(1))}^q \leq A_\epsilon \delta^{-C\epsilon}(\delta^{2d-3}|\mathcal{B}| \cdot \delta^{2d-3}|\mathcal{W}|)^{\frac{q}{2p}} \qquad (39)$$

Proof We start with the following observation concerning rescaling.

Claim Suppose that δ is small enough and that (39) has been proved with δ replaced by $\delta^{1-\epsilon}$. Let Q be a δ^ϵ-cube, and let $\mathcal{B}$ and $\mathcal{W}$ be δ-separated sets of tubes. Then

$$\|\Psi_\theta\|_{L^q(Q)}^q \leq \delta^{\frac{C\epsilon^2}{2}} \cdot A_\epsilon \delta^{-C\epsilon}(\delta^{2d-3}|\mathcal{W}| \cdot \delta^{2d-3}|\mathcal{B}|)^{\frac{q}{2p}}$$

Namely, for each $w \in \mathcal{W}$ let $k(w)$ be the cardinality of the set of tubes $w_1 \in \mathcal{W}$ such that $w_1 \cap Q$ is contained in the double of w; similarly for each $b \in \mathcal{B}$ let $k(b)$ be the cardinality of the set of tubes $b_1 \in \mathcal{B}$ such that $b_1 \cap Q$ is contained in the double of b. Notice that $k(w)$ and $k(b)$ are between 1 and $\delta^{-(d-2)\epsilon}$. Let $\mathcal{W}(\mu) = \{w \in \mathcal{W} : k(w) \in [\mu, 2\mu]\}$, $\mathcal{B}(\nu) = \{b \in \mathcal{B} : k(b) \in [\nu, 2\nu]\}$, and (analogously to the earlier definitions) let

$$\Phi^\mu_{\mathcal{W}} = \sum_{w \in W_\mu} \chi_w,\ \Phi^\nu_{\mathcal{B}} = \sum_{b \in B_\nu} \chi_b$$

$$\Psi^{\mu\nu}_\theta = \min(\Phi^\nu_{\mathcal{B}}, \Phi^\mu_{\mathcal{W}})^\theta \max(\Phi^\nu_{\mathcal{B}}, \Phi^\mu_{\mathcal{W}})^{1-\theta}$$

Then

$$\Psi_\theta \leq \sum_{\mu,\nu} \Psi^{\mu\nu}_\theta \tag{40}$$

where the sum is over dyadic values of μ and ν. This follows from (36).

By (40) and pigeonholing, there are values of μ and ν such that

$$\|\Psi^{\mu\nu}_\theta\|_{L^q(Q)} \gtrsim (\log \frac{1}{\delta})^{-2} \|\Psi_\theta\|_{L^q(Q)}$$

We assume without loss of generality that $\mu \geq \nu$. Now let $\overline{\mathcal{B}}$ (resp. $\overline{\mathcal{W}}$) be subsets of $\mathcal{B}(\nu)$ (resp. $\mathcal{W}(\mu)$) which are maximal with respect to the following property:

($*$): If $b_1, b_2 \in \overline{\mathcal{B}}$ (resp $\overline{\mathcal{W}}$), then $b_1 \cap Q$ is not contained in the double of b_2.

Let $\overline{\Phi}_{\mathcal{B}}$ (resp. $\overline{\Phi}_{\mathcal{W}}$) be the sums of the characteristic functions of the tubes of width $C_0\delta$ coaxial with the tubes in $\overline{\mathcal{B}}$ (resp. $\overline{\mathcal{W}}$), and

$$\overline{\Psi}_\theta = \min(\overline{\Phi}_{\mathcal{B}}, \overline{\Phi}_{\mathcal{W}})^\theta \max(\overline{\Phi}_{\mathcal{B}}, \overline{\Phi}_{\mathcal{W}})^{1-\theta}$$

Then $\Phi^\mu_{\mathcal{W}} \lesssim \mu \overline{\Phi}_{\mathcal{W}}$ and $\Phi^\nu_{\mathcal{B}} \lesssim \nu \overline{\Phi}_{\mathcal{B}}$, pointwise on Q; this follows from maximality of $\overline{\mathcal{W}}$ and $\overline{\mathcal{B}}$ provided C_0 is large enough. Hence also

$$\Psi^{\mu\nu}_\theta \lesssim \mu^\theta \nu^{1-\theta} \overline{\Psi}_\theta$$

by (35). Taking L^q norms we conclude that

$$\|\Psi^{\mu\nu}_\theta\|^q_{L^q(Q)} \lesssim \mu^{\theta q} \nu^{(1-\theta)q} \|\overline{\Psi}_\theta\|^q_{L^q(Q)} \tag{41}$$

Furthermore property ($*$) implies

$$|\overline{\mathcal{B}}| \lesssim \nu^{-1} |\mathcal{B}(\nu)|,\ |\overline{\mathcal{W}}| \lesssim \mu^{-1} |\mathcal{W}(\mu)| \tag{42}$$

We now dilate the situation by a factor $\delta^{-\epsilon}$. This maps Q to a cube Q' of side 1, and maps $\overline{\mathcal{B}}$ and $\overline{\mathcal{W}}$ to $\delta^{1-\epsilon}$-separated families of $C_0\delta^{1-\epsilon}$-tubes. Accordingly we can apply the hypothesis that (39) holds at scale $\delta^{1-\epsilon}$. We conclude that

$$\|\delta^{(1-\epsilon)(d-2)}\overline{\Psi}_\theta(\delta^\epsilon x)\|^q_{L^q(Q',dx)} \lesssim A_\epsilon \delta^{-C\epsilon(1-\epsilon)}(\delta^{(2d-3)(1-\epsilon)}|\overline{\mathcal{W}}| \cdot \delta^{(2d-3)(1-\epsilon)}|\overline{\mathcal{B}}|)^{\frac{q}{2p}}$$

Making the change of variables $x \to \delta^\epsilon x$ and factoring out the powers of δ^ϵ we get

$$\delta^{-d\epsilon}\delta^{-q(d-2)\epsilon}\|\delta^{d-2}\overline{\Psi}_\theta\|^q_{L^q(Q)} \lesssim A_\epsilon\delta^{-C\epsilon(1-\epsilon)}\delta^{-\frac{q}{p}(2d-3)\epsilon}(\delta^{2d-3}|\overline{\mathcal{W}}| \cdot \delta^{2d-3}|\overline{\mathcal{B}}|)^{\frac{q}{2p}}$$

We now substitute in the estimates (41) and (42), obtaining

$$\begin{aligned}\mu^{-\theta q}\nu^{-(1-\theta)q}\delta^{-d\epsilon}\delta^{-q(d-2)\epsilon}\|\delta^{d-2}\Psi^{\mu\nu}_\theta\|^q_{L^q(Q)} \lesssim\ & (\mu\nu)^{-\frac{q}{2p}} \cdot \delta^{-\frac{q}{p}(2d-3)\epsilon} \\ & \cdot A_\epsilon\delta^{-C\epsilon(1-\epsilon)}(\delta^{2d-3}|\mathcal{W}_\mu| \cdot \delta^{2d-3}|\mathcal{B}_\nu|)^{\frac{q}{2p}}\end{aligned}$$

or equivalently

$$\begin{aligned}\|\delta^{d-2}\Psi^{\mu\nu}_\theta\|^q_{L^q(Q)} \lesssim\ & \mu^{q(\theta-\frac{1}{2p})}\nu^{q(1-\theta-\frac{1}{2p})} \cdot \delta^{-\frac{q}{p}(2d-3)\epsilon+d\epsilon+q(d-2)\epsilon} \\ & \cdot A_\epsilon\delta^{-C\epsilon(1-\epsilon)}(\delta^{2d-3}|\mathcal{W}_\mu| \cdot \delta^{2d-3}|\mathcal{B}_\nu|)^{\frac{q}{2p}}\end{aligned}$$

But $\nu \leq \mu \lesssim \delta^{-(d-2)\epsilon}$, and the exponents $q(\theta-\frac{1}{2p})$ and $q(1-\theta-\frac{1}{2p})$ are both nonnegative. Since $\frac{q}{p} \leq \frac{d}{d-1}$, a little juggling of indices shows that therefore

$$\mu^{q(\theta-\frac{1}{2p})}\nu^{q(1-\theta-\frac{1}{2p})}\delta^{-\frac{q}{p}(2d-3)\epsilon+d\epsilon+q(d-2)\epsilon} \lesssim 1$$

It follows that

$$\|\delta^{d-2}\Psi^{\mu\nu}_\theta\|^q_{L^q(Q)} \lesssim \delta^{C\epsilon^2} \cdot A_\epsilon\delta^{-C\epsilon}(\delta^{2d-3}|\mathcal{W}_\mu| \cdot \delta^{2d-3}|\mathcal{B}_\nu|)^{\frac{q}{2p}}$$

and therefore

$$\|\delta^{d-2}\Psi_\theta\|^q_{L^q(Q)} \lesssim (\log\frac{1}{\delta})^{2q}\delta^{\frac{C\epsilon^2}{2}} \cdot \delta^{\frac{C\epsilon^2}{2}}A_\epsilon\delta^{-C\epsilon}(\delta^{2d-3}|\mathcal{W}| \cdot \delta^{2d-3}|\mathcal{B}|)^{\frac{q}{2p}}$$

The factor $(\log\frac{1}{\delta})^{2q}\delta^{\frac{C\epsilon^2}{2}}$ is evidently small for small δ, so the proof of the claim is complete.

We assume now that (39) has been proved for parameter values $\delta > \delta_0$ for a certain δ_0 (the case where δ is large is easy if A_ϵ has been chosen appropriately) and will prove it when $\delta^{1-\epsilon} > \delta_0$. This will evidently establish the lemma.

We use (37), and observe that a bound like (39) with Ψ_θ replaced by T_θ follows from Lemma A.3; the implicit constant in Lemma A.3 is small compared with A_ϵ if A_ϵ has been chosen appropriately. To estimate S_θ, subdivide $Q(1)$ in δ^ϵ-cubes Q. On each fixed Q we can apply the claim to the restriction of S_θ to Q, replacing $\mathcal{W}$ by $\{w \in \mathcal{W} : w \sim Q\}$ and similarly with $\mathcal{B}$.

We obtain for each Q

$$\|\delta^{d-2} S_\theta\|^q_{L^q(Q)} \leq \delta^{\frac{C\epsilon^2}{2}} \cdot A_\epsilon(\delta^{2d-3} n_{\mathcal{W}}(Q) \cdot \delta^{2d-3} n_{\mathcal{B}}(Q))^{\frac{q}{2p}} \tag{43}$$

We now sum over Q concluding that

$$\begin{aligned}
\|\delta^{d-2} S_\theta\|^q_{L^q(Q(1))} &\leq \delta^{\frac{C\epsilon^2}{2}} \cdot A_\epsilon \sum_Q (\delta^{2d-3} n_{\mathcal{W}}(Q) \cdot \delta^{2d-3} n_{\mathcal{B}}(Q))^{\frac{q}{2p}} \\
&\leq \delta^{\frac{C\epsilon^2}{2}} \cdot A_\epsilon (\sum_Q \delta^{2d-3} n_{\mathcal{W}}(Q))^{\frac{q}{2p}} (\sum_Q \delta^{2d-3} n_{\mathcal{B}}(Q))^{\frac{q}{2p}} \\
&\lesssim \delta^{\frac{C\epsilon^2}{2}} \cdot A_\epsilon (C \log \frac{1}{\delta})^{5\frac{q}{p}} (\delta^{2d-3}|\mathcal{W}|)^{\frac{q}{2p}} (\delta^{2d-3}|\mathcal{B}|)^{\frac{q}{2p}}
\end{aligned}$$

The three inequalities followed respectively from (43), from Holder's inequality (recall that $q \geq p$) and from 1. and 2. of Lemma A.1. The factor $\delta^{\frac{C\epsilon^2}{2}} \cdot (C \log \frac{1}{\delta})^{5\frac{q}{p}}$ is small for small δ; the result now follows by combining the last inequality with the preceding bound for $\|T_\theta\|^q_{L^q(Q)}$. □

Lemma A.4 is our main estimate and the rest of the argument is basically just another rescaling argument. This is fairly routine, so we will omit some details. In order to carry out the argument efficiently we first make some further definitions and remarks.

We define a map X^* from functions on $\mathcal{L}$ to functions of $\mathbb{R}^d$ via

$$X^* f(x) = \int_{S^{d-2}} f(\ell(x,\omega)) d\omega$$

This is easily seen to be the adjoint map to X. If c is a spherical cap on S^{d-2}, then define $\mathcal{L}_c$ to be the set of light rays $\ell \in \mathcal{L}$ whose direction is $(\omega, 1)$ for some $\omega \in c$. For given p and r and δ, and a set $Y \subset \mathcal{L}$, define

$$\mathcal{E}^{p,r}_\delta(Y) = \|\chi_{Y_\delta}\|_{L^p(L^r)}$$

where Y_δ is the δ-neighborhood of Y (with respect to a smooth metric on $\mathcal{L}$).

Next, fix a cap c centered at a point $e \in S^{d-2}$ with radius σ. The map T_c in section 5 takes light rays to light rays, so there is an action $T_c : \mathcal{L} \to \mathcal{L}$, which has the following metric properties:

(a): If $Y \subset \mathcal{L}_c$ then $\|\chi_{T_c Y}\|_{L^p(L^r)} \approx \sigma^{-\frac{d-2}{p}-\frac{d}{r}} \|\chi_Y\|_{p,r}$
(b): If $Y \subset \mathcal{L}_c$ then $\mathcal{E}^{p,r}_{\delta}(T_c Y) \lesssim \sigma^{-\frac{d-2}{p}-\frac{d}{r}} \mathcal{E}^{p,r}_{\sigma\delta}(Y)$

(a) is proved as follows: within c, T_c expands distances along S^{d-2} by a factor σ^{-1} (hence volumes by $\sigma^{-(d-2)}$), and if $\omega \in c$ then the action on the fiber $\{x \in \mathbb{R}^d : x \perp (\omega, 1)\}$ expands volumes by roughly $\overbrace{\sigma^{-1} \times \ldots \times \sigma^{-1}}^{d-2\,times} \times \sigma^{-2}$, i.e. by σ^{-d}. Thus $L^p(L^r)$ norms expand by $\sigma^{-\frac{d-2}{p}-\frac{d}{r}}$. Also (b) follows from (a) by observing that T_c maps the $C\sigma\delta$-neighborhood of $Y \subset \mathcal{L}_c$ onto a set which includes the δ-neighborhood of $T_c Y$.

Further, if $Y \subset \mathcal{L}_c$ then

$$X^* \chi_Y(x) \approx \sigma^{d-2} X^* \chi_{T_c Y}(T_c x) \tag{44}$$

This follows from the definition of X^* and the formula for volume expansion along S^{d-2}.

We will now rephrase Lemma A.4 using some of the preceding notation and at the same time will replace it by a somewhat weaker result with a less cumbersome statement.

Lemma A.5 Let $Z \subset \mathcal{L}$, let C be a large constant and let S be a set of points in $\mathbb{R}^d$ with the following properties:

1. The intersection of S with the δ-neighborhood of any given ray $\ell \in Z$ is contained in a cube of side 1.
2. If $x \in S$, then there are two spherical caps c_1 and c_2 on S^{d-2} with width C^{-1} and whose distance apart is at least C^{-1}, such that $\min(X^*(\chi_{\mathcal{L}_{c_1} \cap Z})(x), X^*(\chi_{\mathcal{L}_{c_2} \cap Z})(x)) \geq \mu$.

Then

$$|S| \lesssim \delta^{-\epsilon} \mu^{-q} \mathcal{E}^{p,r}_{\delta}(Z)^q$$

for any fixed $\epsilon > 0$, provided $q \leq 3$, $q \geq p \geq r$, and $\frac{1}{q} \geq \frac{2}{r} - 1$, $\frac{q}{r} \leq \frac{d}{d-1}$.

Proof We first make a couple of reductions. First, it suffices to prove the lemma with the assumption 1. replaced by the stronger assumption that $A \subset Q(1)$. This follows in a standard way using that $q \geq p \geq r$: if the result is proved for A contained in a square of side 1, then one can tile by such squares, take an L^q sum over the squares and use the hypothesis 1. It then also suffices to prove Lemma A.5 when $p = r$, since $\mathcal{E}_\delta^{p,r}(Z)$ increases with p when Z is contained in a fixed compact subset. In addition, it suffices by a simple covering argument to prove the lemma assuming that the caps c_1 and c_2 in 2. are independent of x.

Now define $Z_i = Z \cap \mathcal{L}_{c_i}$, let $\mathcal{W}$ and $\mathcal{B}$ be maximal δ-separated subsets of Z_1 and Z_2 respectively and (for each $w \in \mathcal{W}$) let D_w the δ-disc in $\mathcal{L}$ centered at w. Then $X^*(\chi_{Z_1 \cap D(w)}) \lesssim \delta^{d-2}\chi_w$, where on the right side χ_w is the characteristic function of the δ-neighborhood of the line w. So $X^*(\chi_{Z_1}) \lesssim \sum_w \delta^{d-2}\chi_w = \Phi_{\mathcal{W}}$, where $\Phi_{\mathcal{W}}$ is as in Lemma A.1. Accordingly $\min(X^*(\chi_{\mathcal{L}_{c_1} \cap Z})(x), X^*(\chi_{\mathcal{L}_{c_2} \cap Z})(x)) \lesssim \min(\Phi_{\mathcal{B}}, \Phi_{\mathcal{W}}) \leq \Psi_\theta$. The result now follows from Lemma A.4 using Tchebyshev's inequality and that $(\delta^{2d-3}|\mathcal{B}|\delta^{2d-3}|\mathcal{W}|)^{\frac{1}{2r}} \lesssim \mathcal{E}_\delta^{r,r}(Z)$. □

The point will now be that for appropriate values of the exponents the statement of Lemma A.5 is essentially invariant under the rescaling maps T_c.

Lemma A.6 Assume that $q \leq 3$, $q \geq p \geq r$, $\frac{1}{q} \geq \frac{2}{r} - 1$, $\frac{q}{r} \leq \frac{d}{d-1}$, and

$$\frac{d-2}{p} + \frac{d}{r} \leq d - 2 + \frac{d}{q} \tag{45}$$

Let $Y \subset \mathcal{L}$. Then

$$\|X^*\chi_Y\|_{L^q(Q(1))} \lesssim \delta^{-\epsilon}\mathcal{E}_\delta^{p,r}(Y)$$

Proof A standard argument shows that it will suffice to prove the corresponding distributional estimate

$$|\{x \in Q(1) : X^*\chi_Y(x) \geq \lambda\}| \lesssim \delta^{-\epsilon}\lambda^{-q}\mathcal{E}_\delta^{p,r}(Y)^q \tag{46}$$

in the case where λ is bounded below by a high power of δ, say

$$\lambda \geq \delta^{\frac{B}{2}(d-2)}$$

where B is a large constant depending on d. This is because of the $\delta^{-\epsilon}$ factors and the fact that very small values of λ clearly make a negligible contribution.

To prove (46), let $A = \{x \in Q(1) : X^*\chi_Y(x) \geq \lambda\}$ and define A_σ to be all points x with the property that there are two σ-caps c_1 and c_2 on S^{d-2} whose distance apart is between σ and $C\sigma$ and such that $X^*(\chi_{Y\cap\mathcal{L}_{c_i}}) \geq C^{-1}\delta^{\epsilon}\lambda$ for $i = 1, 2$. We claim that $\cup_\sigma A_\sigma \supset A$; the union here is over dyadic $\sigma \geq \delta^B$.

Namely, if $x \in A$, then take the smallest σ such that $X^*(\chi_{Y\cap\mathcal{L}_c})(x) \geq (C\sigma)^{\frac{\epsilon}{B}}\lambda$ for some cap c of width $C\sigma$. (The lower bound on λ implies that then $\sigma \geq \delta^B$) Consider a covering of σ by caps c_i of width σ. The minimality of σ implies that $X^*(\chi_{Y\cap\mathcal{L}_{c_i}})(x)$ is small compared with $X^*(\chi_{Y\cap\mathcal{L}_c})(x)$ for each fixed i. It follows that the contribution from a fixed finite number of the c_i's is similarly small, and therefore there must be two c_i's, call them c_1 and c_2, which are at distance $\geq \sigma$ apart such that $X^*(\chi_{Y\cap\mathcal{L}_{c_i}})(x) \gtrsim \sigma^{\frac{\epsilon}{B}}\lambda$ for $i = 1$ and 2. This implies the claim.

By pigeonholing we may now choose σ so that

$$|A_\sigma| \geq \delta^{\epsilon}|A| \tag{47}$$

Cover S^{d-2} with a family of $C\sigma$-caps c_i with bounded overlap. This gives a further decompostion

$$A_\sigma = \cup_i A_\sigma^{c_i}$$

where $A_\sigma^{c_i}$ is the set of x for which the two σ-caps c_1 and c_2 in the definition of A_σ may be taken to be contained in c_i.

We now fix one of the c_i's and apply Lemma A.5 to the sets $Z = T_{c_i}(\mathcal{L}_{c_i} \cap Y)$ and $S = T_{c_i}(A_\sigma^{c_i})$. Formula (44) shows that the hypothesis 2. is satisfied with $\mu \approx \sigma^{-(d-2)}\lambda$, and since $A \subset Q(1)$ and T_c preserves lengths in the $(e, 1)$ direction, one can easily see that hypothesis 1. is also satisfied. It follows that

$$\begin{aligned} |T_{c_i}(A_\sigma^{c_i})| &\lesssim \delta^{-\epsilon}(\sigma^{-(d-2)}\lambda)^{-q}\mathcal{E}_\delta^{p,r}(T_{c_i}(\mathcal{L}_{c_i} \cap Y)) \\ &\lesssim (\sigma^{-(d-2)}\lambda)^{-q}\sigma^{-q(\frac{d-2}{p}+\frac{d}{r})}\mathcal{E}_{\sigma\delta}^{p,r}(\mathcal{L}_{c_i} \cap Y)^q \end{aligned}$$

by property (b) above. Thus, using also (25)

$$|A_\sigma| \lesssim \delta^{-\epsilon}\sigma^d(\sigma^{-(d-2)}\lambda)^{-q}\sigma^{-q(\frac{d-2}{p}+\frac{d}{r})}\mathcal{E}_{\sigma\delta}^{p,r}(\mathcal{L}_{c_i} \cap Y)^q$$

which implies that

$$|A_\sigma| \lesssim \delta^{-\epsilon}\mathcal{E}_{\sigma\delta}^{p,r}(\mathcal{L}_{c_i} \cap Y)^q$$

by the assumption (45).

Now observe that the $\sigma\delta$-neighborhoods of the sets $\mathcal{L}_{c_i} \cap Y$ are essentially disjoint (no point $y \in \mathcal{L}$ belongs to more than a bounded number). Accordingly we can sum over c_i to obtain

$$|A_\sigma| \lesssim \delta^{-\epsilon}\mathcal{E}_{\sigma\delta}^{p,r}(Y)^q$$

We now use (47) and the fact that $\mathcal{E}_\epsilon^{p,r}$ increases with ϵ. The result follows. $\square$

Proof of Theorem A.1 Let p, q, r be as in Theorem A.1. Because the statement is obtained by interpolation with (34) we can assume that p has its largest possible value, namely $\frac{5}{3}$ if $d = 3$ and $\frac{d+1}{2}$ if $d \geq 4$. The following relations on the dual exponents will hold:

$$\frac{p'}{r'} \leq \frac{d}{d-1}$$

$$\frac{d-2}{q'} + \frac{d}{r'} \leq d - 2 + \frac{d}{p'}$$

$$\frac{1}{p'} \geq \frac{2}{r'} - 1$$

$$r' \leq q' \leq p' \leq 3$$

Namely the first two are dual to (31) and (32) respectively. The third follows since $p \leq \frac{d+1}{2}$ and r is defined by (31), while the last is most easily checked by using the explicit values of p, q and r. Thus Lemma A.6 is applicable and shows that

$$\|X^* \chi_Y\|_{L^{p'}(Q(1))} \lesssim \delta^{-\epsilon} \mathcal{E}_\delta^{q',r'}(Y) \tag{48}$$

We now pass to the dual estimate. If f is supported in $Q(1)$ then we define $X_\delta f(\ell) = \delta^{-(d-1)} \int_{\ell^\delta} f$, where ℓ^δ is the tube of width δ with axis ℓ.

Fix a nonnegative function f supported in $Q(1)$ with $\|f\|_p = 1$ and consider the quantity $\|X_\delta f\|_{L^q(L^r)}$. By duality there is a function $g : \mathcal{L} \to \mathbb{R}$ such that $\|g\|_{L^{q'}(L^{r'})} = 1$ and

$$\int_{\mathcal{L}} g X_\delta f \gtrsim \|X_\delta f\|_{L^q(L^r)}$$

Since $X_\delta f$ is roughly constant on δ-discs and since values of $X_\delta f$ which are less than a high power of δ make a negligible contribution to the norm, we can then conclude that there is a function $g : \mathcal{L} \to \mathbb{R}$ with $\|g\|_{L^{q'}(L^{r'})} = 1$, with

$$\int_{\mathcal{L}} g X_\delta \chi_E \gtrsim \delta^\epsilon \|X_\delta f\|_{L^q(L^r)}$$

and such that g has the special form

$$g = \mu \chi_Y \tag{49}$$

where μ is a scalar, and the set Y is a union of δ-discs. Note that this implies $\mu \mathcal{E}_{\delta}^{q',r'}(Y) \lesssim 1$. We also let $\tilde{Y}$ be the corresponding union of 2δ-discs.

Letting g be as in (49), we have

$$\begin{aligned} \|X_\delta f\|_{L^q(L^r)} &\lesssim \delta^{-\epsilon} \int_{\mathcal{L}} \mu \chi_Y X_\delta f \\ &\lesssim \delta^{-\epsilon} \int_{\mathcal{L}} \mu \chi_{\tilde{Y}} X f \\ &= \delta^{-\epsilon} \int_{\mathbb{R}^d} X^*(\mu \chi_{\tilde{Y}}) f \end{aligned}$$

Now apply (48) to $\chi_{\tilde{Y}}$ and use Holder's inequality, obtaining

$$\|X_\delta f\|_{L^q(L^r)} \lesssim \delta^{-\epsilon} \|f\|_p \tag{50}$$

since $\mu \mathcal{E}_{\delta}^{q',r'}(\tilde{Y}) \lesssim 1$.

It remains to trade ϵ derivatives for the $\delta^{-\epsilon}$ factor, which is done in the usual way. Suppose that f has $W^{p\epsilon}$-norm 1 and has support in $Q(1)$. If ϕ is an appropriately chosen C_0^∞ function and $\phi_j(x) = 2^{dj}\phi(2^j x)$ then we can express $f = g + \sum_j \phi_j * f_j$, where $\hat{g}$ has compact support, and where $\sum_j 2^{\eta j}\|f_j\|_p \lesssim \|f\|_{p\epsilon}$ for small η. It follows using the smoothing effect of ϕ_j that

$$Xf \lesssim 1 + \sum_j X_{2^{-j}} |f_j|$$

and now the theorem follows by applying (50) with a small enough value of ϵ to the terms in the series. ☐

References

[1] B. Barcelo, *On the restriction of the Fourier transform to a conical surface*, Trans. Amer. Math. Soc. 292 (1985), 321-333.

[2] J. Bourgain, *Besicovitch type maximal operators and applications to Fourier analysis*, Geometric and Functional Analysis 1 (1991), 147-187.

[3] J. Bourgain, *Estimates for cone multipliers*, pp. 41–60 in Geometric aspects of functional analysis (Israel, 1992–1994), Operator Theory: Advances and Applications, vol. 77, Birkhauser 1995.

[4] M. Christ, *Estimates for the k- plane transform*, Indiana Univ. Math. J. 33(1984), 891-910.

[5] S. Drury, *L^p estimates for the x-ray transform*, Ill. J. Math. 27(1983), 125-129.

[6] A. Greenleaf, A. Seeger, *Fourier integral operators with fold singularities*, J. Reine Angew. Math. 455 (1994), 35–56.

[7] A. Greenleaf, A. Seeger, S. Wainger, *On estimates for generalized X-ray transforms and integrals over curves in* $\mathbb{R}^4$, to appear in Proc. Amer. Math. Soc.

[8] A. Greenleaf, G. Uhlmann, *Composition of some singular Fourier integral operators and estimates for restricted X-ray transforms*, Ann. Inst. Fourier (Grenoble) 40 (1990), 443–466.

[9] S. Klainerman, M. Machedon, *Space-time estimates for null forms and the local existence theorem*, Comm. Pure Appl. Math. 46 (1993), 1221–1268.

[10] G. Mockenhaupt, *A note on the cone multiplier*, Proc. Amer. Math. Soc. 117(1993), 145-152.

[11] G. Mockenhaupt, A. Seeger, C. Sogge, *Wave front sets and Bourgain's circular maximal theorem*, Ann. Math. 134 (1992), 207-218.

[12] D. M. Oberlin, E. M. Stein, *Mapping properties of the Radon transform*, Indiana Univ. Math. J. 31(1982), 641-650.

[13] T. Tao, *The weak type endpoint Bochner-Riesz conjecture and related topics*, Indiana Univ. Math. J. 47 (1998), 1097-1124.

[14] T. Tao, A. Vargas, *A bilinear approach to cone multipliers*, I and II, preprints 1998, to appear in Geometric and Functional Analysis.

[15] T. Tao, A. Vargas, L. Vega, *A bilinear approach to the restriction and Kakeya conjectures*, J. Amer. Math. Soc. 11(1998), 967-1000.

[16] H. T. Wang, *L^p estimates for the X-ray transform restricted to line complexes of Kirillov type*, Trans. Amer. Math. Soc. 332 (1992), 793–821.

[17] T. Wolff, *An improved bound for Kakeya type maximal functions*, Revista Math. Iberoamericana 11(1995), 651-674.

[18] T. Wolff, *A mixed norm estimate for the X-ray transform*, Revista Math. Iberoamericana 14(1998), 561-601.

[19] T. Wolff, *Decay of circular means of Fourier transforms of measures*, preprint 1999, to appear in International Mathematics Research Notices.