Annals of Mathematics, 153 (2001), 661–698

A sharp bilinear cone restriction estimate

BY THOMAS WOLFF†

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 $(\tau, x_d)$ with $\tau \in \mathbb{R}^{d-1}$, and let $\Gamma = \{ x : x_d = |\tau| \text{ and } 1 \leq x_d \leq 2 \}$. Let $\Gamma_1$ and $\Gamma_2$ be disjoint conical subsets, i.e.

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

where $\Omega_i$ are disjoint closed subsets of the sphere $S^{d-2}$. Let $f$ and $g$ be two functions on $\Gamma$ whose supports are contained in $\Gamma_1$ and $\Gamma_2$ respectively. We will prove the following estimate, where $\sigma$ is surface measure on $\Gamma$, and $\hat{f} d\sigma$ is the $\mathbb{R}^d$ Fourier transform:

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

$$\| \hat{f} d\sigma \hat{g} d\sigma \|_p \leq C_{p, \Gamma_1, \Gamma_2} \| f \|_2 \| g \|_2.$$  

Bilinear estimates of this general type have been used by several authors; see in particular [11]. The estimate (1) was formulated by Bourgain in [3], and it was proved in [3] when $d = 3$ and $p > 2 - \varepsilon$ for some $\varepsilon > 0$, the case $p = 2$ being easier and implicit in [1]. Tao and Vargas [16] recently obtained the explicit range $p > 2 - \frac{8}{121}$ when $d = 3$, and noted that one can also obtain a range $p > 2 - \varepsilon_0$ 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 [16], where the conjecture that (1) should hold for $d = 3$ and $p \geq \frac{5}{4}$ is attributed to Machedon and Klainerman – and is similarly best possible in higher dimensions; see [7].

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

† Thomas Wolff died tragically on July 31, 2000, after submitting this paper.
follows from the Strichartz inequality. When $d = 4$, Theorem 1 implies (via a rescaling argument as in [17]) 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 $\Gamma$ 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 [12], 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], [12] and [16]. 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 [21].

**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.
- $\chi_E$: 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 $\Gamma_1$ and $\Gamma_2$ as described in the
introduction and will say that a light ray is white (resp. black) if its direction belongs to \( \Gamma_1 \) (resp. \( \Gamma_2 \)). Thus any white and black rays are transverse. We fix a small positive number \( \varepsilon \).

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 \( \mathcal{W} \) (or black line \( \mathcal{B} \)) we associate to \( \mathcal{W} \) (or \( \mathcal{B} \)) the infinite cylinder whose axis is \( \mathcal{W} \) (or \( \mathcal{B} \)) and whose cross section radius is \( \delta \). We will denote these tubes by \( w \) and \( b \). For each tube \( w \) (similarly \( b \)) we define

\[
\phi_w(x) = \min \left( 1, \frac{\delta}{\text{dist}(x,w)} \right)^M
\]

where \( M \) is a large constant depending on \( \varepsilon \). We assume that \( \mathcal{W} \) (similarly \( \mathcal{B} \)) is \( \delta \)-separated; by this we mean the following: if \( D \) is a disc in projective space with radius \( \delta \), 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 \( \delta \).

A \( \mu \)-fold point is a point which belongs to at least \( \mu \) white tubes, and a smooth \( \mu \)-fold point is a point where the quantity

\[
\Phi_W \overset{\text{def}}{=} \sum_{w \in \mathcal{W}} \phi_w
\]

is at least equal to \( \mu \).

We fix a partition of \( Q(1) \) into pairwise disjoint \( \delta \)-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

\[
n_W(Q) = |\{ w \in \mathcal{W} : w \sim Q \}|,
\]

\[
n_B(Q) = |\{ b \in \mathcal{B} : b \sim Q \}|.
\]

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 \( \delta \)-cube containing \( x \) (resp. \( E \)), and we define

\[
\Phi_W(x) = \sum_{w \in \mathcal{W}, w \sim x} \phi_w(x),
\]

\[
\Phi_B(x) = \sum_{b \in \mathcal{B}, b \sim x} \phi_b(x).
\]
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 \( \eta \)-bush to be 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 \( \delta \)-separated. Then there is a relation \( \sim \) between white or black tubes and \( \delta^\varepsilon \)-cubes \( Q \) so that the following hold, where \( C \) depends on \( d \) only; the implicit constants also depend on \( \varepsilon \):

1. \( \sum_Q n_W(Q) \lesssim m \left( \log \frac{1}{\delta} \right)^5 \).
2. \( \sum_Q n_B(Q) \lesssim n \left( \log \frac{1}{\delta} \right)^5 \).
3. The \( \delta \)-entropy of the set \( \{ x \in Q(1) : \tilde{\Phi}_W(x) \geq \mu \text{ and } \Phi_B(x) \geq \nu \} \) is \( \lesssim \delta^{-C\varepsilon \frac{mn}{\mu^2\nu}} \).
4. The \( \delta \)-entropy of the set \( \{ x \in Q(1) : \Phi_W(x) \geq \mu \text{ and } \tilde{\Phi}_B(x) \geq \nu \} \) is \( \lesssim \delta^{-C\varepsilon \frac{mn}{\mu^2\nu}} \).

**Remarks.** 1. It is easy to see that the \( \delta \)-entropy of the points which belong to \( \mu \) white and \( \nu \) 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 property 3 gains a factor of \( \mu \) over the “trivial” bound valid with \( \tilde{\Phi}_W \) replaced by \( \Phi_W \). 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 \( \delta \) in properties 1 and 2 is only logarithmic, or more precisely that it does not involve the specific power \( \delta^{-\varepsilon} \). On the other hand the distinction between \( \mu \)-fold points and smooth \( \mu \)-fold points is purely technical - the functions \( \phi_b \) 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 [19], [20].

Lemma 1.2 below is true because \( \phi_w \) 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 [16].
Lemma 1.2. If \( x \in Q(1) \) is a smooth \( \mu \)-fold point for the white tubes with \( \mu \geq \delta^B \) then \( x \) is a basepoint for an \( \eta \)-bush (of white tubes) with cardinality \( \gtrsim (\log \frac{1}{\delta})^{-1} \mu \left( \frac{\eta}{\delta} \right)^M \) for some \( \eta \leq \delta^{1-\varepsilon} \). Conversely if \( C \) is a large fixed constant and \( x \in Q(1) \) is a basepoint for an \( \eta \)-bush with cardinality \( \gtrsim C \mu \left( \frac{\eta}{\delta} \right)^M \) then \( x \) is a smooth \( \mu \)-fold point.

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

\[
\int_{\Omega} \phi_b \Phi_C \lesssim \delta^{-\varepsilon(d-2)} \delta^d \left( \frac{\eta}{\delta} \right)^{2d-3}.
\]

Proof. First let \( b \) and \( w \) be a black and a white tube respectively. 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

\[
|\{ x \in Q(1) : \min(\phi_b(x), \phi_w(x)) \geq \lambda \}| \lesssim \left( \delta \lambda^{-\frac{1}{M}} \right)^d.
\]

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

\[
\int_{\Omega} \phi_b \phi_w \lesssim \int_{\Omega} \min(\phi_b, \phi_w)
\]

\[
\lesssim \int_0^{(\delta/\Delta(b, w))^M} \left( \delta \lambda^{-\frac{1}{M}} \right)^d d\lambda
\]

\[
\lesssim \delta^{d} \delta^{M-d} \left( \frac{\delta}{\Delta(b, w)} \right).
\]

Now we prove the estimate (3) when \( \eta = \delta \). It is clear from (5) that the contribution to the left side from tubes \( w \in C \) such that \( \Delta(b, w) \gtrsim \delta^\varepsilon \) is small. On the other hand let \( \rho \) be small compared with \( \delta^\varepsilon \), and consider how many tubes \( w \in C \) there can be with \( \Delta(b, w) \leq \rho \). The bush \( C \) 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 \( \Gamma_1 \). If \( b \) contains a certain point \( y \) which lies within \( \rho \) of \( \Gamma_1 \) and is farther than \( \delta^\varepsilon \) from \( p \), then by transversality \( b \) must intersect \( \Gamma \) at a point within \( C\rho \) of \( y \). Thus the number of tubes \( w \) with \( \Delta(b, w) \leq \rho \) is bounded by the \( \delta \)-entropy of the set of lines in \( \Gamma \) which intersect a fixed \( C\rho \)-disc lying at distance farther than \( \delta^\varepsilon \) from the vertex; equivalently,
by the $\delta$-entropy of a $\delta^{-\varepsilon}\rho$-disc on $S^{d-2}$, which is $(\delta^{-\varepsilon}\rho/\delta)^{d-2}$. We conclude using (5) that there is a bound
\[
\sum \left( \frac{\delta^{-\varepsilon}\rho}{\delta} \right)^{d-2} \delta^{d} \left( \frac{\delta}{\rho} \right)^{M-d}
\]
with the sum being over dyadic $\rho \geq \delta$. Thus we get the bound $\delta^{d-\varepsilon(d-2)}$ as claimed.

We now remove the restriction $\eta = \delta$. If $C$ is an $\eta$-bush then, for parameters $\rho$ such that $\rho \geq \eta$ but small compared with $\eta^\varepsilon$, the maximum number of $\eta$-separated lines in $C$ with $\Delta(b, w) \leq \rho$ is bounded by $(\delta^{-\varepsilon}\rho/\eta)^{d-2}$; for this just apply the above argument replacing $\delta$ by $\eta$. The space of light rays is $(2d - 3)$-dimensional, so any fixed light ray can be within $\eta$ of at most $(\eta/\delta)^{2d-3}$ $\delta$-separated ones. It follows that for any $\rho \ll \eta^\varepsilon$ there are $\lesssim (\eta/\delta)^{2d-3}\delta^{-\varepsilon}(\rho + \eta)/\eta)^{d-2}$ tubes $w$ with $\Delta(b, w) \leq \rho$. We now apply (5) as above to bound the left side of (3) by
\[
\sum \left( \frac{\eta}{\delta} \right)^{2d-3} \left( \frac{\delta^{-\varepsilon}(\rho + \eta)}{\eta} \right)^{d-2} \delta^{d} \left( \frac{\delta}{\rho} \right)^{M-d}
\]
plus a negligible error, with the sum being over dyadic $\rho \geq \delta$. Estimate (3) follows from this.

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 $\mu$ (hence the induction argument in the last part of the proof of Lemma 1.1 below).

**Lemma 1.4.** Given a value of $\mu_0$ we can partition $W$ as
\[ W = W_g \cup W_b \]
where

1. $W_g$ has no $\mu_0$-fold points in $Q(1)$, and
2. $W_b = \cup_{i=1}^{R} C_i$ where each $C_i$ 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 $W^i \subset W$, then we let $\kappa(W^i)$ be the maximum possible cardinality for a set of $\delta$-separated $\mu_0$-fold points for $W^i$. We have $\kappa(W) \lesssim \delta^{-d}$ since all the tubes in $W$ are contained in a fixed compact set.

Assume now that $\kappa(W^i) = k$. We will prove: $W^i = W^{i+1} \cup W^i_b$ where $\kappa(W^{i+1}) \leq \frac{k}{2}$, and $W^i_b$ is the union of $\lesssim A \frac{m}{\mu_0} \log \frac{1}{\delta}$ $\delta$-bushes.
Namely, let $\mathcal{R}_i$ be a set of $\delta$-separated $\mu_0$-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 $\mathcal{W}_i^b$ be all tubes $w \in \mathcal{W}_i$ such that $\text{dist}(x, w) < \delta$ for some $x \in \mathcal{R}_i$ and $\mathcal{W}_i^{i+1} = \mathcal{W}_i \setminus \mathcal{W}_i^b$. Evidently $\mathcal{W}_i^b$ is the union of $\lesssim A \frac{m}{\mu_0} \log \frac{1}{\delta}$-bushes; and $\kappa(\mathcal{W}_i^{i+1}) = 0$ since any $\mu$-fold point for $\mathcal{W}_i$ must lie within $\delta$ 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 $\mathcal{W}_i^b$ be the tubes $w \in \mathcal{W}_i$ such that $\text{dist}(x, w) < \delta$ for some $x$ in the random sample, and $\mathcal{W}_i^{i+1} = \mathcal{W}_i \setminus \mathcal{W}_i^b$. Evidently $\mathcal{W}_i^b$ is the union of $\lesssim A \frac{m}{\mu_0} \log \frac{1}{\delta}$-bushes. We will show that with high probability $\kappa(\mathcal{W}_i^{i+1}) \leq \frac{k}{2}$.

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

Now let $\mathcal{R}_{i+1}$ be a maximal set of $\delta$-separated $\mu_0$-fold points for $\mathcal{W}_i^{i+1}$, and let $\mathcal{R}$ be a maximal 2$\delta$-separated subset of $\mathcal{R}_{i+1}$. Consider the quantity

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

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

$$\sum_{w \in \mathcal{W}_i^{i+1}} kP(w) = \sum_{x \in \mathcal{R}_i} |\{w \in \mathcal{W}_i^{i+1} : \text{dist}(w, x) < \delta\}| \geq \mu_0 |\mathcal{R}|.$$ 

The first line follows from the definition by reversing the order of summation, and the second line then follows because every point in $\mathcal{R}$ is within $\delta$ of a point of $\mathcal{R}_i$ and no two points of $\mathcal{R}$ can be within $\delta$ 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 \( W^0 = \mathcal{W} \) and apply the preceding to express \( \mathcal{W}^0 = \mathcal{W}_b^0 \cup \mathcal{W}^1 \). Then apply the preceding to express \( \mathcal{W}^1 = \mathcal{W}_g^1 \cup \mathcal{W}^2 \) 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 \( \kappa(\mathcal{W}) \) 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 \( \mathcal{W}_b = \bigcup_i \mathcal{W}_b^i \). 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. \( \mathcal{W}_b = \bigcup \forall \mathcal{W}_b^k \), and for each \( k \), \( \mathcal{W}_b^k = \bigcup_{i=1}^{R_k} C_i \), where \( C_i \) is a \( 2^k \delta \)-bush with basepoint in \( Q(2) \) and \( R_k \lesssim \frac{m}{2^{m \mu_0} (\log \frac{1}{\delta})} \).

**Proof.** Let \( w^\eta \) be the \( \eta \)-tube with the same axis as \( w \). Notice that Lemma 1.3 is applicable also to the \( w^\eta \)'s (provided \( \log \eta \) is comparable to \( \log \delta \), which will be the case below), since we used \( \delta \)-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 \( \delta \).

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 \( \mu_0 \). If \( k \geq 2 \) and if \( \mathcal{W}_b^j \) have been defined for \( j < k \) then we let \( \mathcal{W}_b^{k-1} = \mathcal{W} \setminus \bigcup_{j=1}^{k-1} \mathcal{W}_b^j \). The following inductive hypothesis will hold:

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

Let \( \eta = 2^k \delta \) and apply Lemma 1.3 to the tubes \( \{ w^\eta : w \in \mathcal{W}_b^{k-1} \} \) replacing \( \mu_0 \) by \( 2^{Mk} \mu_0 \). This decomposes \( \mathcal{W}_b^{k-1} = \mathcal{W}_g^k \cup \mathcal{W}_b^k \) where the tubes \( \{ w^{2^k \delta} : w \in \mathcal{W}_b^k \} \) have no \( 2^{Mk} \mu_0 \)-fold points and \( \mathcal{W}_b^k \) is the union of at most \( \frac{m}{2^{m \mu_0} (\log \frac{1}{\delta})} 2^{k \delta} \)-bushes. The inductive hypothesis (\( \ast \)) 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 \( 2^k < \delta^{-\varepsilon} \), since we have assumed \( \mu_0 \geq \delta^B \). We define \( \mathcal{W}_g \) to be the last \( \mathcal{W}_b^k \).

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^{2k}\delta : w \in \mathcal{W}_g^k\}\}, which is impossible by construction. The lemma now follows by replacing \(\mu_0\) with \((C\log \frac{1}{\delta})^{-1}\mu_0\).

To prove Lemma 1.1 it suffices by symmetry to construct a relation between white tubes and \(\delta^\varepsilon\)-squares so that properties 1 and 3 hold. This will again be done recursively. A remark on terminology: in this argument, when we say that “\(C\) is a \(2^k\delta\)-bush” we mean that \(C\) 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 \(\eta\)-bushes \(\mathcal{C}_1^j\) (thus each \(\mathcal{C}_1^j\) is a \(2^k\delta\)-bush for some \(k\) with \(2^k\delta \leq \delta^{1-\varepsilon}\)). Then we apply Lemma 1.5 to \(\mathcal{W}_g^1\) with \(\mu_0 = \frac{m}{2}\) obtaining \(\mathcal{W}_g^2\) with \(\Phi_{\mathcal{W}_g^2} \leq \frac{m}{4}\) and stage two \(\eta\)-bushes \(\mathcal{C}_2^i\) and continue in this manner, taking \(\mu_0 = \frac{m}{2^j}\) at the \(j\)th stage. We stop the induction at stage \(R\), where \(R\) by definition is the smallest integer such that \(m^{2R} < \delta^R\).

Clearly \(R \lesssim \log \frac{1}{\delta}\). For each \(j_0 \leq R\) we now have a decomposition

\[
\mathcal{W} = \mathcal{W}_g^{j_0} \cup (\bigcup_{j<j_0} \cup_i \mathcal{C}_j^i)
\]

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

For each \(j\) and \(k\) there are \(\lesssim 2^j 2^{-Mk} \left(\log \frac{1}{\delta}\right)^3\) 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 \(~\): 

**Definition.** A tube \(w\) and \(\delta^\varepsilon\)-square \(Q\) are related, \(w \sim Q\), if \(w\) belongs to an \(\eta\)-bush \(\mathcal{C}_i^j\) such that \(p_i^j\) is in \(Q\) or one of its neighbors.

We show first that property 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 \(\Phi_{\mathcal{W}_g^{j-1}} \leq \frac{m}{2^{j-1}}\), 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

\[
\sum_i |\mathcal{C}_i^j| \lesssim \sum_k 2^j 2^{-Mk} \left(\log \frac{1}{\delta}\right)^3 \cdot 2^{Mk} \frac{m}{2^j} \lesssim m \left(\log \frac{1}{\delta}\right)^4.
\]

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

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

\[
(8) \quad |\{ (i, j) : j \leq j_0 \text{ and } C^j_i \text{ is a } 2^k \delta\text{-bush} \}| \lesssim 2^{-Mk} \left( \frac{\log 1}{\delta} \right)^3 \frac{m}{\mu}.
\]

Fix a black tube \( b \), and fix also a choice of \( C^j_i \) with \( j \leq j_0 \). Define \( \Omega_{ij} \) by deleting from \( Q(1) \) the \( \delta\varepsilon \)-square containing \( p^j_i \) and its neighbors. Lemma 1.3 implies that if \( C^j_i \) is a \( 2^k \delta \)-bush then

\[
\int_{\Omega_{ij}} \Phi_b \Phi_{C^j_i} \lesssim \delta^{-(d-2)\varepsilon} \frac{nm}{\mu} \delta d \left( \frac{\log 1}{\delta} \right)^3,
\]

where \( C \) depends on \( d \).

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

\[
\sum_{ij} \int_{\Omega_{ij}} \Phi_b \Phi_{C^j_i} \lesssim n \sum_k 2^{-Mk} \left( \frac{\log 1}{\delta} \right)^3 \frac{m}{\mu} \cdot \delta^{-(d-2)\varepsilon} \frac{nm}{\mu} \delta d \left( \frac{\log 1}{\delta} \right)^3
\]

where the first inequality follows from (8).

Suppose now that \( x \) is a point such that \( \tilde{\Phi}_W(x) \geq \mu \). By the definition of the relation \( \sim \) we have

\[
\tilde{\Phi}_W(x) \leq \Phi_{W^{j_0}}(x) + \sum_{j \leq j_0 \atop x \in \Omega_{ij}} \Phi_{C^j_i}(x).
\]

The first term on the right side is \( \leq \frac{\mu}{2} \), so

\[
\tilde{\Phi}_W(x) \leq 2 \sum_{j \leq j_0 \atop x \in \Omega_{ij}} \Phi_{C^j_i}(x)
\]

whence

\[
\int \Phi_b \tilde{\Phi} \leq 2 \sum_{ij} \int_{\Omega_{ij}} \Phi_b \Phi_{C^j_i} \lesssim \delta^{-(d-2)\varepsilon} \frac{nm}{\mu} \delta d \left( \frac{\log 1}{\delta} \right)^3.
\]

It follows that the measure of the set where \( \Phi_b \geq \nu \) and \( \tilde{\Phi} \geq \mu \) is

\[
\lesssim \delta^{-(d-2)\varepsilon} \frac{nm}{\nu \mu^2} \delta d \left( \frac{\log 1}{\delta} \right)^3.
\]
Using that the functions $\phi_w$ are roughly constant on $\delta$-discs it then follows that the $\delta$-entropy is
\[
\lesssim \delta^{-(d-2)\epsilon} \frac{nm}{\nu \mu^2} \left( \log \frac{1}{\delta} \right)^3
\]
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.

\textbf{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. $\phi$ 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 $\delta$, which are $\delta$-separated in the same sense as before; i.e. the ones whose direction belongs to a given $\delta$-disc in projective space have bounded overlap, and furthermore the axis directions belong to $\Gamma_1$ and $\Gamma_2$ respectively. Let $m = |\mathcal{W}|$, $n = |\mathcal{B}|$. Fix (in addition to $\epsilon$) another small positive $\eta$; the choice of $M$ and the implicit constants below may now also depend on $\eta$. The quantities $\Phi_W$ and $\tilde{\Phi}_W$ are defined in the same way as before, except of course that we use the modified definition of $\phi_w$ via the definition above.

\textbf{Lemma 1.1'}. With the above assumptions there is a relation $\sim$ between white or black tubes $w \in \mathcal{W}$ or $b \in \mathcal{B}$ and $\delta^\epsilon$-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 $\delta$-entropy of the set $\{x \in Q(1) : \Phi_W(x) \geq \mu \text{ and } \Phi_B(x) \geq \nu\}$ is $\lesssim \delta^{-C\epsilon \frac{nm}{\mu \nu^2}}$.

4. The $\delta$-entropy of the set $\{x \in Q(1) : \Phi_W(x) \geq \mu \text{ and } \tilde{\Phi}_B(x) \geq \nu\}$ is $\lesssim \delta^{-C\epsilon \frac{nm}{\mu \nu^2}}$.

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

\(^1\)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 $\delta$-separated.
distance from \( w \) to the origin is less than \( \delta^{-\frac{3}{2}} \). Then properties 1 and 2 follow immediately from properties 1 and 2 of Lemma 1.1, and properties 3 and 4 follow from properties 3 and 4 of Lemma 1.1 using that the contribution to \( \Phi \) from tubes further than \( \delta^{-\frac{3}{2}} \) 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 \( \Gamma \) by a family of “sectors” \( \rho = \rho_c \), where \( \rho_c = \{ x \in \Gamma : \frac{\rho}{\|x\|} \leq c \} \).

We will be using a variant on the square function estimate in [12]. To state it, let \( \{\rho_j\} \) be the sectors \( \rho \) which intersect \( \Gamma_1 \) and let \( \{\tilde{\rho}_k\} \) be the sectors which intersect \( \Gamma_2 \). Let \( f \) 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 \( \text{supp} f_j \) is contained in the \( N^{-1/2} \)-neighborhood of the sector \( \rho = \rho_j \), and likewise \( \text{supp} g_k \) is contained in the \( N^{-1/2} \)-neighborhood of \( \tilde{\rho}_k \). Let \( F = \hat{f}, G = \hat{g} \), and \( SF = (\sum_j |f_j|^2)^{\frac{1}{2}}, SG = (\sum_k |g_k|^2)^{\frac{1}{2}} \).

**Lemma 2.1.** \( \|FG\|_{L^2}^2 \lesssim \min(\mu, \nu) \|(SF)(SG)\|_{L^2}^2 \).

**Proof.** [12] We claim that for a given point \( z \in \mathbb{R}^d \) there are \( \lesssim \min(\mu, \nu) \) pairs \( (j, k) \) such that \( z \in \text{supp} f_j + \text{supp} g_k \).

We will use the following geometrically obvious fact (a consequence of the strict convexity of the sphere): let \( \varepsilon_0 \) be a fixed positive constant and let \( \zeta, \omega_1, \omega_2 \) be points of \( S^{d-2} \) with \( |\omega_i - \zeta| \geq \varepsilon_0 \) for \( i = 1, 2 \). Let \( \ell \) be a line in \( \mathbb{R}^{d-1} \) which passes through the point \( \zeta \) and assume that both \( \omega_1 \) and \( \omega_2 \) are at distance at most \( \delta \) from \( \ell \). Then \( |\omega_1 - \omega_2| \leq C\delta \), where \( C \) depends on \( \varepsilon_0 \).

In order to prove the claim it suffices to show that for fixed \( j \) the set of \( k \) such that \( z \in \text{supp} f_j + \text{supp} g_k \) has bounded cardinality. To this end we fix \( \zeta \) with \( (\zeta, 1) \in \rho_j \), and \( \omega_1 \) and \( \omega_2 \) such that \( (\omega_i, 1) \in \tilde{\rho}_k \) and \( z \in \text{supp} f_j + \text{supp} g_k \), for \( i = 1, 2 \). If we let \( z = (w, t) \) then for suitable \( a, b \in [\frac{1}{2}, 2] \) we have

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

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

and therefore

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

so that

\[
\omega_1 - \zeta = a^{-1}(w - t\zeta) + O(N^{-\frac{1}{2}}).
\]

Estimate (9) says that the distance from \( \omega_1 \) to the line through \( \zeta \) spanned by \( w - t\zeta \) is \( \lesssim N^{-\frac{1}{2}} \). Likewise the distance from \( \omega_2 \) 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 $\Gamma_1$; similarly $\Gamma^{(N)}_1$ is the $\frac{1}{N}$-neighborhood of $\Gamma_1$, etc. Corresponding to the covering of $\Gamma$ 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 $\rho$ to denote one of the latter. Thus $\rho$ is essentially a $1 \times N^{-\frac{1}{2}} \times \ldots \times N^{-\frac{1}{2}} \times N^{-1}$-rectangle. We fix disjoint sets $E_\rho \subset \rho$ with $\cup_\rho E_\rho = \Gamma^{(N)}_1$ and let $\zeta_\rho = \chi_{E_\rho}$.

Let $f$ be a function supported on $\Gamma^{(N)}_1$ with $L^2$ norm 1, $F = \hat{f}$, $F_\rho = \hat{\zeta_\rho} f$, and

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

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 $\rho$ we fix a tiling $\mathcal{F}_\rho$ of $\mathbb{R}^d$ by rectangles $\sigma$ with dimensions $N \times N^\frac{1}{2} \times \ldots \times N^\frac{1}{2}$, the long direction being orthogonal to the light cone $\Gamma$ at points of (the center line of) $\rho$, and we let $\mathcal{F} = \cup_\rho \mathcal{F}_\rho$. We also let $\mathcal{P}_\rho$ be a tiling by $N \times N^\frac{1}{2} \times \ldots \times N^\frac{1}{2} \times 1$ rectangles dual to the sector $\rho$. For each $\rho$ and each $\sigma \in \mathcal{F}_\rho$ we define $F^\sigma_\rho = b_\sigma F_\rho$, where $b_\sigma$ (and also $b_\pi$, $\phi_\sigma$, etc. in the subsequent argument) are as in the definition at the end of Section 1; thus $\sum_\sigma F^\sigma_\rho = F_\rho$. For each $(\rho, \sigma)$ we also further decompose $F^\sigma_\rho$ as $\sum_{\pi \in \mathcal{P}_\rho} F^{\sigma, \pi}_\rho$, where $F^{\sigma, \pi}_\rho = b_\pi F^{\sigma}_\rho$. The following fact (trivial to prove, since $b$ has compact support) will be very important below:

**Lemma 3.1.** The inverse Fourier transforms of the functions $F^\sigma_\rho$ and $F^{\sigma, \pi}_\rho$ are supported in a fixed dilate $\overline{\rho}$ of $\rho$, and in particular are supported in the $C\frac{1}{N}$-neighborhood of $\Gamma$. 


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

\[
\sum_{\rho} \left| \sum_{\sigma \in A^\rho} F^\sigma_{\rho} \right|^2 \lesssim \sum_{\rho} \sum_{\sigma \in A^\rho} |F^\sigma_{\rho}|^2 \phi_{\sigma}^{-2} \lesssim \sum_{\rho, \sigma, \pi} |F^\sigma_{\rho, \pi}|^2 \phi_{\pi}^{-2} \phi_{\sigma}^{-2}.
\]

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 \( \pi_0 \) be a rectangle containing the origin, and let \( \pi \) be a translate of \( \pi_0 \). 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 \( L^\infty \) with norm \( \lesssim |\pi|^{1/2} \), since one can easily show that \( \int |K(x, y)|^2 dy \lesssim |\pi| \) for fixed \( x \).

**Lemma 3.2.** For fixed \( \rho \) and \( \sigma \in F^p \) we have

\[
\sum_{\pi} \|\phi_{\pi}^{-2} \phi_{\sigma}^{-3} F^\sigma_{\rho, \pi}\|_\infty^2 \lesssim N^{-\frac{q}{2}} \|\phi_{\sigma}^{-4} F^\sigma_{\rho}\|_2^2.
\]

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

\[
|\phi_{\pi}^{-2}(x) \phi_{\sigma}^{-3}(x) F^\sigma_{\rho, \pi}(x)| \lesssim \int K(x, y)|\phi_{\pi}^{-4}(y) \phi_{\sigma}^{-4}(y) F^\sigma_{\rho, \pi}(y)| dy
\]

where

\[
K(x, y) = |\rho| \phi_{\pi}(x)^{-2} \phi_{\sigma}(x)^{-3} \phi_{\pi_0}(x - y)^{100} \phi_{\sigma_0}(x - y)^{100} \phi_{\pi}(y)^4 \phi_{\sigma}(y)^4
\]

\[
\lesssim |\rho| \phi_{\pi}(x)^{-2} \phi_{\pi_0}(x - y)^{100} \phi_{\pi}(y)^4.
\]

We have seen that the norm of this kernel from \( L^2 \) to \( L^\infty \) is \( \lesssim |\pi|^{1/2} |\rho| \approx |\rho|^{1/2} \approx N^{-\frac{q}{2}} \). Accordingly

\[
\sum_{\pi} \|\phi_{\pi}^{-2} \phi_{\sigma}^{-3} F^\sigma_{\rho, \pi}\|_\infty^2 \lesssim N^{-\frac{q}{2}} \sum_{\pi} \|\phi_{\pi}^{-4} \phi_{\sigma}^{-4} F^\sigma_{\rho, \pi}\|_2^2
\]

\[
= N^{-\frac{q}{2}} \sum_{\pi} \|(\phi_{\pi}^{-4} b_{\pi}) \phi_{\sigma}^{-4} F^\sigma_{\rho}\|_2^2
\]

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

For each \( \rho \) and each \( \sigma \in F^p \) we define a parameter

\[
h(\sigma) = \left( N^{-\frac{q}{2}} \|\phi_{\sigma}^{-4} F^\sigma_{\rho}\|_2 \right)^{1/2}.
\]
We think of $h(\sigma)$ as being essentially the $L^2$ average of $F_\rho$ on $\sigma$. We group the $\sigma$’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

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

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

If $g$ is a function supported on $\Gamma_2^{(N)}$ with $L^2$ norm 1 we will likewise denote $\hat{g} d\sigma$ by $G$, etc. Thus we obtain also functions $G_\rho, G_\rho^\sigma, G_\rho^\sigma, G_h$, and families of tubes $G, G_\rho, G(h)$. 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 $\mathcal{F}$ and $\mathcal{G}$ be subsets of $\mathcal{F}(h_1)$ and $\mathcal{G}(h_2)$ respectively and let $\mu$ and $\nu$ be the maximum values on the square $Q$ of the functions $\Phi_\mathcal{F}$ and $\Phi_\mathcal{G}$. Then

$$\int_Q \left| \left( \sum_{\rho} \sum_{\sigma \in \mathcal{F}_\rho} F_\rho^\sigma \right) \left( \sum_{\rho_2} \sum_{\sigma_2 \in \mathcal{G}_{\rho_2}} G_{\rho_2}^{\sigma_2} \right) \right|^2 \lesssim h_1^2 h_2^2 \mu \nu \min(\mu, \nu) N^{\frac{d}{2}}.$$ 

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

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

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

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

$$\|\phi_\sigma\|_\infty \lesssim 2^{-k}.$$ 

(12)
This follows from the rapid decay of $\phi$ and the fact that $\sigma$ contains a translate of $Q$. Hence also $\|\phi\sigma b_Q\|_\infty \lesssim 2^{-k}$. Furthermore, from the definition of $\mu$ and $\nu$, we have
\begin{equation}
|\tilde{F}(k)| \lesssim 2^k \mu \quad \text{and} \quad |\tilde{G}(\ell)| \lesssim 2^\ell \nu.
\end{equation}
The left side of (11) is $\lesssim \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} 2^{k+\ell} A(k, \ell)$, where
\begin{equation}
A(k, \ell) = \left| \left( \sum_{\rho} b_Q^3 \sum_{\sigma \in \tilde{F}(k) \cap \mathcal{F}} F_{\rho}^\sigma \right) \left( \sum_{\rho_2} b_Q^3 \sum_{\sigma_2 \in \tilde{G}(\ell) \cap \mathcal{G}^\rho_2} G_{\rho_2}^{\sigma_2} \right) \right|^2.
\end{equation}
Using Lemma 3.1 and that $\hat{b}$ has compact support, one sees that the Fourier transform of the function $b_Q^3 \sum_{\sigma \in \tilde{F}(k) \cap \mathcal{F}} F_{\rho}^\sigma$ is supported in the $CN^{-\frac{d}{2}}$-neighborhood of the sector $\rho$; and similarly with the second factor in (14). Lemma 2.1 is therefore applicable and implies that
\begin{equation}
A(k, \ell) \lesssim \min(|\tilde{F}(k)|, |\tilde{G}(\ell)|) \int \sum_{\rho} b_Q^4 \sum_{\sigma \in \tilde{F}(k) \cap \mathcal{F}} F_{\rho}^\sigma \left| \sum_{\rho_2} b_Q^3 \sum_{\sigma_2 \in \tilde{G}(\ell) \cap \mathcal{G}^\rho_2} G_{\rho_2}^{\sigma_2} \right|^2.
\end{equation}
It follows by (10) that $A(k, \ell)$ is
\begin{equation}
\lesssim \min(|\tilde{F}(k)|, |\tilde{G}(\ell)|) \int b_Q^4 \sum_{\rho, \sigma, \pi} \sum_{\rho_2, \sigma_2, \pi_2} |F_{\rho}^\sigma \pi| |G_{\rho_2}^{\sigma_2 \pi_2}|^2 \phi_{\rho}^{-2} \phi_\sigma^{-2} \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-2}.
\end{equation}
We claim next that for each pair $(\pi, \pi_2)$ we have
\begin{equation}
\int b_Q^4 \phi_{\pi_2}^2 \phi_{\pi_2}^2 \lesssim N^{\frac{d-2}{4}}.
\end{equation}
Namely, $\pi$ and $\pi_2$ each have one “short” direction in which the width is 1, and these directions lie in $\Gamma_1$ and $\Gamma_2$ 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
\begin{equation}
\int_{\pi \cap \pi_2} b_Q^4 \lesssim N^{\frac{d-2}{4}}.
\end{equation}
Estimate (16) is just a version of (17) incorporating Schwartz tails, and is proved by estimating $\phi_{\pi}^2$ by an appropriate sum of constants times characteristic functions of translates of $\pi$ (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 $(\rho, \sigma, \pi)$ and $(\rho_2, \sigma_2, \pi_2)$ we have
\begin{equation}
\int b_Q^4 |F_{\rho}^\sigma \pi| |G_{\rho_2}^{\sigma_2 \pi_2}|^2 \phi_{\rho}^{-2} \phi_\sigma^{-2} \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-2} \lesssim \|b_Q^2 \phi_{\rho}^{-2} \phi_\sigma^{-1} F_{\rho}^\sigma \pi\|_{\infty} \|b_Q^2 \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-1} G_{\rho_2}^{\sigma_2 \pi_2}\|_{\infty} \int b_Q^4 \phi_{\rho}^2 \phi_{\pi_2}^2 \lesssim N^{\frac{d-2}{4}} \|b_Q^2 \phi_{\rho}^{-2} \phi_\sigma^{-1} F_{\rho}^\sigma \pi\|_{\infty} \|b_Q^2 \phi_{\pi_2}^{-2} \phi_{\sigma_2}^{-1} G_{\rho_2}^{\sigma_2 \pi_2}\|_{\infty}.$
by (16). It then follows that

\begin{equation}
\int b_{Q}^{12} |F_{\rho}^{\sigma, \pi}|^2 |C_{\rho_2}^{\sigma_2, \pi_2}|^2 \\
\lesssim N^{\frac{4d}{s}} \|\phi_{\sigma}^2 b_{Q}\|_{\infty}^2 \|\phi_{\sigma_2}^2 b_{Q}\|_{\infty}^2 \|\phi_{\rho}^{-3} \phi_{\pi}^{-2} F_{\rho}^{\sigma, \pi}\|_{\infty}^2 \|\phi_{\rho_2}^{-3} \phi_{\pi_2}^{-2} C_{\rho_2}^{\sigma_2, \pi_2}\|_{\infty}^2 \\
\lesssim 2^{-4k-4\ell} N^{\frac{4d}{s}} \|\phi_{\rho}^{-3} \phi_{\pi}^{-2} F_{\rho}^{\sigma, \pi}\|_{\infty}^2 \|\phi_{\rho_2}^{-3} \phi_{\pi_2}^{-2} C_{\rho_2}^{\sigma_2, \pi_2}\|_{\infty}^2.
\end{equation}

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

\[
\min(|\tilde{F}(k)|, |\tilde{G}(\ell)|) 2^{-4k-4\ell} N^{-\frac{4d}{s}} \sum_{\rho, \rho_2} \sum_{\sigma \in \tilde{F}(k) \cap \tilde{F}_\rho} \sum_{\sigma_2 \in \tilde{G}(\ell) \cap \tilde{G}_\rho} \|\phi_{\sigma}^{-4} F_{\rho}^{\sigma} \|_{2}^2 \|\phi_{\sigma_2}^{-4} C_{\rho_2}^{\sigma_2} \|_{2}^2
\]

which by definition of $h_1$ and $h_2$ is

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

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

\[
2^{-3k-3\ell} \mu \nu \min(2^{k+1}, 2^\ell) N^{\frac{d}{s}} h_1^2 h_2^2.
\]

Summing over $k$ and $\ell$ gives the lemma.

Fix $\varepsilon > 0$ and then $\eta > 0$ and partition $Q(N)$ in nonoverlapping $N^{1-\varepsilon}$-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

\[
F_{h_1} = F_g + F_b \quad \text{and} \quad F_b = \sum_R F_b^R
\]

\[
G_{h_2} = G_g + G_b \quad \text{and} \quad G_b = \sum_R G_b^R
\]

where $\text{supp} F_R^b \subset R$, $\text{supp} G_R^b \subset R$, and the following estimates hold.

1. $\int_{Q(N)} |F_g G_g|^2 + |F_b G_g|^2 + |F_g G_b|^2 \lesssim N^{-\frac{4d}{s} + C\varepsilon}$

2. For each $R$ we have $F_R^b = \alpha_R \tilde{f}_R$ and $G_R^b = \beta_R \tilde{g}_R$, where $\alpha_R$ and $\beta_R$ are supported on $R$ and have $L^\infty$ norm $\leq 1$, and $f_R$ and $g_R$ are supported on the $N^{-(1-\varepsilon)}$-neighborhoods of $\Gamma_1$ and $\Gamma_2$ respectively, and

\[
\sum_R \|f_R\|_2^2 + \|g_R\|_2^2 \leq C_\eta N^{-\varepsilon + \eta}.
\]
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 $\delta$ in Lemma 1.1 is $N^{-\frac{d}{2}}$; and we also set $\varepsilon$ in Lemma 1.1' equal to twice the present $\varepsilon$.

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

$$F^R_b = \begin{cases} \sum_{\sigma \in W} F^\sigma_{\rho} & \text{on } R \\ 0 & \text{elsewhere} \end{cases}$$

$$G^R_b = \begin{cases} \sum_{\sigma \in B} G^\sigma_{\rho} & \text{on } R \\ 0 & \text{elsewhere}. \end{cases}$$

Define $F_b$ to be equal to $F^R_b$ 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

$$\int_{Q(N)} |F_g G|^2 \lesssim N^{\varepsilon} h_1^{\frac{d+2}{4}} + C \varepsilon.$$

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

$$\int_{Q(N)} |F_g G|^2 \lesssim h_1^2 h_2^2 \mu^2 \nu N^{\varepsilon}.$$

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

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

which is $\lesssim N^{C \varepsilon} \cdot N^{\varepsilon} \cdot N^{\frac{d+2}{4}}$ by Lemma 3.3. We can clearly estimate $\int_{Q(N)} |F_g|^2$ and $\int_{Q(N)} |F_g G_g|^2$ in the same way, and it follows that property 1 holds.

We have the following almost orthogonality estimate:

$$\left\| \sum_{\sigma \sim R} F^\sigma_{\rho} \right\|_2^2 \lesssim h_1^2 N^{\frac{d+2}{4}} |\{\sigma : \sigma \sim R\}|.$$

Namely, for fixed $\rho$ we have

$$\left\| \sum_{\sigma \sim R} F^\sigma_{\rho} \right\|_2^2 \lesssim \sum_{\sigma \sim R} \left\| F^\sigma_{\rho} \right\|_2^2 \lesssim h_1^2 N^{\frac{d+1}{4}} |\{\sigma \in \mathcal{F}^\rho : \sigma \sim R\}|$$

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

Using Lemma 3.1 again we see that, on each fixed \( N^{1-\varepsilon} \) square \( R \), \( F_{b} = \sum_{\sigma \sim R} F^{\sigma}_{\rho} \) agrees with the Fourier transform of a function \( f^{0}_{R} \) supported on the \( C_{N} \)-neighborhood of \( \Gamma_{1} \). We have

\[
\sum_{R} \| f^{0}_{R} \|_{2}^{2} = \sum_{R} \left\| \sum_{\sigma \sim R} F^{\sigma}_{\rho} \right\|_{2}^{2} \lesssim h_{1}^{2} N^{d \frac{d+1}{2}} \sum_{R} \left| \{ \sigma : \sigma \sim R \} \right|
\]

\[
\lesssim h_{1}^{2} N^{d \frac{d+1}{2}} | \mathcal{F}(h_{1}) | N^{n}
\]

\[
\lesssim N^{n}.
\]

The first inequality follows from (21), the second inequality follows from property 1 of Lemma 1.1' and the last inequality follows from Lemma 3.3. Now fix \( R \) and take a suitable Schwartz function \( \kappa \) supported in \( D(0, \frac{1}{2}) \) and whose Fourier transform is \( \geq 1 \) on a large disc centered at the origin. Let \( \kappa_{R}(x) = e^{ik \cdot x} N^{d(1-\varepsilon)} \kappa(N^{1-\varepsilon} x) \) for an appropriate \( k \); if \( k \) is chosen correctly then \( \hat{\kappa}_{R} \geq 1 \) on \( R \). Define \( \alpha_{R} = \frac{1}{\hat{\kappa}_{R}} \) on \( R \) and zero otherwise, and \( f_{R} = \kappa_{R} \ast f^{0}_{R} \). Then \( F^{R}_{b} = \alpha_{R} \hat{f}_{R} \). 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 \( \text{supp} f \) intersects every disc of radius \( r \) in measure \( \leq \gamma r \), then

\[
\| s \ast f \|_{2} \lesssim \gamma^{\frac{1}{2}} \| f \|_{2}.
\]

We apply this with \( s = \kappa_{R}, f = f^{0}_{R}, r \approx N^{1-\varepsilon}, \gamma \approx N^{-\varepsilon} \), which is justified since \( f^{0}_{R} \) is supported on the \( \frac{C_{N}}{N} \)-neighborhood of \( \Gamma \). It follows that

\[
\| f_{R} \|_{2} \lesssim N^{-\varepsilon} \| f^{0}_{R} \|_{2},
\]

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], [16]). In order to prove Theorem 1 it suffices to prove that

\[
\int_{Q(N)} | \hat{f} d\sigma \hat{g} d\sigma |^{p} \lesssim N^{n} \| f \|_{2}^{p} \| g \|_{2}^{p}
\]
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 [16]. 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 $\Gamma_1$ and $\Gamma_2$ respectively, then

\begin{equation}
\int_{Q(N)} |\hat{f} \hat{g}|^p \lesssim N^{-p+\gamma}
\end{equation}

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

\begin{equation}
N^p \int_{Q(N)} |\hat{f} \hat{g}|^p
\end{equation}

over functions $f$ and $g$ with $L^2$ norm 1 which are supported in the $\frac{1}{N}$-neighborhoods of $\Gamma_1$ and $\Gamma_2$ respectively. Fix a sufficiently small $\varepsilon$ and then a much smaller $\eta$; we will show that

\begin{equation}
\phi(N) \leq C(1 + N^\eta \phi(N^{1-\varepsilon}))
\end{equation}

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)

\[ \phi(N) \lesssim (\log N)^{2p} \int_{Q(N)} (|F_b G_g|^p + |F_g G_b|^p + |F_g G_g|^p) + (\log N)^{2p} \sum_R \int_R |\hat{f}_R \hat{g}_R|^p. \]

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

\[ N^{-(1-\varepsilon)p} \phi(N^{1-\varepsilon}) \|f_R\|_2 \|g_R\|_2^p. \]
Making these estimates we conclude that
\[ N^{-p} \phi(N) \lesssim (\log N)^{2p} \left( \int_{Q(N)} (|FG_g|^2 + |F_gG|^2 + |F_gG_g|^2) \right)^{p-1} \]
\[ + (\log N)^{2p} N^{-(1-\varepsilon)p} \phi(N^{1-\varepsilon}) \sum_R \|f_R\|_2^p \|g_R\|_2^p. \]

We now use Hölder’s inequality on the sum over \( R \) and then insert the estimates in Lemma 3.5; this gives
\[ N^{-p} \phi(N) \lesssim (\log N)^{2p} \left( \int_{Q(N)} (|FG_g|^2 + |F_gG|^2 + |F_gG_g|^2) \right)^{p-1} \]
\[ + (\log N)^{2p} N^{-(1-\varepsilon)p} \phi(N^{1-\varepsilon}) \left( \sum_R \|f_R\|_2^2 \right)^{\frac{p}{2}} \left( \sum_R \|g_R\|_2^2 \right)^{\frac{p}{2}} \]
\[ \lesssim (\log N)^{2p} N^{(p-1)(C\varepsilon-\frac{d+2}{2})} \]
\[ + (\log N)^{2p} N^{-(1-\varepsilon)p} \phi(N^{1-\varepsilon}) \cdot N^{p(-\varepsilon+\eta)}. \]

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 \( \eta \) by \( \frac{d-1}{d+1} \), say.

If \( \gamma \) is given and if we take \( \eta \) 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.

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 \( \Gamma \) splitting the \( S_{d-2} \) and radial variables:
\[ \|f\|_{L^p(L^q)} \overset{\text{def}}{=} \left( \int_{S_{d-2}} \left( \int_{1}^{2} |f(t\omega)|^q dt \right)^{\frac{p}{q}} d\omega \right)^{\frac{1}{p}}. \]

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 \( \Gamma \). Then \( \|f d\sigma\|_p \leq C_p \|f\|_{L^p(L^2)}. \)

**Proof.** This is the same as the proof of Theorem 2.2 in [17]; see also [16], 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 $\Gamma_c = \{ x \in \Gamma : x.e \in c \}$. 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:

$$\det T_c = N^d$$

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 $\Gamma$ with $L^p(L^2)$ norm 1 which are supported on $\Gamma_{c_1}$ and $\Gamma_{c_2}$ respectively. Define $\hat{f}d\sigma$ and $\hat{g}d\sigma$ to be the measures obtained by pushing forward $f d\sigma$ and $g d\sigma$ by the map $T_c$. Then $\hat{f}$ and $\hat{g}$ are functions on $\Gamma$ 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

$$\hat{f}d\sigma = \hat{f}d\sigma \circ T_c^{-1}$$
$$\hat{g}d\sigma = \hat{g}d\sigma \circ T_c^{-1}$$

and therefore, by (25) and Theorem 1,

$$\int |\hat{f}d\sigma \hat{g}d\sigma|^\frac{p}{2} = N^d \int |\hat{f}d\sigma \hat{g}d\sigma|^\frac{p}{2}$$

$$\lesssim N^{d-p\frac{d-2}{2}} = N^{d-(p-1)(d-2)}$$

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 $\Gamma$ with $\text{supp} f_j \subset \Gamma_{c_j}$. By applying the preceding estimate and summing over $j$ we obtain

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

The exponent of $N$ is negative if $p > 2 + \frac{4}{d}$. The result now follows exactly as in [17], since the supports of the Fourier transforms of the functions $\chi_j \hat{f}d\sigma \chi_k \hat{f}d\sigma$ have finite overlap if $N$ is fixed and $\text{dist}(c_j, c_k) \approx N^{-1}$.

We now consider the Mockenhaupt square function

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

where $F = \hat{f}d\sigma$ with $f$ supported on $\Gamma$, $f = \sum_\rho f_\rho$ with $f_\rho$ supported in the sector $\rho$ of width about $N^{-\frac{1}{2}}$ and $F_\rho = f_\rho d\sigma$. The following simple result
appears natural in higher dimensions where the expected critical exponent is $2 + \frac{4}{d - 2}$; 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 [16].

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

$$
\|F\|_p \lesssim N^{\left(\frac{1}{2} - \frac{1}{p}\right) \frac{d - 2}{2} + \varepsilon}\|SF\|_p
$$

for any $\varepsilon > 0$.

**Proof.** We introduce a “weaker” square function $\tilde{S}$ defined as follows: let $F = \widehat{f d\sigma}$ be as above, let $\Delta$ run through a covering of $\Gamma$ by discs of radius $N^{-\frac{1}{2}}$, suppose that $f_{\Delta}$ is supported in $\Delta$ and $f = \sum_{\Delta} f_{\Delta}, F_{\Delta} = \widehat{f_{\Delta} d\sigma}$ and

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

To prove (26) 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 $F = \widehat{f d\sigma}, G = \widehat{g d\sigma}$, and will show that

$$
\|FG\|_{L^p_{\mathbb{R}}(Q)} \lesssim N^{-\frac{1}{2} + \frac{d}{2} + \varepsilon} \left(\|\tilde{S}F\|_p^2 + \|\tilde{S}G\|_p^2\right).
$$

Namely, we have

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

when $p > 2 + \frac{4}{d}$. This follows by applying (22) (with $N$ replaced by $N^{\frac{d}{2}}$ 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

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

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 Hölder’s inequality we obtain

$$
\|FG\|_{L^p_{\mathbb{R}}(Q)} \lesssim N^{-\frac{d}{2} - \frac{1}{2} \frac{d - 2}{2} + \varepsilon} \|b_Q \tilde{S}F\|_2 \|b_Q \tilde{S}G\|_2
$$

$$
\lesssim N^{-\frac{d}{2} - \frac{1}{2} \frac{d - 2}{2} + \varepsilon} \cdot N^{\frac{d}{2} (1 - \frac{2}{p} + \varepsilon)} \|b_Q \tilde{S}F\|_p \|b_Q \tilde{S}G\|_p
$$

$$
\leq N^{\frac{1}{2} - \frac{1}{p}} \frac{d - 2}{2} + \varepsilon (\|b_Q \tilde{S}F\|_p^2 + \|b_Q \tilde{S}G\|_p^2).
$$

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

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

$$
\|FG\|_{\mathbb{R}} \lesssim N^{\frac{1}{2} - \frac{1}{p}} \frac{d - 2}{2} + \varepsilon (\|SF\|_p^2 + \|SG\|_p^2).
$$
In the case of $S$, since the maps $T_c$ essentially take sectors contained in $\Gamma_c$ to sectors one can pass from the estimate (28) to the “linear” one (i.e. (26)) by rescaling, just as in [16] 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 [12] 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

\begin{equation}
\| \hat{f} \|_{L^{p'}(L^p)} \lesssim \| f \|_{p'}
\end{equation}

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

\begin{equation}
\forall \varepsilon \exists C_\varepsilon : \| \hat{f} \|_{L^{p'}(L^p)} \leq C_\varepsilon \lambda^\varepsilon \| f \|_{p'}
\end{equation}

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

This statement would easily imply the restriction conjecture for the sphere $S^{d-2}$. Namely, suppose that $f \in L^p(\mathbb{R}^d)$ with $p'$ as above and that $f$ is supported in $Q(N)$, and apply (29) to the function $f(\tau)e^{2\pi i x \cdot \phi(\frac{\tau}{N})}$ where $\phi$ is a suitable bump function. (If one assumes instead (30) then this argument still works using Tao’s $\varepsilon$-removal lemma, see [16] 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 [15], and is therefore also weaker than the so-called local smoothing conjecture [13]. 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 [15]; 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 [15, formula (33)] (and that $p' \leq 2$!), that if $\text{supp} f \subset D(0, \lambda) \subset \mathbb{R}^d$ then

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

Thus if [15, (33)] were true for all $\alpha > -\frac{d-1}{p}$ as is conjectured in [15] 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.

**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], [8], [9], [10], [16] and [18]. This leads to a family of mixed norm estimates which we formulate as Theorem A.1 below.

Let $L$ be the space of light rays with the integral defined by

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

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

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

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

$$\frac{d}{p} - \frac{d - 1}{r} = 1.$$ 

See e.g. [4] and [9]. Furthermore, the maps $T_\epsilon$ used in Section 5 give the condition

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

Again see [9]. Another condition can be obtained by considering the example $f = \chi_E$ where $E$ is the $\delta$-neighborhood of the cone segment $\Gamma$. This takes the form

$$\frac{1}{p} \leq \frac{d}{2r}$$
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 \( W^{p,\varepsilon}(Q(1)) \) to \( L^q(L^r) \) for any \( \varepsilon > 0 \).

By \( W^{p,\varepsilon}(Q(1)) \) we mean functions supported in \( Q(1) \) with
\[
\|f\|_{p,\varepsilon} \overset{\text{def}}{=} \|(1 - \Delta)^{\frac{\varepsilon}{2}} f\|_p < \infty.
\]

There is an obvious bound on \( L^1 \), namely, by Fubini’s theorem
\[(34) \quad \|Xf\|_{L^\infty(L^1)} = \|f\|_1.
\]
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 \( W^{p,\varepsilon}(Q(1)) \) to \( L^q(L^r) \) for any \( \varepsilon > 0 \).

This would imply all local \( W^{p,\varepsilon} \rightarrow L^q(L^r) \) 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 [14] and [4]) for the full X-ray transform.

2. Consider the case \( d = 3 \). In this case, the angular parameter \( \omega \) 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” [8], [9]. If \( d = 3 \) and \( q = r \), then the estimate in Theorem A.1 is an estimate from \( W^{2,\varepsilon} \) to \( L^2 \). The latter estimate is known, actually in the sharper form where \( \varepsilon = 0 \) – cf. [18] (I thank Allan Greenleaf for this reference) and [8] – and a dual formulation of this same estimate is used in [16]. However, Theorem A.1 is new also in the three dimensional case if \( p > \frac{3}{2} \). [Note added in proof: some higher-dimensional versions of these results have since been obtained in [6].]
3. It may be possible to obtain a scale invariant result (i.e. $\varepsilon = 0$) 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 $W^p_{p\varepsilon}$ 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 $d + \frac{3}{2} - 3$, as may be seen by applying the restricted X-ray bound to the indicator function of a cylinder over the $\delta$-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.

\begin{align}
\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} \\
\min(\sum_j a_j \sum_k b_k)^\theta \max(\sum_j a_j, \sum_k b_k)^{1-\theta} &\leq \sum_j \min(a_j, b_k)^\theta \max(a_j, b_k)^{1-\theta}
\end{align}

**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

$$
\min(ax, by)^\theta \max(ax, by)^{1-\theta} = \left(\frac{by}{ax}\right)^{2\theta-1} a^\theta x^\theta b^{1-\theta} y^{1-\theta}
$$

$$
\leq a^\theta x^\theta b^{1-\theta} y^{1-\theta}
$$

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

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 \( W \) and \( B \) be \( \delta \)-separated sets of white and black \( \delta \)-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 \( n_W(\mathcal{Q}) \) defined there. Let

\[
\Phi_W(x) = \sum_{w \in W} \chi_w(x), \quad \Phi_B(x) = \sum_{b \in B} \chi_b(x)
\]

\[
\Phi^*_W(x) = \sum_{w \in W \sim x} \chi_w(x), \quad \Phi^*_B(x) = \sum_{b \in B \sim x} \chi_b(x)
\]

\[
\tilde{\Phi}_W = \Phi_W - \Phi^*_W, \quad \tilde{\Phi}_B = \Phi_B - \Phi^*_B.
\]

**Lemma A.1.** The following hold, where \( C \) depends only on \( d \); the implicit constants also depend on \( \epsilon \), and \( Q \) runs over a partition of \( Q(1) \) into \( \delta^5 \)-squares:

1. \( \sum_Q n_W(\mathcal{Q}) \lesssim |W| \left( \log \frac{1}{\delta} \right)^5 \).
2. \( \sum_Q n_B(\mathcal{Q}) \lesssim |B| \left( \log \frac{1}{\delta} \right)^5 \).
3. \( |\{x \in Q(1) : \Phi_W(x) \geq \mu \text{ and } \Phi_B(x) \geq \nu\}| \lesssim \delta^{-C\epsilon \left( |W|/|B| \right)} \delta^d \).
4. \( |\{x \in Q(1) : \Phi_B(x) \geq \nu \text{ and } \Phi_W(x) \geq \mu\}| \lesssim \delta^{-C\epsilon \left( |W|/|B| \right)} \delta^d \).

The rough idea now is to regard properties 3 and 4 of Lemma A.1 as a “virtual” \( L^{2\infty} \) to \( L^3 \) estimate and to interpolate between this and an \( L^1 \) to \( L^1 \) estimate, namely the following:

**Lemma A.2.**

\[
|\{x \in Q(1) : \Phi_W(x) \geq \mu \text{ and } \Phi_B(x) \geq \nu\}| \lesssim \delta^{d-1} \min \left( \frac{|W|}{\mu}, \frac{|B|}{\nu} \right).
\]
Proof. It is clear that \( \| \sum_{w \in W} \chi_{w} \|_{L^1(Q(1))} \lesssim |W|^{d-1} \), hence the measure of the \( \mu \)-fold points is \( \lesssim \frac{|W|}{\mu} \delta^{d-1} \), which implies the lemma.

Fix \( \theta \in \left[ \frac{1}{2}, 1 \right] \) and define
\[
\Psi_{\theta} = \min(\Phi_{|B|}, \Phi_{|W|})^\theta \max(\Phi_{|B|}, \Phi_{|W|})^{1-\theta},
\]
\[
S_{\theta} = \min(\Phi_{|B|}, \Phi^*_{|W|})^\theta \max(\Phi_{|B|}, \Phi^*_{|W|})^{1-\theta},
\]
\[
T_{\theta} = (\Phi_{|B|})^{\theta} \Phi_{|W|}^1 + (\Phi_{|W|})^{\theta} \Phi_{|B|}^{1-\theta}.
\]

We will use below that
\[
(37) \quad \Psi_{\theta} \lesssim S_{\theta} + T_{\theta}.
\]
This is a consequence of the numerical inequality
\[
\min(a + b, c + d)\theta \max(a + b, c + d)^{1-\theta} \lesssim a^\theta(c + d)^{1-\theta} + c^\theta(a + b)^{1-\theta}
\]
which follows for example from (36).

We now estimate \( T_{\theta} \) for appropriate \( \theta \) by interpolation between Lemmas A.1 and A.2.

**Lemma A.3.** Let \( p \) and \( q \) satisfy \( 1 \leq q \leq 3 \) and \( \frac{1}{q} \geq \frac{2}{p} - 1 \). Let \( \theta = \frac{1}{4}(3 - \frac{1}{q}) \). Then
\[
(38) \quad \| \delta^{d-2}T_{\theta} \|_{L^q(Q(1))} \lesssim \delta^{-C\varepsilon} \left( \delta^{2d-3} |B| \cdot \delta^{2d-3} |W| \right)^{\frac{2}{pq}}.
\]

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

Define \( Y(\mu, \nu) \) to be the set where \( \tilde{\Phi}_{|W|} \geq \mu \) and \( \phi_{|B|} \geq \nu \). Lemmas A.1 and A.2 give
\[
|Y(\mu, \nu)| \lesssim \delta^{-C\varepsilon} \min \left( \frac{|W| |B|}{\mu^2 \nu} \delta^d, \frac{|W|}{\mu} \delta^{d-1}, \frac{|B|}{\nu} \delta^{d-1} \right)
\]
and therefore also
\[
|Y(\mu, \nu)| \lesssim \delta^{-C\varepsilon} \left( \frac{|W| |B|}{\mu^2 \nu} \delta^d \right)^{\frac{2}{p} - 1} \left( \frac{|W|}{\mu} \delta^{d-1} \right)^{1 - \frac{2}{pq}} \left( \frac{|B|}{\nu} \delta^{d-1} \right)^{1 - \frac{2}{pq}}
\]
\[
= \delta^{-C\varepsilon} \left( \frac{|B| |W|}{\mu^2 \nu} \delta^{d-2 + \frac{2}{p}} \right)^{\frac{2}{pq} \cdot \delta^{d-2 + \frac{2}{p}}}
\]
where we used the value of \( \theta \) to obtain the last line. Summing over dyadic levels for \( \mu \) and \( \nu \) between 1 and a negative power of \( \delta \) gives
\[
\| \Phi_{|B|}^{1-\theta} \phi_{|W|}^\theta \|_{L^q} \lesssim \delta^{-C\varepsilon} (|B| |W|)^{\frac{2}{pq} \cdot \delta^{d-2 + \frac{2}{p}}}.
\]
This and the analogous estimate with the roles of $\mathcal{B}$ and $\mathcal{W}$ reversed imply
\[
\|T_\theta\|_q^q \lesssim \delta^{-C\varepsilon}(|\mathcal{B}| |\mathcal{W}|) \frac{2}{q} \delta^{d-2+\frac{2}{p}}
\]
which is equivalent to (38) when $\frac{1}{q} = \frac{2}{p} - 1$.

We will now pass to a similar estimate for $\Psi_\theta$. We will use a rescaling argument and induction on $\delta$ like the final argument in [19] or [20]. 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 $\Psi_\theta$. It is possible that the slightly stronger result obtained by considering $\Psi_\theta$ could prove useful, but the main reason we use $\Psi_\theta$ 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 $\varepsilon > 0$ there is a constant $A_\varepsilon$ making the following estimate valid; here $\theta = \frac{1}{4}(3 - \frac{1}{q})$:
\[
\|\delta^{d-2}\Psi_\theta\|_{L^q(Q(1))} \leq A_\varepsilon \delta^{-C\varepsilon} \left( \delta^{2d-3} |\mathcal{B}| \cdot \delta^{2d-3} |\mathcal{W}| \right)^{\frac{2}{2p}}.
\]

Proof. We start with the following observation concerning rescaling.

Claim. Suppose that $\delta$ is small enough and that (39) has been proved with $\delta$ replaced by $\delta^{1-\varepsilon}$. Let $Q$ be a $\delta^{1-\varepsilon}$-cube, and let $\mathcal{B}$ and $\mathcal{W}$ be $\delta$-separated sets of tubes. Then
\[
\|\Psi_\theta\|_{L^q(Q)}^q \leq \delta^{\frac{C\varepsilon^2}{2} - \varepsilon} \cdot A_\varepsilon \delta^{-C\varepsilon} \left( \delta^{2d-3} |\mathcal{B}| \cdot \delta^{2d-3} |\mathcal{W}| \right)^{\frac{2}{2p}}.
\]

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 $\delta^{-(d-2)\varepsilon}$. 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
\[
\Phi_\mathcal{W}^\mu = \sum_{w \in \mathcal{W}(\mu)} \chi_w, \quad \Phi_\mathcal{B}^\nu = \sum_{b \in \mathcal{B}(\nu)} \chi_b,
\]
\[
\Psi_\theta^{\mu\nu} = \min(\Phi_\mathcal{B}^\nu, \Phi_\mathcal{W}^\mu)^\theta \max(\Phi_\mathcal{B}^\nu, \Phi_\mathcal{W}^\mu)^{1-\theta}.
\]
Then
\[
\Psi_\theta \leq \sum_{\mu, \nu} \Psi_\theta^{\mu\nu}
\]
where the sum is over dyadic values of $\mu$ and $\nu$. This follows from (36).
By (40) and pigeonholing, there are values of \( \mu \) and \( \nu \) such that
\[
\| \Psi_{\theta}^{\mu \nu} \|_{L^q(Q)} \gtrsim (\log \frac{1}{\delta})^{-2} \| \Psi_\theta \|_{L^q(Q)}.
\]
We assume without loss of generality that \( \mu \geq \nu \). Now let \( \overline{B} \) (resp. \( \overline{W} \)) be subsets of \( B(\nu) \) (resp. \( W(\mu) \)) which are maximal with respect to the following property:

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

Let \( \Phi_B \) (resp. \( \Phi_W \)) be the sums of the characteristic functions of the tubes of width \( C_0 \delta \) coaxial with the tubes in \( \overline{B} \) (resp. \( \overline{W} \)), and
\[
\Psi_\theta = \min \{ \Phi_B, \Phi_W \}^\theta \max \{ \Phi_B, \Phi_W \}^{1-\theta}.
\]
Then \( \Phi_B^\mu \lesssim \mu \Phi_W \) and \( \Phi_W^\nu \lesssim \nu \Phi_B \), pointwise on \( Q \); this follows from maximality of \( \overline{W} \) and \( \overline{B} \) provided \( C_0 \) is large enough. Hence also
\[
\Psi_\theta^{\mu \nu} \lesssim \mu^{\theta \nu^{1-\theta}} \Psi_\theta
\]
by (35). Taking \( L^q \) norms we conclude that
\[
\| \Psi_\theta^{\mu \nu} \|_{L^q(Q)} \lesssim \mu^\theta \nu^{(1-\theta)q} \| \Psi_\theta \|_{L^q(Q)}.
\]
Furthermore property (\ast) implies
\[
| \overline{B} | \lesssim \nu^{-1} | B(\nu) |, \quad | \overline{W} | \lesssim \mu^{-1} | W(\mu) |.
\]
We now dilate the situation by a factor \( \delta^{-\varepsilon} \). This maps \( Q \) to a cube \( Q' \) of side 1, and maps \( \overline{B} \) and \( \overline{W} \) to \( \delta^{1-\varepsilon} \) separated families of \( C_0 \delta^{1-\varepsilon} \) tubes. Accordingly we can apply the hypothesis that (39) holds at scale \( \delta^{1-\varepsilon} \). We conclude that
\[
\| \delta^{(1-\varepsilon)(d-2)} \Psi_\theta (\delta^\varepsilon x) \|_{L^q(Q', dx)} \lesssim A_\varepsilon \delta^{-C \varepsilon (1-\varepsilon) (\delta^{2d-3}(1-\varepsilon) | \overline{W} | \cdot \delta^{(2d-3)(1-\varepsilon)} | \overline{B} |)^{\frac{q}{2p}}.
\]
Making the change of variables \( x \to \delta^\varepsilon x \) and factoring out the powers of \( \delta^\varepsilon \) we get
\[
\delta^{-d \varepsilon} \delta^{q (d-2) \varepsilon} \| \delta^{d-2} \overline{\Psi}_\theta \|_{L^q(Q)} \lesssim A_\varepsilon \delta^{-C \varepsilon (1-\varepsilon) \delta^{-\frac{q}{p} (2d-3) \varepsilon} (\delta^{2d-3} | \overline{W} | \cdot \delta^{2d-3} | \overline{B} |)^{\frac{q}{2p}}.
\]
We now substitute in the estimates (41) and (42), obtaining
\[
\mu^{-\theta q \nu^{-(1-\theta)q} \delta^{-d \varepsilon} \delta^{q (d-2) \varepsilon}} \| \delta^{d-2} \overline{\Psi}_\theta \|_{L^q(Q)} \lesssim (\mu \nu)^{-\frac{q \theta}{2p}} \cdot \delta^{-\frac{q}{p} (2d-3) \varepsilon} \cdot A_\varepsilon \delta^{-C \varepsilon (1-\varepsilon)} (\delta^{2d-3} | W_\mu | \cdot \delta^{2d-3} | B_\nu |)^{\frac{q}{2p}}
\]
or equivalently
\[
\| \delta^{d-2} \overline{\Psi}_\theta \|_{L^q(Q)} \lesssim \mu^{q (\theta - \frac{1}{2p}) \nu^{1-\theta - \frac{1}{2p}}} \cdot \delta^{-\frac{q}{p} (2d-3) \varepsilon + d \varepsilon + q (d-2) \varepsilon}
\]
\[
\cdot A_\varepsilon \delta^{-C \varepsilon (1-\varepsilon)} (\delta^{2d-3} | W_\mu | \cdot \delta^{2d-3} | B_\nu |)^{\frac{q}{2p}}.
\]
But \( \nu \leq \mu \leq \delta^{-(d-2)\varepsilon} \), 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

\[
\mu^{q/(\theta - \frac{1}{2p})} L^q(1 - \theta - \frac{1}{2p} - \frac{d}{2p}) (2d-3)\varepsilon + d\varepsilon + q(d(2d-2)) \leq 1.
\]

It follows that

\[
\|\delta^{d-2} \Psi_{\theta} \|_{L^q(Q)} \lesssim \delta^{C\varepsilon^2} A \delta^{-C\varepsilon} (2d-3)\varepsilon W_{\mu} \cdot \delta^{2d-3} |B_{\mu}| \frac{\delta}{\varepsilon}
\]

and therefore

\[
\|\delta^{d-2} \Psi_{\theta} \|_{L^q(Q)} \lesssim \left( \log \frac{1}{\delta} \right)^{2q} \delta^{C\varepsilon^2} \delta^{2d-3} A \delta^{-C\varepsilon} \varepsilon W_{\mu} \cdot \delta^{2d-3} |B_{\mu}| \frac{\delta}{\varepsilon}.
\]

The factor \( \left( \log \frac{1}{\delta} \right)^{2q} \delta^{C\varepsilon^2} \) is evidently small for small \( \delta \), 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 \( \delta_0 \) (the case where \( \delta \) is large is easy if \( A_\varepsilon \) has been chosen appropriately) and will prove it when \( \delta^{1-\varepsilon} > \delta_0 \). This will evidently establish the lemma.

We use (37), and observe that a bound like (39) with \( \Psi_{\theta} \) replaced by \( T_{\theta} \) follows from Lemma A.3; the implicit constant in Lemma A.3 is small compared with \( A_\varepsilon \) if \( A_\varepsilon \) has been chosen appropriately. To estimate \( S_\theta \), subdivide \( Q(1) \) in \( \delta^2 \)-cubes \( Q \). On each fixed \( Q \) we can apply the claim to the restriction of \( S_\theta \) to \( Q \), replacing \( W \) by \( \{ w \in W : w \sim Q \} \) and similarly with \( B \).

We obtain for each \( Q \)

\[
(43) \quad \|\delta^{d-2} S_\theta \|_{L^q(Q)} \lesssim \delta^{C\varepsilon^2} A \delta^{-C\varepsilon} (\delta^{2d-3} n_W(Q) \cdot \delta^{2d-3} n_B(Q)) \frac{\delta}{\varepsilon}.
\]

We now sum over \( Q \) concluding that

\[
\|\delta^{d-2} S_\theta \|_{L^q(Q(1))} \lesssim \delta^{C\varepsilon^2} \cdot A_\varepsilon \sum_Q (\delta^{2d-3} n_W(Q) \cdot \delta^{2d-3} n_B(Q)) \frac{\delta}{\varepsilon}
\]

\[
\lesssim \delta^{C\varepsilon^2} \cdot A_\varepsilon (\sum_Q \delta^{2d-3} n_W(Q)) \frac{\delta}{\varepsilon} (\sum_Q \delta^{2d-3} n_B(Q)) \frac{\delta}{\varepsilon}
\]

\[
\lesssim \delta^{C\varepsilon^2} \cdot A_\varepsilon (C \log \frac{1}{\delta} \cdot \delta^{2d-3} |W| \cdot (\delta^{2d-3} |B|) \frac{\delta}{\varepsilon}.
\]

The three inequalities followed respectively from (43), from Hölder’s inequality (recall that \( q \geq p \)) and from properties 1 and 2 of Lemma A.1. The factor \( \delta^{C\varepsilon^2} \cdot (C \log \frac{1}{\delta} \cdot \delta^{2d-3} |B|) \) is small for small \( \delta \); the result now follows by combining the last inequality with the preceding bound for \( \|T_\theta \|_{L^q(Q)} \).

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 $\delta$, and a set $Y \subset \mathcal{L}$, define

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

where $Y_{\delta}$ is the $\delta$-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 $\sigma$. 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 $\|\chi_{T_c Y}\|_{L^p(L^r)} \approx \sigma^{-\frac{d-2}{p}} - \frac{d}{r} \|\chi_Y\|_{L^p(L^r)}$.

(b) If $Y \subset \mathcal{L}_c$ then $E_{\delta}^{p,r}(T_c Y) \lesssim \sigma^{-\frac{d-2}{p}} - \frac{d}{r} E_{\sigma \delta}^{p,r}(Y)$.

Property (a) is proved as follows: within $c$, $T_c$ expands distances along $S^{d-2}$ by a factor $\sigma^{-1}$ (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

$$\sigma^{-1} \times \ldots \times \sigma^{-1} \times \sigma^{-2},$$

i.e. by $\sigma^{-d}$. Thus $L^p(L^r)$ norms expand by $\sigma^{-\frac{d-2}{p}} - \frac{d}{r}$. Also property (b) follows from property (a) by observing that $T_c$ maps the $C\sigma \delta$-neighborhood of $Y \subset \mathcal{L}_c$ onto a set which includes the $\delta$-neighborhood of $T_c Y$.

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

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

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 $\delta$-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

$$\min(X^*(\chi_{c_1 \cap Z})(x), X^*(\chi_{c_2 \cap Z})(x)) \geq \mu.$$
Then
\[ |S| \lesssim \delta^{-\varepsilon} \mu^{-q} E^{p,r}_{\delta}(Z)^q \]
for any fixed \( \varepsilon > 0 \), provided \( q \leq 3 \), \( q \geq p \geq r \), and \( \frac{1}{q} \geq \frac{2}{r} - 1 \), \( \frac{2}{q} \leq \frac{d}{d-1} \).

**Proof.** We first make a couple of reductions. First, it suffices to prove the lemma with 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 hypothesis 1. It then also suffices to prove Lemma A.5 when \( p = r \), since \( E^{p,r}_{\delta}(Z) \) 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 hypothesis 2 are independent of \( x \).

Now define \( Z_i = Z \cap Lc_i \), let \( W \) and \( B \) be maximal \( \delta \)-separated subsets of \( Z_1 \) and \( Z_2 \) respectively and (for each \( w \in W \)) let \( D_w \) be the \( \delta \)-disc in \( L \) centered at \( w \). Then
\[ X^*(\chi_{Z_1 \cap Lc_1}(x)) \lesssim \sum_w \delta^{d-2} \chi_w = \Phi_W, \]
where on the right side \( \chi_w \) is the characteristic function of the \( \delta \)-neighborhood of the line \( w \). So
\[ X^*(\chi_{Z_1}) \lesssim \sum_w \delta^{d-2} \chi_w = \Phi_W, \]
where \( \Phi_W \) is as in Lemma A.1. Accordingly \( \min(X^*(\chi_{c_1 \cap Z})(x), X^*(\chi_{c_2 \cap Z}(x))) \lesssim \min(\Phi_B, \Phi_W) \leq \Psi_\theta \). The result now follows from Lemma A.4 using Tchebyshev’s inequality and that
\[ (\delta^{2d-3}|B|^{2d-3}|W|)^{\frac{1}{2d}} \lesssim \mathcal{E}^{p,r}_{\delta}(Z). \]

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 \( q \leq 3 \), \( q \geq p \geq r \), \( \frac{1}{q} \geq \frac{2}{r} - 1 \), \( \frac{2}{q} \leq \frac{d}{d-1} \), and
\[ \frac{d-2}{p} + \frac{d}{r} \leq d - 2 + \frac{d}{q}. \]
Let \( Y \subset L \). Then
\[ \|X^*\chi_Y\|_{L^q(Q(1))} \lesssim \delta^{-\varepsilon} \mathcal{E}^{p,r}_{\delta}(Y). \]

**Proof.** A standard argument shows that it will suffice to prove the corresponding distributional estimate
\[ \{x \in Q(1) : X^*\chi_Y(x) \geq \lambda\} \lesssim \delta^{-\varepsilon} \lambda^{-q} \mathcal{E}^{p,r}_{\delta}(Y)^q \]
in the case where \( \lambda \) is bounded below by a high power of \( \delta \), say
\[ \lambda \geq \delta^{\frac{d-2}{2}(d-2)} \]
where \( B \) is a large constant depending on \( d \). This is because of the \( \delta^{-\varepsilon} \) factors and the fact that very small values of \( \lambda \) clearly make a negligible contribution.
To prove (46), let \( A = \{ x \in Q(1) : X^*(x) \geq \lambda \} \) and define \( A_\sigma \) to be all points \( x \) with the property that there are two \( \sigma \)-caps \( c_1 \) and \( c_2 \) on \( S^{d-2} \) whose distance apart is between \( \sigma \) and \( C\sigma \) and such that \( X^*(x) \geq C^{-1}\delta^\varepsilon \lambda \) for \( i = 1, 2 \). We claim that \( \cup_i A_\sigma \supseteq A \); the union is over dyadic \( \sigma \geq \delta^B \).

Namely, if \( x \in A \), then take the smallest \( \sigma \) such that \( X^*(x) \geq (C\sigma)^{\varepsilon} \lambda \) for some cap \( c \) of width \( C\sigma \). (The lower bound on \( \lambda \) implies that then \( \sigma \geq \delta^B \).) Consider a covering of \( \sigma \) by caps \( c_i \) of width \( \sigma \). The minimality of \( \sigma \) implies that \( X^*(x) \) is small compared with \( X^*(x) \) 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 \( X^*(x) \geq \sigma^{\varepsilon} \lambda \) for \( i = 1 \) and 2. This implies the claim.

By pigeonholing we may now choose \( \sigma \) so that

\[
|A_\sigma| \geq \delta^\varepsilon |A|.
\]

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

\[
A_\sigma = \cup_i A_{\sigma i}^c
\]

where \( A_{\sigma i}^c \) is the set of \( x \) for which the two \( \sigma \)-caps \( c_1 \) and \( c_2 \) in the definition of \( A_\sigma \) 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 \( S = T_{c_i}(A_{\sigma i}^c) \). Formula (44) shows that 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

\[
|T_{c_i}(A_{\sigma i}^c)| \lesssim \delta^{-\varepsilon} (\sigma^{-(d-2)} \lambda)^{-q} \mathcal{E}_{\sigma}^{p,r}(T_{c_i}(\mathcal{L}_{c_i} \cap Y)) \lesssim (\sigma^{-(d-2)} \lambda)^{-q} \sigma^{-q(d-\frac{d}{p} + 2)} \mathcal{E}_{\sigma}^{p,r}(\mathcal{L}_{c_i} \cap Y)^q
\]

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

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

which implies that

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

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

\[
|A_\sigma| \lesssim \delta^{-\varepsilon} \mathcal{E}_{\sigma\delta}^{p,r}(Y)^q.
\]

We now use (47) and the fact that \( \mathcal{E}_{\varepsilon}^{p,r} \) increases with \( \varepsilon \). The result follows.
**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 $\frac{d}{3}$ if $d = 3$ and $\frac{d+1}{2}$ if $d \geq 4$. The following relations on the dual exponents will hold:

\[
\begin{align*}
\frac{p'}{r'} & \leq \frac{d}{d-1}, \\
\frac{d-2}{q'} + \frac{d}{r'} & \leq d - 2 + \frac{d}{p'}, \\
\frac{1}{p'} & \geq \frac{2}{r'} - 1, \\
r' & \leq q' \leq p' \leq 3.
\end{align*}
\]

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

\[
\|X^* \chi Y\|_{L^p(Q(1))} \lesssim \delta^{1-\varepsilon} \mathcal{E}^{q',r'}_\delta(Y).
\]

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 $\ell^\delta$ is the tube of width $\delta$ with axis $\ell$.

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 $\delta$-discs and since values of $X_\delta f$ which are less than a high power of $\delta$ 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

\[
\int_{\mathcal{L}} g X_\delta \chi E \gtrsim \delta^{1-\varepsilon} \|X_\delta f\|_{L^q(L^r)}
\]

and such that $g$ has the special form

\[
g = \mu \chi_Y
\]

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

Letting $g$ be as in (49), we have

\[
\|X_\delta f\|_{L^q(L^r)} \gtrsim \delta^{-\varepsilon} \int_{\mathcal{L}} \mu \chi_Y X f
\]

\[
\lesssim \delta^{-\varepsilon} \int_{\mathcal{L}} \mu \chi_{\tilde{Y}} X f
\]

\[
= \delta^{-\varepsilon} \int_{\mathbb{R}^d} X^* (\mu \chi_{\tilde{Y}}) f.
\]
Now apply (48) to $\chi_{\tilde{Y}}$ and use Hölder’s inequality, obtaining
\begin{equation}
\|X_\delta f\|_{L^q(r)} \lesssim \delta^{-\epsilon} \|f\|_p
\end{equation}
since $\mu E_{q', r'}(\tilde{Y}) \lesssim 1$.

It remains to trade $\epsilon$ 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 $\phi$ is an appropriately chosen $C_0^\infty$ 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 $\sum_j 2^{nj} \|f_j\|_p \lesssim \|f\|_{p\epsilon}$ for small $\eta$. It follows using the smoothing effect of $\phi_j$ 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 $\epsilon$ to the terms in the series.

California Institute of Technology, Pasadena, CA

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, Geom. Funct. Anal. 1 (1991), 147–187.
[3] , Estimates for cone multipliers, in Geometric Aspects of Functional Analysis (Israel, 1992–1994), Operator Theory Adv. Appl. 77, 41–60, Birkhäuser Basel, 1995.
[4] M. Christ, Estimates for the $k$-plane transform, Indiana Univ. Math. J. 33 (1984), 891–910.
[5] S. W. Drury, $L^p$ estimates for the X-ray transform, Illinois J. Math. 27 (1983), 125–129.
[6] M. B. Erdogan, Mixed norm estimates for the X-ray transform restricted to a rigid well-curved line complex in $R^4$ and $R^5$, preprint.
[7] D. Foschi and S. Klainerman, Homogeneous $L^2$ bilinear estimates for wave equations, Les Ann. Sci. l’Ecole Norm. Sup. 33 (2000), 211–274.
[8] A. Greenleaf and A. Seeger, Fourier integral operators with fold singularities, J. Reine Angew. Math. 455 (1994), 35–56.
[9] A. Greenleaf, A. Seeger, and S. Wainger, On estimates for generalized X-ray transforms and integrals over curves in $R^4$, Proc. Amer. Math. Soc., to appear.
[10] A. Greenleaf and G. Uhlmann, Composition of some singular Fourier integral operators and estimates for restricted X-ray transforms, Ann. Inst. Fourier (Grenoble) 40 (1990), 443–466.
[11] S. Klainerman and M. Machedon, Space-time estimates for null forms and the local existence theorem, Comm. Pure Appl. Math. 46 (1993), 1221–1268.
[12] G. Mockenhaupt, A note on the cone multiplier, Proc. Amer. Math. Soc. 117 (1993), 145–152.
[13] G. Mockenhaupt, A. Seeger, and C. Sogge, Wave front sets and Bourgain’s circular maximal theorem, Ann. of Math. 134 (1992), 207–218.
[14] D. M. Oberlin and E. M. Stein, Mapping properties of the Radon transform, Indiana Univ. Math. J. 31 (1982), 641–650.
[15] T. Tao, The weak type endpoint Bochner-Riesz conjecture and related topics, *Indiana Univ. Math. J.* **47** (1998), 1097–1124.

[16] T. Tao and A. Vargas, A bilinear approach to cone multipliers, I and II, *Geom. Funct. Anal.* **10** (2000), 185–215, 216–258.

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

[18] 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.

[19] T. Wolff, An improved bound for Kakeya type maximal functions, *Rev. Math. Iberoamericana* **11** (1995), 651–674.

[20] ———, A mixed norm estimate for the X-ray transform, *Rev. Math. Iberoamericana* **14** (1998), 561–601.

[21] ———, Decay of circular means of Fourier transforms of measures, *Internat. Math. Res. Not.* **10** (1999), 547–567.

(Received April 9, 1999)