Sparse Bounds in Harmonic Analysis and Semiperiodic Estimates

by Alexander Barron

May 2019
This dissertation by Alex Barron is accepted in its present form by the Department of Mathematics as satisfying the dissertation requirement for the degree of Doctor of Philosophy

Date_________ ________________________________

Jill Pipher, Ph.D., Advisor

Recommended to the Graduate Council

Date_________ ________________________________

Benoit Pausader, Ph.D., Reader

Date_________ ________________________________

Sergei Treil, Ph.D., Reader

Approved by the Graduate Council

Date_________ ________________________________

Andrew G. Campbell, Dean of the Graduate School
Vitae

Alexander Barron graduated from Colby College with a B.A. in Mathematics in May 2013 (summa cum laude, Phi Beta Kappa). He also earned an M.S. in mathematics from Brown University in May 2016. At Brown he was a teaching assistant and instructor for multiple sections of Calculus I and II, and also for multivariable calculus. He has given research talks at various conferences and institutions, including Georgia Tech, the University of Virginia, the University of New Mexico, and the American Institute of Math (AIM). Starting in August 2019 he will be serving as a Doob Visiting Assistant Professor at the University of Illinois Urbana-Champaign.

Below is a list of publications and preprints by the author.


Preface

Part I of this thesis studies some results related to sparse bounds in harmonic analysis. We provide a brief overview of sparse bounds in Chapter 1. In Chapter 2 we study sparse bounds for certain ‘rough’ bilinear singular integrals with limited smoothness assumptions. In Chapters 3 and 4 we study the extent to which the sparse theory applies in the setting of operators with a bi- or multi-parameter structure. In particular in Chapter 4 we show that the simplest one-parameter sparse bound (for the dyadic Hardy-Littlewood maximal function) does not directly generalize to the bi-parameter setting. Chapter 2 is based on [1]. Chapter 3 is based on [2] and is joint work with Jill Pipher. Chapter 4 is based on [3] and is joint work with José Conde Alonso, Yumeng Ou, and Guillermo Rey.

In Part II we study certain global-in-time Strichartz-type estimates for solutions to the linear semiperiodic Schrödinger equation (on manifolds of the type $\mathbb{R}^n \times \mathbb{T}^d$). In Chapter 5 we survey the related theory on $\mathbb{R}^n$ and $\mathbb{T}^d$. In Chapter 6 we discuss some earlier results on $\mathbb{R} \times \mathbb{T}$ and then prove a certain generalization on $\mathbb{R}^n \times \mathbb{T}^d$ (see Chapter 6 for a precise statement). A large part of Chapter 6 is based on [69].
Acknowledgments

This thesis would not have been possible without the generous help and support of my advisor Jill Pipher. I also benefited greatly from discussions with other faculty at Brown, including José Conde Alonso, Francesco Di Plinio, Benoit Pausader and Sergei Treil. The same can be said of all the students in analysis that I’ve had the privilege of talking to during my time at Brown: Amalia Culiuc, Milen Ivanov, Spyridon Kakaroumpas, Linhan Li, and Yumeng Ou.
# Contents

I Sparse Bounds in Harmonic Analysis .................................................. 5

1 A Brief Overview of Sparse Bounds ................................................. 7

2 Sparse Bounds for Rough Bilinear Operators .................................. 11

2.0.1 Structure of the Chapter ....................................................... 15

2.0.2 Notation and Definitions ....................................................... 15

2.1 Abstract Sparse Theorem in Multilinear Setting .......................... 16

2.1.1 Remark on Truncations ....................................................... 18

2.1.2 The Abstract Theorem ....................................................... 19

2.1.3 Some Remarks on Theorem 2.2 ........................................... 21

2.2 Multilinear Calderón-Zygmund Operators .................................. 23

2.2.1 Single-Scale Estimates ....................................................... 25

2.2.2 Cancellation Estimates ....................................................... 28

2.3 Rough Bilinear Singular Integrals .............................................. 33

2.3.1 Interpolation Lemmas ....................................................... 34

2.3.2 The Sparse Bound .............................................................. 38

2.4 Weighted Estimates ................................................................. 40

2.4.1 Single Weight - Proof of Corollary 2.2 ................................ 40

2.4.2 Multiple Weights ............................................................... 42

2.5 Proof of Theorem The Abstract Theorem .................................. 44
2.5.1 Construction of the Sparse Collection ................................................. 44
2.5.2 Multiplication Operators ................................................................. 49

3 Sparse Bounds in the Multi-Parameter Setting ............................................. 53

3.1 Some Difficulties ................................................................................. 54
3.2 An Approach Using Square Functions .................................................... 56
3.2.1 Remarks on Theorem 3.1 ................................................................. 58
3.3 The Lacey-Ou Example ......................................................................... 60
3.4 The Bi-Parameter Martingale Transform ............................................... 62
3.4.1 The Geometric Proof of Theorem 3.1 .................................................. 63
3.5 Bi-Parameter Dyadic Shifts and Singular Integrals ................................. 68
3.5.1 Definitions. ..................................................................................... 68
3.5.2 The Sparse Bound ........................................................................... 69
3.5.3 Bi-Parameter Singular Integrals ......................................................... 72
3.6 Weighted Estimates for Bi-Parameter Martingale Transform .................. 73
3.7 Weighted Estimates for the Shifted Square Function ............................... 78
3.7.1 Preliminary Results ........................................................................ 79
3.7.2 The Sparse Bound ........................................................................... 82

4 A Counterexample for the Strong Maximal Function .................................. 93

4.1 Lower bounds for $M_S$ and upper bounds for sparse forms .................. 96
4.1.1 Lower bound for $M_S$ ................................................................. 99
4.1.2 Upper bound for sparse forms ....................................................... 100
4.2 Construction of the extremal sets ......................................................... 104
4.2.1 Construction of $\mathcal{P}$ and the proof of Theorem 4.2 .................... 105
4.2.2 Construction of $\mathcal{Z}$ ................................................................ 107
4.2.3 Proof of lemma 4.5 and (Z.5) ......................................................... 109
4.3 Extensions and open questions ............................................................... 112
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.1 Failure of sparse bound for bi-parameter martingale transform</td>
<td>114</td>
</tr>
<tr>
<td>4.3.2 Failure of sparse bound in higher dimensions</td>
<td>116</td>
</tr>
<tr>
<td>4.3.3 Other sparse forms</td>
<td>119</td>
</tr>
<tr>
<td>II Semiperiodic Strichartz Estimates</td>
<td>123</td>
</tr>
<tr>
<td>5 Strichartz Estimates on $\mathbb{R}^n$ and $\mathbb{T}^d$</td>
<td>125</td>
</tr>
<tr>
<td>5.0.1 Strichartz Estimates on Euclidean Space</td>
<td>125</td>
</tr>
<tr>
<td>5.0.2 Strichartz Estimates on the Torus</td>
<td>127</td>
</tr>
<tr>
<td>6 The Semiperiodic Case</td>
<td>131</td>
</tr>
<tr>
<td>6.1 Local Estimates on $\mathbb{R} \times \mathbb{T}$</td>
<td>132</td>
</tr>
<tr>
<td>6.2 Global Estimates on $\mathbb{R}^n \times \mathbb{T}^d$</td>
<td>137</td>
</tr>
<tr>
<td>6.3 Preliminary Results</td>
<td>142</td>
</tr>
<tr>
<td>6.4 The Decoupling Argument</td>
<td>147</td>
</tr>
<tr>
<td>6.4.1 Main Lemmas</td>
<td>148</td>
</tr>
<tr>
<td>6.4.2 Main Argument</td>
<td>150</td>
</tr>
<tr>
<td>6.4.3 Proof of Decoupling Lemma 6.7</td>
<td>152</td>
</tr>
<tr>
<td>6.5 The $\epsilon$-Removal Argument</td>
<td>156</td>
</tr>
<tr>
<td>6.5.1 The Local-In-Time Case</td>
<td>156</td>
</tr>
<tr>
<td>6.5.2 The Global Case</td>
<td>160</td>
</tr>
<tr>
<td>6.6 Some Applications</td>
<td>167</td>
</tr>
<tr>
<td>6.6.1 Function Spaces</td>
<td>167</td>
</tr>
<tr>
<td>6.6.2 The Quintic Equation on $\mathbb{R} \times \mathbb{T}$</td>
<td>169</td>
</tr>
<tr>
<td>6.6.3 The Cubic Equation on $\mathbb{R}^2 \times \mathbb{T}$</td>
<td>176</td>
</tr>
<tr>
<td>6.7 Additional Remarks</td>
<td>178</td>
</tr>
</tbody>
</table>
Part I

Sparse Bounds in Harmonic Analysis
Chapter 1

A Brief Overview of Sparse Bounds

Over the past several years there has been a lot of activity concerning the sparse domination of various operators arising in harmonic analysis. Given an operator $T$ and suitable test functions $f, g$ on $\mathbb{R}^n$ one typically shows that there exists a sparse collection of cubes $S$ such that

$$Tf(x) \lesssim \sum_{Q \in S} \langle f \rangle_{Q} 1_{Q}(x)$$

or such that

$$|\langle Tf, g \rangle| \leq C \sum_{Q \in S} \langle f \rangle_{r, Q} \langle g \rangle_{s, Q} |Q|$$

(1.1)

for some $1 \leq r, s < \infty$, where $\langle f \rangle_{r, Q} = \left( \frac{1}{|Q|} \int_{Q} |f(x)|^{r} \, dx \right)^{1/r}$ and $\langle f \rangle_{Q} = \langle f \rangle_{1, Q}$ (if $T$ is multilinear one needs to work with a more general multilinear version of this estimate, see below). The sparse collection $S$ is almost disjoint in the following sense: there exists $\eta \in (0, 1)$ such that for each $Q \in S$ there is a subset $E_{Q} \subset Q$ such that $|E_{Q}| \geq \eta |Q|$, and moreover $E_{Q} \cap E_{Q'} = \emptyset$ for distinct $Q, Q' \in S$.

The sparse theory offers a new approach to the analysis of many classical (and non-classical) operators in harmonic analysis, and indeed most of the major results in the area were proved within the last five years. We now know that bounds of the type (1.1) hold for a wide variety of operators: smooth and rough Calderón-Zygmund operators, singular integrals along curves, the
spherical maximal function, the variational Carleson operator, the bilinear Hilbert transform, certain Bochner-Riesz multipliers, pseudodifferential operators, singular operators associated to elliptic operators on manifolds, and various discrete analogues of classical operators (this list is far from exhaustive, see for example [53], [48], [4], [6], [19], [1] and the references therein).

Sparse bounds typically recover the full range of known strong and weak-type $L^p$ bounds for $T$, and also yield vector-valued estimates as a corollary. Moreover, a bound of type (1.1) implies a range of quantitative weighted estimates for $T$ with respect to Muckenhoupt and reverse Hölder weight classes, and these bounds are often sharp in terms of the weight characteristic (e.g. the sparse bound for Calderón-Zygmund operators yields a simple proof of the sharp weighted bounds originally proved in [41]). Note that the operators $T$ listed above are oscillatory or non-local, yet the sparse bound (1.1) gives control of $T$ by an operator that is positive and localized without losing information about the mapping properties of $T$. The sparse arguments also generalize to the probabilistic setting where $T$ is a discrete or continuous martingale transform, as long as one replaces averages over cubes with conditional expectations adapted to iterated stopping times [27].

In subsequent chapters we will work with positive sparse forms. Let $S$ be a sparse collection of cubes in $\mathbb{R}^d$ and suppose $\vec{p} = (p_1, p_2, \ldots, p_{m+1})$ is an $(m+1)$-tuple of Lebesgue exponents. We define the form $PSF_{S}^{\vec{p}}(f_1, \ldots, f_{m+1}) := \sum_{Q \in S} |Q| \prod_{i=1}^{m+1} (f_i)_{p_i,Q}$, where

$$(f)_{q,Q} = |Q|^{-1/q} \|f1_Q\|_q.$$ 

These operators were initially studied in [6], [24], [25], motivated by earlier pointwise estimates in the sparse domination theory (see [48], [53], [54], and [55] for some examples). It is important to note that there are examples of operators that are bounded by sparse forms, but which may not admit a pointwise sparse bound (for example, the bilinear Hilbert transform [24]). It is therefore typically easier to prove sparse-form estimates in place of pointwise sparse bounds,
and in practice sparse form bounds are not much weaker than pointwise estimates. Working with sparse forms also helps to avoid certain technical obstacles (for example, one can avoid the reliance on maximal truncation estimates present in [48], [54]). We refer the reader to [19], [47], and [51] for further discussion and examples of recent developments.

In the next chapter we study sparse-form bounds for a certain class of rough bilinear singular integrals. In Chapters 3 and 4 we turn our attention to sparse bounds in the setting of multi-parameter operators, where very little is known.
Chapter 2

Sparse Bounds for Rough Bilinear Operators

The study of rough singular integral operators dates back to Calderón and Zygmund’s classic papers [10] and [11]. In [11] the authors proved that if $\Omega \in L \log L(S^{n-1})$ with $\int_{S^{n-1}} \Omega \ d\sigma = 0$, then the operator

$$R_\Omega f(x) := \text{p.v.} \int_{\mathbb{R}^n} \frac{\Omega(y/|y|)}{|y|^n} f(x - y)dy$$

is bounded on $L^p(\mathbb{R}^n)$ for $1 < p < \infty$. Hoffman [38] and Christ and Rubio de Francia [15], [16] proved such operators are bounded from $L^1 \to L^{1,\infty}$ for small dimensions, and weak-(1,1) boundedness in arbitrary dimensions was later proved by Seeger [63] and Tao [67] (in a more general setting). The weighted theory for such operators was developed during this time as well; for some examples, see the work by Duoandikoetxea [29] and Watson [68]. More recently, Conde-Alonso, Culiuc, Di Plinio, and Ou [19] have shown that the bilinear form associated to a rough operator $R_\Omega$ can be bounded by positive sparse forms, proving quantitative weighted estimates for $R_\Omega$ as an easy corollary. Also see the recent work by Hytönen, Roncal, and Tapiola [44], where the authors establish quantitative estimates using a different method.

We are interested in the bilinear analogues of the operators $R_\Omega$. The study of these operators originates with work by Coifman and Meyer [17]. Suppose $\Omega \in L^q(S^{2d-1})$ for some $q > 1$ with
\[ \int_{S^{2d-1}} \Omega \, d\sigma = 0, \] and define the rough bilinear operator

\[ T_\Omega(f_1, f_2)(x) = \text{p.v.} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f_1(x - y_1) f_2(x - y_2) \frac{\Omega((y_1, y_2)/((y_1, y_2)))}{|(y_1, y_2)|^{2d}} \, dy_1 dy_2. \]  

(2.1)

Grafakos, He, and Honzík [34] have proved using a wavelet decomposition that if \( \Omega \in L^\infty(S^{2d-1}) \) then

\[ \|T_\Omega\|_{L^{p_1}(\mathbb{R}^d) \times L^{p_2}(\mathbb{R}^d) \to L^p(\mathbb{R}^d)} < \infty \]

when \( 1 < p_1, p_2 < \infty \) and \( \frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p} \). Also see [34] for references to earlier work. In this chapter we develop a weighted theory for the rough bilinear operators \( T_\Omega \), using a multilinear generalization of the sparse domination theory from [19] along with the results by Grafakos, He, and Honzík. The main result is the following theorem.

**Theorem 2.1.** Suppose \( T_\Omega \) is the rough bilinear singular integral operator defined in (2.1), with \( \Omega \in L^\infty(S^{2d-1}) \) and \( \int_{S^{2d-1}} \Omega = 0 \). Then for any \( 1 < p < \infty \), there is a constant \( C_p > 0 \) so that

\[ |\langle T_\Omega(f_1, f_2), f_3 \rangle| \leq C_p \|\Omega\|_{L^\infty(S^{2d-1})} \sup_S PSF_{S}^{(p,p,p)}(f_1, f_2, f_3). \]

Here the supremum is taken over all sparse collections \( S \) with some fixed sparsity constant \( \eta \) that does not depend on the functions. This theorem is a consequence of a more general multilinear sparse domination result, which is stated in Section 2. As an application of Theorem 2.1 we derive weighted estimates for \( T_\Omega \). Recall the \( A_p \) class of weights \( w \), where \( w \in A_p \) for \( 1 < p < \infty \) if \( w > 0, w \in L^1_{\text{loc}}, \) and

\[ [w]_{A_p} := \sup_{Q \subset \mathbb{R}^d, \text{cubes}} \left( \frac{1}{|Q|} \int_Q w \, dx \right) \left( \frac{1}{|Q|} \int_Q w^{-\frac{1}{p-1}} \, dx \right)^{p-1} < \infty. \]

Below we write \( L^p(w) \) for the space \( L^p \) with measure \( w(x)dx \).

**Corollary 2.1.** Fix \( 1 < p_1, p_2 < \infty \) and \( \frac{1}{2} < p < \infty \) such that \( \frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p} \). Then for all weights \( (w_1^{p_1}, w_2^{p_2}) \) in \( (A_{p_1}, A_{p_2}) \), there is a constant \( C \) depending on \([w_1^{p_1}]_{A_{p_1}}, [w_2^{p_2}]_{A_{p_2}}, d, p_1, p_2 \) such
that
\[
\|T_\Omega(f_1, f_2)\|_{L^p(w_1^p w_2^p)} \leq C \|f_1\|_{L^p_1(w_1^{p_1})} \|f_2\|_{L^p_2(w_2^{p_2})}
\]
for all \( f_i \in L^p_i(w_i^{p_i}) \).

These estimates were originally proved by Cruz-Uribe and Naibo in [23] with a different technique. Our proof uses the sparse domination from Theorem 2.1 along with methods from [24], [55], [56] and extrapolation. Note the following special case of Corollary 2.1 in the single-weight case.

**Corollary 2.2.** Suppose \( 1 < q < \infty \) and \( \Omega \in L^\infty(S^{2d-1}) \) with \( \int_{S^{2d-1}} \Omega \, d\sigma = 0 \). Then if \( w \in A_q \), there is a constant \( C = C(w, q, \Omega) \) such that
\[
\|T_\Omega(f, g)\|_{L^{q/2}(w)} \leq C \|f\|_{L^q(w)} \|g\|_{L^q(w)}
\]
for all \( f, g \in L^p(w) \).

We provide a separate proof of this corollary in Section 5.1 that indicates how to track the dependence of \( C \) on \([w]_{A_p}\). The proof of Corollary 2.2 is again a consequence of sparse domination and extrapolation.

We can also prove weighted estimates with respect to the more general multilinear Muckenhoupt classes. In particular, suppose \( \sum_{i=1}^3 \frac{1}{q_i} = 1 \) with \( 1 < q_1, q_2, q_3 < \infty \) and let \( v_1, v_2, v_3 \) be strictly positive functions such that
\[
\prod_{i=1}^3 v_i^{\frac{1}{q_i}} = 1.
\]
Define
\[
[\vec{v}]_A^{\vec{q}} := \sup_Q \prod_{i=1}^3 \left( \frac{1}{|Q|} \int_\Omega v_i^{\frac{p_i}{q_i}} \right)^{\frac{1}{p_i} - \frac{1}{q_i}}
\]
for any tuple \( \vec{p} = (p_1, p_2, p_3) \) with \( 1 \leq p_i < q_i < \infty \) for \( i = 1, 2, 3 \). Notice that we assume \( v_i > 0 \) for each \( i \), but we do not assume that \( v_i \in L^1_{loc} \). This multilinear class was originally introduced in [56] with \( \vec{p} = (1, 1, 1) \), and used in [24] for more general \( \vec{p} \). In the case when \( \vec{p} = (1, 1, 1) \), it
is the natural weight class associated to the maximal operator

\[
\mathcal{M}(f)(x) := \sup_{x \in Q} \prod_{i=1}^{2} \frac{1}{|Q|} \int_{Q} |f_i(y_i)| dy_i.
\]

We will prove the following in Section 5.2.

**Corollary 2.3.** Suppose the tuples \( \vec{p}, \vec{q} \) and the functions \( v_1, v_2, v_3 \) are defined as above, with \( p_i > 1 \) for \( i = 1, 2, 3 \). Also let

\[
\sigma = v_3^{-q_3/q_3} = v_1^{-q_2/q_1+q_2} v_2^{-q_1/q_1+q_2},
\]

where \( q_3' \) is the conjugate of \( q_3 \). Then if \( [v]_{A_q^{p,q}} < \infty \) and \( \Omega \in L^\infty(S^{2d-1}) \) with \( \int_{S^{2d-1}} \Omega = 0 \), we have

\[
\|T_\Omega(f_1, f_2)\|_{L^{q_3'}(\sigma)} \leq C_{\Omega, d, p_i, q_i} [v]_{A_q^{p,q}}^{\max\{q_i/p_i\}} \|f_1\|_{L^{q_1}(v_1)} \|f_2\|_{L^{q_2}(v_2)}
\]

for all \( f_i \in L^{q_i}(v_i) \), \( i = 1, 2 \).

We will see in Section 5.2 below that Corollary 2.1 and Corollary 2.3 can both be deduced from the same lemma, which is a consequence of the sparse domination. However, the class \( [v]_{A_q^{p,q}} \) is in general strictly larger than \( A_{q_1} \times A_{q_2} \) since, for example, \( v_1 \) and \( v_2 \) do not have to be in \( L^1_{loc} \) (see [56] for some examples when \( \vec{p} = (1, 1, 1) \)).

**Sample Application.** We note that the Calderón commutator

\[
\mathcal{C}(a, f)(x) = \text{p.v.} \int_{\mathbb{R}} \frac{A(x) - A(y)}{(x - y)^2} f(y) dy,
\]

where \( a \) is the derivative of \( A \), is an example of the rough bilinear operators considered in this paper [17]. Let \( e(t) = 1 \) if \( t > 0 \) and \( e(t) = 0 \) if \( t < 0 \). Then we can write this operator as

\[
\text{p.v.} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x - y, x - z) f(y) a(z) dy dz
\]

where

\[
K(x - y, x - z) = \mathcal{M}(f)(x) := \sup_{x \in Q} \prod_{i=1}^{2} \frac{1}{|Q|} \int_{Q} |f_i(y_i)| dy_i.
\]
with \( K(y, z) = \frac{\phi(z) - \phi(z-y)}{y^d} =: \Omega((y,z)/|(y,z)|) \). Moreover, \( \Omega \) is odd and bounded, so Theorem 2.1 and the weighted corollaries all apply to \( C(a,f) \).

2.0.1 Structure of the Chapter

The proof of Theorem 2.1 is broken up into two parts. We begin by formulating an abstract sparse domination theorem in the multilinear setting (Theorem 2.2 below), a result that generalizes Theorem C in [19] by Conde-Alonso, Culiuc, Di Plinio, and Ou. The proof of this theorem is similar to their result, so it is deferred to Section 2.5. In the second part of the paper we use the deep results of Grafakos, He, and Honzík to show that the assumptions of Theorem 2.2 are satisfied by the rough bilinear operators \( T_\Omega \). Along the way we also prove a sparse domination result for multilinear Calderón-Zygmund operators with a standard smoothness assumption, as defined by Grafakos and Torres in [36]. This is the content of Theorem 2.3 below. This sparse domination result for multilinear Calderón-Zygmund operators uses different methods than the recent paper by K. Li [57], and in particular avoids reliance on maximal truncations. As a corollary we can partially recover the weighted estimates from [55]. Finally, in Section 2.4 we prove the weighted estimates outlined in the corollaries above.

Theorem 2.2 is independently interesting, and due to its generality we expect it to be useful for analyzing other multilinear operators in the future. We also note that Theorem 2.1 only applies when \( \Omega \in L^\infty(S^{2d-1}) \), since we need some analogue of the classical Calderón-Zygmund size condition. Sparse domination when \( \Omega \in L^q(S^{2d-1}) \) for other values of \( q \), or when \( \Omega \) is in some Orlicz-Lorentz space, is still an open problem. In fact, the full range of boundedness of \( T_\Omega \) when \( \Omega \in L^q(S^{2d-1}) \) for \( 2 < q < \infty \) is not known. It is also unknown whether \( T_\Omega \) is bounded anywhere for \( q < 2 \). See [34] for more details.

2.0.2 Notation and Definitions

Given a dyadic cube \( L \), we let \( s_L = \log_2(\text{length}(L)) \) and let \( \hat{L} \) be the \( 2^5 \)-fold dilate of \( L \). Throughout the paper we fix a dyadic lattice \( D \) in \( \mathbb{R}^d \). A collection of disjoint cubes \( \mathcal{P} \subset D \) will

15
be called a *stopping collection* with top $Q$ if its elements are contained in $3Q$ and satisfy the following separation properties:

1. If $L, R \in \mathcal{P}$ and $|s_L - s_R| \geq 8$ then $7L \cap 7R = \emptyset$

2. $\bigcup_{L \in \mathcal{P}, 3L \cap 2Q \neq \emptyset} 9L \subset \bigcup_{L \in \mathcal{P}} L$.

This definition is taken from [19] (the particular constants here are chosen for technical reasons related to the proof of Theorem 2.2 below).

Throughout the paper we use $c_\alpha$ to represent a positive constant depending on the parameter $\alpha$ that may change line to line. We often write $A \lesssim B$ to mean $A \leq cB$, where $c$ is a positive constant depending on the dimension $d$, the multilinearity term $m$, or relevant exponents. For $E \subset \mathbb{R}^d$ we let $|E|$ denote its Lebesgue measure and $1_E$ its indicator function. Finally, we will use $M_p(f)(x)$ to denote the $p$-th Hardy-Littlewood maximal function

$$M_p(f)(x) = \sup_{x \in Q} \left( \frac{1}{|Q|} \int_Q |f(y)|^p dy \right)^{1/p}$$

(here the supremum is taken over cubes $Q \subset \mathbb{R}^d$ containing $x$). Recall that $M_p$ is bounded on $L^r(\mathbb{R}^d)$ when $r > p$. We also write $M_p^w(f)$ to denote the $p$-th maximal function associated to a weight $w$. This operator satisfies the same boundedness properties as the standard maximal operator when $w$ is doubling (and in particular when $w \in A_q$ for some $q$).

### 2.1 Abstract Sparse Theorem in Multilinear Setting

In this section we formulate an abstract sparse domination result which we will apply in Section 2.3 to prove Theorem 1. Let $T$ be a bounded $m$-linear operator mapping $L^{r_1} \times \ldots \times L^{r_m} \to L^\alpha$ for some $r_i, \alpha \geq 1$ with $\frac{1}{r_1} + \ldots + \frac{1}{r_m} = \frac{1}{\alpha}$, and assume $T$ is given by integration against a kernel $K(x_1, \ldots, x_{m+1})$ away from the diagonal. We will assume the kernel of $T$ has a decomposition

$$K(x_1, \ldots, x_m, x_{m+1}) = \sum_{s \in \mathbb{Z}} K_s(x_1, \ldots, x_m, x_{m+1}), \quad (2.2)$$
such that in the support of $K_s$ we have $|x_k - x_l| \leq 2^s$ for all $l, k$. We also define

$$[K]_p := \sup_{s \in \mathbb{Z}} \sup_{y \in \mathbb{R}^d} \left( \|K_s(y, \cdot + y, \cdot, \ldots, \cdot + y)\|_{L^p(\mathbb{R}^md)} + \|K_s(\cdot + y, y, \cdot + y, \cdot, \ldots, \cdot + y)\|_{L^p(\mathbb{R}^md)} + \ldots \right)$$

(the last ‘...’ indicates the other possible symmetric terms), and require that

$$[K]_p < \infty. \quad (2.3)$$

This is an abstract analogue of the basic size estimate for multilinear Calderón-Zygmund kernels, see [36] and Section 2.2 below. We will refer to conditions (2.2) and (2.3) as the $(S)$ (single-scale) properties.

Since the operators under consideration are $m$-linear, we will have to deal with $(m+1)$-linear forms of type

$$\int_{\mathbb{R}^n} T(f_1, \ldots, f_m)(x)f_{m+1}(x_{m+1}) \, dx_{m+1}.$$ 

We define $\Lambda^\nu_\mu(f_1, \ldots, f_{m+1})$ to be the form

$$\int_{\mathbb{R}^{(m+1)d}} \sum_{\mu < s < \nu} K_s(x_1, \ldots, x_{m+1})f_1(x_1)\ldots f_{m+1}(x_{m+1}) \, dx_1 \ldots dx_{m+1},$$

and always assume $\mu > 0$ and $\nu < \infty$. Finally, we will assume the following uniform estimate on the truncations: given $\frac{1}{r_1} + \ldots + \frac{1}{r_m} = \frac{1}{\alpha}$ as above,

$$C_T(r_1, \ldots, r_m, \alpha) := \sup \|\Lambda^\nu_\mu\|_{L^{r_1} \times \ldots \times L^{r_m} \times L^{\nu'} \to \mathbb{C}} < \infty.$$ 

Here the supremum is taken over all finite truncations. From now on the truncation bounds $\mu, \nu$ will be omitted from the notation unless explicitly needed, and we will write $C_T$ in place of $C_T(r_1, \ldots, r_m, \alpha)$. 

17
2.1.1 Remark on Truncations

We assume below that our operator is truncated to finitely many scales, and prove estimates that are uniform in the number of truncations. The justification for this assumption is sketched below. The argument is a straightforward generalization of a result that can be found in [65] Ch. 1, section 7.2.

Let $T^\epsilon$ denote our operator truncated at scales larger than $\epsilon$ and smaller than $1/\epsilon$. Fix $f_i \in L^{r_i}$ with compact support, $1 \leq i \leq m$. By the uniform bound assumption on $C_T$, after passing to a subsequence we can assume $T^\epsilon(f_1, ..., f_m)$ converges weakly in $L^\alpha$ to some $T_0(f_1, ..., f_m)$. It is clear from the weak convergence that $T_0$ must be multilinear. We claim that if $Q_i$ is a cube in $\mathbb{R}^n$ then

$$
(T - T_0)(f_1 \mathbf{1}_{Q_1}, ..., f_m \mathbf{1}_{Q_m})(x_{m+1})
$$

$$
= \mathbf{1}_{Q_1}(x_{m+1})...\mathbf{1}_{Q_m}(x_{m+1})(T - T_0)(f_1, ..., f_m)(x_{m+1}) \text{ a.e.}
$$

In fact, the support restriction on the kernel of $T - T^\epsilon$ shows that the integral vanishes if $x_{m+1} \notin Q_i$ for any $i$, provided $\epsilon$ is small enough. Then the identity follows from the weak convergence of $T^\epsilon$. Using multilinearity we can extend (2.4) to simple functions, and then using the boundedness of $T$ we see that (2.4) holds with $g_i \in L^{r_i}$ in place of $\mathbf{1}_{Q_i}$. In particular, let $E^j_i$ be an increasing sequence of open sets that exhaust $\mathbb{R}^d$, and suppose $f_i \in L^{r_i}$ with support in $E^j_i$. Then we must have

$$
(T - T_0)(f_1, ..., f_m)(x_{m+1})
$$

$$
= f_1(x_{m+1})...f_m(x_{m+1})(T - T_0)(\mathbf{1}_{E^j_1}, ..., \mathbf{1}_{E^j_m})(x_{m+1}) \text{ a.e.}
$$

It follows that $T$ differs from the limit $T_0$ by a multiplication operator, provided

$$
(T - T_0)(\mathbf{1}_{E^j_1}, ..., \mathbf{1}_{E^j_m})(x_{m+1}) =: \phi(x_{m+1})
$$
forms a coherent function with respect to the $E^j$. This is clear as in the linear case. Moreover, $\phi \in L^\infty$ since there exists $c_T > 0$ such that

$$c_T > \sup_{f_i \in L^r_{\text{c.p. supp}}} \frac{\|(T - T_0)(f_1, \ldots, f_m)\|_{L^\alpha}}{\|f_1\|_{L^r_1} \ldots \|f_m\|_{L^r_m}} = \sup_{f_i \in L^r_{\text{c.p. supp}}} \frac{\|\hat{\phi} \cdot f_1 \ldots f_m\|_{L^\alpha}}{\|f_1\|_{L^r_1} \ldots \|f_m\|_{L^r_m}} \geq \|\phi\|_{L^\infty}.$$

Therefore, as long as we can prove admissible bounds for multiplication operators of the form $A_\phi(f_1, \ldots, f_m) = \phi \cdot f_1 \ldots f_m$, with $\phi \in L^\infty$, we are justified in working with a finite (but otherwise arbitrary) number of scales.

### 2.1.2 The Abstract Theorem

Assume we are given some stopping collection of cubes $\mathcal{P}$ with top $Q$. We will use the space $\mathcal{Y}_p = \mathcal{Y}_p(\mathcal{P})$ from [19], with norm $\| \cdot \|_{\mathcal{Y}_p}$ defined by

$$\|h\|_{\mathcal{Y}_p} := \begin{cases} \max \left( \|h1_{\mathbb{R}^d \setminus \bigcup_{L \in \mathcal{P}} L}\|_\infty, \sup_{L \in \mathcal{P}} \inf_{x \in L} M_p h(x) \right) & \text{if } p < \infty \\ \|h\|_\infty & \text{if } p = \infty \end{cases} \quad (2.5)$$

(recall from above that $\hat{L}$ is the $2^5$-fold dilate of $L$). We let $\|b\|_{\mathcal{X}_p}$ denote the $\mathcal{Y}_p$-norm of $b$ when $b = \sum_{L \in \mathcal{P}} b_L$ with $b_L$ supported on $L \in \mathcal{P}$, and use $\mathcal{X}_p$ to signal that $\int b_L = 0$ for each $L$. Observe that these norms are increasing in $p$, a property we use many times below. Also recall that if $h \in \mathcal{Y}_p$ there is a natural Calderón-Zygmund decomposition of $h$ associated to the stopping collection $\mathcal{P}$: we can split $h = g + b$, where

$$b = \sum_{L \in \mathcal{P}} \left( h - \frac{1}{|L|} \int_L h \right) 1_L \quad (2.6)$$

and

$$\|g\|_{\mathcal{Y}_\infty} \leq 2^{5d}\|h\|_{\mathcal{Y}_p}, \quad b \in \mathcal{X}_p, \quad \|b\|_{\mathcal{X}_p} \leq 2^{5d+1}\|h\|_{\mathcal{Y}_p}.$$
Given a cube $L$, define

$$
\Lambda_L(f_1, \ldots, f_m, f_{m+1}) := \Lambda_{\text{min}(s_L, \text{top trunc})}(f_11_L, f_2, \ldots, f_m1_{3L}, f_{m+1}1_{3L}),
$$

meaning the truncation from above never exceeds the scale of $L$. The second equality follows from the support condition on the kernel imposed by (2.2). For simplicity we often write $\Lambda_{s_L}$ to indicate that all truncations are at or below level $s_L$. We will work with

$$
\Lambda_P(f_1, \ldots, f_m, f_{m+1}) := \Lambda_Q(f_1, \ldots, f_{m+1}) - \sum_{L \in P} \Lambda_L(f_1, \ldots, f_{m+1})
$$

(2.7)

(as above, $Q$ is the top cube of the stopping collection $P$). Note that $\Lambda_P$ is not symmetric in all of its arguments, due to some lack of symmetry in definitions. However, we will see below that $\Lambda_P$ is ‘almost’ symmetric, meaning it is symmetric up to a controllable error term.

We can now state the abstract sparse domination theorem.

**Theorem 2.2.** Let $T$ be an $m$-linear operator with kernel $K$ as above, such that $K$ can be decomposed as in (2.2) and $C_T < \infty$. Also let $\Lambda$ be the $(m+1)$-linear form associated to $T$. Assume there exist $1 \leq p_1, \ldots, p_m, p_{m+1} \leq \infty$ and some positive constant $C_L$ such that the following estimates hold uniformly over all finite truncations, all dyadic lattices $D$, and all stopping collections $P$:

$$
|\Lambda_P(b, g_2, g_3, \ldots, g_{m+1})| \leq C_L|Q||b||\chi_{p_1}||g_2||y_{p_2}||g_3||y_{p_3}||g_{m+1}|y_{p_{m+1}}
$$

$$
|\Lambda_P(g_1, b, g_3, \ldots, g_{m+1})| \leq C_L|Q||g_1||y_\infty||b||\chi_{p_2}||g_3||y_{p_3}||g_{m+1}|y_{p_{m+1}}
$$

(2.8)

$$
|\Lambda_P(g_1, g_2, b, g_4, \ldots, g_{m+1})| \leq C_L|Q||g_1||y_\infty||g_2||y_\infty||b||\chi_{p_3}||g_4||y_{p_4}||g_{m+1}|y_{p_{m+1}}
$$

$$
\vdots
$$

$$
|\Lambda_P(g_1, g_2, \ldots, g_m, b)| \leq C_L|Q||g_1||y_\infty||g_2||y_\infty||\ldots||g_m||y_\infty||b||\chi_{p_{m+1}}.
$$

Also let $\vec{p} = (p_1, \ldots, p_{m+1})$. Then there is some constant $c_d$ depending on the dimension $d$ such
that
\[ \sup_{\mu,\nu} |\Lambda_{\mu,\nu}(f_1, ..., f_m, f_{m+1})| \leq c_d [C_T + C_L] \sup_S \text{PSF}_{S,\mathbb{P}}(f_1, ..., f_m, f_{m+1}) \]
for all \( f_j \in L^{p_j}(\mathbb{R}^d) \) with compact support, where the supremum is taken with respect to all sparse collections \( S \) with some fixed sparsity constant that depends only on \( d, m \).

### 2.1.3 Some Remarks on Theorem 2.2

We will prove in Section 2.5.1 that the multiplication operators \( A_\phi \) considered above in Remark 2.1 satisfy admissible PSF bounds. Therefore Theorem 2.2 implies
\[ |\Lambda(f_1, ..., f_m, f_{m+1})| \leq c_d [C_T + C_L] \sup_S \text{PSF}_{S,\mathbb{P}}(f_1, ..., f_m, f_{m+1}) \]
when \( f_j \in L^{\infty}(\mathbb{R}^d) \) with compact support. If \( \Lambda \) extends boundedly to \( L^{q_1}(\mathbb{R}^d) \times \ldots \times L^{q_{m+1}}(\mathbb{R}^d) \), we can use standard density arguments to lift the PSF bound to the case where \( f_i \in L^{q_i}(\mathbb{R}^d) \).

We also provide some more motivation for the estimates (2.8). Let \( \mathcal{P} \) be a stopping collection of dyadic cubes with top \( Q \) and suppose \( b = \sum_{L \in \mathcal{P}} b_L \) with \( b_L \) supported in \( L \). Also assume \( g_1, g_2, ..., g_{m+1} \) are functions supported in \( 3Q \). If we fix \( L \in \mathcal{P} \) with scale \( s_L \), then by definition \( \Lambda_{\mathcal{P}}(b_L, g_2, ..., g_{m+1}) \) splits as
\[ \Lambda_{\mathcal{P}}(b_L, g_2, ..., g_{m+1}) = \Lambda^{s_Q}(b_L, g_2, ..., g_{m+1}) - \Lambda^{s_L}(b_L, g_2 1_{3L}, ..., g_{m+1} 1_{3L}). \]

Now let \( s \) be a fixed scale with \( s \leq s_L \). Then the piece of \( \Lambda_{\mathcal{P}}(b_L, g_2, ..., g_{m+1}) \) corresponding to this scale can be written as
\[ \int_{\mathbb{R}^{(m+1)d}} K_s b_L(x_1) (g_2(x_2) ... g_{m+1}(x_{m+1}) - g_2 1_{3L}(x_2) ... g_{m+1} 1_{3L}(x_{m+1})) \, d\vec{x} \]
\[ = \int_{\mathbb{R}^{(m+1)d}} K_s b_L (g_2 1_{3L} ... g_{m+1} 1_{3L} - g_2 1_{3L} ... g_{m+1} 1_{3L}) \, d\vec{x} \]
\[ = 0. \]
Here we’ve used the truncation of the kernel: since \( x_1 \in L \) and \( |x_1 - x_i| \leq 2^s \leq 2^{s_L} \) for any \( i \), we must have \( x_i \in 3L \) for each \( i \). Therefore all scales \( s \) entering into \( \Lambda_P(b_L, g_2, ..., g_{m+1}) \) satisfy \( s > s_L \), and as a consequence we have the decomposition

\[
\Lambda_P(b_L, g_2, ..., g_{m+1}) = \int_{\mathbb{R}^{(m+1)d}} \sum_{l \geq 1} K_{s_L + l} b_L(x_1) g_2(x_2) ... g_{m+1}(x_{m+1}) d\vec{x}.
\]

Summing over \( L \in P \) then gives

\[
\Lambda_P(b, g_2, ..., g_{m+1}) = \sum_{L \in P} \int_{\mathbb{R}^{(m+1)d}} \sum_{l \geq 1} K_{s_L + l} b_L(x_1) g_2(x_2) ... g_{m+1}(x_{m+1}) d\vec{x}
\]

\( (2.9) \)

where

\[
b_{s-l} = \sum_{\substack{L \in P \\ s_L = s-l}} b_L.
\]

The representation (2.9) often enables one to verify the estimates (2.8) in practice. We will see this in the case of multilinear Calderón-Zygmund operators below in Section 2.2. In the Appendix we prove the following result about the adjoint forms:

**Proposition 2.1.** Let \( b, g_1, ..., g_{m+1} \) be defined as above, fix \( 1 < p \leq \infty \), and suppose \([K]_{p'} < \infty\). If we set

\[
b^{in} = \sum_{\substack{L \in P \\ 3L \cap \partial Q = \emptyset}} b_L,
\]

then

\[
\Lambda_P(g_1, b, g_3, ..., g_{m+1}) = \int_{\mathbb{R}^{(m+1)d}} \sum_{s} \sum_{l \geq 1} K_{s} b^{in}_{s-l}(x_2) g_1(x_1) g_3(x_3) ... g_{m+1}(x_{m+1}) d\vec{x} + \phi,
\]

22
where $\phi$ is an error term that satisfies the estimate

$$|\phi| \lesssim [K]_{p'} |Q| b \|x_1\|g_1\|y_2\|g_2\|y_3\|g_3\|\cdots\|g_{m+1}\|y_p.$$ 

An analogous decomposition holds for each of the other forms

$$\Lambda_P(g_1, g_2, b, g_4, \ldots, g_{m+1}), \ldots, \Lambda_P(g_1, \ldots, g_{m-1}, b, g_{m+1}), \Lambda_P(g_1, \ldots, g_m, b),$$

with error terms satisfying the same bounds (with obvious changes in indices).

In practice the error term $\phi$ does not cause any issues. This quantifies the ‘almost symmetry’ of $\Lambda_P$ mentioned above.

The proof of Theorem 2.2 is similar to the proof of the abstract Theorem C in [19], so we postpone the argument to Section 2.5 below.

## 2.2 Multilinear Calderón-Zygmund Operators

In this section we use Theorem 2.2 to prove sparse bounds for multilinear Calderón-Zygmund (CZ) operators that satisfy a certain smoothness condition (defined below). For the theory of these operators see [33] or [36]. In the process we prove an estimate which is necessary for the proof of Theorem 2.1 given in Section 2.3 below. The reader who wishes to skip to the proof of Theorem 2.1 only needs to be familiar with the statement of Proposition 2.4 and the list of inequalities immediately following its proof.

Fix an $m$-linear CZ operator $T$ with kernel $K$. We have to suitably decompose the kernel to verify the initial assumptions of Theorem 2.2. Recall that we have the basic size estimate

$$|K(x_1, \ldots, x_m, x_{m+1})| \leq \frac{C_K}{(\max_{i \neq j} |x_i - x_j|)^{md}},$$

where $1 \leq i, j \leq m + 1$. This estimate motivates the following decomposition. Let $\vec{x} = \ldots$
\((x_1, \ldots, x_m, x_{m+1})\) with \(x_i \in \mathbb{R}^d\), and define

\[ M_{\vec{x}} = \left( \max_{i \neq j} |x_i - x_j| \right)^{md} \]

and

\[ M_{ij} = \{ \vec{x} \in \mathbb{R}^{(m+1)d} : M_{\vec{x}} = |x_i - x_j| \}. \]

Note that there are \(m(m-1)/2\) such sets \(M_{ij}\). Also define \(M_{ij}^*\) be the open subset of \(\mathbb{R}^{(m+1)d}\)

where

\[ |x_i - x_j| > \frac{1}{2} |x_k - x_l| \]

for all \(k, l\). Let \(\psi_{ij}\) be a smooth partition of unity of \(\mathbb{R}^{(m+1)d}\setminus\{0\}\) subordinated to the open cover \(\{M_{ij}^*\}\), such that \(\psi_{ij} = 1\) on \(M_{ij}\). Let \(K_{ij}^* = \psi_{ij} K\), so that our operator splits as

\[ T(f_1, \ldots, f_m)(x_{m+1}) = \sum_{i,j} \int_{\mathbb{R}^{md}} K_{ij}^*(x_1, \ldots, x_m, x_{m+1}) f_1(x_1) \ldots f_m(x_m) d\vec{x}. \]

Now choose the integer \(l\) so that \(2^{l-1} < m(m-1)/2 \leq 2^l\) and set

\[ A_s = \{ \vec{x} \in \mathbb{R}^{(m+1)d} : 2^{s-l-1} < \sum_{i=1}^m \sum_{j=i+1}^m |x_i - x_j| \leq 2^{s-1} \}. \]

Let \(\phi\) be a smooth radial function supported in \(A_0\) such that \(\sum_s \phi(2^{-s} \vec{x}) = 1\) for \(\vec{x} \neq 0\), and write \(\phi_s(\vec{x}) = \phi(2^{-s} \vec{x})\). Finally, let

\[ K_{ij}^s(\vec{x}) = \psi_{ij}(\vec{x}) \phi_s(\vec{x}) K(\vec{x}). \]

Notice that if \(\vec{x} \in M_{ij} \cap A_s\), then \(|x_i - x_j| \sim 2^s\). This leads to the decomposition

\[ K_{ij}(x_1, \ldots, x_m, x_{m+1}) = \sum_s K_{ij}^s(x_1, \ldots, x_m, x_{m+1}) = \sum_s K_{ij}^s(x_1, \ldots, x_m, x_{m+1}) \phi_s(\vec{x}). \]
Observe that if $|x_i - x_j| \sim 2^s$ in the region $M_{ij}^*$ then necessarily $|x_k - x_l| \leq 2^s$ for all other pairs, since $|x_k - x_l| \leq 2|x_i - x_j|$ in $M_{ij}^*$. Given $1 \leq p < \infty$ we can easily prove the single-scale size estimate

$$\left( \int_{\mathbb{R}^md} |K^{ij}(\vec{x})|^p \right)^{1/p} \lesssim 2^{-mds/p}$$

for each $i, j$. This bound holds uniformly in the free variable. One also has

$$\|K_s^{ij}\|_{L^\infty(\mathbb{R}^md)} < 2^{-mds},$$

with a similar interpretation for the free variable. Therefore $[K]_p < \infty$ in this setting, for $1 \leq p \leq \infty$. We are also free to assume that $T$ satisfies the uniform truncation bound $C_T(r_1, ..., r_m, \alpha) < \infty$ for some collection of exponents with $\frac{1}{r_1} + ... + \frac{1}{r_m} = \frac{1}{\alpha}$ (see [33], [36]).

Assume for the rest of the section that $\mathcal{P}$ is a fixed stopping collection of cubes with top $Q$. We will work towards a proof of the estimates (2.8) needed to apply Theorem 2.2 in this context. We begin by proving a single-scale estimate in a slightly more abstract setting, which will be useful for the rest of the paper.

### 2.2.1 Single-Scale Estimates

Suppose $b = \sum_{L \in \mathcal{P}} b_L$ with $b_L$ supported in $L$. Also pick functions $g_1, ..., g_{m+1}$ supported in $3Q$. The following lemma applies to all kernels $K = \sum_{s \in \mathbb{Z}} K_s$ satisfying the (S) properties from section 2.

**Lemma 2.2.** Suppose $b, g_1, g_2, ..., g_{m+1}$ are defined as above and $1 < \beta \leq \infty$. Let $T$ be an m-linear operator with kernel $K$, such that $K$ satisfies the (S) properties with $[K]_{\beta'} < \infty$. For a
fixed \( l \geq 1 \) we have

\[
\int_{\mathbb{R}^{m+1}} \sum_{s} |K_s| \cdot |b_{s-l}|(x_1)|g_2(x_2)|...|g_{m+1}(x_{m+1})| \, d\vec{x}
\]

\[
\lesssim [K]_{\beta'} |Q||b||x_1||y_\beta ...||g_{m+1}||y_\beta.
\]

Symmetric estimates also hold for the other tuples \((g_1, b, g_3, ..., g_{m+1}), ..., (g_1, ..., g_m, b)\).

**Proof.** From (2.9), we see that it is enough to show that if \( L \in \mathcal{P} \) and \( s = s_L + l \), then

\[
\int_{\mathbb{R}^{m+1}} |K_s||b_L(x_1)||g_2(x_2)|...|g_{m+1}(x_{m+1})| \, d\vec{x}
\]

\[
\lesssim [K]_{\beta'} |L||b||x_1||g_2||y_\beta ...||g_{m+1}||y_\beta.
\]

We can then sum over \( L \) and use disjointness to complete the proof. Begin by fixing \( x_1 \in L \), with the goal of proving an \( L_{x_1}^\infty \) bound for

\[
\int_{\mathbb{R}^{m+1}} |K_s(x_1, ..., x_{m+1})||g_2(x_2)|...|g_{m+1}(x_{m+1})| \, dx_2...dx_{m+1}.
\]

We change variables and set \( z_i = x_i - x_1 \) for \( i = 2, ..., m+1 \). Notice that the support of \( K_s \) implies \( |z_i| \lesssim 2^s \) for each \( i \). Now for all such \( x_1 \),
\[
\int_{\mathbb{R}^{m+d}} |K_s(x_1, x_2, \ldots, x_{m+1})| |g_2(x_2)| \cdots |g_{m+1}(x_{m+1})| dz_2 \cdots dz_{m+1}.
\]

\[
\leq \left( \int_{\mathbb{R}^{m+d}} |K_s(x_1, z_2 + x_1, \ldots, z_{m+1} + x_1)|^{\beta'} d z_2 \cdots d z_{m+1} \right)^{1/\beta'}
\times \prod_{i=2}^{m+1} \left( \int_{B(x_1, 2^{s+10})} |g_i|^\beta d z_i \right)^{1/\beta}
\leq 2^{-\frac{mds}{\beta'} [K]_{\beta'}} \prod_{i=2}^{m+1} \left( \int_{B(x_1, 2^{s+10})} |g_i|^\beta d z_i \right)^{1/\beta}
\lesssim [K]_{\beta'} \prod_{i=2}^{m+1} \inf_L M^\beta (g_i)
\lesssim [K]_{\beta'} \|g_2\|_{Y_\beta} \cdots \|g_{m+1}\|_{Y_\beta}.
\]

Note that \( s > s_L \) and \( x_1 \in L \), so \( \tilde{L} \) is contained in \( B(x_1, 2^{s+10}) \). This justifies adding the infimums in the fourth line. As a consequence, we get

\[
\int_{\mathbb{R}^{(m+1)d}} |K_s||b_L(x_1)||g_2(x_2)| \cdots |g_{m+1}(x_{m+1})| d\vec{x}.
\]

\[
\lesssim [K]_{\beta'} |L||b||x_1||g_2||Y_\beta \cdots ||g_{m+1}||Y_\beta,
\]

since \( \|b_L\|_{L^1} \lesssim |L||b||_{X_1} \). This completes the proof of the main estimate. The relevant estimates for the tuples \((g_1, b, g_3, \ldots, g_{m+1})\), etc., can be proved in just the same way using the representations from Proposition 2.1. Notice that the error term \( \phi \) from this proposition has an acceptable contribution.

If \( \Lambda \) is the \((m + 1)\)-linear form associated to an \( m \)-linear CZ operator \( T \), then from (2.9) we see that

\[
\Lambda_{\mathcal{P}}(b, g_2, \ldots, g_{m+1}) = \int_{\mathbb{R}^{(m+1)d}} \sum_{ij} \sum_s \sum_{l \geq 1} K_{sq}^{ij} b_{s-l}(x_1) g_2(x_2) \cdots g_{m+1}(x_{m+1}) d\vec{x}.
\]
The number of pairs $i, j$ is bounded by a constant depending only on $m$, so it is clear that the result of Lemma 2.2 applies to this operator as well (with a possibly different constant).

### 2.2.2 Cancellation Estimates

Let $T$ be an $m$-linear Calderón-Zygmund operator as defined above, such that

$$|K(x_1, x_2, ..., x_{m+1}) - K(x'_1, x_2, ..., x_{m+1})| \leq \frac{A|x_1 - x'|^\epsilon}{(\max_{i \neq j}|x_i - x_j|)^{md+\epsilon}}$$

(2.12)

for some $0 < \epsilon \leq 1$ whenever $|x_1 - x'| \leq \frac{1}{m+1}(|x_2 - x_1| + ... + |x_{m+1} - x_1|)$, and suppose that symmetric estimates hold with respect to the other variables $x_i$ (with the same constant). If $C_K$ is the constant from the basic Calderón-Zygmund size estimate, we also assume that $C_K \leq A$. We then call (2.12) the ‘$\epsilon$-smoothness’ property of $T$, and $A$ the ‘smoothness’ constant. We will use (2.12) to sum over $l$ in the context of Lemma 2.2, and as a result prove the required estimates (2.8) for the sparse bound. It is possible that this smoothness condition may be relaxed, but it is all we need below for the proof of Theorem 2.1.

Suppose $b = \sum_{L \in P} b_L$ and $g_1, ..., g_{m+1}$ are as in the last section, with the additional property that $\int b_L = 0$ for each $L$. We will use the cancellation of $b$ and (2.12) to estimate

$$\Lambda_P(b, g_2, ..., g_{m+1}), ..., \Lambda_P(g_1, ..., g_m, b).$$

Before proceeding, we need to prove a technical lemma related to the truncation.

**Lemma 2.3.** Suppose $|x_1 - x'| \leq 2^{s-l}$ for some integer $l \geq 1$, and let

$$\psi_{ij}^s(x_1, ..., x_{m+1}) = \psi^{ij}(x_1, ..., x_{m+1})\phi_s(x_1, ..., x_{m+1}).$$
Then

$$
\left| \sum_{i,j} K(x_1, x_2, ..., x_{m+1}) \psi_s^{ij}(x_1, x_2, ..., x_{m+1}) \\
- K(x', x_2, ..., x_{m+1}) \psi_s^{ij}(x', x_2, ..., x_{m+1}) \right| \\
\lesssim \sum_{i,j} I_{M_{ij}}(x_1, ..., x_{m+1}) \left| K(x_1, ..., x_{m+1}) - K(x', ..., x_{m+1}) \right| \\
+ 2^{-l} \| \nabla \phi_0 \|_\infty \left| K(x', ..., x_{m+1}) \right|.
$$

(2.13)

Symmetric estimates also hold for the other variables under suitable assumptions.

**Proof.** We will prove the lemma for the pair \(x_1, x'\). The proof for the other variables is identical. For notational simplicity, below we will write \(K = K(x_1, ..., x_{m+1})\) and \(\tilde{K} = K(x', x_2, ..., x_{m+1})\). After adding and subtracting \(\tilde{K} \cdot \psi_s^{ij}\) to each term of the sum, we can estimate (2.13) by

$$
\sum_{ij} |K - \tilde{K}| \psi_s^{ij}(x) + \sum_{ij} (\psi_s^{ij}(x', x_2, ...x_{m+1}) - \psi_s^{ij}(x_1, ..., x_{m+1})) \cdot |\tilde{K}|.
$$

So we have to show that

$$
\left| \sum_{ij} (\psi_s^{ij}(x', x_2, ...x_{m+1}) - \psi_s^{ij}(x_1, ..., x_{m+1})) \right| \lesssim 2^{-l} \| \nabla \phi_0 \|_\infty
$$

when \(|x_1 - x'| \leq 2^{s-l}|x' - x_1|\). Note that under this hypothesis

$$
\left| \phi_s(x_1, x_2, ..., x_{m+1}) - \phi_s(x', x_2, ..., x_{m+1}) \right| \lesssim \| \nabla \phi_0 \|_\infty 2^{-s}|x' - x_1| \\
\lesssim 2^{-l} \| \nabla \phi_0 \|_\infty,
$$

by the scaling and smoothness of \(\phi_s\). Summing over \(i, j\) and applying this estimate proves the lemma.

We can now prove the main result of the section.
Proposition 2.4. Let $K$ be an $m$-linear Calderón-Zygmund kernel such that the $\epsilon$-smoothness condition (2.12) holds with constant $A_\epsilon$. Fix $1 < p \leq \infty$. Then

$$
|\sum_{ij} \sum_{l \geq 1} \int_{\mathbb{R}^{(m+1)d}} K^i_s \cdot b_{s-l}(x_1)g_2(x_2)\ldots g_{m+1}(x_{m+1})d\vec{x}|
$$

$$
\lesssim \epsilon A_\epsilon |Q|\|b\|_X_1 \|g_2\|_Y_p \ldots \|g_{m+1}\|_Y_p.
$$

Analogous estimates also hold for $\Lambda(g_1, b, g_3, \ldots, g_{m+1}), \ldots, \Lambda(g_1, \ldots, g_m, b)$.

Proof. By (2.9), it suffices to prove

$$
\sum_{l \geq c_m} \left| \int_{\mathbb{R}^{(m+1)d}} \sum_{i,j} K^i_{s_L+l}(x_1)g_2(x_2)\ldots g_{m+1}(x_{m+1})d\vec{x} \right|
$$

$$
\lesssim \epsilon A_\epsilon |L|\|b\|_X_1 \|g_2\|_Y_p \ldots \|g_{m+1}\|_Y_p,
$$

where $c_m$ is some constant depending on the multilinearity parameter $m$, to be determined below. We can then use disjointness of the collection to sum over $L$, and finitely many applications of Lemma 2.2 to handle the cases where $l < c_m$. Let $x' = c_L$, the center of $L$. Note that if $s = s_L + l$ then $|x_1 - x'| \leq 2^{s-l}$ since $x_1 \in L$. We use the mean-zero condition on $b_L$ to replace

$$
\sum_{i,j} K^i_{s_L+l}(x_1, x_2, \ldots, x_{m+1})
$$

by

$$
\sum_{i,j} \left( K^i_{s_L+l}(x_1, x_2, \ldots, x_{m+1}) - K^i_{s_L+l}(x', x_2, \ldots, x_{m+1}) \right),
$$

and then estimate (2.15) using Lemma 2.3. To complete the proof, we can argue as in the proof of Lemma 2.2, using the cancellation estimate (2.12) in place of the basic size estimate for $|K^i_s|$ (with $s = s_L + l$).

We now sketch the rest of the proof. Set $F = K - \overline{K}$ and observe that the first term from
the estimate in (2.13) contributes

$$\sum_{ij} \int_{\mathbb{R}^{(m+1)d}} 1_{M_{ij}^s}(|x_j|) \int \frac{dM^s_*}{dx_1^1} \left( m^1 \right) d\bar{x}.$$ 

We now fix $x_1 \in L$, with the goal of proving an $L^\infty$ bound for

$$\int_{\text{supp} K_{ij}^s} |F(x_1, \ldots, x_{m+1})||g_2(x_2)||g_{m+1}(x_{m+1})|dx_2 \ldots dx_{m+1}.$$ 

Once again make the change of variables $z_i = x_i - x_1$ for $i = 2, \ldots, m+1$, so in particular $|z_i| \lesssim 2^s$. Then we proceed as in Lemma 2.2, with $F(x_1, x_1 + z_2, \ldots, x_1 + z_2)$ in place of the $K_s$ term. The only change is in the estimate involving the kernel term. Since we are working at scale $\sim s$, in any region $M_{ij}^s$ there is some uniform constant $\alpha_m$ depending only on $m$ such that $|x_i - x_j| > 2^{s-\alpha_m}$. Therefore we must have either $|x_1 - x_i| \geq 2^{s-\alpha_m-1}$ or $|x_1 - x_j| \geq 2^{s-\alpha_m-1}$. It follows that we can choose some uniform $c_m$ such that if $l \geq c_m$ and $|x_1 - x'| \leq 2^{s-l}$, then

$$|x_1 - x'| \leq \frac{1}{m+1} (|z_2| + \ldots + |z_{m+1}|).$$ 

Hence we can apply the cancellation estimate (2.12) for $|F(x_1, x_1 + z_2, \ldots, x_1 + z_{m+1})|$. A straightforward calculation shows that for this choice of $s$ and any $i, j$ and $x_1 \in L$, we have

$$(\int_{\text{supp} K_{ij}^s} |x_1 - x'|^p d_{x_2} \ldots d_{x_{m+1}})^{1/p'} \lesssim 2^{-cl} 2^{-\frac{md}{p'}}.$$ (2.16)

This allows us to proceed as in the proof of Lemma 2.2, with an addition of the $2^{-cl}$ term. The second term from Lemma 2.3 can be handled using the same methods from Lemma 2.2. Since the number of pairs $i, j$ is bounded by some $c_m$, we ultimately see that the left side of (2.14) is
bounded by

\[ c_m \sum_{l \geq 1} (2^{-l} + 2^{-l} \| \nabla \phi_0 \|_\infty) A_\epsilon |L| \| b \|_{\dot{X}_1} \| g_2 \|_{\gamma_p} \cdots \| g_{m+1} \|_{\gamma_p} \leq c A_\epsilon |L| \| b \|_{\dot{X}_1} \| g_2 \|_{\gamma_p} \cdots \| g_{m+1} \|_{\gamma_p}, \]

and summing over \( L \) proves the desired estimate. To prove the remaining part of the proposition, use the representations from Proposition 2.1 and repeat the argument just given with respect to the new variables.

As a consequence of the preceding proposition and (2.9), we see that

\[ |\Lambda_P (b, g_2, g_3, \ldots, g_{m+1})| \leq A_\epsilon |Q| \| b \|_{\dot{X}_1} \| g_2 \|_{\gamma_p} \| g_3 \|_{\gamma_p} \cdots \| g_{m+1} \|_{\gamma_p}, \]

where \( A_\epsilon \) is the smoothness constant from (2.12). These estimates hold uniformly over finite truncations and stopping collections. We can therefore apply Theorem 2.2 to prove the following.

**Theorem 2.3.** Let \( T \) be an \( m \)-linear Calderón-Zygmund operator satisfying the \( \epsilon \)-smoothness condition (2.12) with constant \( A_\epsilon \). Suppose \( T \) is bounded from \( L^{r_1}(\mathbb{R}^d) \times \cdots \times L^{r_m}(\mathbb{R}^d) \rightarrow L^\alpha(\mathbb{R}^d) \), such that \( C_T(r_1, \ldots, r_m, \alpha) < \infty \). Also fix \( 1 < p \leq \infty \). If \( \Lambda \) is the \( (m+1) \)-linear form associated to \( T \), then for any \( f_1, \ldots, f_{m+1} \) with \( f_1 \in L^1(\mathbb{R}^d) \) and \( f_i \in L^p(\mathbb{R}^d) \) for \( i = 2, \ldots, m+1 \), we have

\[ |\Lambda(f_1, f_2, \ldots, f_{m+1})| \leq c_d [C_T + A_\epsilon] \sup_S PSF_{S}^{(1, p, \ldots, p)}(f_1, f_2, \ldots, f_{m+1}). \]

As we mentioned in the introduction, sparse domination results for multilinear Calderón-Zygmund operators have been known for a few years. The main novelty here is that we do not appeal
to local mean oscillation \[55\] or maximal truncation estimates \[57\]. We obtain a subset of the known weighted estimates for multilinear Calderón-Zygmund operators as an easy corollary of Theorem 2.3. See \[9\] and \[55\] for some examples.

**Remark.** The same sparse domination result will hold if we assume an abstract Dini condition for the kernel, as in Lemma 3.2 in \[19\]. Given \(1 < p \leq \infty\), let

\[
\Gamma_p(h) = \left(\|K_s(x, x + \cdot, ..., x + \cdot) - K_s(x + h, x + \cdot, ..., x + \cdot)\|_p + ...\right).
\]

The last ‘...’ before the end of the parenthesis indicates the other possible symmetric terms. Now define

\[
\omega_{j,p}(K) := \sup_{s \in \mathbb{Z}} \sup_{x \in \mathbb{R}^d} \sup_{h \in \mathbb{R}^d} \sup_{\|h\|_{\infty} < 2^{k-j-cm}} \Gamma_p(h).
\]

Then if \([K]_p < \infty\) and

\[
[K]_{1,p} := \sum_{j=1}^{\infty} \omega_{j,p}(K) < \infty,
\]

we can prove an analogue of Proposition 2.4 and show that the assumptions of Theorem 2.2 are satisfied. The argument is similar to the proof of Lemma 3.2 in \[19\]. However, note that we would have to do essentially the same amount of work as above to check that this condition is satisfied by the \(\epsilon\)-smooth Calderón-Zygmund operators considered in this section.

### 2.3 Rough Bilinear Singular Integrals

In this section we use our sparse domination theorem to prove Theorem 2.1. Recall that

\[
T_{\Omega}(f_1, f_2)(x) = \text{p.v.} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f_1(x - y_1) f_2(x - y_2) \frac{\Omega((y_1, y_2) / |(y_1, y_2)|)}{|(y_1, y_2)|^{2d}} \, dy_1 dy_2,
\]

with \(\Omega \in L^{\infty}(S^{2d-1})\) and \(\int_{S^{2d-1}} \Omega = 0\). We will use the results from the paper \[34\] by Grafakos, He, and Honzík. We quickly review their initial decomposition of the kernel \(K\) of \(T_{\Omega}\). Let \(\{\beta_j\}_{j \in \mathbb{Z}}\)
be a smooth partition of unity on $\mathbb{R}^d \setminus \{0\}$, with $\beta_j$ adapted to the annulus $\{2^{j-1} < |z| < 2^{j+1}\}$.

Let $\Delta_j$ denote the standard Littlewood-Paley operator localizing frequencies at scale $2^j$. Then define $K^i_j = \Delta_{j-i} \beta_i K$ and decompose $K = \sum_{j \in \mathbb{Z}} K_j$, with

$$K_j = \sum_{i \in \mathbb{Z}} K^i_j.$$ 

Write $T^j$ for the operator associated to the kernel $K_j$.

**Proposition 2.5** (Prop. 3 and 5 in [34]). There exists $0 < \delta < 1$ so that

$$\|T^j(f_1, f_2)\|_{L^1} \leq C 2^{-\delta j} \|\Omega\|_\infty \|f_1\|_{L^2} \|f_2\|_{L^2}$$

when $j \geq 0$, and

$$\|T^j(f_1, f_2)\|_{L^1} \leq C 2^{-(1-\delta)j} \|\Omega\|_\infty \|f_1\|_{L^2} \|f_2\|_{L^2}$$

when $j < 0$.

**Proposition 2.6** (Lem. 10 in [34]). Fix any $j \in \mathbb{Z}$ and $0 < \eta < 1$. The kernel of $T_j$ is a bilinear Calderón-Zygmund kernel that satisfies the $\eta$-smoothness condition (2.12) with constant $A_{j,\eta} \leq C_{d,\eta} \|\Omega\|_\infty 2^{j|\eta|}$.

By Proposition 2.6 we can apply Proposition 2.4 at each scale $j$, with resulting constant $A_{j,\eta} \leq C_{d,\eta} \|\Omega\|_\infty 2^{j|\eta|}$. The $\eta$ can be arbitrarily small, so we will be able to interpolate with the bound from Proposition 2.5 and then sum over $j$ to get the required estimates for the sparse-form bound of $T^j$.

### 2.3.1 Interpolation Lemmas

Assume below that a stopping collection $\mathcal{P}$ with top $Q$ has been fixed. The following lemmas allow us to interpolate bounds between various $\mathcal{X}(\mathcal{P})$ and $\mathcal{Y}(\mathcal{P})$ spaces in the trilinear setting. A more general interpolation theorem for these spaces should be available, but we only prove two particular results needed for the proof of Theorem 2.1.
Lemma 2.7. Fix any $0 < \epsilon < \frac{1}{8}$ and suppose $p = 1 + 4\epsilon$. Let $\Lambda$ be a (sub)-trilinear form such that

$$|\Lambda(b, g, h)| \leq A_1 \|b\|_{\tilde{X}_1} \|g\|_{Y_p} \|h\|_{Y_p}$$

and

$$|\Lambda(b, g, h)| \leq A_2 \|b\|_{\tilde{X}_2} \|g\|_{Y_2} \|h\|_{Y_\infty}.$$

Then if $q = p + 4\epsilon$ and $q < r < \infty$, we have

$$|\Lambda(f_1, f_2, f_3)| \leq (A_1)^{1-\epsilon}(A_2)^{\epsilon} \|f_1\|_{\tilde{X}_q} \|f_2\|_{Y_q} \|f_3\|_{Y_r}.$$

Proof. We can assume $A_2 < A_1$. We also make the normalizations $A_1 = 1$ and

$$\|f_1\|_{\tilde{X}_q} = \|f_2\|_{Y_q} = \|f_3\|_{Y_r} = 1.$$

It will now be enough to prove that

$$|\Lambda(f_1, f_2, f_3)| \lesssim A_2^\epsilon.$$

Pick $\lambda \geq 1$ and define $f_{>\lambda} = f1_{|f|>\lambda}$. We decompose $f_1 = b_1 + g_1$, where

$$b_1 := \sum_{R \in \mathcal{P}} \left( (f_1)_{>\lambda} - \frac{1}{|R|} \int_{R} (f_1)_{>\lambda} \right) 1_R.$$

For $i = 2, 3$ we also decompose $f_i = b_i + g_i$ with $b_i = (f_i)_{>\lambda}$. Using the normalization assumption and the definition of the $Y$ spaces, it is straightforward to verify the following estimates:
\[ \|g_1\|_{\dot{X}_p} \lesssim 1, \quad \|b_1\|_{\dot{X}_p} \lesssim \lambda^{1-q} \]
\[ \|g_2\|_{Y_p} \lesssim \lambda^{1-\frac{q}{p}}, \quad \|b_2\|_{Y_p} \lesssim \lambda^{1-\frac{q}{p}} \]
\[ \|g_3\|_{Y_\infty} \lesssim \lambda, \quad \|b_3\|_{Y_p} \lesssim \lambda^{1-\frac{r}{p}}. \]

Note that \( g_1 \in \dot{X}_2 \) since \( f_1 \) and \( b_1 \) are supported on cubes in \( P \), with mean zero on each cube.

We now estimate \( |\Lambda(f_1, f_2, f_3)| \) by the sum of eight terms

\[ |\Lambda(g_1, g_2, g_3)| + |\Lambda(b_1, g_2, g_3)| + |\Lambda(g_1, b_2, g_3)| + |\Lambda(g_1, g_2, b_3)| \]
\[ + |\Lambda(b_1, b_2, g_3)| + |\Lambda(b_1, g_2, b_3)| + |\Lambda(g_1, b_2, b_3)| + |\Lambda(b_1, b_2, b_3)|. \]

Applying the estimates in (2.17), we get

\[ |\Lambda(g_1, g_2, g_3)| \lesssim A_2 \|g_1\|_{\dot{X}_2} \|g_2\|_{Y_p} \|g_3\|_{Y_\infty} \lesssim A_2 \lambda^{3-q}. \]

For the remaining seven terms, we use the first assumption \( |\Lambda(b, g, h)| \leq \|b\|_{\dot{X}_1} \|g\|_{Y_p} \|h\|_{Y_p} \), along with suitable estimates from (2.17). Here it is crucial that \( g_1, b_1 \in \dot{X}_2 \). Ultimately we conclude

\[ |\Lambda(f_1, f_2, f_3)| \lesssim \lambda^{1-q} + \lambda^{1-\frac{q}{p}} + \lambda^{1-\frac{r}{p}} + \lambda^{1-q} \lambda^{1-\frac{q}{p}} + \lambda^{1-q} \lambda^{1-\frac{r}{p}} + \lambda^{1-\frac{q}{p}} \lambda^{1-\frac{r}{p}} \]
\[ + \lambda^{1-q} \lambda^{1-\frac{q}{p}} \lambda^{1-\frac{r}{p}} + A_2 \lambda^3 - q. \]

Let \( \alpha = 1 - \frac{q}{p} \) and notice that \(-4\epsilon < \alpha < -\epsilon \) and \( \alpha = \max(1 - q, 1 - \frac{q}{p}, 1 - \frac{r}{p}) \). We can simplify
the previous estimate to get

\[ |\Lambda(f_1, f_2, f_3)| \lesssim 3\lambda^\alpha + 3\lambda^{2\alpha} + \lambda^{3\alpha} + A_2\lambda^{3-q} \]

\[ \lesssim \lambda^\alpha (7 + A_2\lambda^{3-q-\alpha}) \]

\[ \lesssim \lambda^{-\epsilon} (7 + A_2\lambda^{3-q+4\epsilon}). \]

We set \( \lambda = A_2^{-3+q-4\epsilon} \). Since \( A_2 < 1 \) and \( q < 2 \) this is admissible and we conclude that

\[ |\Lambda(f_1, f_2, f_3)| \lesssim A_2^{-3+q-4\epsilon} \lesssim A_2^{3\epsilon + 4\epsilon^2 - q\epsilon} \lesssim A_2^\epsilon. \]

This completes the proof.

**Lemma 2.8.** Fix any \( 0 < \epsilon < \frac{1}{8} \) and set \( p = 1 + 4\epsilon \). Let \( \Lambda \) be a (sub)-trilinear form such that

\[ |\Lambda(g, h, b)| \leq A_1 \|g\|_{Y_p} \|h\|_{Y_p} \|b\|_{\dot{X}_1} \]

and

\[ |\Lambda(g, h, b)| \leq A_2 \|g\|_{Y_2} \|h\|_{Y_2} \|b\|_{\dot{Y}_\infty}. \]

Then if \( q = p + 4\epsilon \) and \( q < r < \infty \), we have

\[ |\Lambda(f_1, f_2, f_3)| \leq (A_1)^{1-\epsilon} (A_2)^\epsilon \|f_1\|_{Y_q} \|f_2\|_{Y_q} \|f_3\|_{\dot{X}_r}. \]

**Proof.** The argument is almost identical to the proof of Lemma 2.7. In this case we decompose \( f_3 = g_3 + b_3 \) with

\[ b_3 = \sum_{R \in \mathcal{P}} \left( (f_3)_{>\lambda} - \frac{1}{|R|} \int_R (f_3)_{>\lambda} \right) 1_R. \]

For \( i = 1, 2 \) we let \( f_i = g_i + b_i \) with \( g_i = (f_i)_{\leq \lambda} \), and then proceed as in Lemma 2.7. \( \square \)
2.3.2 The Sparse Bound

We can now prove Theorem 2.1. Let $T_\Omega$ be a rough bilinear operator defined as above, with $\Omega \in L^\infty(S^{2d-1})$. Such operators $T_\Omega$ satisfy the standard Calderón-Zygmund size estimate, so the single-scale properties (S) hold. The proof that $C_{T_\Omega} < \infty$ for some tuple of exponents is also straightforward. In particular, inspection of the proofs of Propositions 2.5 and 2.6 in [34] shows that the same estimates hold for $T_j$ if one replaces the kernel of $T_j$ by

$$\tilde{K}_j := \sum_{i=-\infty}^{\infty} a_i \cdot K^i_j,$$

with $a_i \in \{0, 1\}$ but otherwise arbitrary. Moreover, these estimates are uniform over the choice of $a_i$. Any truncation of the kernel of $T$ leads to a special case of this small modification, and therefore the $L^2 \times L^2 \to L^1$ estimate of Proposition 2.5 still holds. Likewise, the Calderón-Zygmund smoothness estimate from Proposition 2.6 holds uniformly over the choice of $a_i$. This is once again obvious from the proof in [34]. It follows that $C_{T_\Omega}(2, 2, 1) < \infty$. We must now verify the estimates (2.8) for $(q_1, q_2, q_3) = (r, r, r)$, where $r > 1$.

Fix $0 < \eta < 1$. Since $T^j$ is a bilinear Calderón-Zygmund kernel satisfying the $\eta$-smoothness condition (2.12) with constant $A \leq C_{d,\eta,\|\Omega\|_\infty} 2^{j\eta}$, we see from Proposition 2.4 that

$$|\Lambda^j(b, g, h)| \leq C_{d,\eta,\|\Omega\|_\infty} 2^{j\eta} |Q| \|b\|_{X_1} \|g\|_{Y_2} \|h\|_{Y_\infty},$$

for any $1 < p \leq \infty$. From Proposition 2.5 we also have

$$|\Lambda^j(b, g, h)| \lesssim 2^{-\epsilon j} \|\Omega\|_\infty |Q| \|b\|_{X_2} \|g\|_{Y_2} \|h\|_{Y_\infty},$$

where $c$ is some positive constant independent of $j$. Interpolating via Lemma 2.7, we find that
for any $0 < \epsilon < \frac{1}{8}$ there are $1 < q < r$ so that

$$|\Lambda^j(b, g, h)| \lesssim_{\epsilon, \eta} (2^{\eta j})^{1-\epsilon}(2^{-c j})^\epsilon |Q||\Omega||b||\hat{\chi}_q ||g||_{\nu} ||h||_{\nu}.$$  

$$\lesssim_{\epsilon, \eta} 2^{\eta j} 2^{-\epsilon c j} 2^{-c \epsilon j} |Q||\Omega||b||\hat{\chi}_q ||g||_{\nu} ||h||_{\nu}.$$  

Suppose we have chosen $\eta$ so that $\eta < \frac{c}{8}$. If we let $\epsilon = \frac{\eta}{c}$ then $0 < \epsilon < \frac{1}{8}$, and therefore

$$|\Lambda^j(b, g, h)| \lesssim_{\epsilon, \eta} 2^{-\eta j} |Q||\Omega||b||\hat{\chi}_q ||g||_{\nu} ||h||_{\nu}.$$  

This is summable over $j \in \mathbb{Z}$. Hence if $\Lambda_\mathcal{P}$ is the form associated to $T_\Omega$, truncated to some finite number of scales, we can conclude that

$$|\Lambda_\mathcal{P}(b, g, h)| \lesssim |Q||\Omega||b||\hat{\chi}_q ||g||_{\nu} ||h||_{\nu}.$$  

(2.18)

By symmetry, the same argument also yields

$$|\Lambda_\mathcal{P}(g, b, h)| \lesssim |Q||\Omega||b||\hat{\chi}_q ||g||_{\nu} ||h||_{\nu}.$$  

(2.19)

These estimates are uniform over all finite truncations and stopping collections. Finally, we argue as above and interpolate using Lemma 2.8 to prove

$$|\Lambda_\mathcal{P}(g, h, b)| \lesssim |Q||\Omega||g||\nu_q ||h||_{\nu} ||b||_{\nu}.$$  

(2.20)

which is once again uniform over truncations and stopping collections.

The estimates (2.18), (2.19), (2.20) show that the form associated to a rough bilinear operator $T_\Omega$ satisfies assumption (2.8) of Theorem 2.2 with tuple $(r, r, r)$, since the $\nu_q$ norm is increasing in $q$. It is clear that we can take any $r > 1$, so this completes the proof of Theorem 2.1.


2.4 Weighted Estimates

We now prove the weighted estimates claimed in Corollary 2.2 and Corollary 2.1. We assume that the reader is familiar with basic results from the theory of $A_p$ weights, for example the openness property

$$ [w]_{A_p - \eta} \lesssim [w]_{A_p} \quad \text{when} \quad \eta = c_{d,p}[w]_{A_p}^{1-p'} $$

(2.21) (see [32] or [65] for a proof). Below we always assume that $\Omega \in L^\infty(S^{2d-1})$ with mean zero.

2.4.1 Single Weight - Proof of Corollary 2.2

The argument is a combination of known results from the sparse domination theory. It suffices to prove the desired estimate for fixed $p > 2$. We can then apply the multilinear extrapolation theory due to Grafakos and Martell, in particular Theorem 2 in [35]. Fix $2 < p < \infty$ and $w \in A_p$, with the goal of showing that

$$ \|T_1(f,g)\|_{L^{p/2}(w)} \leq C_{p,\Omega} c_w \|f\|_{L^p(w)} \|g\|_{L^p(w)}. $$

Define

$$ \sigma = w^{-\frac{2}{p-2}} $$

and notice that

$$ w^2 \sigma^{\frac{p-2}{p}} = 1. $$

Let $r = (p/2)' = \frac{p}{p-2}$, and choose $q_i = 1+\epsilon_i$ for $\epsilon_i > 0$, such that $q_i < r$ and $q_i < p$. By Theorem 2.1 and duality considerations, it will be enough to show that for any sparse collection $S$,

$$ \text{PSF}_S^{(q_1,q_2,q_3)}(f,g,h) \lesssim_p c_w \|f\|_{L^p(w)} \|g\|_{L^p(w)} \|h\|_{L^r(\sigma)}. $$

(2.22)
We begin by quoting the estimate

\[ \text{PSF}_{S}^{(1,1,q_3)}(g_1, g_2, g_3) \lesssim_p \gamma_w \max(p^{-1}, q_3) \| g_1 \|_{L^p(w)} \| g_2 \|_{L^p(w)} \| g_3 \|_{L^r(\sigma)} \]

with

\[ \gamma_w = \sup_Q \left( \frac{1}{|Q|} \int_{Q} w^{1-\frac{1}{p-1}} \right)^{\frac{q_3}{p}} \left( \frac{1}{|Q|} \int_{Q} \sigma^{\frac{q_3}{q_3-r}} \right)^{\frac{1}{q_3} - \frac{1}{r}}. \]

This can be proved using techniques from [24] or [55] (see, for example, Theorem 3 in [24] and its proof). Now write the term involving \( \sigma \) as

\[ \left( \frac{1}{|Q|} \int_{Q} w^{1+\alpha} \right)^{\frac{1}{q_3} - \frac{1}{r}}, \quad 1 + \alpha = \frac{1}{1 - \frac{1}{q_3}} - \frac{1}{r}. \]

We can apply the reverse Hölder inequality if we choose \( \epsilon_3 \) correctly (as we are free to do), and after some straightforward calculation this leads to the estimate

\[ \gamma_w \lesssim_p c_{w,p} [w]_{A_p}^\frac{2}{q}, \]

where \( c_{w,p} \) is a power of the constant appearing in the reverse Hölder inequality. Hence

\[ \text{PSF}_{S}^{(1,1,q_3)}(g_1, g_2, g_3) \lesssim_p c_{w,p} [w]_{A_p}^{\frac{2}{q}} \max(p^{-1}, q_3) \| g_1 \|_{L^p(w)} \| g_2 \|_{L^p(w)} \| g_3 \|_{L^r(\sigma)}, \quad (2.23) \]

and \( c_{w,p} \) can be computed explicitly in terms of \([w]_{A_p}\) (see, for example, Chapter 7.2 in [32]).

We can lift (2.23) to a bound for \( \text{PSF}_{S}^{q_1,q_2,q_3}(f, g, h) \) using an inequality due to Di Plinio and Lerner ([26], Proposition 4.1):

\[ (f)_{1+\epsilon,Q} \leq (f)_{1,Q} + 2^d \epsilon (M_1 + \epsilon f)_{1,Q}. \quad (2.24) \]

Recall also that

\[ \| M_q \|_{L^p(w) \to L^p(w)} \lesssim [w]_{A_p}^{\frac{q}{p-q}}. \]
when \( p > q \) (see [8]). Using the openness of the \( A_t \) classes, we see that if \( \epsilon_1, \epsilon_2 \) are chosen properly then in fact
\[
\| M_q \|_{L^p(w) \to L^p(w)} \lesssim [w]_{A_p}^{\left( \frac{q_1}{p-q_1} \right) \left( \frac{1}{1-p} \right)}
\] (2.25)
for \( i = 1, 2 \). Now apply (2.24) to the \( q_1 \)- and \( q_2 \)-averages occurring in the form \( \text{PSF}_{S_{q_1-q_2-q_3}}(f, g, h) \) to get
\[
\text{PSF}^{(q_1, q_2, q_3)}(f, g, h) \lesssim \text{PSF}^{(1, 1, q_3)}(f, g, h) + c'_w \text{PSF}^{(1, 1, q_3)}(f, M_1+\epsilon_2 g, h)
\]
\[+ c'_w \text{PSF}^{(1, 1, q_3)}(M_1+\epsilon_1 f, g, h)
\]
\[+ c'_w \text{PSF}^{(1, 1, q_3)}(M_1+\epsilon_1 f, M_1+\epsilon_2 g, h),
\]
with \( c'_w \) appearing from (2.24) and the openness estimate. Then (2.23) and (2.25) imply the claimed boundedness. If \( \alpha_w \) is the constant appearing in (2.23), we see as a consequence that \( T: L^p(w) \times L^p(w) \to L^{p/2}(w) \) with constant
\[
C_w \lesssim (\alpha_w + c'_w) \cdot [w]_{A_p}^{\left( \frac{q_1}{p-q_1} \right) \left( \frac{1}{1-p} \right)} [w]_{A_p}^{\left( \frac{q_2}{p-q_2} \right) \left( \frac{1}{1-p} \right)}.
\]
Note that \( c'_w \) can also be computed explicitly in terms of \( [w]_{A_p} \), after solving for the \( \epsilon_1, \epsilon_2 \) used above and using (2.21).

### 2.4.2 Multiple Weights

We prove Corollaries 2.1 and 2.3. Suppose \( v_1, v_2, v_3 \) are strictly positive functions such that
\[
\prod_{i=1}^{3} \frac{1}{v_i^{\frac{1}{q_i}}} = 1
\]
for some \( q_i \geq 1 \) with \( \sum_{i=1}^{3} \frac{1}{q_i} = 1 \). Let \( \vec{p} = (p_1, p_2, p_3) \) be any tuple of exponents with \( 1 < p_i < q_i \), and recall that
\[
[w]_{A_{\vec{p}}} = \sup_Q \prod_{i=1}^{3} \left( \frac{1}{|Q|} \int_Q v_i^{\frac{p_i}{q_i-p_i}} \right)^{\frac{1}{p_i} - \frac{1}{q_i}},
\]
with the supremum taken over cubes $Q \subset \mathbb{R}^d$. We can argue as in the proof of Lemma 6.1 in [24] to prove that for any sparse collection $S$ we have

$$\text{PSF}_S^\mathcal{E}(f_1, f_2, f_3) \leq \left( c_{\vec{p}, \vec{q}}(\vec{v}) \right) \prod_{i=1}^3 \|f_i\|_{L^{q_i}(v_i)}$$

when $[\vec{v}]_{A^{\mathcal{E}}_q}$ is finite. Applying Theorem 2.1 yields the following result.

**Lemma 2.9.** Let $v_1, v_2, v_3$ and $\vec{p}, \vec{q}$ be as above. Also assume $[\vec{v}]_{A^{\mathcal{E}}_q} < \infty$. Then

$$|\langle T_\Omega(f_1, f_2), f_3 \rangle| \leq C_{\Omega, \vec{p}, \vec{q}}(\vec{v}) \prod_{i=1}^3 \|f_i\|_{L^{q_i}(v_i)}$$

for $f_i \in L^{q_i}(v_i)$.

The reader can consult [24] for the explicit value of the constant. Corollary 2.3 immediately follows from this lemma by duality. We now show that Corollary 2.1 is also a consequence.

By using multilinear extrapolation techniques from [30], [35], it is enough to prove the claimed estimate of Corollary 2.1 when $q_i = 3$ for each $i = 1, 2, 3$. Write $\vec{3} = (3, 3, 3)$, and suppose $\vec{p} = (1 + \epsilon, 1 + \epsilon, 1 + \epsilon)$ for some small $\epsilon > 0$. Observe that if $t > 0$ is chosen properly, then there is some absolute $C > 0$ such that

$$[v_i^{1+t}]_{A^3} \leq C[v_i]_{A^3}, \quad i = 1, 2. \quad (2.26)$$

This estimate can be proved using the openness property and reverse Hölder estimates as in the last section; see [22], Section 3.7, for a more general version of this inequality. Now Lemma 2.9 implies that

$$T_\Omega : L^3(v_1) \times L^3(v_2) \to L^{3/2}(v_3^{-1/2})$$

with operator norm bounded by $C_{\epsilon}[\vec{v}]_{A^{\mathcal{E}}_q}^3$. Hence it will be enough to prove

$$[\vec{v}]_{A^{\mathcal{E}}_q} \leq [v_1^{1+t}]_{A^3}^{1/3} \cdot [v_2^{1+t}]_{A^3}^{1/3} \cdot [v_3^{1+t}]_{A^3}^{1/3} \quad (2.27)$$

43
with \( 1 + t = \frac{2(1+\epsilon)}{2-\epsilon} \), since we can then apply (2.26) after choosing \( \epsilon \) correctly (as we are free to do). One can prove (2.27) by using Hölder’s inequality; we omit the details.

2.5 Proof of Theorem

2.5.1 Construction of the Sparse Collection

Suppose we are given a form \( \Lambda \) as in the statement of Theorem 2.2 with \( C_T = C_T(r_1, ..., r_m, \alpha) \), and \( \Lambda \) truncated to a finite (but otherwise arbitrary) number of scales. Given \( f_i \in L^{p_i}(\mathbb{R}^d) \) with compact support, we would like to construct a sparse collection of cubes \( S \) so that

\[
|\Lambda(f_1, ..., f_{m+1})| \leq c_d [C_L + C_T] \sum_{R \in S} |R| \prod_{i=1}^{m+1} (f_i)_{p_i, R}.
\]

We will generalize the iterative argument from [19]. The following lemma is crucial for the induction step.

**Lemma 2.10.** Let \( Q \) be a fixed dyadic cube and \( P \) a stopping collection with top \( Q \). If the estimates (2.8) hold, then

\[
|\Lambda^Q(h_1 1_Q, h_2 1_{3Q}, ..., h_{m+1} 1_{3Q})| \leq C_d |Q| \prod |h_i| \prod |y_{p_i}| \prod |y_{q_i}| \prod |y_{r_{m+1}}|
\]

\[
+ \sum_{L \in P} |\Lambda^L(h_1 1_L, h_2 1_{3L}, ..., h_{m+1} 1_{3L})|.
\]

**Proof.** We can assume \( \text{supp} h_1 \subset Q \) and \( \text{supp} h_i \subset 3Q \) for \( 2 \leq i \leq m + 1 \). By definition of \( \Lambda_P \) we
have

\[
\Lambda^{sQ}(h_1, h_2, \ldots, h_{m+1}) = \Lambda_P(h_1, h_2, \ldots, h_{m+1}) + \sum_{L \in P} \Lambda^{sL}(h_1 \mathbf{1}_L, h_2 \mathbf{1}_{3L}, \ldots, h_{m+1} \mathbf{1}_{3L}).
\]

so it is enough to estimate \( \Lambda_P(h_1, h_2, \ldots, h_{m+1}) \). For each \( j \) perform the Calderón-Zygmund (CZ) decomposition (2.6) of \( h_j \) with respect to \( P \). Then \( h_j = b_j + g_j \) with

\[
b_j = \sum_{L \in P} b_{jL}, \quad b_{jL} := (h_j - (h_j)_L) \mathbf{1}_L
\]

and

\[
\|g_j\|_{L^\infty} \lesssim \|h_j\|_{Y^{p_j}}, \quad \|b_j\|_{L^{p_j}} \lesssim \|h_j\|_{Y^{p_j}}.
\]

We now decompose \( \Lambda_P(h_1, h_2, \ldots, h_{m+1}) \) using the CZ decomposition. If we let \( f_{h_i}(0) = g_i \) and \( f_{h_i}(1) = b_i \), we see that the form breaks up into the \( 2^{m+1} \) terms

\[
\sum_{0 \leq k_1, \ldots, k_{m+1} \leq 1} \Lambda_P(f_{h_1}(k_1), f_{h_2}(k_2), \ldots, f_{h_{m+1}}(k_{m+1})). \tag{2.28}
\]

From the definition of \( \Lambda_P \) and the pairwise disjointness of elements of \( P \), we see that finiteness of \( C_T \) implies

\[
|\Lambda_P(g_1, g_2, \ldots, g_{m+1})| \leq C_T \|g_1\|_{L^{r_1}} \cdots \|g_m\|_{L^{r_m}} \|g_{m+1}\|_{L^{\alpha'}}
\]

\[
+ C_T \sum_{L \in P} \|g_1 \mathbf{1}_L\|_{L^{r_1}} \cdots \|g_{m+1} \mathbf{1}_{3L}\|_{L^{\alpha'}}
\]

\[
\leq c_dC_T|Q| \|g_1\|_{Y^{r_1}} \|g_2\|_{Y^{r_2}} \cdots \|g_{m+1}\|_{Y^{\alpha'}}
\]

\[
\leq c_dC_T|Q| \|h_1\|_{Y^{r_1}} \|h_2\|_{Y^{r_2}} \cdots \|h_{m+1}\|_{Y^{\alpha_{m+1}}}.
\]
Here we’ve used
\[ \frac{1}{r_1} + \ldots + \frac{1}{r_m} + \frac{1}{\alpha'} = \frac{1}{\alpha} + \frac{1}{\alpha'} = 1, \]
the disjointness of \( L \in \mathcal{P} \), and the \( \| g_j \|_{y\infty} \) bound. We can now use (2.8) and the CZ estimates to control the remaining terms by the desired quantity. For example,

\[
|\Lambda_{\mathcal{P}}(b_1, g_2, \ldots, g_{m+1})| \leq C_L |Q| \| b_1 \|_{\hat{y}_{p_1}} \| g_2 \|_{y_{p_2}} \ldots \| g_{m+1} \|_{y_{p_{m+1}}} \\
\leq c_d C_L |Q| \| h_1 \|_{y_{p_1}} \| h_2 \|_{y_{p_2}} \ldots \| h_{m+1} \|_{y_{p_{m+1}}},
\]
by the first estimate in (2.8) and the CZ properties. The other estimates are similar. We use the last estimate in (2.8) to control the single term \( \Lambda_{\mathcal{P}}(g_1, g_2, \ldots, g_m, b_{m+1}) \):

\[
|\Lambda_{\mathcal{P}}(g_1, g_2, \ldots, g_m, b_{m+1})| \leq C_L |Q| \| g_1 \|_{y\infty} \| g_2 \|_{y\infty} \ldots \| g_m \|_{y\infty} \| b_{m+1} \|_{\hat{y}_{p_{m+1}}}
\leq c_d C_L |Q| \| h_1 \|_{y_{p_1}} \| h_2 \|_{y_{p_2}} \ldots \| h_m \|_{y_{p_{m}}} \| h_{m+1} \|_{y_{p_{m+1}}},
\]

To bound the remaining terms, we choose from the first \( m \) inequalities in (2.8). If there are more than one \( b_i \) terms in the particular piece of (2.28) we wish to control, we let \( b_n \) be the first mean-zero term that appears (ordered from left to right) and then apply the \( n \)-th estimate in the list (2.8), ordered from top to bottom. Then applying the CZ properties as above yields the desired estimates.

The construction of the collection \( \mathcal{S} \) now proceeds as in [19], with minor changes to account for the addition of more functions. For this reason we sketch the proof here, and send the reader to [19] for the finer details.

We will construct the required sparse collection \( \mathcal{S} \) iteratively by decomposing exceptional sets of the form
\[
E_Q := \left\{ x \in 3Q : \max_{i=1,\ldots,m+1} \frac{M_{p_i}(f_i 1_{3Q})(x)}{(f_i)_{p_i,3Q}} \geq C_d \right\},
\]
where $C_d$ is some constant depending on the dimension. Notice that if $C_d$ is chosen large enough, then by the maximal theorem we can assume

$$|E_Q| \leq 2^{-cd}|Q| \quad (2.30)$$

for some uniform $c > 0$.

We begin the argument by fixing $f_i \in L^p_j(\mathbb{R}^d)$ with compact support, $i = 1, \ldots, m + 1$. We may assume we have chosen our dyadic lattice $\mathcal{D}$ so that there is $Q_0 \in \mathcal{D}$ with $\text{supp}(f_1) \subset Q_0$ and $\text{supp}(f_i) \subset 3Q_0$ for $i = 2, \ldots, m + 1$, such that $s_{Q_0}$ is bigger than the largest scale occurring in the truncation of $\Lambda$. We let $S_0 = \{Q_0\}$ and $E_0 = 3Q_0$, and then define (using definition (2.29))

$$E_1 := E_{Q_0}, \quad S_1 := \text{ maximal cubes } L \in \mathcal{D} \text{ such that } 9L \subset E_1.$$

It is easy to verify that $\mathcal{P}_1(Q_0) := S_1$ is a stopping collection with top $Q_0$, such that $|Q_0 \setminus E_1| \geq (1 - 2^{-cd})|Q_0|$. By maximality and the definition of the exceptional set $E_{Q_0}$ we have

$$\|f_i\|_{Y_{p_j}(\mathcal{P}_1(Q_0))} \leq C_d(f_i)_{p_i,3Q_0}, \quad i = 1, \ldots, m + 1,$$

so applying Lemma 2.10 yields

$$|\Lambda(f_1, f_2, \ldots, f_{m+1})| = |\Lambda^{s_{Q_0}}(f_11_{Q_0}, f_21_{3Q_0}, \ldots, f_{m+1}1_{3Q_0})| \leq C_d|Q_0| \prod_{i=1}^{m+1} (f_i)_{p_i,3Q_0}$$

$$+ \sum_{L \in \mathcal{P}_1(Q_0) \setminus Q_0} |\Lambda^{s_L}(f_11_L, f_21_{3L}, \ldots, f_{m+1}1_{3L})|.$$
above argument with $L$ in place of $Q_0$, defining

$$E_2 = \bigcup_{L \in \mathcal{P}_1(Q_0)} E_L$$

and

$$S_2 := \text{maximal cubes } R \in \mathcal{D} \text{ such that } 9R \subset E_2.$$ 

For each $L \in \mathcal{P}_1(Q_0)$ we also define

$$\mathcal{P}_2(L) = \{ R \in S_2 : R \subset 3L \},$$

which can be shown to be a stopping collection with top $L$ (the argument is the same as in [19]). Now we apply Lemma 2.10 to estimate each piece of the sum

$$\sum_{\substack{L \in \mathcal{P}_1(Q_0) \\ L \subset Q_0}} |\Lambda^s_L(h_1 1_L, h_2 1_{3L}, ..., h_{m+1} 1_{3L})|,$$

using the stopping collection $\mathcal{P}_2(L)$ in the application to the term

$$|\Lambda^s_L(f_1 1_L, f_2 1_{3L}, ..., f_{m+1} 1_{3L})|.$$ 

We then repeat the process just described at the next level.

Since we are working with finitely many scales the process eventually terminates with the desired sparse form bound. The sparse collection $S$ consists of all cubes chosen in the various stopping collections constructed along the way. Condition (2.30) at each level guarantees the sparsity of these cubes.
2.5.2 Multiplication Operators

Let $A_\phi : L^{r_1} \times \ldots \times L^{r_m} \to L^\alpha$ be the multiplication operator described in Remark 2.1, with $\phi \in L^\infty$ and $\frac{1}{r_1} + \ldots + \frac{1}{r_m} = \frac{1}{\alpha}$. We let

$$A(f_1, \ldots, f_m, f_{m+1}) = \langle A_\phi(f_1, \ldots, f_m), f_{m+1} \rangle.$$

Using a simpler version of the stopping-time argument from the previous section, it is easy to see that

$$|A(f_1, \ldots, f_m, f_{m+1})| \lesssim \sup_S \text{PSF}_{S}^{(1, \ldots, 1)}(f_1, \ldots, f_{m+1})$$

when $f_i \in L^{r_i} \cap L^\infty$ with compact support. Let $P$ be any stopping collection with top $Q$ and $b = \sum_{L \in P} b_L$ with $b_L$ supported on $L$, and observe that if

$$A_P(b, g_2, \ldots, g_{m+1}) := A(b1_{3Q}, g_21_{3Q}, \ldots, g_{m+1}1_{3Q})$$

$$- \sum_{R \in P \subset Q} A(b1_{3R}, g_21_{3R}, \ldots, g_{m+1}1_{3R})$$

then in fact

$$A_P(b, g_2, \ldots, g_{m+1}) = 0.$$

Similarly $A_P(g_1, b, g_3, \ldots, g_{m+1}), \ldots, A_P(g_1, \ldots, g_m, b)$ all vanish. Hence we can repeat the argument given in the proof of Theorem 2.2, avoiding most of the technical complications. In particular, if $f_i = g_i + b_i$ with $g_i$ the ‘good’ terms from the Calderón-Zygmund decomposition relative to $P$, then we can argue as in the proof of Lemma 2.10 to show

$$|A_P(g_1, \ldots, g_{m+1})| \lesssim \|\phi\|_{L^\infty} |Q| \|f_1\|_{Y_1} \|f_2\|_{Y_1} \ldots \|f_{m+1}\|_{Y_1}.$$

We just saw that all other terms of $A_P(g_1 + b_1, \ldots, g_{m+1} + b_{m+1})$ vanish, so we can easily run the stopping-time argument given in the last section to prove the claimed $\text{PSF}_{S}^{(1, \ldots, 1)}$ bound.
Appendix: Adjoint Forms

We prove Proposition 2.1. The argument is almost identical to the proof of the linear variant from [19], so we only provide a sketch. Below we will call two dyadic cubes \( L, R \) neighbors and write \( L \sim R \) if \( 7L \cap 7R \neq \emptyset \) and \( |s_L - s_R| < 8 \). By separation property (i), if \( L, R \in \mathcal{P} \) are distinct cubes with \( 7L \cap 7R \neq \emptyset \), \( L \sim R \).

Let \( \mathcal{P} \) be a stopping collection with top \( Q \). Let \( b = \sum_{L \in \mathcal{P}} b_L \) as in the last sections. We want to show that

\[
\Lambda_{\mathcal{P}}(g_1, b, g_3, \ldots, g_{m+1}), \ldots, \Lambda_{\mathcal{P}}(g_1, \ldots, g_m, b)
\]

can be decomposed in the same way as \( \Lambda_{\mathcal{P}}(b, g_2, \ldots, g_{m+1}) \), up to a controllable error term. We first analyze \( \Lambda_{\mathcal{P}}(g_1, \ldots, g_m, b) \). Below we assume \( g_1 \) is supported in \( Q \) and \( g_i \) is supported in \( 3Q \) for \( i \geq 2 \).

As in the appendix in [19], split \( b = b^{\text{in}} + b^{\text{out}} \) with

\[
b^{\text{in}} = \sum_{\substack{L \in \mathcal{P} \colon s_L \notin 3L \cap 2Q \neq \emptyset}} b_L.
\]

Fix any \( R \subset Q \). Using the support of the kernel, it is easy to see that if \( s < s_R \) then

\[
\int_{\mathbb{R}^{(m+1)d}} K_s(x_1, \ldots, x_m) g_1(x_1) 1_R(x_1) g_2(x_2) \ldots g_m(x_m) b^{\text{out}}(x_{m+1}) dx_1 \ldots dx_{m+1}
\]

is identically zero. This is because \( \text{dist}(\text{supp } b^{\text{out}}, R) \geq l(R)/2 \), since \( b^{\text{out}} \) is supported on \( L \) with \( 3L \cap 2Q = \emptyset \), but \( R \subset Q \). Therefore

\[
\Lambda_{\mathcal{P}}(g_1, \ldots, g_m, b^{\text{out}}) = \int_{\mathbb{R}^{(m+1)d}} \sum_{i,j} K_{s_Q}^{ij}(\vec{x}) g_1(x_1) \ldots g_m(x_m) b^{\text{out}}(x_{m+1}) d\vec{x}
\]

\[
- \sum_{R \in \mathcal{P}} \int_{\mathbb{R}^{(m+1)d}} \sum_{i,j} K_{s_R}^{ij}(\vec{x}) g_1 1_R(x_1) \ldots g_m(x_m) b^{\text{out}}(x_{m+1}) d\vec{x}
\]

50
There is only one scale in the kernel in each integral, so we can estimate each term as in Lemma 2.2 and sum over $R$ to see that

$$|\Lambda_P(g_1, \ldots, g_m, b^{\text{out}})| \lesssim [K] |Q| \|b\| x_i \|g_1\| y_p \ldots \|g_m\| y_p.$$  

We now have to analyze $\Lambda_P(g_1, \ldots, g_m, b^{\text{in}})$. As in the appendix of [19], it is enough to show that

$$\Lambda_P(g_1, \ldots, g_m, b^{\text{in}}) = \left( \Lambda^s_Q(g_1, \ldots, g_m, b^{\text{in}}) - \sum_{L \in \mathcal{P}} \Lambda^s_L(g_1, \ldots, g_m, b_L) \right) + \phi(g_1, \ldots, g_m, b),$$  

(2.31)

where $\phi$ is some error term satisfying the same estimate as $|\Lambda_P(g_1, \ldots, g_m, b^{\text{out}})|$. Then we can decompose the term in parenthesis as in Section 2.2, and use symmetry of the kernel to prove the desired estimates for $\Lambda_P(g_1, \ldots, g_m, b)$.

The proof of (2.31) is almost the same as the linear case in [19]. The main observation is that

$$\sum_{R \in \mathcal{P}} \Lambda^s_R(g_11_R, g_2, \ldots, g_m, b^{\text{in}}) = \sum_{R \in \mathcal{P}} \sum_{L \in \mathcal{P}} \Lambda^s_R(g_11_R, g_2, \ldots, g_m, b_L),$$  

(2.32)

since $\Lambda_R(g_11_R, g_2, \ldots, g_m, b_L) = 0$ if $3L \cap 3R = \emptyset$. In particular this implies that $L \sim R$, so if $R$ is fixed then the number of terms in the second sum in (2.32) is bounded by a universal dimensional constant. It is easy to show that

$$|\Lambda^s_R(g_11_R, g_2, \ldots, g_m, b_L) - \Lambda^s_L(g_11_R, g_2, \ldots, g_m, b_L)| \lesssim |L| [K] |Q| \|g_1\| y_p \ldots \|g_m\| y_p \|b\| x_i,$$

since $|s_L - s_R| < 8$ (apply single-scale estimates, noting that the number of such estimates will be bounded by a dimensional constant). With the help of the separation properties we then can
replace each $\Lambda^s_r$ in (2.32) by $\Lambda^s_l$, up to an admissible error term $\phi$. Now repeat the rest of the argument from [19], with trivial changes to account for the addition of more functions, to show that the remaining terms are of the form (2.31).

Also observe that there was nothing special about the choice of the position of $b$ in the above argument, so the same reasoning applies to $\Lambda_P(g_1, b, g_3, ..., g_{m+1}), \Lambda_P(g_1, g_2, b, g_4, ..., g_{m+1})$, etc. This proves Proposition 2.1, up to the trivial steps we have omitted.
Chapter 3

Sparse Bounds in the
Multi-Parameter Setting

Despite all of the recent activity surrounding sparse domination of operators in harmonic analysis, little is known about how to appropriately generalize 'sparse bounds' to the product setting where $T$ is a singular operator that commutes with a multi-parameter family of dilations. Some simple examples of these operators include the bi-parameter Hilbert transform

$$H \otimes H f(x, y) = \text{p.v. } f \ast \frac{1}{xy},$$

the strong maximal function

$$\mathcal{M}_S f(x) = \sup_{x \in R} \frac{1}{|R|} \int_{R} |f(y)|dy$$

(where $R$ are axis-parallel rectangles), and multi-parameter Haar multipliers (see below for a precise definition). The analysis of this family of operators dates back to work of Chang, Fefferman, Journé, and Stein (see for example [12], [45]). In recent years, multi-parameter harmonic analysis and its generalizations have been used to study problems related to discrepancy theory (see Chapter 4), PDE (bi-parameter Leibniz estimates, hypoelliptic operators [59], [66]), and
problems in several complex variables [61]. A major theme in the field is that one cannot simply ‘iterate’ results from the one-parameter theory to fully understand multi-parameter operators. In particular, the one-parameter sparse theory says essentially nothing about the possible (or impossible) extensions of bounds of type (1.1) to the multi-parameter setting, even in the product setting on low-dimensional Euclidean space.

This chapter is devoted to some attempts at understanding the extent to which sparse bounds are possible for multi-parameter operators. In Chapter 4 we provide a counterexample demonstrating that the (dyadic) strong maximal function does not admit a sparse bound with respect to $L^1$-type forms.

### 3.1 Some Difficulties

As a first step one could hope that it would be possible to ‘iterate’ the one-parameter sparse estimates to obtain some positive results for tensor-product type operators. Unfortunately there does not appear to be a straightforward way to carry out this argument. Suppose that

$$Tf(x_1, x_2) = \text{p.v.} \int_{\mathbb{R}^2} \frac{f(y_1, y_2)}{(x_1 - y_1)(x_2 - y_2)} dy_1 dy_2,$$

so that $T = H_1 \otimes H_2$ with $H_1$ a Hilbert transform in the $x_1$ direction and $H_2$ a Hilbert transform in the $x_2$ direction. If we fix the variable $x_1$ then we can apply the one-parameter pointwise sparse bound for $H$ to show that there is a sparse collection of intervals $S_{x_1}$ such that

$$|Tf(x_1, x_2)| \lesssim \sum_{Q \in S_{x_1}} (|H_1 f(x_1, \cdot)|)_Q 1_Q(x_2)$$

for almost every $x_2$. However, the collection of intervals depends on $x_1$ and so there is no obvious way to iterate this estimate to get a sparse bound for the full operator $H_1 \otimes H_2$.

We therefore need to find a more direct approach that does not rely on the one-parameter results. It is clear that any analogue of the one-parameter sparse bound must involve collections
of rectangles rather than cubes, due to the underlying geometry of bi-parameter singular integrals like $H_1 \otimes H_2$. At this point one encounters substantial difficulties adapting the one-parameter methods. For example, to prove a sparse-form estimate of type (1.1) we construct the sparse collection of cubes using a stopping-time argument that is intimately related to the Calderón-Zygmund decomposition and the Hardy-Littlewood maximal operator. In the bi-parameter setting the natural maximal operator to work with is the strong maximal function

$$M_S f(x) = \sup_{x \in R} \frac{1}{|R|} \int_R |f(y)| dy,$$

where the supremum is taken over axis-parallel rectangles containing $x$. However, this operator lacks the martingale structure that enables the type of stopping-time arguments used in the one-parameter setting. In the bi-parameter setting, for both singular integral theory and martingale theory, it turns out that the square function is a more natural operator to work with. For example, in [5] Bernard used the square function to obtain an atomic decomposition for bi-parameter Hardy spaces and to prove $H^1 - BMO$ duality. These results were later extended to the continuous setting of bi-parameter singular integrals, Hardy spaces, and product $BMO$ in [14] by Chang and R. Fefferman. Once again square functions played a key role in their arguments.

There is another related geometric problem one encounters: in the bi-parameter setting, it is also not immediately clear what the proper definition of a ‘sparse collection of rectangles’ should be. In this chapter and the next we will work with the following two notions of ‘sparse collection,’ which turn out to be equivalent:

**Definition 3.1.** A collection $S$ of rectangles in $\mathbb{R}^n$ is said to be *sparse in the disjoint-pieces sense* if there is some $\eta > 0$ such that for all $R \in S$ there is $E_R \subset R$ with $|E_R| > \eta |R|$, and such that if $R \neq R'$ then $E_R \cap E_{R'} = \emptyset$.

**Definition 3.2.** A collection $S$ of rectangles in $\mathbb{R}^n$ satisfies the *Carleson packing condition* if
there is some $\Lambda > 0$ such that for all open sets $U \subset \mathbb{R}^n$,

$$\sum_{R \in S \atop R \subset U} |R| \leq \Lambda |U|.$$  

The structure of the packing condition in definition 3.2 is natural in light of the fact that the definition of bi-parameter $BMO$ requires a similar packing condition on Haar or wavelet coefficients relative to rectangles contained in open sets (see the next section for more on bi-parameter $BMO$ spaces). Both definitions are equivalent for collections of cubes [55] and much more generally, as shown by Dor [28]. We include a proof of this equivalence in an appendix at the end of the chapter.

### 3.2 An Approach Using Square Functions

Let $S$ be the dyadic bi-parameter square function given by

$$Sf(x)^2 = \sum_{R \in \mathcal{D}} |\langle f, h_R \rangle|^2 \frac{1_R(x)}{|R|},$$

where $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$ is the collection of dyadic rectangles in $\mathbb{R}^2$ (relative to two grids $\mathcal{D}_1$ and $\mathcal{D}_2$ in $\mathbb{R}$), and $h_R$ is the bi-parameter Haar function associated to $R$. Recall that if $R = I \times J$ then

$$h_R(x_1, x_2) = h_I(x_1)h_J(x_2), \quad h_I = \frac{1}{|I|^{1/2}} (1_{I_l} - 1_{I_r}),$$

where $I_l$ and $I_r$ are the left and right children of $I$. Also recall that the functions $h_R$ form a basis of $L^2(\mathbb{R})$ for any $\mathcal{D}$. Given the role of square functions in the bi-parameter theory, it is natural to attempt to prove a sparse form bound of the type

$$|\langle Tf, g \rangle| \lesssim \sum_{R \in S} |R|(Sf)_R(Sg)_R,$$  \hspace{1cm} (3.1)
where $S$ is a collection of rectangles that satisfies either the disjoint-pieces or Carleson packing condition.

We begin by studying the bi-paramater martingale transform

$$Tf = \sum_{R \in \mathcal{D}} \epsilon_R \langle f, h_R \rangle h_R, \quad \sup_R |\epsilon_R| \leq C.$$ 

The following theorem can be proved by a simple argument using Cauchy-Schwarz and the sparse bound for the identity operator (and in fact this proof gives a collection of cubes, not just rectangles).

**Theorem 3.1.** Let $T$ be the bi-parameter martingale transform defined above, and suppose $f$ and $g$ are functions with finitely many Haar coefficients. Then there exists a collection of rectangles $S$ that is sparse in the sense of Definitions 3.1 and 3.2 such that

$$|\langle Tf, g \rangle| \lesssim (\sup_R |\epsilon_R|) \sum_{R \in \mathcal{S}} |R|(Sf)_R(Sg)_R.$$  \hspace{1cm} (3.2)

The implicit constant does not depend on $f$ or $g$.

In the next section we will present a geometric proof of this theorem based on the the underlying bi-parameter geometry. The hope is that this argument is more robust and will be useful in contexts where $T$ is more complicated than a Haar multiplier, and where a simple Cauchy-Schwarz argument does not suffice.

The methods used to prove Theorem 3.1 generalize to the case where $T$ is a bi-parameter dyadic shift as long as we replace the dyadic square functions $S$ by certain shifted square functions $S^{i,j}$. Since the statement of this result is somewhat technical, we defer the detailed definitions until Section 3.5. As a consequence of Martikainen’s representation theorem [58], we are then able to deduce a type of sparse bound for paraproduct-free bi-parameter singular integrals belonging to the Journé class. See Corollary 3.1 for a precise statement of this result.

The one-parameter theory indicates that we should be able to easily prove weighted estimates once we have established a sparse bound. This is still the case for the square-function sparse form.
estimate (3.1), and we derive weighted corollaries of our main results in Sections 3.6 and 3.7. It is also straightforward to track the dependence of the constants on the $A_p$ characteristic of the weight. However, due to the addition of the square functions $S$ and some extra complications related to the strong maximal function, this approach does not give weighted estimates that are sharp in terms of the $A_p$ characteristic (see Section 3.6 for definitions). Nevertheless, our sparse bounds provide an alternative approach to proving $A_p$ estimates for bi-parameter martingale transforms and cancellative dyadic shifts (see [39] for another recent method).

3.2.1 Remarks on Theorem 3.1

(1) For simplicity the results and the proofs are stated for $\mathbb{R} \times \mathbb{R}$, but our methods all extend directly to the product space $\mathbb{R}^n \times \mathbb{R}^m$ once suitable modifications are made to the definition of the Haar functions. We also do not see any obstacles to carrying out the arguments below in the multi-parameter setting.

(2) The sparse bound (3.1) is true in the one-parameter setting when we are working with intervals (or cubes), but in a stronger sense. That is, (3.1) holds with localized square functions, so that

$$|\langle Tf, g \rangle| \lesssim \sum_{I \in S} |I|(S_I f)_I(S_I g)_I.$$ 

Here the square function $S_I$ only involves dyadic intervals $J$ contained in $I$ (see Theorem 15 in [4]). This localized square-function sparse bound cannot hold in the bi-parameter setting, as observed by M. Lacey and Y. Ou [50]. We prove this claim in the following section.

(3) It is clear from our proofs of the weighted corollaries in Sections 3.6 and 3.7 that there are still significant obstacles to overcome if we wish to develop a sharp weighted theory for multi-parameter operators by using sparse domination (see, for example, the comments after the proof of Theorem 3.3 and the first appendix). A different notion of ‘sparse operator’ in the multi-parameter setting may be needed, possibly one that allows us to circumvent the obstructions caused by the strong maximal function.
3.3 The Lacey-Ou Example

Let $T$ be a bi-parameter Haar multiplier, and let $S_Rf$ be the localized bi-parameter dyadic square function

$$S_Rf(x)^2 = \sum_{L \subseteq R} |\langle f, h_L \rangle|^2 \frac{1_L(x)}{|L|}.$$ 

Here $R$ and $L$ are dyadic rectangles with $h_L$ the associated product Haar function. We show that in general there can be no sparse bound of the type

$$|\langle Tf, g \rangle| \leq C \sum_{R \in S} |R|(S_Rf)_R(S_Rg)_R,$$

where the collection $S$ is $\eta$-sparse with respect to some fixed but arbitrary $0 < \eta < 1$. We learned of this example from Michael Lacey and Yumeng Ou. Before continuing it will be helpful to review the definitions of two bi-parameter BMO spaces. Recall that dyadic $BMO_{rect}$ is the space of all functions such there is some $\Lambda > 0$ with

$$\sum_{L \subseteq R} |\langle f, h_L \rangle|^2 \leq \Lambda |R|$$

for all dyadic rectangles $R$, with $\|f\|_{BMO_{rect}}$ the smallest such $\Lambda$. Also recall that dyadic $BMO_{prod}$ is the space of functions such that there is some $\Lambda' > 0$ with

$$\sum_{L \subseteq \Omega} |\langle f, h_L \rangle|^2 \leq \Lambda' |\Omega|$$

for all open sets $\Omega$, with $\|f\|_{BMO_{prod}}$ the smallest such $\Lambda'$. One has strict containment $BMO_{prod} \subsetneq BMO_{rect}$, as can be shown by using the function considered below.

We will assume that $T$ is just the identity operator (though suitable modifications show that this argument works for any Haar multiplier $Tf = \sum_R \epsilon_R \langle f, h_R \rangle h_R(x)$ with $\epsilon_R \in [\lambda^{-1}, \lambda]$ for some fixed $\lambda > 0$.) Recall that for each $N > 1$ we can construct a collection $C$ of rectangles in the unit cube $Q$ with the following properties:
• $\sum_{L \in C} |L| = 1$

• If $R_0$ is any dyadic rectangle then $\sum_{L \subseteq R_0} |L| \leq |R_0|$

• For $\Omega = \bigcup_{L \in C} L$ we have $|\Omega| \sim 1/N$

This is the Carleson ‘quilt’ example that shows there is a distinction between rectangle and product BMO (see for example Chapter 3 in [60] for a construction). Now assume for a contradiction that the localized-square-function sparse bound holds, with some uniform constant $C$ and sparse parameter $\eta$. We will also make the extra assumption that the rectangles from the sparse collection are contained in $3Q$. Note that in the one-parameter setting one typically gets a sparse collection contained in some dilate of the support of the functions. It is, however, possible to remove this assumption; see below.

For an arbitrary $N > 1$ we take

$$f = \sum_{L \in C} |L|^{1/2} h_L(x),$$

a normalized Haar series adapted to $C$ with $|\Omega| \sim 1/N$ (this is the same function that demonstrates the difference between $BMO_{\text{rect}}$ and $BMO_{\text{prod}}$). If the localized sparse bound held we would have

$$1 = \langle f, f \rangle \lesssim \sum_{R \in S} |R|(S_R f)_R^2$$

$$\lesssim \sum_{R \in S} \begin{cases} |R|(S_R f)_R^2 & |R \cap \Omega| \geq \alpha |R| \\ |R|(S_R f)_R^2 & |R \cap \Omega| < \alpha |R| \end{cases}$$

$$:= I + II,$$

where $0 < \alpha < 1$ is some parameter to be chosen later. To estimate $I$, observe that by Cauchy-
Schwarz and the definition of $S_R$ we must have

$$I \lesssim \sum_{R \in S} \sum_{|L| \leq \alpha|R| \cap \Omega} |L| \lesssim \sum_{R \in S} |R|,$$

the second inequality following from the rectangle packing property of $C$. Hence by the sparseness of the collection $S$ we conclude

$$I \lesssim |\{M(1_\Omega) > \alpha\}| \lesssim \alpha^{-1} \log(\alpha^{-1})|\Omega|.$$

To estimate $II$, we first define $C_R = \bigcup_{L \subset R \cap \Omega} L$, and then observe that by Cauchy-Schwarz (as before)

$$II \lesssim \sum_{R \in S} \frac{|C_R|}{|R|} \sum_{L \subset R \cap \Omega} |L| \lesssim \sum_{R \in S} |C_R|.$$

But for each $R \in S$ we have $C_R \subset R \cap \Omega$, and therefore $|C_R| < \alpha|R|$ in the above sum. Since the collection $S$ is sparse and contained in $3Q$ we conclude

$$II \lesssim \alpha \sum_{R \in S} |R| \lesssim \alpha|Q| \lesssim \alpha.$$

In summary we have shown that

$$1 \lesssim \alpha^{-1} \log(\alpha^{-1})|\Omega| + \alpha,$$

which translates to

$$N \lesssim \alpha^{-1} \log(\alpha^{-1}) + \alpha N$$

for any $0 < \alpha < 1$. Now take $\alpha = N^{-1/2}$ to get

$$N \lesssim N^{1/2} \log(N^{1/2}) + N^{1/2}.$$
We can assume that the implicit constant is negligible compared to $N$ (by taking $N$ large enough in the beginning), so this is the desired contradiction.

To remove the assumption that rectangles from $S$ are contained in $3Q$, one can further decompose the sum $II$ according to pieces with both $|R \cap \Omega| \leq \alpha |R|$ and $|R \cap Q| \sim 2^{-k}|R|$ for $k \geq 0$. In this case $C_R \subset R \cap Q \cap \Omega = R \cap \Omega$, so that

$$|C_R| \leq |R \cap \Omega| \leq |R \cap \Omega|^{1/4}|R \cap Q|^{3/4} \leq 2^{-3k/4}\alpha^{1/4}|R|.$$  

Hence these extra terms contribute a term $c \cdot \alpha^{1/4}$ to the final estimate (after summing over $k \geq 1$). This gives

$$1 \lesssim \alpha^{-1} \log(\alpha^{-1})|\Omega| + \alpha + \alpha^{1/4},$$

or

$$N \lesssim \alpha^{-1} \log(\alpha^{-1}) + \alpha N + \alpha^{1/4}N.$$  

Taking $\alpha = N^{-1/2}$ as before gives

$$N \lesssim N^{1/2} \log(N^{1/2}) + N^{1/2} + N^{7/8},$$

which is once again a contradiction.

### 3.4 The Bi-Parameter Martingale Transform

Fix two dyadic lattices $\mathcal{D}_1, \mathcal{D}_2$ in $\mathbb{R}$ and let $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$ be the associated dyadic rectangles in $\mathbb{R}^2$. We give a geometric proof of the square-function sparse form bound claimed in Theorem 3.1 for the bi-parameter martingale transform

$$Tf = \sum_{R \in \mathcal{D}} \epsilon_R \langle f, h_R \rangle h_R,$$  

(3.3)
where as above \( \sup_R |e_R| \leq C \). The argument begins by decomposing the form \( \langle Tf, g \rangle \) according to the Chang-Fefferman variant of the Calderón-Zygmund decomposition from [13]. We then select a certain sparse collection of rectangles using the Córdoba-Fefferman algorithm from [20], and further decompose the operator in terms of these rectangles. The structure of the square function allows us to absorb the ‘error’ terms (i.e., the rectangles not belonging to the sparse collection).

There are a few similarities between our basic approach and the standard sparse domination scheme in the one-parameter setting. For example, the bi-parameter analogue of the Calderón-Zygmund decomposition plays an important role in the first step. Additionally, we select the sparse collection of rectangles via a covering lemma that is equivalent to the boundedness of the strong maximal function; in the one-parameter setting, sparse cubes are typically chosen via a similar covering lemma associated to the Hardy-Littlewood maximal function.

### 3.4.1 The Geometric Proof of Theorem 3.1

We fix two test functions \( f, g \) on \( \mathbb{R}^2 \) with support in some large cube \( Q_0 \), and assume there are only finitely many dyadic rectangles \( R \) with \( \langle f, h_R \rangle \) or \( \langle g, h_R \rangle \) nonzero. Let \( \alpha_f = c \cdot (Sf)_{Q_0} \) and \( \alpha_g = c \cdot (Sg)_{Q_0} \), where \( c \) is some large constant. Define

\[
\Omega_0 = \{ x \in Q_0 : Sf(x) > \alpha_f \} \cup \{ x \in Q_0 : Sg(x) > \alpha_g \},
\]

and assume \( c \) has been chosen so that \( |\Omega_0| \leq \frac{1}{2} |Q_0| \). Let \( R_0 \) be the collection of rectangles \( R \) such that \( |R \cap \Omega_0| < \frac{1}{2} |R| \), and for positive integers \( k \) define

\[
\Omega_k = \{ x \in Q_0 : Sf(x) > 2^k \alpha_f \} \cup \{ x \in Q_0 : Sg(x) > 2^k \alpha_g \}.
\]

Also set

\[
F_k = \{ R : |R \cap \Omega_k| > \frac{1}{2} |R| \text{ and } |R \cap \Omega_{k+1}| \leq \frac{1}{2} |R| \}.
\]
We begin with the case where $\epsilon_R = 1$ for all $R$. We wish to estimate

$$\sum_R \alpha_R = \sum_{R \in \mathcal{R}_0} \alpha_R + \sum_k \sum_{R \in \mathcal{F}_k} \alpha_R$$

by a sparse form (with square function averages). Observe that since $|\Omega_k| \to 0$ as $k \to \infty$ there are only finitely many $\mathcal{F}_k$ that contribute to the sum (recall that $f, g$ have only finitely many nonzero Haar coefficients). Therefore it suffices to fix a large $N$ and bound

$$\sum_{R \in \mathcal{R}_0} \alpha_R + \sum_{k=0}^N \sum_{R \in \mathcal{F}_k} \alpha_R := I + II$$

by a sparse form, provided all constants are independent of $N$. Note that $I$ corresponds to the ‘good’ piece in the Chang-Fefferman variant of the Calderón-Zygmund decomposition, and $II$ corresponds to the ‘bad’ piece. The estimate for $I$ is straightforward:

$$\left| \sum_{R \in \mathcal{R}_0} \langle f, h_R \rangle \langle g, h_R \rangle \right| \leq \sum_{R \in \mathcal{R}_0} \int_{R \cap \Omega_0^0} |\langle f, h_R \rangle \langle g, h_R \rangle| \frac{1_{R \cap \Omega_0^0}(y)}{|R \cap \Omega_0^0|} \, dy \leq 2 \sum_{R \in \mathcal{R}_0} \int_{R \cap \Omega_0^0} |\langle f, h_R \rangle \langle g, h_R \rangle| \frac{1_{R}(y)}{|R|} \, dy \leq |Q_0|(Sf)_{Q_0}(Sg)_{Q_0}. \tag{3.4}$$

The last inequality follows from Cauchy-Schwarz and the definition of $\Omega_0$. To handle the remaining term $II$, we construct a sparse collection of rectangles using the Córdoba-Fefferman selection algorithm from [20], and decompose $II$ in terms of these rectangles.

Fix $\beta \in (0, 1)$ and begin at level $N$. Order the rectangles $\{R_i\}$ in $\mathcal{F}_N$ according to size (for example), and set $R^*_1 = R_1$. Proceeding inductively, choose those $R^*_k$ such that

$$|R^*_k \cap \bigcup_{j<k} R^*_j| < \beta |R^*_k|,$$

64
and such that \( R_k^* \) is minimal with this property relative to the initial order. Relabel the collection \( \{ R_k^* \} \) as \( \{ R_k^{(N)} \} \) (the rectangles in the collection at level \( N \)). Now suppose we have added rectangles to the collection up until level \( l + 1 \). Let \( \Lambda^{l+1} \) denote the union of all rectangles added to this point. Order the rectangles in \( \mathcal{F}_l \) as before, and let \( R_1^{(l)} \) be the first rectangle relative to this order such that

\[
|R_1^{(l)} \cap \Lambda^{l+1}| < \beta |R_1^{(l)}|.
\]

Inductively, choose \( R_k^{(l)} \) such that

\[
|R_k^{(l)} \cap (\bigcup_{j<k} R_j^{(l)} \cup \Lambda^{l+1})| < \beta |R_k^{(l)}|,
\]

and such that \( R_k^{(l)} \) is minimal with this property (relative to the initial order). The resulting collection \( \{ R_{j(m)} \}_{m,j} \) is sparse in the disjoint-pieces sense, with sparse parameter \( 1 - \beta \). In particular, for \( R = R_k^{(l)} \) we can choose \( E_R = R \setminus (\bigcup_{j<k} R_j^{(l)} \cup \Lambda^{l+1}) \). By construction \( |E_R| \geq (1 - \beta)|R| \), and clearly \( E_R \cap E_{R'} = \emptyset \) for distinct \( R, R' \).

It remains to be shown that

\[
\left| \sum_{k=0}^{N} \sum_{R \in \mathcal{F}_k} \alpha_R \right| \leq \sum_{m,j} |R_{j(m)}^{(m)}|(Sf)_{R_{j(m)}^{(m)}}(Sg)_{R_{j(m)}^{(m)}}.
\]

Break up this sum as

\[
\sum_{k=0}^{N} \sum_{R=R_i^{(k)}} \alpha_R + \sum_{\text{rest}} \alpha_R := A + B.
\]

To estimate \( A \), first observe that if \( \alpha_R = \alpha_{R_i^{(k)}} \) then
\[ \alpha_R = \int_{R_i^{(k)} \cap \Omega_{k+1}^c} \alpha_R \cdot \frac{1_{R_i^{(k)} \cap \Omega_{k+1}^c}(y)}{|R_i^{(k)} \cap \Omega_{k+1}^c|} \, dy \]
\[ \leq 2 \int_{R_i^{(k)} \cap \Omega_{k+1}^c} \alpha_R \frac{1_{R}(y)}{|R|} \, dy \]
\[ \leq 2 \int_{R_i^{(k)} \cap \Omega_{k+1}^c} Sf(x)Sg(x) \, dx. \]

Now recall that \( Sf \lesssim 2^k(Sf)Q_0 \) and \( Sg \lesssim 2^k(Sg)Q_0 \) in \( \Omega_{k+1}^c \). Moreover, by construction we must have either \( Sf \gtrsim 2^k(Sf)Q_0 \) or \( Sg \gtrsim 2^k(Sg)Q_0 \) in more than a quarter of \( R_i^{(k)} \). Without loss of generality suppose \( Sf \gtrsim 2^k(Sf)Q_0 \). Then

\[ \int_{R_i^{(k)} \cap \Omega_{k+1}^c} Sf(x)Sg(x) \, dx \lesssim \left(2^k(Sf)Q_0 \cdot |R_i^{(k)}|\right) \frac{1}{|R_i^{(k)}|} \int_{R_i^{(k)}} Sg(y) \, dy \]
\[ \lesssim |R_i^{(k)}| \langle Sf \rangle_{R_i^{(k)}} \langle Sg \rangle_{R_i^{(k)}}, \quad (3.5) \]

and we can ultimately conclude that

\[ A \lesssim \sum_k \sum_i |R_i^{(k)}| \langle Sf \rangle_{R_i^{(k)}} \langle Sg \rangle_{R_i^{(k)}}. \quad (3.6) \]

We now turn to the term \( B \). Observe that a rectangle \( R \in F_l \) contributes to the sum \( B \) if \( R \) was never chosen in the Córdoba-Fefferman selection process. It follows that for such an \( R \in F_l \) we must have

\[ |R \cap \bigcup_{k \geq l} \bigcup_i R_i^{(k)} \cap \Omega_{l+1}^c| \geq (\beta - 1/2)|R|, \]

provided we have chosen \( \beta > 1/2 \). This is because

\[ |R \cap \Omega_{l+1}^c| \geq \frac{1}{2}|R| \]
and

\[ |R \cap \bigcup_{k \geq l} \bigcup_i R_i^{(k)}| \geq \beta |R|. \]

We then have

\[
B \leq (\beta - 1/2)^{-1} \sum_l \sum_{R \in \mathcal{F}_l} \alpha_R \frac{|R \cap \bigcup_{k \geq l} \bigcup_i R_i^{(k)} \cap \Omega_{k+1}^c|}{|R|}
\]

\[
\leq (\beta - 1/2)^{-1} \sum_l \sum_{R \in \mathcal{F}_l} \alpha_R \sum_{k \geq l} \sum_i \frac{|R \cap R_i^{(k)} \cap \Omega_{k+1}^c|}{|R|}
\]

\[
\leq (\beta - 1/2)^{-1} \sum_l \sum_{R \in \mathcal{F}_l} \int_{R_i^{(k)} \cap \Omega_{k+1}^c} \left( \sum_R \alpha_R \frac{1_R}{|R|} \right) \, dy.
\] (3.7)

We’ve used the fact that \( \Omega_{i+1}^c \subset \Omega_{k+1}^c \) when \( k \geq l \) in the second line. We can now finish the estimate by applying Cauchy-Schwarz and using the properties of the \( \Omega_{k+1}^c \), as in (3.5). Hence

\[
B \lesssim \sum_k \sum_i |R_i^{(k)}|(Sf)_{R_i^{(k)}}(Sg)_{R_i^{(k)}},
\]

as well, completing the proof of the case where \( \epsilon_R = 1 \) for all \( R \).

For the general martingale transform, we simply remark that we have not used any cancellation in the above argument. Hence the argument is exactly the same if we replace \( \alpha_R \) with \( |\alpha_R| \), and as a consequence we may repeat the above argument with \( (\sup_R |\epsilon_R|)|\alpha_R| \) in place of \( \alpha_R \). The sparse bound for \( \sum_R \epsilon_R \langle f, h_R \rangle h_R \) follows.

### 3.5 Bi-Parameter Dyadic Shifts and Singular Integrals

The argument from Section 3.4 generalizes to the case where \( T \) is a cancellative bi-parameter dyadic shift if one replaces the usual square function by certain shifted variants. This ultimately leads to a type of sparse bound for paraproduct-free bi-parameter singular integrals via Martikainen’s representation theorem [58].
3.5.1 Definitions.

We let $D_1$ and $D_2$ be two dyadic grids in $\mathbb{R}$ (not necessarily the standard grids), and let $D = D_1 \times D_2$ be the collection of dyadic rectangles relative to these grids. If $I$ is a dyadic interval and $k \in \mathbb{N}$, we let $(I)_k$ denote the children of $I$ at level $k$, so that $J \in (I)_k$ if and only if $J \subset I$ and $|J| = 2^{-k}|I|$. We also denote the bi-parameter Haar wavelets by $h_R = h_{R_1} \otimes h_{R_2}$ for rectangles $R = R_1 \times R_2$, and use $\hat{f}(R)$ to denote the Haar coefficient of a function $f$.

Given tuples of non-negative integers $i = (i_1, i_2)$ and $j = (j_1, j_2)$, we define the cancellative bi-parameter dyadic shift of complexity $(i, j)$ by

$$T^{i,j}f(y) = \sum_{R_1 \in D_1} \sum_{P_1 \in (R_1)_{i_1}} \sum_{Q_1 \in (R_1)_{i_2}} a_{PQR} \cdot \hat{f}(P) h_Q(y)$$

Here $P = P_1 \times P_2$, $Q = Q_1 \times Q_2$ and $R = R_1 \times R_2$ are dyadic rectangles, and $a_{PQR}$ is a constant satisfying the bound

$$|a_{PQR}| \leq \frac{\sqrt{|P_1||Q_1|\sqrt{|P_2||Q_2|}}}{|R_1||R_2|} = 2^{-\frac{1}{2}(i_1+j_1+i_2+j_2)}.$$

(3.9)

We also define the dyadic shifted square function adapted to the shift parameters $i, j$ by

$$(S^{i,j}f(y))^2 = \sum_{R_1 \in D_1} \sum_{P_1 \in (R_1)_{i_1}} \sum_{Q_1 \in (R_1)_{i_2}} |\hat{f}(P)|^2 \left( \sum_{Q_2 \in (R_2)_{j_2}} \frac{1}{|Q_2|} \otimes \frac{1}{|Q_2|} \right) (y).$$

(3.10)

This clearly depends on the choice of $D$, but we omit this dependence from the notation since our bounds will be independent of $D$. Also note that this definition is not symmetric in $i, j$. The same is true for the bi-parameter shift of complexity $(i, j)$. (The definition (3.10) is taken from the paper [39] by Holmes, Petermichl, and Wick).
3.5.2 The Sparse Bound

We will now adapt the argument from Section 3.4 to prove the following sparse bound.

**Theorem 3.2.** Let $\mathcal{D}$ be an arbitrary system of dyadic rectangles. Let $T^{i,j}$ be the cancellative shift defined above (using rectangles from $\mathcal{D}$), and fix test functions $f, g$. Then there exists a sparse collection $S$ of $\mathcal{D}$-dyadic rectangles such that

$$|\langle T^{i,j}f, g \rangle| \lesssim 2^{-(i_1+i_2+j_1+j_2)} \sum_{R \in S} |R| (S^{i,j}f)_{R} (S^{j,i}g)_{R},$$

The collection $S$ depends on $f, g$ and $(i, j)$, but the implicit constant does not.

Note that the order of $i, j$ is switched in the term containing $g$. From [39] we know that

$$\|S^{i,j}f\|_{L^p(w)} \lesssim c_{w} 2^{ \frac{1}{2}(i_1+i_2+j_1+j_2)} \|f\|_{L^p(w)},$$

so we need the factor in front of the sparse form for applications to weighted estimates. In the last section we will show that in the case $p = 2$ we can at least take $c_{w} = \left[w\right]_A^5 A_2$.

The proof of the theorem is similar to what we have seen above. Begin by assuming that $f, g$ are supported in some cube $L$. Let $\alpha_f = c \cdot (S^{i,j}f)_{L}$ and $\alpha_g = c \cdot (S^{j,i}g)_{L}$, where $c$ is some large constant. Define

$$\Omega_0 = \{x \in L : S^{i,j}f(x) > \alpha_f\} \cup \{x \in Q_0 : S^{j,i}g(x) > \alpha_g\}.$$

Notice that

$$|\Omega_0| \leq \frac{1}{\alpha_f} \int_{L} S^{i,j}f + \frac{1}{\alpha_g} \int_{L} S^{j,i}g \leq \frac{2}{c} |L|,$$

so we can assume $c$ has been chosen independent of $(i, j)$ such that $|\Omega_0| \leq \frac{1}{2} |L|$. For positive integers $k$, also define

$$\Omega_k = \{x \in L : S^{i,j}f(x) > 2^k \alpha_f\} \cup \{x \in L : S^{j,i}g(x) > 2^k \alpha_g\},$$
and let \( \mathcal{R}_0 \) be the collection of rectangles \( R \) such that \( |R \cap \Omega_0| < \frac{1}{2} |R| \). Finally, for \( k \geq 0 \) define

\[
\mathcal{F}_k = \{ R : |R \cap \Omega_k| > \frac{1}{2} |R| \text{ and } |R \cap \Omega_{k+1}| \leq \frac{1}{2} |R| \}.
\]

We first estimate the ‘good’ part of our form corresponding to the rectangles in \( \mathcal{R}_0 \). To further simplify the notation, if \( R = R_1 \times R_2 \) is a dyadic rectangle and \( P = P_1 \times P_2 \) is a dyadic rectangle contained in \( R \), we write \( P \in R_1 \) to mean \( P_1 \in (R_1)_{i_1} \) and \( P_2 \in (R_2)_{i_2} \).

**Lemma 3.1.** Let \( R \) be a dyadic rectangle. Then

\[
\frac{1_R(y)}{|R|} = 2^{-\frac{1}{2}(i_1+i_2+j_1+j_2)} \left( \sum_{P \in R_1} 1_{P_1} \otimes 1_{P_2}(y) \right)^{1/2} \left( \sum_{Q \in R_2} 1_{Q_1} \otimes 1_{Q_2}(y) \right)^{1/2} \frac{1}{(|P_1||P_2|)^{1/2}(|Q_1||Q_2|)^{1/2}}.
\]

**Proof.** This is a simple consequence of the fact that \( R \) is a disjoint union of all rectangles \( P \) such that \( P \in R_1 \), and similarly \( R \) is a disjoint union of all rectangles \( Q \) such that \( Q \in R_2 \). Since \( |R|^{1/2} = 2^{\frac{1}{2}(i_1+i_2)}(|P_1||P_2|)^{1/2} \) and \( |R|^{1/2} = 2^{\frac{1}{2}(j_1+j_2)}(|Q_1||Q_2|)^{1/2} \) the identity follows. \( \square \)

Now let

\[
\langle T^{i,j} f, g \rangle_{\text{good}} = \sum_{R \in \mathcal{R}_0} \sum_{P \in R_1} \sum_{Q \in R_2} a_{PQR} \hat{f}(P) \hat{g}(Q).
\]

Arguing as in the beginning of the proof of Theorem 3.1 we find that

\[
|\langle T^{i,j} f, g \rangle_{\text{good}}| \leq 2^{-\frac{1}{2}(i_1+i_2+j_1+j_2)} \sum_{R \in \mathcal{R}_0} \sum_{P \in R_1} \sum_{Q \in R_2} \hat{f}(P) \hat{g}(Q) | \frac{1_{R \cap \Omega_0}(y)}{|R \cap \Omega_0|} dy \]

\[
= 2^{-\frac{1}{2}(i_1+i_2+j_1+j_2)} \sum_{R \in \mathcal{R}_0} \int_{R \cap \Omega_0} \sum_{P \in R_1} \sum_{Q \in R_2} \hat{f}(P) \hat{g}(Q) | \frac{1_{R \cap \Omega_0}(y)}{|R \cap \Omega_0|} dy
\]

\[
\lesssim 2^{-\frac{1}{2}(i_1+i_2+j_1+j_2)} \int_{\Omega_0} \sum_{R \in \mathcal{R}_0} \sum_{P \in R_1} \sum_{Q \in R_2} \hat{f}(P) | \hat{g}(Q) | \frac{1_{R}(y)}{|R|} dy \]

\[
\lesssim 2^{-\frac{1}{2}(i_1+i_2+j_1+j_2)} \int_{\Omega_0} \sum_{R \in \mathcal{R}_0} \sum_{P \in R_1} \sum_{Q \in R_2} \hat{f}(P) | \hat{g}(Q) | dy \]

\[
\lesssim 2^{-(i_1+i_2+j_1+j_2)} \int_{\Omega_0} S^{i,j} f(y) \cdot S^{j,i} g(y) dy.
\]
To get to the last line, we applied Lemma 3.1 and then used Cauchy-Schwarz and the definition of the shifted square functions. But the integral is over $\Omega^c_0$, so we can conclude that

$$|\langle T^{i,j} f, g \rangle_{\text{good}}| \lesssim 2^{-(i_1+i_2+j_1+j_2)}|L|(S^{i,j} f)_L(S^{j,i} g)_L.$$ 

It remains to estimate

$$\langle T^{i,j} f, g \rangle_{\text{bad}} = \langle T^{i,j} f, g \rangle - \langle T^{i,j} f, g \rangle_{\text{good}}.$$ 

As in Section 3.4, this is where the sparse collection enters into the picture. We will assume that $\hat{f}(R)$ and $\hat{g}(R)$ are nonzero for only finitely many $R$, and let $N$ denote the largest integer such that $\hat{f}(R)$ and $\hat{g}(R)$ are nonzero for some $R \in \mathcal{F}_N$. All bounds will be independent of $N$, so density arguments will allow us to extend the results to more general $f, g$. We construct the sparse collection using the Córdoba-Fefferman selection algorithm as before. The only change is that our exceptional sets $\Omega_k$ now depend on the shifted square function $S^{i,j}$, but otherwise the construction proceeds in exactly the same way as in the case of the martingale transform. We omit the details since the argument would be a copy of what appears in Section 3.4. Let $\{R_n^{(k)}\}$ denote the resulting collection, with $R_n^{(k)} \in \mathcal{F}_k$. We can break up the ‘bad’ part of the form as

$$\langle T^{i,j} f, g \rangle_{\text{bad}} = \sum_{R=R_n^{(k)}} \sum_{P \in R_i} a_{PQR} \hat{f}(P) \hat{g}(Q) + \sum_{\text{rest}} a_{PQR} \hat{f}(P) \hat{g}(Q) := A + B.$$

71
Estimating A. Fix $R = R_n^{(k)}$ and observe that since $R \in \mathcal{F}_k$ we have

$$\left| \sum_{P \in R_i} a_{PQR} \hat{f}(P) \hat{g}(Q) \right| \leq 2^{-\frac{1}{2}(i_1+i_2+j_1+j_2)} \int_{R \cap \Omega^c_{k+1}} \sum_{P \in R_i} \sum_{Q \in R_j} |\hat{f}(P)\hat{g}(Q)| \frac{|1_{\Omega_{k+1}^c}(y)|}{|R \cap \Omega^c_{k+1}|} dy$$

$$\lesssim 2^{-\frac{1}{2}(i_1+i_2+j_1+j_2)} \int_{R \cap \Omega^c_{k+1}} |\hat{f}(P)| \sum_{Q \in R_j} |\hat{g}(Q)| \frac{1_{\Omega_{k+1}^c}(y)}{|R|} dy$$

$$\lesssim 2^{-(i_1+i_2+j_1+j_2)} \int_{R \cap \Omega^c_{k+1}} S_{i,j}^i f(y) S_{j,i}^j g(y) dy,$$

applying Lemma 3.1 as above to get to the last line. Now argue as in (3.5) and sum over all $R_n^{(k)}$ to get

$$|A| \lesssim 2^{-(i_1+i_2+j_1+j_2)} \sum_{n,k} |R_n^{(k)}| (S_{i,j}^i f)_{R_n^{(k)}} (S_{j,i}^j g)_{R_n^{(k)}}.$$

Estimating B. The term $B$ involves a sum of $\sum_{P \in R_i} a_{PQR} \hat{f}(P) \hat{g}(Q)$ over all $R$ not chosen in the Córdoba-Fefferman selection process. For any such $R \in \mathcal{F}_l$ we must have

$$|R \cap \bigcup_{n \geq l} \bigcup_{k \geq n} R_n^{(k)} \cap \Omega^c_{k+1}| \geq (\beta - 1/2) |R|$$

as in Section 3.4, provided we have chosen $\beta > \frac{1}{2}$.

The proof now proceeds as in (3.7). One repeats the argument given in (3.7) and makes modifications similar to what we’ve seen the the proof of A in order to insert the shifted operators $S_{i,j}^i f$, $S_{j,i}^j g$. It follows that

$$|B| \lesssim \beta \ 2^{-(i_1+i_2+j_1+j_2)} \sum_{n,k} |R_n^{(k)}| (S_{i,j}^i f)_{R_n^{(k)}} (S_{j,i}^j g)_{R_n^{(k)}},$$

completing the proof of Theorem 3.2.
3.5.3 Bi-Parameter Singular Integrals

We can now use Martikainen’s representation theorem [58] to show that if $T$ is a paraproduct-free bi-parameter singular integral belonging to the Journé class, then $T$ can be estimated by an average of sparse forms of the type appearing in Theorem 3.2. Loosely speaking, $T$ is a bi-parameter Journé operator on $\mathbb{R} \times \mathbb{R}$ if it has a kernel $K(x_1, x_2, y_1, y_2)$ on $\mathbb{R}^2 \times \mathbb{R}^2$ that satisfies analogues of the Calderón-Zygmund kernel conditions in each variable separately, along with mixed Hölder and size conditions involving the two parameters. We send the reader to Martikainen’s paper [58] for the precise definition of a Journé operator $T$. We say that such a $T$ is paraproduct-free if the bi-parameter version $T(1) = T^*(1) = 0$ holds, meaning there are only cancellative shifts in the dyadic representation from [58].

We briefly recall one more definition. Let $D_0$ denote the standard dyadic intervals in $\mathbb{R}$, and for every $\eta = (\eta_j)_{j \in \mathbb{Z}} \in \{0, 1\}^\mathbb{Z}$ define the shifted grid

$$D_\eta := \{I + \eta : I \in D_0\}, \quad \text{with} \ I + \eta := I + \sum_{2^{-k} < |I|} 2^{-k} \eta_k.$$ 

We assign $\{0, 1\}^\mathbb{Z}$ the natural Bernoulli(1/2) product measure. This gives us a probability measure on the space of shifted grids, and hence a probability measure on the space of shifted dyadic rectangles $D_\eta \times D_\eta'$ (see [41] or [58] for more properties of these random grids).

**Corollary 3.1.** Let $T$ be a paraproduct-free bi-parameter Journé singular integral on $\mathbb{R}^2$ and suppose $f, g$ are test functions with finitely many (bi-parameter) Haar coefficients. For each pair of tuples of non-negative integers $i, j$ set $\tau_{i,j} = 2^{-(i_1 + i_2 + j_1 + j_2)}$. Also let $S^{i,j}_\omega$ be the shifted square function (3.10) defined with respect to the random dyadic system $D_\omega = D_{\omega_1} \times D_{\omega_2}$. Then there exists $\delta > 0$ and sparse collections of rectangles $\Lambda_{i,j}$ (depending on $f, g$) such that

$$|\langle Tf, g \rangle| \lesssim \mathbb{E}_{\omega_1} \mathbb{E}_{\omega_2} \sum_{i,j \geq 0} 2^{-\max(i_1, j_1)\delta/2} 2^{-\max(i_2, j_2)\delta/2} \tau_{i,j} \sum_{R \in \Lambda_{i,j}} |R|(S^{i,j}_\omega f)_R(S^{j,i}_\omega g)_R.$$ 

**Proof.** Given a system of shifted dyadic rectangles $D_\omega = D_{\omega_1} \times D_{\omega_2}$, we let $T^{i,j}_\omega$ denote the shift
operator (3.8) defined with respect to rectangles from $D_{\omega}$. Martikainen proved that if $T$ is a paraproduct-free operator in the Journé class then

$$
\langle T f, g \rangle = \mathbb{E}_{\omega_1} \mathbb{E}_{\omega_2} \sum_{i,j \geq 0} 2^{-\max(i_1, j_1)\delta/2} 2^{-\max(i_2, j_2)\delta/2} \langle T_{i,j} \omega f, g \rangle.
$$

The claimed result now follows by applying Theorem 3.2 to each form $\langle T_{i,j} \omega f, g \rangle$ (recall that Theorem 3.2 applies for any dyadic system $D_{\omega_1} \times D_{\omega_2}$).

It should be possible to extend the result of Corollary 3.1 to arbitrary Journé operators $T$, although the weighted estimates that follow would be far from optimal. We briefly outline one approach. It would be sufficient to prove analogues of Theorem 3.2 for the various paraproducts that show up in the dyadic representation of $T$. To this end, one can work with the mixed operators $SM$ and $MS$, where $S$ and $M$ are one-parameter square and maximal functions in different directions. By combining methods from [39] or [59] with our sparse domination scheme, it should be possible to prove bounds of the type

$$
|\langle \Pi f, g \rangle| \lesssim \sum_{R \in S} |R| (SMf)_R (SMg)_R,
$$

(3.11)

where $\Pi$ is a bi-parameter paraproduct. One may also have to work with shifted variants of the mixed operators (see [39]), and prove analogues of (3.11) involving these operators.

### 3.6 Weighted Estimates for Bi-Parameter Martingale Transform

Let $w(x_1, x_2)$ be a positive, locally integrable weight on $\mathbb{R} \times \mathbb{R}$. Recall that $w$ is a two-parameter $A_p(\mathbb{R} \times \mathbb{R})$ weight for $1 < p < \infty$ if and only if

$$
[w]_{A_p(\mathbb{R} \times \mathbb{R})} := \sup_{R} \left( \frac{1}{|R|} \int_R w(x_1, x_2) \, dx \right) \left( \frac{1}{|R|} \int_R w(x_1, x_2)^{1-p'} \, dx \right)^{p-1} < \infty.
$$

(3.12)
By the Lebesgue differentiation theorem this condition is equivalent to $w(\cdot, x_2) \in A_p(\mathbb{R})$ uniformly in $x_2$ and $w(x_1, \cdot) \in A_p(\mathbb{R})$ uniformly in $x_1$, and in fact

$$[w]_{A_p(\mathbb{R} \times \mathbb{R})} \lesssim \max(\|[w(\cdot, x_2)]_{A_p(\mathbb{R})}\|_{L^\infty_{x_2}}, \|[w(x_1, \cdot)]_{A_p(\mathbb{R})}\|_{L^\infty_{x_1}}).$$

Write $[w]_{A_p} = [w]_{A_p(\mathbb{R} \times \mathbb{R})}$. As we noted in the introduction, in the one-parameter setting sparse bounds lead to weighted estimates that are sharp in terms of the $A_p$ characteristic. Here we use our sparse bound (3.1) to derive $A_p$ estimates in terms of the bi-parameter characteristic. Unfortunately, the square-function sparse bound we have proved does not seem to imply estimates that are sharp. By using known methods, for example the arguments in [39], one can prove

$$\|Tf\|_{L^p(w)} \lesssim [w]_{A_p(\mathbb{R} \times \mathbb{R})}^8 \|f\|_{L^p(w)}$$

for a Journé-type operator [62], whereas our methods yield a power that is much worse. On the other hand, our method of proof simplifies the somewhat technical arguments that currently exist in the literature.

**Lemma 3.2.** Let $S$ be the bi-parameter square function with respect to some fixed dyadic grid $D$, and let $M$ be the strong maximal function. Then if $w \in A_2$,

$$\|Sf\|_{L^2(w)} \lesssim [w]_{A_2}^2 \|f\|_{L^2(w)}$$

and

$$\|Mf\|_{L^2(w)} \lesssim [w]_{A_2}^2 \|f\|_{L^2(w)}$$

for all $f \in L^2(w)$.

**Proof.** Recall that the dyadic one-parameter square function satisfies the weighted estimate

$$\|S_1(f)\|_{L^2(w)} \lesssim [w]_{A_2(\mathbb{R})} \|f\|_{L^2(w)}$$

75
and the Hardy-Littlewood maximal operator satisfies the estimate

\[ \|M_1(f)\|_{L^2(w)} \lesssim [w]_{A_2(\mathbb{R})} \|f\|_{L^2(w)}. \]

Both of the claimed estimates follow by iterating the one-parameter results, using the pointwise bound \( Mf \leq M_1(M_2f) \) for the strong maximal function (here \( M_1 \) is the Hardy-Littlewood operator in the direction \( x_1 \), and \( M_2 \) is the Hardy-Littlewood operator in the direction \( x_2 \)).

\[ \square \]

**Proposition 3.3.** Let \( \epsilon_R \) be a uniformly bounded sequence indexed over dyadic rectangles with \( \sup_R |\epsilon_R| \leq C \epsilon \), and let \( Tf \) be the following bi-parameter martingale transform:

\[ Tf(x) = \sum_R \epsilon_R \langle f, h_R \rangle h_R(x). \]

Then for all \( w \in A_2 = A_2(\mathbb{R} \times \mathbb{R}) \) and \( f \in L^2(w) \) we have

\[ \|Tf\|_{L^2(w)} \lesssim C_\epsilon [w]_{A_2}^8 \|f\|_{L^2(w)}. \]

**Proof.** The estimate follows from sparse domination. Let \( \sigma = w^{-1} \). By duality it is enough to show that for all \( g \in L^2(\sigma) \) we have

\[ |\langle Tf, g \rangle| \lesssim C_\epsilon [w]_{A_2}^8 \|f\|_{L^2(w)} \|g\|_{L^2(\sigma)}. \]

We know from above that there is a sparse collection of rectangles \( S \) so that

\[ |\langle Tf, g \rangle| \lesssim C_\epsilon \sum_{R \in S} |R|(Sf)_R(Sg)_R. \]

We now repeat a version of the standard argument from the one-parameter theory, using the
strong maximal function in place of the Hardy-Littlewood maximal function. We have

$$\sum_{R \in S} |R|(Sf)_R(Sg)_R \lesssim \sum_{R \in S} |E_R| \left( \inf_{x \in R} M(Sf)(x) \right) \left( \inf_{x \in R} M(Sg)(x) \right)$$

$$\lesssim \sum_{R \in S} \int_{E_R} M(Sf)(x) M(Sg)(x) w^{1/2}(x) \sigma^{1/2}(x) \, dx$$

$$\lesssim \int_{\mathbb{R}^2} M(Sf)(x) M(Sg)(x) w^{1/2}(x) \sigma^{1/2}(x) \, dx$$

$$\lesssim \|M(Sf)\|_{L^2(w)} \|M(Sg)\|_{L^2(\sigma)}$$

$$\lesssim [w]_{A_2}^4 \|Sf\|_{L^2(w)} \|Sg\|_{L^2(\sigma)}$$

$$\lesssim [w]_{A_2}^8 \|f\|_{L^2(w)} \|g\|_{L^2(\sigma)},$$

as desired.

By passing through a square function, it is not too hard to show that the bi-parameter martingale transform satisfies the $A_2$ bound

$$\|Tf\|_{L^2(w)} \lesssim [w]_{A_2}^3 \|f\|_{L^2(w)},$$

hence the constants in Proposition 3.3 are far from optimal. We do not know if the power of 8 appearing in Proposition 3.3 can be pushed down further using our methods. In the one-parameter setting, the usual argument that produces the sharp $A_2$ bound invokes the weighted maximal operator

$$M_1^\mu f(x) = \sup_{x \in I} \frac{1}{\mu(I)} \int_I |f(y)| \mu(y) \, dy,$$

(3.13)

which is bounded on $L^2(\mu)$ for any positive function $\mu$, with norm independent of $\mu$ (this follows from the Besicovitch covering lemma or martingale theory, see [82] for example). However, the bi-parameter analogue of (3.13) is in general not bounded on $L^2(\mu)$, due to the more complicated geometry. R. Fefferman proved in [31] that $\mu \in A_\infty(\mathbb{R} \times \mathbb{R})$ is sufficient for the strong weighted
maximal function $M^\mu$ to be bounded on $L^2(\mu)$, but the sharp dependence of the operator norm on $[\mu]_{A_\infty}$ is unclear from his argument. It is somewhat surprising that if we trace the dependence in his argument and use some recent sharp results related to $A_\infty$ ([75], [42]), we uncover a dependence that is exponential in the $A_\infty$ characteristic. Recall that $w$ is in the bi-parameter weight class $A_\infty(\mathbb{R} \times \mathbb{R})$ if $w$ is in the one-parameter class $A_\infty(\mathbb{R})$ uniformly in each variable.

**Proposition 3.4.** Suppose $w \in A_p(\mathbb{R} \times \mathbb{R})$ and let $M^w$ be the two-dimensional weighted strong maximal function

$$M^w(f)(y) = \sup_{y \in R} \frac{1}{w(R)} \int_{R} |f(x)| \ w(x)dx.$$ 

Then for all $1 < p < \infty$ we have $\|M^w\|_{L^p(w) \to L^p(w)} \lesssim_p [w]_{A_p} e^{c[w]_{A\infty}}$.

We prove this proposition in the first appendix.

The sparse bounds from Section 3.5 also allow us to derive weighted estimates for dyadic shifts and paraproduct-free Journé operators. The argument is almost the same as the proof of Theorem 3.3, but in this case we have to work with weighted estimates for the shifted square functions $S^{i,j}$. We know from [39] that if $w \in A_p(\mathbb{R} \times \mathbb{R})$ there is some $c_w > 0$ such that

$$\|S^{i,j}f\|_{L^p(w)} \leq 2^{(i+j)/2}c_w \|f\|_{L^p(w)}$$

for all $f \in L^p(w)$, but we would like to track the dependence of $c_w$ on $[w]_{A_p}$. This is the content of the next section.

### 3.7 Weighted Estimates for the Shifted Square Function

It was proved in [40] that the one-parameter shifted square function

$$S^{i,j}_1f(x) = \sum_{R \in \mathcal{D}} \left( \sum_{P \in (R)_i} |\hat{f}(P)| \right)^2 \sum_{Q \in (R)_j} \frac{1}{|Q|}$$

for all $f \in L^p(w)$. The next section will focus on the dependence of $c_w$ on $[w]_{A_p}$. This is the content of the next section.
satisfies the weighted estimate \( \|S^{i,j} f\|_{L^2(w)} \leq 2^{(i+j)/2} C_w \|f\|_{L^2(w)} \) for \( w \in A_2 \). In particular, the argument in [40] gives \( C_w \leq [w]_A^2 \). In this section we prove a type of sparse bound for \( S^{i,j} \) that allows us to show \( C_w \leq [w]_{A_2}^{1/2} [w]_{A_\infty}^{1/2} \). An iteration argument then shows that the bi-parameter analogue of \( S^{i,j} \) satisfies a weighted bound with constant \( c_w \lesssim [w]_A^{4} [w]_{A_\infty}^{4} \). The method of proof is an adaptation of the scalar case of the argument by Hytönen, Petermichl, and Volberg in [43].

### 3.7.1 Preliminary Results

Fix an arbitrary (one-parameter) dyadic lattice \( D \).

**Lemma 3.5.** Suppose \( f_k \) is a sequence of functions such that \( S^{i,j}(f_k) \) is defined for each \( k \). Then \( S^{i,j}(\sum_k f_k) \leq \sum_k S^{i,j}(f_k) \).

**Proof.** Fix an arbitrary \( x \in \mathbb{R} \). The lemma is a simple consequence of Minkowski’s inequality for the weighted space \( \ell^2(1_R(x)/|R|) \), where \( \|\{\alpha_R\}\|_{\ell^2(1_R/|R|)} = \sum_R \alpha_R^2 1_R/|R| \). Let \( F_{k,(R)} = \sum_{P \in (R)} |\hat{f}_k(P)| \). Then

\[
S^{i,j}(\sum_k f_k) = 2^{i/2} \sum_k \|F_{k,(R)}\|_{\ell^2(1_R/|R|)} \\
\leq 2^{i/2} \sum_k \|F_{k,(R)}\|_{\ell^2(1_R/|R|)} \\
= 2^{i/2} \sum_k \left( \sum_R (F_{k,(R)})^2 1_R/|R| \right)^{1/2} = \sum_k S^{i,j}(f_k).
\]

**Proposition 3.6.** The operator \( S^{i,j}_1 \) maps \( L^1(\mathbb{R}) \) into \( L^{1,\infty}(\mathbb{R}) \) with \( \|S^{i,j}_1\|_{L^1 \rightarrow L^{1,\infty}} \lesssim \lambda^{2(i+j)/2} \).

**Proof.** The argument is a variation of the standard approach via the Calderón-Zygmund decomposition. Fix \( f \in L^1(\mathbb{R}) \) and \( \lambda, \alpha > 0 \). Choose maximal dyadic intervals \( J \) such that \( \frac{1}{|J|} \int_J |f| > \alpha \lambda \) and let \( \Omega \) denote the union of such intervals. Then \( f = g + b \), with \( g = f 1_{\Omega^c} + \sum_J (f)_J 1_J \) and \( b = \sum_J (f - (f)_J) 1_J \). Moreover \( \|g\|_{L^\infty} \lesssim \alpha \lambda \) and \( \|b_J\|_{L^1} \lesssim \alpha \lambda |J| \).
By scaling invariance we may normalize $\|f\|_{L^1} = 1$. By Lemma 3.5 we have

$$|\{S_{i,j}^1 f > \lambda\}| \leq |\{S_{i,j}^1 g > \lambda/2\}| + |\{S_{i,j}^1 b > \lambda/2\}|.$$

Using the $L^2$-boundedness of $S_{i,j}^1$ we can immediately conclude that

$$|\{S_{i,j}^1 g > \lambda/2\}| \lesssim \lambda^{-2} 2^{i+j} \|g\|_{L^2}^2 \lesssim 2^{i+j} \alpha \lambda.$$

Let $E = \bigcup J5J$ and note $|E| \lesssim \alpha^{-1} \lambda^{-1}$. We also claim that

$$|\{x \in E^c : S_{i,j}^1 b(x) > \lambda/2\}| \lesssim \frac{1}{\lambda} \cdot 2^{(i+j)/2}. \quad (3.14)$$

To prove the last claim, we first note that if $J \subsetneq P$ then $\hat{b}_J = 0$ since $\int b_J = 0$ and $h_P$ is constant on dyadic sub-intervals of $P$. Moreover, only the intervals $R$ with $R \supset J$ contribute to $S_{i,j}^1(b_J)(x)$ when $x \in E^c$, since $E^c = (\bigcup J5J)^c$ (and if $J \cap R = \emptyset$ and $P \in (R)_i$, then $\hat{b}_J(P) = 0$).

It follows that

$$\int_{E^c} S_{i,j}^1 (b_J)(x) \, dx = 2^{i/2} \int_{E^c} \left[ \sum_{R \supset J} \left( \sum_{P \in (R)_i, P \subset J} |\hat{b}_J(P)| \right)^2 \frac{1_{R}(x)}{|R|} \right]^{1/2} dx.$$

Now since

$$2^{i/2} \int_{E^c} \left[ \sum_{R \supset J} \left( \sum_{P \in (R)_i, P \subset J} |\hat{b}_J(P)| \right)^2 \frac{1_{R}(x)}{|R|} \right]^{1/2} dx \leq 2^{i/2} \int_{E^c} \sum_{R \supset J} \sum_{P \in (R)_i, P \subset J} |\hat{b}_J(P)| \frac{1_{R}(x)}{|R|^{1/2}} dx.$$
we get

\[
\int_{E^c} S_1^{i,j}(b)(x) dx \leq \sum_J \int_{E^c} S_1^{i,j}(b_J)(x) dx \\
\leq 2^{i/2} \sum_J \int_{E^c} \sum_{P \in (R)_i} \sum_{P \subset J} \left| b_J(P) \right| \left| \frac{1_{R}(x)}{|R|^{1/2}} \right| dx \\
\leq 2^{i/2} \sum_J \sum_{P \in (R)_i} 2^{i/2} \left| P \right|^{1/2} \left| \hat{b}_J(P) \right| \\
\leq 2^{(i+j)/2} \sum_J \sum_{P \in (R)_i} \sum_{P \subset J} \int \left| b_J(x) \right| \left| P \right|^{1/2} \left| h_P(x) \right| dx \\
\leq 2^{(i+j)/2} \sum_J \sum_{P \subset J} \left\| b_J \right\|_{L^1}
\]

using additionally the fact that \( \sum_{P \in (R)_i} \left| P \right|^{1/2} \left| h_P(x) \right| \) is bounded independent of \( i \) (due to the disjointness of \( P \in (R)_i \)). Now for each \( J \) there are no more than \( i \) dyadic intervals \( R \) such that \( R \supset J \) and \( 2^i |J| \geq |R| \), and therefore we conclude

\[
\int_{E^c} S_1^{i,j}(b)(x) dx \lesssim i 2^{(i+j)/2} \sum_J \left\| b_J \right\|_{L^1} \lesssim i 2^{(i+j)/2},
\]

which implies (3.14).

In summary, we have shown

\[
\left| \left\{ S_1^{i,j} f > \lambda \right\} \right| \lesssim 2^{i+j} \frac{\alpha}{\lambda} + \frac{1}{\alpha \lambda} + i 2^{(i+j)/2} \frac{1}{\lambda}
\]

for arbitrary \( \alpha > 0 \). Setting \( \alpha = 2^{-(i+j)/2} \) yields

\[
\left| \left\{ S_1^{i,j} f > \lambda \right\} \right| \lesssim i 2^{(i+j)/2} \frac{1}{\lambda} \left\| f \right\|_{L^1},
\]

completing the proof.
3.7.2 The Sparse Bound

The weak bound for $S^{i,j}_{1}$ allows us to mimic the sparse domination scheme from [43]. A simple computation shows that

$$\|S^{i,j}_{1} f\|_{L^2(w)}^2 = 2^j \sum_{R \in D} \left( \sum_{P \in (R)} |\hat{f}(P)| \right)^2 (w)_R. \quad (3.15)$$

We will estimate the term on the right by the norm of a sparse operator. We assume there are only finitely many Haar coefficients of $f$, so there is some large interval $J$ that contains every interval contributing to the sum in (3.15). Fix large constants $C_1, C_2 > 0$ to be determined below, and begin by choosing maximal dyadic intervals $L$ such that either

$$\sum_{R \supset L} \left( \sum_{P \in (R)} |\hat{f}(P)| \right)^2 2^{2j} |R| > 2^{i+j} C_1 (|f|)_J^2 \quad (3.16)$$

or

$$(w)_L > C_2 (w)_J. \quad (3.17)$$

Let $S'_1$ denote the collection of maximal intervals from (3.16), and let $S''_1$ denote the collection of maximal intervals from (3.17). The initial collections are $S_0 = \{J\}$ and $S_1 = S'_1 \cup S''_1$. We have

$$2^j \sum_R \left( \sum_{P \in (R)} |\hat{f}(P)| \right)^2 (w)_R = 2^j \sum_{R \text{ s.t. } \forall L \in S_1} \left( \sum_{P \in (R)} |\hat{f}(P)| \right)^2 (w)_R$$

$$+ 2^j \sum_{R \text{ s.t. } \exists L \in S_1} \left( \sum_{P \in (R)} |\hat{f}(P)| \right)^2 (w)_R$$

$$:= A + B.$$
To estimate $A$ we use the stopping conditions (3.16) and (3.17):

$$A = 2^j \sum_{R \text{ s.t. } \forall L \in S_1 \text{}} \left( \sum_{P \in (R)_i} |\hat{f}(P)| \right)^2 (w)_R,$$

$$\leq 2^j C_2 \sum_{R \text{ s.t. } \forall L \in S_1 \text{}} \left( \sum_{P \in (R)_i} |\hat{f}(P)| \right)^2 (w)_J,$$

$$\leq C_2 \sum_{R \text{ s.t. } \forall L \in S_1 \text{}} \left( \sum_{P \in (R)_i} |\hat{f}(P)| \right)^2 2^j \frac{|J|}{|R|} (w)_J,$$

$$\leq 2^{i+j} C_1 C_2 |J| (|f|)^2_J (w)_J.$$

The term $B$ may be handled by recursion, by decomposing it as a sum of operators of type (3.15) localized to each $L$. The same selection process is used at each iteration, with the same constants $C_1, C_2$. It remains to check that the stopping intervals actually form a sparse collection if we choose $C_1$ and $C_2$ correctly, and also that we can choose $C_1, C_2$ independent of $i, j$.

We claim that all of the intervals $L$ chosen in (3.16) are contained in \{$(S_1^{i,j} f)^2 1_{L}(x) > 2^{i+j} C_1 (|f|)_J^2$\}. In fact,

$$(S_1^{i,j} f(x))^2 1_{L}(x) \geq 2^j \sum_{R \supset L} \left( \sum_{P \in (R)_i} |\hat{f}(P)| \right)^2 1_{R(x)} \frac{1_L(x)}{|R|} > 2^{i+j} C_1 (|f|)_J^2 \cdot 1_{L}(x)$$

by selection, which proves the claim. It follows from Proposition 3.6 that we can choose $C_1 \sim i^2$ such that $\sum_{L \in S_1^j} |L| \leq \frac{1}{i} |J|$. For the intervals chosen in (3.17), we directly estimate the following sum:

$$C_2 \sum_{L \in S_1''} |L| \leq \sum_{L \in S_1''} (w)_J^{-1} \int_L w(x) dx \leq |J|,$$

using the disjointness of the $L \in S_1''$. Hence if $C_2 = 4$ then $\sum_{L \in S_1''} |L| \leq \frac{1}{4} |J|$, and as a consequence $\sum_{L \in S_1} |L| \leq \frac{1}{2} |J|$. Moreover, we can choose $C_1$ and $C_2$ independently of $j$ and the dependence on $i$ is only quadratic. The same choice of $C_1, C_2$ at each iteration guarantees that
the collection is sparse. We have proved the following:

**Proposition 3.7.** Suppose \( f \) has finitely many Haar coefficients and fix non-negative integers \( i, j \). Then there exists a sparse collection \( S \) of dyadic intervals such that

\[
\|S_i^j f\|_{L^2(w)}^2 \lesssim 2^{i+j} \sum_{J \in S} |J|(w)_J(|f|)^2_J. \tag{3.18}
\]

The implicit constant is independent of \( i, j \) and \( f, w \).

**Corollary 3.2.** Suppose \( f \in L^2(w) \) with \( w \in A_2 \) and fix non-negative integers \( i, j \). Then

\[
\|S_i^j f\|_{L^2(w)} \lesssim |w|^{1/2} [w]_{A_2}^{1/2} 2^{(i+j)/2} \|f\|_{L^2(w)}.
\]

**Proof.** We use a special case of the general argument outlined in [43]. Note that we have estimated \( \|S_i^j f\|_{L^2(w)} \) by the (scalar version of the) same sparse object appearing in that paper.

By plugging \( w^{-1/2} f \) into the sparse bound (3.18), we get

\[
\|S_i^j (w^{-1/2} f)\|_{L^2(w)}^2 \lesssim 2^{i+j} \sum_{J \in S} |J|(w)_J(|f|w^{-1/2})^2_J.
\]

Hence it will be enough to show that the sum on the right above is no more than \( C |w|_{A_2} [w]_{A_\infty} \|f\|_{L^2(R)}^2 \).

Let \( \delta = \frac{1}{c|w|_{A_\infty}} \) and \( r = 2(1 + \delta) \). By Hölder’s inequality

\[
(|f|w^{-1/2})^2_J \leq (|f|^r)^2_J (w^{-r(1 + \delta)})^{1/r}_J,
\]

hence the reverse Hölder inequality yields

\[
(|f|w^{-1/2})^2_J \lesssim (|f|^r)^{2/r}_J (w^{-1})_J.
\]
It follows that
\[
\sum_{J \in S} |J(\omega) J(|f| w^{-1/2})_J^2_2 \lesssim \sum_{J \in S} |J(|f|^r')_{J}^2 / r'(w^{-1})_J(\omega)_J
\]
\[
\lesssim [w]_{A_2} \sum_{J \in S} |J(|f|^r')_{J}^2 / r'
\]
\[
\lesssim [w]_{A_2} \int_{\mathbb{R}} M(|f|^r')^{2/r'}(x) dx,
\]
using the sparsity of the collection to get the integral over $\mathbb{R}$ in the last line. Note that $r' < 2$, hence
\[
\|M^{r'}(f)\|_{L^2(\mathbb{R})}^2 \lesssim ((2/r')')^{2/r'} \|f\|_{L^2(\mathbb{R})}^2
\]
with $(2/r')'$ the dual exponent to $2/r'$. Using the definition of $r$ we see that $((2/r')')^{2/r'} \lesssim [w]_{A_\infty}$, and as a consequence we can conclude that
\[
\sum_{J \in S} |J(\omega) J(|f| w^{-1/2})_J^2_2 \lesssim [w]_{A_2} [w]_{A_\infty} \|f\|_{L^2(\mathbb{R})}^2
\]
as desired.

Remark. It is now straightforward to prove weighted estimates for bi-parameter dyadic shifts and the type of Journé operators considered in Corollary 3.1, although these estimates are far from sharp. In particular, by using Corollary 3.2 and the iteration argument from Section 3 in [39], one can show that the bi-parameter shifted square function $S_{i,j}$ satisfies the $A_2$ bound $\|S_{i,j}f\|_{L^2(\omega)} \lesssim [w]_{A_2}^4 [w]_{A_\infty}$ (the extra powers come from passing through a martingale transform). Then argue as in the proof of Theorem 3.3, with the $S_{i,j}$ replacing the simpler square functions $S$. Note that the dependence on $[w]_{A_2}$ would be improved if we could prove a sharper weighted estimate for $S_{i,j}$. 85
Appendix: The Weighted Strong Maximal Function

Here we prove Proposition 3.4. As in [31], the idea is to bootstrap the boundedness of the maximal function in dimension one with the help of the $A_\infty$ property of the weight. We follow R. Fefferman’s argument and also use some recent ‘weighted Solyanik estimates’ due to P. Hagelstein and I. Parissis [75], which rely on the sharp reverse Hölder estimates in [42].

Fix $1 < p < \infty$. By the covering-lemma argument in [20], it is enough to show that if $R_1, R_2, ...$ is a sequence of rectangles with sides parallel to the axes, then there is a subcollection \{$\tilde{R}_j$\} of \{$R_j$\} such that

$$\int_{\bigcup \tilde{R}_j} w(x)dx \geq C_1 \int_{\bigcup R_j} w(x)dx$$

(3.19)

and

$$\| \sum_j 1_{\tilde{R}_j} \|_{L^{p'}(w)} \leq C_2 \left( \int_{\bigcup R_j} w(x)dx \right)^{1/p'}$$

(3.20)

where $p'$ is the dual exponent to $p$. In this case one has

$$w(\{M^w f > \alpha\})^{1/p} \leq C_1^{-1} C_2 \frac{\| f \|_{L^p(w)}}{\alpha},$$

so that $\| M^w \|_{L^p(w) \to L^{p,\infty}(w)} \leq C_1^{-1} C_2.$

By monotone convergence we may assume the initial sequence \{$R_j$\} is finite. We choose the subcollection \{$\tilde{R}_j$\} using the Córdoba-Fefferman selection algorithm from [20]. Assume the rectangles \{$R_j$\} have been ordered with decreasing sidelengths in the $x_2$ direction, and take $\tilde{R}_1 = R_1$. Proceeding inductively, let $\tilde{R}_j$ be the first $R_k$ occurring after $\tilde{R}_{j-1}$ so that

$$|R_k \cap \bigcup_{l<k} R_l^*| < (1 - \epsilon)|R_k|,$$
where \( \epsilon \in (0, e^{-c[w]_{A_\infty}}) \). Then arguing as in the proof of Cor. 5.3 from [75], we conclude that
\[
\begin{align*}
  w(\bigcup_j R_j) &\leq (1 + ce^{c[w]_{A_\infty}})^{-1}w(\bigcup_k \tilde{R}_k) \\
  &\leq A \cdot w(\bigcup_k \tilde{R}_k),
\end{align*}
\]
with \( A \) independent of \( w \) (we’ve used the assumed upper bound on \( \epsilon \)). Therefore we can take \( C_1 \) independent of \( w \) in (3.19).

Now take a point \( \bar{x} = (\alpha, \beta) \) inside a rectangle \( R_k \) which does not occur among the \( \tilde{R}_j \). Let \( \{T_i\} \) denote the intervals obtained by slicing the two-dimensional rectangles \( \{R_i\} \) with a line perpendicular to the \( x_2 \) axis at height given by \( \beta \) (the \( x_2 \)-coordinate of \( \bar{x} \)). Given any rectangle \( R = I \times J \), write \( R^* = I \times 3J \). We claim that for each such \( R_i \sim T_i \times J_i \) we must have
\[
|T_i \cap \bigcup \tilde{T}^*_j| \geq (1 - \epsilon)|T_i|, \quad (3.21)
\]
with \( \tilde{T}^*_j \) the slices corresponding to \( \tilde{R}^*_j \). This follows from the assumption about decreasing sidelengths. In fact, we may assume all \( \tilde{T}^*_j \) appearing in the union correspond to \( \tilde{R}_j \) that intersect \( R_i \) and were chosen before \( R_i \) relative to the initial order (the full union is only larger). By the selection criterion
\[
|R_i \cap (\bigcup \tilde{R}^*_j)| \geq (1 - \epsilon)|R_i| = (1 - \epsilon)|T_i||J_i|.
\]
But the sidelengths of the \( \tilde{R}_j \) parallel to the \( x_2 \) axis are longer than \( J_i \), and in particular their three-fold dilates contain \( J_i \). It follows that \( R_i \cap (\bigcup \tilde{R}^*_j) = (T_i \cap \bigcup \tilde{T}^*_j) \times J_i \). This implies (3.21), since we must have \( \frac{|R_i \cap (\bigcup \tilde{R}^*_j)|}{|J_i|} = |T_i \cap \bigcup \tilde{T}^*_j| \).

Next, observe that if \( E_j = \tilde{T}_j - \bigcup_{t < j} \tilde{T}^*_t \) then by arguing as above (and using the selection criterion) we see that \( |\tilde{T}_j \cap \bigcup_{t < j} \tilde{T}^*_t| \leq (1 - \epsilon)|\tilde{T}_j| \), and therefore \( |E_j|/|\tilde{T}_j| > \epsilon \). Since \( w \in A_p \) we
can conclude that
\[ \epsilon^p \leq \left( \frac{|E_j|}{|T_j|} \right)^p \leq [w]_{A_p} \frac{w(E_j)}{w(T_j)} \]
uniformly in the free variable. Hence for any \( f \in L^p(wdx_1) \) with \( \|f\|_p \leq 1 \) we have

\[
\int \sum_j 1_{\tilde{T}_j}(x_1) f(x_1)w(x)dx_1 \leq \epsilon^{-p}[w]_{A_p} \int_{E_j} w(x)dx_1 \cdot \frac{1}{\int_{\tilde{T}_j} w(x)dx_1} \int_{\tilde{T}_j} f(x_1)w(x)dx_1
\]

\[
\leq \epsilon^{-p}[w]_{A_p} \int_{\bigcup \tilde{T}_j} M^w_1(f)(x_1)w(x)dx_1
\]

\[
\leq \epsilon^{-p}[w]_{A_p} \|M^w_1 f\|_{L^p(wdx_1)} \left( \int_{\bigcup \tilde{T}_j} w(x)dx_1 \right)^{1/p'}
\]

\[
\lesssim \epsilon^{-p}[w]_{A_p} \left( \int_{\bigcup \tilde{T}_j} w(x)dx_1 \right)^{1/p'}
\]

using the fact that the weighted one-dimensional maximal operator is bounded independent of \( w \). We also used the disjointness of the sets \( E_j \) to sum. After integrating in \( x_2 \) it follows that we can take \( C_2 = \epsilon^{-p}[w]_{A_p} \) in (3.20).

Combining the above results yields

\[
\|M^w\|_{L^p(w) \to L^{p,\infty}(w)} \lesssim \epsilon^{-p}[w]_{A_p}
\]

for any \( \epsilon \in (0, e^{-c[w]_{A_\infty}}) \). Hence \( \|M^w\|_{L^p(w) \to L^{p,\infty}(w)} \lesssim [w]_{A_p} e^{c[w]_{A_\infty}} \) as claimed.

We do not know if there is an alternative approach the the boundedness of \( M^w \) that yields a smaller dependence on \([w]_{A_\infty}\). Note, however, that Fefferman’s covering lemma is equivalent to the boundedness of \( M^w \) on \( L^p(w) \), up to constants (see [20]).
Appendix: The Equivalence of Disjoint-Parts and Packing Conditions

We prove that Definitions 3.1 and 3.2 are equivalent for rectangles. This proof is contained in a much more general argument due to L. Dor [28]. We owe the presentation here to Jose Condé Alonso, Yumeng Ou, and Jill Pipher. Note that it follows almost from the definition that that disjoint-parts condition implies the packing condition. The main theorem is then the following:

**Theorem 3.3.** Let \( \mathcal{R} \) be a \( \Lambda \)-Carleson collection of axis-parallel rectangles. Then for each \( R \in \mathcal{R} \) there exists a subset \( E_R \subset R \) such that \( |E_R| \geq \Lambda^{-1} |R| \) and such that the sets in the family \( \{ E_R \}_{R \in \mathcal{R}} \) are disjoint.

We state and prove this theorem for Lebesgue measure, but in fact it remains true for any measure \( \mu \) that has no atoms. Note that even in this simplified setting the proof still requires the axiom of choice.

*Proof.* We can assume the collection \( \mathcal{R} \) is finite as long as our estimates are independent of the number of rectangles, so suppose \( \mathcal{R} = \{ R_j \}_{j=1}^N \). Let \( \Sigma = \sigma(\mathcal{R}) \) be the \( \sigma \)-algebra generated by the collection \( \mathcal{R} \). Then \( \Sigma \) is finite and has a finite number of atoms, which we denote by \( \{ B_i \}_{i=1}^L \). Finally we let \( \Omega = \bigcup_j R_j \).

We will study families \( \mathcal{A} = \{ A_j \}_{j=1}^N \) of sets such that

\[
A_j \subset R_j \quad \text{for all } j, \quad \Omega = \bigcup_j A_j, \quad A_j \cap A_{j'} = \emptyset \quad \text{if } j \neq j'. \tag{3.22}
\]

Begin by assuming there exists a partition \( \mathcal{A} \) as in (3.22) that maximizes the quantity

\[
c_\mathcal{A} = \min_{1 \leq j \leq N} \frac{|A_j|}{|R_j|}. \tag{3.23}
\]

We will show that if \( \mathcal{A} \) is such a maximizer then \( c_\mathcal{A} \geq \Lambda^{-1} \), which implies the desired result. We then prove that maximizers actually exist.
We first claim that if $A$ maximizes (3.23) then there must be two different indices $j_0$ and $j_1$ such that

$$
c_A = \frac{|A_{j_0}|}{|R_{j_0}|} = \frac{|A_{j_1}|}{|R_{j_1}|}.
$$

(3.24)

Indeed, suppose this is not the case and (after relabeling)

$$
c_A = \frac{|A_1|}{|R_1|} < \frac{|A_j|}{|R_j|}, \quad 2 \leq j \leq N.
$$

If $c_A = 1$ we have nothing to prove, and otherwise there must be some $j$ with $A_j$ intersecting $R_1$ in a set of positive measure. But then we can modify $A_1 \rightarrow \tilde{A}_1$ and $A_j \rightarrow \tilde{A}_j$ by removing some of $A_j$ and adding it to $A_1$ in such a way that

$$
\frac{|\tilde{A}_1|}{|R_1|} > c_A, \quad \frac{|\tilde{A}_1|}{|R_1|} < \frac{|\tilde{A}_j|}{|R_j|},
$$

$$
\frac{|\tilde{A}_1|}{|R_1|} < \frac{|A_k|}{|R_k|} \quad k \neq 1, j.
$$

This contradicts the assumption that $A$ maximized $c_A$, and so in our case (3.24) must hold.

Next, we define a set

$$\mathcal{I} = \mathcal{I}_A = \{ j : \frac{|A_j|}{|R_j|} = c_A \}.$$

Note that $|\mathcal{I}| \geq 2$. We assume our maximizer $A$ has been chosen so that $|\mathcal{I}_A|$ is as small as possible. In this case we see that if $j \in \mathcal{I}$ and $k \notin \mathcal{I}$ then $|R_j \cap A_k| = 0$, since otherwise we could modify $A_j$ and $A_k$ as in the last paragraph and obtain a maximizer $\tilde{A}$ with $|\mathcal{I}_{\tilde{A}}|$ smaller. In particular, since $A_j \subset R_j$ we can conclude that

$$\bigcup_{j \in \mathcal{I}} R_j = \bigcup_{j \in \mathcal{I}} A_j,$$
which follows from the identity
\[
\bigcup_{j \in I} R_j = \bigcup_{j \in I} \bigcup_{k \in I} (R_j \cap A_k).
\]

Then using the packing condition we have
\[
\frac{1}{\Lambda} \sum_{j \in I} |R_j| \leq \big| \bigcup_{j \in I} R_j \big| = \big| \bigcup_{j \in I} A_j \big| = \sum_{j \in I} |A_j| = c_A \sum_{j \in I} |R_j|,
\]
as desired.

It remains to show that maximizers \( \mathcal{A} \) actually exist. To each measurable \( A_j \subset R_j \) we associate a point \( v_{R_j} \in \mathbb{R}^L \) whose components are
\[
v_{R_j}(i) = |A_j \cap B_i|, \quad 1 \leq i \leq L.
\]

Of course given such a point \( v_{R_j} \) there exist uncountably many subsets of \( R_j \) associated to \( v_{R_j} \).

Note, however, that as long as \( v_j \in \mathbb{R}^L \) satisfies
\[
0 \leq v_j(i) \leq |R_j \cap B_i|, \quad 1 \leq i \leq L,
\]
there exists some subset \( A_{v_j} \subset R_j \) associated to \( v_j \) (since Lebesgue measure is non-atomic).

Since we are looking for ‘disjoint parts’ we will only consider families of \( N \) points \( v_1, ..., v_N \) in \( \mathbb{R}^L \) such that
\[
\sum_{j=1}^{N} v_j(i) = |B_i|, \quad 1 \leq i \leq L, \quad v_j(i) \geq 0, \quad 1 \leq i \leq L, \quad 1 \leq j \leq N.
\]

(3.25)

Given a collection of points \( \{v_j\}_{j=1}^{N} \) satisfying (3.25) we claim that we may associate a partition \( \{A_{v_j}\}_{j=1}^{N} \) satisfying (3.22). Indeed, suppose we have such a collection with \( 0 \leq v_j(i) \leq
\[ |R_j \cap B_i| \]. Using (3.25) we split each \( B_i \) into disjoint pieces \( b_j^i \) such that

\[ |b_j^i| = v_j(i) \quad \text{and} \quad b_j^i \subset R_j, \]

with \( b_j^i = \emptyset \) if \( v_j(i) = 0 \). We then set

\[ A_{v_j} = \bigcup_i b_j^i \]

for each \( j \), and note that these are disjoint unions. Clearly \( A_{v_j} \subset R_j \), and moreover

\[ \bigcup_j A_{v_j} = \bigcup_j b_j^i = \bigcup_i B_i = \Omega. \]

This proves the claim. The collection \( \{A_{v_j}\} \) is non-unique, but this will not affect the argument; we only need some collection with the given properties associated to each tuple of points \((v_1, \ldots, v_N) \subset \mathbb{R}^{L_N}\).

Now assume (by the axiom of choice...) that we fix one collection \( \{A_{v_j}\}_{j=1}^N \) for each tuple of points \((v_1, \ldots v_N)\) satisfying (3.25). The collection of such tuples in \( \mathbb{R}^{L_N} \) define a compact set \( M \subset \mathbb{R}^{L_N} \). We define the function \( f : M \to \mathbb{R}_+ \) by

\[ f(v_1, \ldots, v_N) = \min_{1 \leq j \leq N} \frac{\sum_{i=1}^L v_j(i)}{|R_j|}, \]

and note that \( f \) is continuous. It follows that \( f \) attains a maximum on \( M \) and therefore there must exist a partition \( \mathcal{A} \) as in (3.22) that maximizes (3.23). This completes the proof.

\[ \square \]
Chapter 4

A Counterexample for the Strong Maximal Function

In this chapter we continue to study analogues of sparse bounds in the multi-parameter setting. We will focus on the bi-parameter analogue of the dyadic maximal function, the *dyadic strong maximal function*

\[ \mathcal{M}_S f(x) := \sup_{R \ni x} \langle |f| \rangle_R. \quad (4.1) \]

Here supremum is taken over all dyadic rectangles containing \( x \) with sides parallel to the axes. The strong maximal function is one of the most important operators in the theory of multi-parameter singular integrals, and also perhaps the simplest to analyze since it is positive.

We note that axis-parallel rectangles are the natural geometric objects to consider if one wishes to develop some kind of sparse bound for (4.1), since \( \mathcal{M}_S \) will in general be large on axis-parallel rectangles of arbitrary eccentricity. It is then reasonable to ask if the following sparse estimate holds:

\[ |\langle \mathcal{M}_S f, g \rangle| \lesssim \sum_{R \in \mathcal{S}} \langle |f| \rangle_R \langle |g| \rangle_R |R|, \quad (4.2) \]

where the collection \( \mathcal{S} \) consists of dyadic axis-parallel rectangles. Recall that the direct analogue
of (4.2) for the dyadic Hardy-Littlewood maximal function is true (in fact this operator satisfies a stronger pointwise bound with respect to a sparse collection of dyadic cubes). It is hopeful to note that the strong maximal function is indeed dominated by sparse forms when restricted to a single point mass. Indeed, one can make sense of (4.1) when applied to finite positive measures, and then it is easy to see that

$$M_S(\delta_0)(x, y) \leq \frac{1}{|x||y|}. \quad (4.3)$$

This function can then be dominated by a sparse operator: if we define $$I_m = [0, 2^m)$$ and $$J_m = [2^{m-1}, 2^m)$$ then

$$\frac{1}{|x||y|} = \sum_{m,n \in \mathbb{Z}} \frac{1}{|x||y|} 1_{J_m \times J_n}(x, y) \leq \sum_{m,n \in \mathbb{Z}} 2^{2-m-n} 1_{J_m \times J_n}(x, y) = 4 \sum_{m,n \in \mathbb{Z}} \langle \delta_0 \rangle_{I_m \times I_n} 1_{I_m \times I_n}(x, y).$$

Note that the collection $$S = \{I_m \times I_n : m, n \in \mathbb{Z}\}$$ is $$\frac{1}{4}$$-sparse since for every $$R = I_m \times I_n$$ we can define $$E(R) = J_m \times J_n$$.

This example is relevant because (approximate) point masses are extremal examples for the weak-type behavior of $$M_S$$ and the Hardy-Littlewood maximal functions near the $$L^1$$ endpoint. Moreover, we know from the one-parameter theory that there is a connection between sparse bounds and weak-type endpoint estimates; for example, a sparse bound of type (1.1) with $$(r,s) = (1,1)$$ implies that $$T$$ maps $$L^1$$ into weak $$L^1$$ (see the appendix in [24]). We also note that there is no immediate contradiction implied by a bound of type (4.2); indeed, an estimate of the type

$$|\langle Tf, g \rangle| \lesssim \sum_{R \in S} |f_R||g_R||R|$$

implies that $$T$$ maps $$L \log L$$ into weak $$L^1$$, but known arguments do not imply that $$T$$ maps $$L^1$$ into weak $$L^1$$. The proof is similar to the argument in the appendix of [24], though for
completeness we include the details in an appendix at the end of the chapter.

The main result of this chapter answers the question raised above in the negative: there can be no domination by positive sparse forms of the type (4.2) for the strong maximal function.

**Theorem 4.1.** For every $C > 0$ and $0 < \eta < 1$ there exist a pair of compactly supported integrable functions $f$ and $g$ such that

$$|\langle M_S f, g \rangle| \geq C \sum_{R \in S} \langle |f| \rangle_R \langle |g| \rangle_R |R|,$$

for all $\eta$-sparse collections $S$ of dyadic rectangles with sides parallel to the axes.

The proof of Theorem 4.1 is based on the construction of pairs of extremal functions for which the sparse bound cannot hold. These extremal examples take advantage of the behavior of $M_S$ when applied to sums of several point masses. When we apply $M_S$ to a single point mass we get level sets that look like dyadic stairs, and we will describe a way to place many point masses sufficiently far from each other in such a way that these stairs are packed too tightly to be sparse. Since one needs to ‘cover’ these stairs in order to control $\langle M_S f, g \rangle$, it will follow that no sparse form is large enough to bound $\langle M_S f, g \rangle$ for all $f, g$.

In particular, we will construct a set of points $P$ which are maximally separated with respect to a quantity that captures natural bi-parameter geometry of the problem. Our first extremal function $f$ is then the sum of point masses at the points in $P$. As we said above, this will guarantee that $M_S f$ is uniformly large on the unit square. We complement the construction of the set $P$ with another set, $Z$, with the following structure: for each point $p \in P$ there exist a large amount of points $z \in Z$ that are both near $p$ and far away from all the other points in $P$. Moreover, the family of points in $Z$ associated with a given $p \in P$ are not clustered together, but are instead spread out over the boundary of a hyperbolic ball centered at $p$. Our second extremal function is then the normalized sum of point masses at the points in $Z$. We will show that the placement of the points in $Z$ implies that if a form given by a family $S$ dominates the pair $\langle M_S f, g \rangle$ then the rectangles in $S$ need to cover the portion of the stair centered at $p$ that
contains each $z$, for all $z \in \mathcal{Z}$. But then the resulting collection of rectangles cannot be sparse.

We remark that our result actually extends to the situation in which we take slightly larger averaged norms of $f$. In particular our methods allow us to push Theorem 4.1 to the case of Orlicz $\phi(L)$-norms with $\phi(x) = x \log(x)^\alpha$ for $\alpha < 1/2$. We carry out this extension in Theorem 4.6.

We have learned that our construction of $\mathcal{P}$ is closely connected to discrepancy theory. In particular it is a special construction of a low-discrepancy sequence. Our results show that these kind of sets are particularly well-suited to the study of the strong maximal function (see also [7] for more connections between discrepancy theory and bi-parameter harmonic analysis). It would be interesting to see if tools related to discrepancy theory can be used to analyze other types of maximal operators. In particular, Zygmund-type maximal operators (defined like the strong maximal function, but with rectangles with fewer “parameters” than the ambient dimension) could be amenable to this connection.

The rest of the chapter is devoted to the proof of Theorem 4.1 and some extensions. In Section 4.1 we prove lower bounds for the form associated to $\mathcal{M}_S$ and also upper bounds for the sparse form 4.2 when $f, g$ are the approximate uniform measures described above. We assume the existence of the key sets of points $\mathcal{P}$ and $\mathcal{Z}$, which are needed to construct $f$ and $g$. We then construct the sets $\mathcal{P}$ and $\mathcal{Z}$ in detail in Section 4.2, completing the proof of Theorem 4.1. Finally, in Section 4.3 we show how our results extend to higher dimensions by a tensor product argument. We also show that the strong maximal function is in a sense supercritical for $L^1$-sparse forms, which allows us to slightly strengthen our result.

### 4.1 Lower bounds for $\mathcal{M}_S$ and upper bounds for sparse forms

We start by introducing a quantity which is intimately related with the geometry of the bi-parameter setting: for every pair of points $p, q$ in $[0,1)^2$ we set

$$\text{dist}_H(p,q) = \inf\{|R|^{1/2} : R \in \mathcal{D} \times \mathcal{D} \text{ is a dyadic rectangle containing } p \text{ and } q\}.$$
One can also define a distance between two (dyadic) rectangles:

\[ \text{dist}_{H}(R_1, R_2) := \inf_{p \in R_1, q \in R_2} \text{dist}_H(p, q). \]

Note that \( \text{dist}_{H} \) is not a distance function, and not even a quasidistance, as the triangle inequality is completely false in general. We will, however, refer to it as the distance between two points.

Using this language, the (dyadic) strong maximal function of a point mass (4.3) can now be written as

\[ M_S(\delta_0)(p) = \frac{1}{\text{dist}_H(p, 0)^2}. \]

We are going to study the precise behavior of \( M_S \) when applied to uniform probability measures concentrated on certain point sets: given a finite set of points \( F \) let

\[ \mu = \mu_F = \frac{1}{\#(F)} \sum_{p \in F} \delta_p. \]  

(4.4)

If the points in \( F \) are ‘uniformly’ spread out, then at large enough scales one expects the averages

\[ \langle \mu \rangle_R = \frac{\mu(R)}{|R|} = \frac{|R \cap F|}{\#(F)|R|} \]

to be relatively small. At smaller scales we will be able to exploit some of the key properties of the collections \( P, Z \) described in the following two theorems. We postpone their proofs to Section 4.2.

**Theorem 4.2.** For every \( m \geq 0 \) there exists a collection of points \( P \subset [0, 1)^2 \) such that

\[ \text{dist}_{H}(p, q) \geq 2^{-m}, \quad \forall p \neq q \text{ in } P, \]  

(P.1)

and

\[ \#(P) = 2^{2m+1}. \]  

(P.2)

**Theorem 4.3.** For every \( k \ll m \) and every \( 2^{-m} \)-separated set \( P \subset [0, 1)^2 \) there exists a set
\( Z \subset [0,1)^2 \) satisfying the following properties:

(Z.1) \( \#(Z) \gtrsim m2^m \).

(Z.2) For every \( z \in Z \) there exists exactly one point \( p(z) \in P \) such that

\[
\text{dist}_H(p(z), z)^2 < 2^{-2m-1}.
\]

(Z.3) For every \( z \in Z \),

\[
\text{dist}_H(p(z), z)^2 \sim 2^{-2m-k}.
\]

(Z.4) For every dyadic rectangle \( R \) with \( |R| \leq 2^{-2m-2} \) that intersects \( P \) we have

\[
\#(R \cap Z) \leq k.
\]

(Z.5) For every dyadic rectangle \( R \) with \( |R| \geq 2^{-2m-1} \) we have

\[
\#(R \cap Z) \lesssim 2^{2m}mk|R|.
\]

The implied constants are independent of \( k \) and \( m \).

Remark 4.1.1. A simple pigeonholing argument shows that if \( P \) is the collection from Theorem 4.2 and \( R \) is a dyadic rectangle contained in the unit cube with \( |R| \geq 2^{-2m-1} \), then \( \#(R \cap P) = 2^{2m+1}|R| \). Indeed, if \( |R| = 2^{-2m-1} \) then we may tile the unit cube with \( R \) and \( 2^m \) translates \( \tilde{R} \). But no rectangle in this collection can contain more than one point since the area of each is too small; since there are \( 2^{2m+1} \) rectangles in the tiling then every rectangle must contain exactly one point from \( P \), and in particular \( R \) must contain exactly one point from \( P \). Now if \( |R| > 2^{-2m-1} \) the claim follows by dyadically sub-dividing \( R \) into \( 2^{2m+1}|R| \) distinct rectangles of area \( 2^{-2m-1} \).

Remark 4.1.2. Above, we deliberately omitted the precise dependence of \( m \) on \( k \). It turns out
that any \( m \geq k + C \) is admissible as one can check from the proofs in Section 4.2. This justifies our choices of \( m \) and \( k \) in the next subsections.

### 4.1.1 Lower bound for \( M_S \)

For the rest of this section \( k \ll m \) will be some fixed large numbers, and \((P, Z)\) will be the sets given by Theorems 4.2 and 4.3. Also, \( \mu = \mu_P \) and \( \nu = \nu_Z \) will always denote the associated uniform probability measures introduced in (4.4). The proofs of Theorems 4.3 and 4.3 in the next section show that \( \mu \) and \( \nu \) can be arbitrarily well-approximated by \( L^1 \) functions, in particular because we will be able to choose the points from \( P \) and \( Z \) in arbitrarily small cubes. Because of this we will prove Theorem 4.1 with \( f \) and \( g \) replaced by the measures \( \mu \) and \( \nu \), respectively, and the full result can then be recovered by the limiting argument found at the end of the section in Lemma 4.3.

**Proposition 4.1.** Under these conditions we have

\[
\langle M_S(\mu), \nu \rangle \gtrsim 2^k. \tag{4.5}
\]

**Proof.** The positivity of \( M_S \) makes the proof almost trivial since we do not need to care about possible interactions among the points in \( P \). In particular for all \( z \in Z \) let \( p(z) \) be the point guaranteed by (Z.2) of Theorem 4.3. Then by ((Z.3)),

\[
M_S(\mu)(z) \geq \frac{1}{\#(P)\text{dist}_H(p(z), z)^2} \approx \frac{2^{2m+k}}{\#(P)}.
\]

Now by (P.2) of Theorem 4.2,

\[
M_S(\mu)(z) \gtrsim 2^k
\]

and the claim follows from the fact that \( \nu \) is a uniform probability measure over the set \( Z \). \( \square \)
4.1.2 Upper bound for sparse forms

We now prove upper bounds for sparse forms when acting on $\mu$ and $\nu$. We recall from the last chapter the definition of Carleson sequences:

\textit{Definition 4.1.} Let $\alpha$ be a non-negative function defined on all rectangles that is zero for all but a finite collection of rectangles. We say that $\alpha$ is $\Lambda$-Carleson if for all open sets $\Omega$ we have

$$\sum_{R \subseteq \Omega} \alpha_R |R| \leq \Lambda |\Omega|,$$

where the sum is taken over all rectangles contained in $\Omega$.

We call a collection of dyadic rectangles $S$ a $\Lambda$-Carleson collection if the sequence

$$\alpha_R = \begin{cases} 
1 & \text{if } R \in S, \\
0 & \text{otherwise}
\end{cases}$$

is a $\Lambda$-Carleson sequence. The notions of sparse and Carleson collections are equivalent as we saw in the last chapter: a $\Lambda$-Carleson collection is $\Lambda^{-1}$-sparse and vice-versa.

\textbf{Proposition 4.2.} \textit{For every $\Lambda$-Carleson collection $S$ and all $k$ and $m$ we have}

$$\sum_{R \in S} \langle \mu \rangle_R \langle \nu \rangle_R |R| \lesssim \Lambda k \left( 1 + \frac{2^k}{m} \right). \quad (4.6)$$

\textit{Proof.} For any rectangle $R$ let $\overline{R}$ be the intersection of $R$ with the unit square $[0,1]^2$. We will show that

$$\langle \mu \rangle_R \langle \nu \rangle_R \lesssim k \left( 1 + \frac{2^k}{m} \right) \left( \frac{|\overline{R}|}{|R|} \right)^2. \quad (4.7)$$

Assume first that $R$ is large, i.e.: $|\overline{R}| \geq 2^{-2m-1}$. Then by the $2^{-m}$-separation of $\mathcal{P}$

$$\langle \mu \rangle_R = \frac{1}{2^{2m+1}} \frac{\#(\mathcal{P} \cap \overline{R})}{|R|} = \frac{|\overline{R}|}{|R|}$$

100
Similarly, for \( \nu \) we have by (Z.5):

\[
\langle \nu \rangle_R \lesssim k \frac{|R|}{|R|}.
\]

Therefore, for large rectangles we have

\[
\langle \mu \rangle_R \langle \nu \rangle_R \lesssim k \left( \frac{|R|}{|R|} \right)^2.
\]

If instead \( R \) is small, i.e.: \(|R| < 2^{-2m-1}\), then we have to be more careful. If

\[
\langle \mu \rangle_R \langle \nu \rangle_R \neq 0
\]

then \( \overline{R} \) must contain at least one point \( p \in \mathcal{P} \). In fact, since \( \overline{R} \) is sufficiently small, there exists exactly one \( p \in \mathcal{P} \cap \overline{R} \). Hence

\[
\langle \mu \rangle_{\overline{R}} \lesssim \frac{1}{2^{2m}|R|}.
\]

Since \( \langle \nu \rangle_R \neq 0 \) the rectangle \( \overline{R} \) must contain at least one point from \( \mathcal{Z} \). This implies a lower bound on the size of \( \overline{R} \). To see this observe that if \( z \) is any point in \( R \cap \mathcal{Z} \), then by the smallness of \( R \) and (Z.2) we must have \( p = p(z) \), and hence \( \text{dist}_H(p, z) \sim 2^{-2m-k} \) by (Z.3), therefore

\[
|R| \gtrsim 2^{-2m-k}.
\]

Now, by (Z.4) we have

\[
\#(\overline{R} \cap \mathcal{Z}) \lesssim k,
\]

so

\[
\langle \mu \rangle_{\overline{R}} \lesssim 2^k \quad \text{and} \quad \langle \nu \rangle_{\overline{R}} \lesssim \frac{k2^k}{m},
\]

from which inequality (4.7) follows.
Now we can finish the proof by splitting $\mathcal{S}$ into families $\mathcal{S}_j$ as follows:

$$\mathcal{S}_j = \{ R \in \mathcal{S} : 2^{-j-1}|R| \leq |R| < 2^{-j}|R| \}.$$  

Note that the rectangles in $\mathcal{S}_j$ are contained in

$$\Omega_j = \{ p \in \mathbb{R}^2 : M_S(1_{[0,1]^2}) \gtrsim 2^{-j} \},$$

which, by the weak-type boundedness of the strong maximal function, satisfies

$$|\Omega_j| \lesssim j2^j. \quad (4.8)$$

Then by estimate (4.7), the Carleson condition, and (4.8),

$$\sum_{R \in \mathcal{S}} \langle \mu \rangle_R \langle \nu \rangle_R |R| = \sum_{j=0}^{\infty} \sum_{R \in \mathcal{S}_j} \langle \mu \rangle_R \langle \nu \rangle_R |R|
\lesssim k \left(1 + \frac{2^{2k}}{m}\right) \sum_{j=0}^{\infty} 2^{-2j} \sum_{R \in \mathcal{S}_j} |R|
\leq k \left(1 + \frac{2^{2k}}{m}\right) \sum_{j=0}^{\infty} 2^{-2j} \Lambda|\Omega_j|
\lesssim \Lambda k \left(1 + \frac{2^{2k}}{m}\right) \sum_{j=0}^{\infty} j2^{-j}
\lesssim \Lambda k \left(1 + \frac{2^{2k}}{m}\right).$$

\[\square\]

\textit{Remark 4.1.3.} An examination of the proof above shows that we do not use the full power of the Carleson condition. In particular, the collection $\mathcal{S}$ can be assumed to satisfy the $\Lambda$-Carleson packing condition with respect to the unit cube and the level sets $\Omega_j$, but it does not need to be $\Lambda$-Carleson with respect to any other open sets (and in particular need not be $\Lambda$-Carleson at
small scales).

We can now formally conclude the proof of Theorem 4.1 conditionally on Theorems 4.2 and 4.3.

**Proof of Theorem 4.1.** Choose \( m = k 2^{2k} \). Proposition 4.1 implies that

\[
\langle M_S(\mu), \nu \rangle \gtrsim 2^k,
\]

so Proposition 4.2 forces \( \Lambda \to \infty \) if we make \( k \to \infty \), which leads to a contradiction. \( \square \)

Finally we resolve the issue of approximating the measures \( \mu, \nu \).

**Lemma 4.3.** Let \( \mu, \nu \) be the uniform measures described above. Then if there is no sparse bound of the form

\[
\langle M_S \mu, \nu \rangle \leq C \sum_{R \in \mathcal{S}} |R| R \langle \mu \rangle R \langle \nu \rangle R
\]

there can be no sparse bound of the form

\[
\langle M_S f, g \rangle \leq C \sum_{R \in \mathcal{S}} |R| R \langle f \rangle R \langle g \rangle R
\]

for all \( f, g \in L^1 \).

**Proof.** Fix a sparse parameter \( \eta \in (0, 1) \) and fix \( C > 0 \). We construct \( \mu, \nu \) as described above for fixed \( m, k \). From the proofs of Theorems 4.2 and 4.3 in the next section we see that we may vary points \( p \in \mathcal{P} \) and \( z \in \mathcal{Z} \) in a small neighborhood and still preserve the properties of the theorems.

For each \( p \in \mathcal{P} \) and \( z \in \mathcal{Z} \) we let \( Q_{n,p} \) and \( Q_{n,z} \) be descending sequences of dyadic cubes contained in these neighborhoods, with \( p \in Q_{n,p} \) and \( z \in Q_{n,z} \). We then set

\[
f_n = \frac{1}{\# \mathcal{P}} \sum_{p \in \mathcal{P}} \frac{1_{Q_{n,p}}}{|Q_{n,p}|}
\]
and

\[ g_n = \frac{1}{\# \mathcal{Z}} \sum_{z \in \mathcal{Z}} \chi_{Q_{n,z}}. \]

Then we have the uniform bound

\[ \sum_{R \in \mathcal{S}} \langle f_n \rangle_R \langle g_n \rangle_R |R| \leq \sum_{R \in \mathcal{S}} \langle \mu \rangle_R \langle \nu \rangle_R |R|. \]

On the other hand

\[ \langle M_s f_n, g_n \rangle = \frac{1}{\# \mathcal{Z}} \sum_{z \in \mathcal{Z}} \frac{1}{|Q_{n,z}|} \int_{Q_{n,z}} M_s f_n(y) \, dy, \]

and from the construction of \( \mathcal{Z} \) below we see that for any \( y \in Q_{n,z} \) there exists \( p \in \mathcal{P} \) with

\[ \text{dist}_{2,1}(y, x) \geq 2^{-2m-k} \]

for any \( x \in Q_{n,z} \). By arguing as in Proposition 4.1 we see that

\[ \langle M_s f_n, g_n \rangle \geq c \frac{2^{2m+k}}{\# \mathcal{P}} \sim 2^k \]

(for each \( z \) test against the smallest rectangle containing both \( Q_{n,p} \) and \( Q_{n,z} \)). Therefore if we had an \( \eta \)-sparse form bound with constant \( C \) for all \( L^1 \) functions, it would imply that

\[ c 2^k \leq C \sum_{R \in \mathcal{S}} \langle \mu \rangle_R \langle \nu \rangle_R |R|. \]

Arguing as above shows that this forces \( C \geq c' 2^k \eta \), and we may let \( k \to \infty \) to obtain a contradiction. This works for any \( \eta \in (0, 1) \) and so we are done. \( \square \)

### 4.2 Construction of the extremal sets

In this section, we will sometimes need to work with the projections of a rectangle \( R \) onto the axes. To that end, if \( R = I \times J \) we denote

\[ \pi_1(R) := I, \quad \pi_2(R) := J. \]
For dyadic intervals $I$ we let $\hat{I}$ denote the dyadic parent of $I$. We also let $I^{(j)}$ denote the $j$-fold dyadic dilation of $I$, so that $I^{(j)}$ is the dyadic interval containing $I$ with $|I| = 2^j |I|$. 

### 4.2.1 Construction of $P$ and the proof of Theorem 4.2

The following observation will be useful in the construction:

**Lemma 4.4.** Let $R_1 = I_1 \times J_1, R_2 = I_2 \times J_2$ be two dyadic rectangles such that $I_1 \cap I_2 = J_1 \cap J_2 = \emptyset$. Then

$$\text{dist}_H(p_1, p_2) = \text{dist}_H(q_1, q_2), \quad \forall p_1, q_1 \in R_1, p_2, q_2 \in R_2.$$  

**Proof.** Given any $p_1 \in R_1, p_2 \in R_2$, it suffices to show that any dyadic rectangle $R = I \times J$ containing both $p_1$ and $p_2$ needs to contain $R_1$ and $R_2$. Since $I \cap I_i \neq \emptyset, i = 1, 2$, and $I_1 \cap I_2 = \emptyset$, one has $I \supset I_i, i = 1, 2$. The same holds for $J$: $J \supset J_i$ for $i = 1, 2$. 

**Proof of theorem 4.2.** We are going to show the following claim by induction: for every non-negative integer $m$ there exist $2^{2m+1}$ dyadic squares $Q^m_1, \ldots, Q^m_{2^{2m+1}}$ in $[0, 1)^2$ satisfying

1. $\ell(Q^m_i) = 2^{-2^m-1}$ for all $i$,
2. $\text{dist}_H(Q^m_i, Q^m_j) \geq 2^{-m}$ for all $i \neq j$.

Assuming the claim, it is enough to take

$$P = \{p_Q : Q \in \{Q^m_1, \ldots, Q^m_{2^{2m+1}}\}\},$$

where $p_Q$ denotes the center of the cube $Q$ —in fact, any point of $Q$ would work—. Then, by Lemma 4.4 we immediately get (P.1) and (P.2) and we end the proof.

We turn to the proof of the claim. The case $m = 0$ is easy: it suffices to take $Q^0_1 = [0, 1/2)^2$ and $Q^0_2 = [1/2, 1)^2$. Assume now by induction that the theorem is true for $m - 1$. Scale the family obtained in that step by 1/2 in both dimensions and place a translated copy of the result
in each square $Q \in D_1([0,1)^2)$. Denote the cubes so constructed by $P^Q_1, \ldots, P^Q_{2^{2m-1}}$. By the
dilation invariance of $\text{dist}_H$ we have

$$\ell(P^Q_i) = 2^{-1} \cdot 2^{-2(m-1)} = 2^{-2m}.$$

Therefore, it is enough to choose a first-generation child from each of these squares $\{P^Q_i\}_{i,Q}$ in
such a way that (2) is satisfied. Write $D_1([0,1)^2) = \{Q^0_0, Q^1_0, Q^0_1, Q^1_1\}$, where the children are
listed in the order of upper left, upper right, lower left, and lower right. We first consider $Q^0_0$ and
$Q^1_1$. For each $P^Q_i$ with $Q \in \{Q^0_0, Q^1_1\}$ we choose an arbitrary first-generation child and call
it $Q^Q_i$. By induction we have

$$\text{dist}_H(Q^Q_i, Q^Q_j) \geq \text{dist}_H(P^Q_i, P^Q_j) \geq 2^{-1} \cdot 2^{-(m-1)} = 2^{-m} \ \text{for all} \ i \neq j,$$

while the distance between any square in $Q^0_0$ and any other square in $Q^1_1$ must be exactly 1
according to Lemma 4.4.

Now we must choose the children from the squares in the other diagonal. This choice is
more delicate since the squares from the first diagonal could be much closer. By the pigeonhole
principle, for each $P = P^Q_i, Q \in \{Q^1_0, Q^0_1\}$, there exist exactly two other squares, $P^Q_0_j$ and $P^Q_1_k$,
which are at distance at most $2^{-m}$. Indeed, according to Lemma 4.4, the distance condition
forces $P^Q_0_j, P^Q_1_k$ to be in the same row (or column) as $P^Q_i$, while each row (or column) of $Q^0_0$ (or
$Q^1_1$) contains exactly one square by induction hypothesis. In fact, the distance between $P^Q_i$ and
$P^Q_0_j$ is precisely equal to 0, and the same holds for $P^Q_1_k$.

We now project $Q^Q_j$ and $Q^Q_k$, the children chosen from $P^Q_0_j$ and $P^Q_1_k$ onto $P$. Note that
their projection leaves exactly one first-generation child of $P$ untouched, which we select as $Q^Q_i$.
It suffices to show that the distance from $Q^Q_i$ to any square in $Q^0_0$ is larger than $2^{-m}$, since the
case of $Q^1_1$ is symmetric and any other combination can be dealt with in the same way as above.
To see this, if the square in $Q^0_0$ does not come from the $P^Q_0_j$ chosen above, the distance has to
be larger than $2^{-m}$ by Lemma 4.4. Otherwise, the square is exactly $Q^Q_0_j$. By the construction
above, both its horizontal projection and its vertical projection and the respective ones of $Q^2_i$ are disjoint, which allows one to apply Lemma 4.4 once more to conclude the proof.

### 4.2.2 Construction of $\mathcal{Z}$

There are two aspects of the construction of $\mathcal{Z}$ that will require special care: each point in $\mathcal{Z}$ has to be close to exactly one point from $\mathcal{P}$ (but not too close), while two different points in $\mathcal{Z}$ cannot be too close to each other. We begin by constructing certain special rectangles inside which we shall place the points forming $\mathcal{Z}$.

**Definition 4.2.** We say that a rectangle $R$ is a standard rectangle if it belongs to the collection

$$
\mathcal{E} = \{ R \subseteq [0,1)^2 : R \text{ dyadic, } |R| = 2^{-2m-2}, \text{ and } R \cap \mathcal{P} \neq \emptyset \}.
$$

If $R \in \mathcal{E}$ we let $p_R$ denote the unique point in $R \cap \mathcal{P}$. Also let $\mathcal{E}_p$ denote the collection of standard rectangles containing $p \in \mathcal{P}$.

One can obtain $\mathcal{E}_p$ by area-preserving dilations of one fixed $R \in \mathcal{E}_p$. Indeed, for any dyadic $R = I \times J$ and any $j \in \mathbb{Z}$ define

$$
\text{Dil}_j R = I' \times J',
$$

where $I' \times J'$ is the unique dyadic rectangle in $\mathcal{E}_p$ of dimensions $2^j|I| \times 2^{-j}|J|$. Then it is easy to see that for each standard rectangle $R$, $\mathcal{E}_p = \{ \text{Dil}_j R \subseteq [0,1)^2 \}$. In particular, for any $p \in \mathcal{P}$ there are approximately $m$ distinct standard rectangles in $\mathcal{E}_p$. Given $x,y \in \mathbb{R}$, let

$$
\delta(x,y) = \inf\{|Q| : Q \in \mathcal{D} \text{ is a dyadic interval containing } x \text{ and } y\}.
$$

Also, for dyadic intervals $K$ define

$$
\delta(x,K) := \inf_{y \in K} \delta(x,y).
$$

Now suppose $R = I \times J \in \mathcal{E}$, with $p_R = p = (x,y)$. Given $k \geq 1$, we define $I_{(k)}$ to be the largest
dyadic interval $I'$ such that 

$$\delta(x, I') = 2^{-k+1}|I|,$$

and define $J_{(k)}$ similarly. Note that $I_{(1)}$ is the child of $I$ that does not contain $x$, and for all $k \geq 1$, $I_{(k)}$ is the unique dyadic offspring of $I$ of generation $k$ satisfying $x \notin I_{(k)}$ and $x \in \overline{I_{(k)}}$ (the dyadic parent of $I_{(k)}$).

The following result is the main technical lemma needed for our construction.

**Lemma 4.5.** Fix an integer $k \geq 1$ with $k \ll m$. For any dyadic rectangle $R = I \times J \in \mathcal{E}_p$ there exists a sub-rectangle $R^* \subset I_{(1)} \times J_{(k)}$ such that for all $q \neq p$ in $\mathcal{P}$:

$$\text{dist}_E(q, R^*)^2 \geq 2^{-2m-1}.$$  

We will apply this lemma to construct the points $Z$ used in the proof of the main theorem as follows: given $R = I \times J \in \mathcal{E}$, we choose one point $z_R$ in $R^* \subset I_{(1)} \times J_{(k)}$ and let $Z = \bigcup_{R \in \mathcal{E}} z_R$.

Note that this immediately guarantees properties (Z.1), (Z.2) and (Z.3) of Theorem 4.3. This placing also allows us to prove the remaining two properties. We start by proving directly (Z.4):

**Lemma 4.6.** For every $R \in \mathcal{E}$ we have

$$\#(R \cap Z) \lesssim k.$$  

**Proof.** First observe that if $z_{R'} \in R$ with $R, R' \in \mathcal{E}$ then by Lemma 4.5 we must have $p_{R'} = p_R = p$. So let $Z_R$ be defined by

$$Z_R = \{T \in \mathcal{E}_p : z_T \in R\}.$$  

It suffices to show that $\#(Z_R) \lesssim k$.

Observe that any $T \in Z_R$ must have $\pi_1(T) \subseteq \pi_1(R)$. Indeed, if $\pi_1(T) \supsetneq \pi_1(R)$, by definition $z_T \in \pi_1(T)_{(1)} \times \pi_2(T)_{(k)}$ and one has $\pi_1(R) \subset \pi_1(T)_{(1)}$. But this is impossible since $\pi_1(T)_{(1)} \subseteq$
\( \pi_1(T) \setminus \pi_1(R) \). In addition, there must hold

\[ |\pi_2(T)| \leq 2^{k-1}|\pi_2(R)|, \]

as the smallest interval containing both \( \pi_2(p_T) \) and \( \pi_2(z_T) \) is \( \pi_2(T)(k) \). The desired estimate then follows from the fact that there are \( O(k) \) such standard rectangles.

4.2.3 Proof of lemma 4.5 and (Z.5)

In this subsection, given two dyadic rectangles \( R_1, R_2 \) that are not contained in one another but which do intersect at some nontrivial set, we say that \( R_2 \) intersects \( R_1 \) horizontally if

\[ \pi_2(R_2) \subsetneq \pi_2(R_1), \]

and similarly \( R_2 \) intersects \( R_1 \) vertically if

\[ \pi_1(R_2) \subsetneq \pi_1(R_1). \]

One can see that these two options are mutually exclusive for \( R_1 \) and \( R_2 \). Also, one can see that if \( R_2 \) intersects \( R_1 \) horizontally then \( \pi_1(R_1) \subsetneq \pi_1(R_2) \), and if \( R_2 \) intersects \( R_1 \) vertically then \( \pi_2(R_1) \subsetneq \pi_2(R_2) \). We start by recording the following information that will be used later:

Lemma 4.7. Let \( R = I \times J \in \mathcal{E} \) be such that \( I \times J^{(\ell+1)} \subseteq [0,1)^2, \ell \geq 1 \). Then

\[ I^{(\ell)} \times J^{(\ell+1)} \]

contains exactly one \( q \in \mathcal{P}, q \neq p_R \), and furthermore

\[ q \in I^{(\ell)} \times (J^{(\ell+1)} \setminus J^{(\ell)}). \]

Proof. We have that \( |\widehat{I^{(\ell)}} \times J^{(\ell+1)}| = 2^{-\ell+1}2^{\ell+1}|I||J| = 2^{-2m}, \) so there are exactly two points
from $\mathcal{P}$ in $\widehat{I}_\ell \times J^{(\ell+1)}$. One of these points must be $p_R$, so the other one is $q$. To prove
the second part of the assertion, note that we cannot have $q \in (\widehat{I}_\ell \setminus I_\ell) \times J^{(\ell+1)}$ or $q \in I_\ell \times J^\ell$, since in each case we would have $p_R$ and $q$ contained in a dyadic rectangle of area smaller or equal than $2^{-2m-1}$, violating (P.1).

Lemma 4.8. Let $R = I \times J \in \mathcal{E}$. For any $\ell \geq 1$ with $2^{\ell+1}|J| \leq 1$, there exists exactly one
$T = W \times H \in \mathcal{E}$ intersecting $R$ vertically such that $p_T \neq p_R$, $W \subset I_\ell$, and $|W| = 2^{-1}|I_\ell|$.

Proof. By Lemma 4.7 the rectangle $I_\ell \times J^{(\ell+1)}$ contains exactly one $q \in \mathcal{P}$ and $q \neq p_R$. Moreover, we know that $q \in I_\ell \times (J^{(\ell+1)} \setminus J^\ell)$. We define $W$ as the child of $I_\ell$ that contains
$\pi_1(q)$, and choose $H$ so that $\pi_2(q) \in H$ and $T = W \times H$ has area $2^{-2m-2}$. In particular, we
have $H = J^{(\ell+1)}$. By construction, $T \in \mathcal{E}$.

We now claim that $T$ is the unique standard rectangle intersecting $R$ vertically with $p_T \neq p_R$, $W \subset I_\ell$, and $|W| = \frac{1}{2}|I_\ell|$. By contradiction, suppose there existed another $T' = W' \times H'$ satisfying the same properties. Then since $W, W' \subset I_\ell$ with $|W| = |W'|$ we know that $T$ and $T'$ are disjoint and therefore $p_T \neq p_{T'}$. Since $T'$ is a standard rectangle and $T, T'$ both intersect $R$ vertically, we also know that $H' = H = J^{(\ell+1)}$. But this contradicts Lemma 4.7: that there can only be one $q \in \mathcal{P}$ in $I_\ell \times J^{(\ell+1)}$ with $q \neq p_R$, so $p_T = p_{T'}$, which is a contradiction.

Lemma 4.8 holds for rectangles intersecting $R$ horizontally with exactly the same proof, which we omit for brevity. Assume now that the rectangle $R = I \times J$ has been fixed. Let

$$\mathcal{E}^{(\ell,v)}_R = \{ T \in \mathcal{E} : T \text{ intersects } R \text{ vertically and } \pi_1(T) \subset I_\ell \}.$$ 

and

$$\mathcal{E}^{(\ell,h)}_R = \{ T \in \mathcal{E} : T \text{ intersects } R \text{ horizontally and } \pi_2(T) \subset J_\ell \}.$$ 

We will analyze the vertical rectangles $\mathcal{E}^{(\ell,v)}_R$; essentially the same arguments apply to the hori-
zontal collection $\mathcal{E}^{(\ell,h)}_R$.

If $S$ and $T$ are in $\mathcal{E}^{(\ell,v)}_R$, we say that $S \preceq T$ if $\pi_1(S) \subset \pi_1(T)$ (an analogous order can be
defined in $\mathcal{E}_R^{\ell,v}$. Denote by $\mathcal{E}_R^{\ell,v,*} = \{T_1, T_2, \ldots\}$ the set of maximal elements with respect to the ordering $\preceq$. Assume they are ordered so that the sequence $a_i := |\pi_1(T_i)|$ is non-increasing.

**Lemma 4.9.** The sequence $a_i$ satisfies

$$a_i = 2^{-i}|I(\ell)|, \quad 1 \leq i \leq m - \ell.$$

**Proof.** By Lemma 4.8 we find that $T_1$ satisfies $\pi_1(T_1) \subset I(\ell)$ and $a_1 = 2^{-1}|I(\ell)|$. Moreover, we know that there can be no other $T_j$ with $a_j = a_1$, and so in particular $a_2 \leq 2^{-2}|I(\ell)|$.

We assume inductively that the desired result holds for $T_1, \ldots, T_{i-1}$, and aim to show that $a_i = 2^{-i}|I(\ell)|$. Note that $\{\pi_1(T_1), \ldots, \pi_1(T_{i-1})\}$ are pairwise disjoint. Then by induction

$$\pi_1(T_i) \subset I(\ell) \setminus \bigcup_{j=1}^{i-1} \pi_1(T_j),$$

and we also have that $a_j = 2^{-1}a_{j-1}$ and $a_1 = 2^{-1}|I(\ell)|$. Therefore, $\pi_1(T_i)$ must be contained in some interval $K$ with $|K| = 2^{-i+1}|I(\ell)|$. Moreover, by induction $K$ and $\pi_1(T_{i-1})$ must have the same dyadic parent.

We first show that $a_i < |K|$. To see this, suppose by contradiction that $\pi_1(T_i) = K$. Then

$$a_i = a_{i-1}$$

and

$$\pi_1(T_i) = \pi_1(T_{i-1}).$$

Since both $T_{i-1}$ and $T_i$ must also have the same height, their union is a dyadic rectangle of area $2^{-2m-1}$ which contains both $p_{T_{i-1}}$ and $p_{T_i}$, which is impossible by (P.1).

Finally, we have to show that $a_i \geq 2^{-1}a_{i-1}$ or, equivalently, that there exists $T \in \mathcal{E}_R^{\ell,v,*}$ with $|\pi_1(T)| = 2^{-1}a_{i-1}$. To that end, let $\tilde{T}_{i-1} \in \mathcal{E}_{p_{T_{i-1}}}$ be the unique rectangle with $\pi_1(\tilde{T}_{i-1}) = \hat{K}$. If $2^2|\pi_2(\tilde{T}_{i-1})| \leq 1$, one can apply Lemma 4.8 with $\ell = 1$ to $\tilde{T}_{i-1}$ to conclude that there is exactly one $T \in \mathcal{E}$ intersecting $\tilde{T}_{i-1}$ vertically with $p_T \neq p_{T_{i-1}}$ and

$$|\pi_1(T)| = 2^{-1}|\hat{K}_{(1)}| = 2^{-i}|I(\ell)|.$$
$T_i = T$ by maximality, and this closes the induction. If on the other hand $2^2|\pi_2(\tilde{T}_{i-1})| > 1$, then we claim that $i > m - \ell$ and we have finished. Indeed, if $T_i$ exists in this case, one has

$$|\pi_1(T_i)| < |K| = 2^{-1}|\pi_1(\tilde{T}_{i-1})|.$$  

Therefore,

$$|\pi_2(T_i)| > 2|\pi_2(\tilde{T}_{i-1})| \geq 2^2|\pi_2(\tilde{T}_{i-1})| > 1,$$

which means it is impossible for $T_i$ to be contained in $[0,1)^2$.

Proof of lemma 4.5. It is enough to show the following slightly stronger claim: for any pair of integers $k, \ell \geq 1$, any $R = I \times J \in \mathcal{E}$ contains a non-empty rectangle $E \times F$ such that

$$E \times F \subseteq I(\ell) \times J(k) \setminus \bigcup_{R' \in \mathcal{E} \atop p_R \neq p_{R'}} R'.$$

Any $R' \in \mathcal{E}$ which intersects $I(\ell) \times J(k)$ must be in one of the two collections $\mathcal{E}_R^{\ell,v}$, $\mathcal{E}_R^{k,h}$, when $p_R \neq p_{R'}$.

We start by constructing $E$, so now we only need to take rectangles belonging to $\mathcal{E}_R^{\ell,v,*}$ into account. From Lemma 4.9 we know that $a_i = |\pi_1(T_i)|$ satisfies $a_i = 2^{-i}|I(\ell)|$. It thus follows that $|\pi_2(T_i)|$ must increase in size exponentially, and therefore the total number of $T_i \in \mathcal{E}_R^{\ell,v,*}$ is bounded by some constant that depends on $R$ (since all rectangles are contained in $[0,1)^2$).

Since there are only finitely many $\pi_1(T_i) \subseteq I(\ell)$, the estimate on their sizes from Lemma 4.9 implies that there must be some dyadic subinterval $E \subseteq I(\ell)$ with

$$\pi_1(T_i) \cap E = \emptyset$$

for all $i$,

as desired. The same procedure yields an interval $F \subseteq J(k)$ which no rectangles $T' \in \mathcal{E}_R^{k,h,*}$ can intersect.

Finally, we turn to the proof of (Z.5), which is the content of the next lemma.
Lemma 4.10. For every dyadic rectangle $R \subset [0,1]^2$ with $|R| \geq 2^{-2m-1}$ we have

$$\#(R \cap \mathcal{Z}) \lesssim 2^{2m} mk |R|.$$ 

Proof. First, we may assume that $|R| = 2^{-2m-1}$ and show

$$\#(R \cap \mathcal{Z}) \lesssim mk.$$ 

Indeed, if it is larger we can just write it as a disjoint union of dyadic rectangles of area $2^{-2m-1}$ and use the estimate for rectangles of that size. Since $|R| = 2^{-2m-1}$ we know that it contains exactly one point $p_0 \in \mathcal{P}$, and from Lemma 4.6 we know $R$ contains about $k$ points $z \in \mathcal{Z}$ such that $p(z) = p_0$.

For every $z \in R \cap \mathcal{Z}$ such that $p(z) \neq p_0$, there exists $T \in \mathcal{E}$ for which $z_T = z$, which we denote by $T_z$. Since both $R$ and $T_z$ are dyadic rectangles and $|T_z| < |R|$ we must have

$$\mathcal{Z} \cap R = \mathcal{V} \sqcup \mathcal{H} \sqcup \mathcal{O},$$

where $\mathcal{V}$ consists of those points $z \in \mathcal{Z} \cap R$ such that $T_z$ intersects $R$ vertically, $\mathcal{H}$ consists of those points $z \in \mathcal{Z} \cap R$ such that $T_z$ intersects $R$ horizontally, and $\mathcal{O}$ is the collection of points $z \in \mathcal{Z}$ whose associated point $p(z) = p_0$. We will only estimate the size of $\mathcal{V}$, since the argument for $\mathcal{H}$ is similar.

By the construction of $\mathcal{Z}$ and the previous arguments, for each $z \in \mathcal{V}$ there exists $T \in \mathcal{E}_R^{\ell,v,*}$ containing $z$. By Lemma 4.6 it suffices to show that there are no more than some constant times $m$ such rectangles.

Let $\tilde{R} = \tilde{I} \times \tilde{J}$ be the unique rectangle in $\mathcal{E}$ satisfying

$$p_R \in \tilde{R} \subset R \text{ and } \pi_1(\tilde{R}) = \pi_1(R).$$

Observe that any rectangle $T \in \mathcal{E}_R^{\ell,v,*}$ intersecting $R$ vertically and whose $z_T$ is inside $R$ must
intersect \((\tilde{I})_{(1)} \times \tilde{J}\) vertically (since otherwise \(z_T\) would be too close to \(p_0\)). According to Lemma 4.8 with \(\ell = 1\), the maximal rectangle with shortest height must have height at least \(4|\tilde{J}| = |\pi_2(R)|\) and the heights of these maximal rectangles increase exponentially. Therefore the number of maximal rectangles going through \((\tilde{I})_{(1)}\) is at most

\[
\log_2(2^{2m}|\pi_2(R)|) \leq 2m + \log_2(|\pi_2(R)|) \leq 2m.
\]

\[\square\]

The proof of Theorem 4.3 is complete.

### 4.3 Extensions and open questions

#### 4.3.1 Failure of sparse bound for bi-parameter martingale transform

In this subsection, we extend our main theorem to the bi-parameter martingale transform, which is a 0-complexity dyadic shift that resembles the behavior of the bi-parameter Hilbert transform. We recall the definition of such operators from the last chapter: given a sequence \(\sigma = \{\sigma_R\}_{R \in D \times D}\) satisfying \(|\sigma_R| \leq 1\), the corresponding martingale transform is defined as

\[
T_\sigma(f) := \sum_{R \in D} \sigma_R \langle f, h_R \rangle h_R.
\]

We have the following lower bound result.

**Theorem 4.4.** Let measures \(\mu, \nu\) be the same as above. Then there exists a martingale transform \(T_\sigma\) such that

\[
|\langle T_\sigma(\mu), \nu \rangle| \gtrsim 2^k.
\]

Recalling the upper estimate of sparse forms obtained in Subsection 4.1.2, one immediately gets the following corollary:
Corollary 4.1. There cannot be any $(1,1)$ sparse domination for all bi-parameter martingale transforms.

Proof of Theorem 4.4. We first assume that $\sigma_R = 0$ unless $R \subset [0,1]^2$. For any $R = I \times J \in \mathcal{E}$ that gives rise to some point $z \in \mathcal{Z}$ we let $T_R := \overline{I(\ell)} \times \overline{J(k)}$. This is the smallest rectangle that contains both $z$ and $p(z) \in P$. We also denote the standard rectangle that gives rise to $z \in \mathcal{Z}$ by $R_z$. Now for any fixed $z \in \mathcal{Z}$,

$$T_\sigma(\mu)(z) = \frac{1}{\#P} \sum_{p \in P} \sum_{R \subset [0,1]^2} \sigma_R h_R(p) h_R(z).$$

Set $\sigma_T = 0$ unless $T = T_R$ for some $R \in \mathcal{E}$. In particular, $\{T_R\}_{R \in \mathcal{E}}$ and $\{R\}_{R \in \mathcal{E}}$ are one-to-one, and each $T_R$ contains exactly one $p_R \in P$. Hence,

$$T_\sigma(\mu)(z) \sim 2^{-2m} \sum_{R \in \mathcal{E}} \sigma_{T_R} h_{T_R}(p_R) h_{T_R}(z).$$

Note that $z$ cannot be contained in $T_R$ unless $R = R_z$ according to Lemma 4.11 below, and therefore

$$T_\sigma(\mu)(z) \sim 2^{-2m} \sigma_{T_{R_z}} h_{T_{R_z}}(p_{R_z}) h_{T_{R_z}}(z) = 2^{-2m} \sigma_{T_{R_z}} |T_{R_z}|^{-1} \xi_{T_{R_z}},$$

where $\xi_{T_{R_z}} = \pm 1$. Choosing $\sigma_{T_{R_z}} = \xi_{T_{R_z}}$, one has

$$T_\sigma(\mu)(z) \sim 2^{-2m} |T_{R_z}|^{-1} = 2^{-2m} \left(2^{-(k+1)|R_z|}\right)^{-1} \sim 2^k,$$

and the desired estimate follows immediately. \qed

In the proof above, we have used the key observation that even though $R \in \mathcal{E}$ can contain many points $z \in \mathcal{Z}$ with $p_R = p_{R_z}$, $T_R$ can only contain one $z_R \in \mathcal{Z}$. This is justified by the following result.

Lemma 4.11. For any integers $k, \ell \geq 1$ and $R = I \times J \in \mathcal{E}$, let $z_R \in \mathcal{Z}$ be the point chosen in...
as in Subsection 4.2.2, and \( T_R = \widehat{I_\ell} \times \widehat{J_k} \). Then

\[
    z_T \notin T_R, \quad \forall R, T \in \mathcal{E}, R \neq T.
\]

**Proof.** Fix \( R = I \times J \in \mathcal{E} \) and denote \( z = z_R \). Given any other \( R' = I' \times J' \in \mathcal{E} \), our goal is to show \( z_{R'} \notin T_R \). Obviously, if \( p_{R'} \neq p_R \), then by Lemma 4.5 \( z_{R'} \) is not even contained in \( R \). It thus suffices to assume \( p_{R'} = p_R = p \). Suppose \( |I| > |I'| \), i.e. \( I = I^{(j)} \) for some \( j \geq 1 \), then \( |J| = 2^{-j} |J'| \) and \( J \subset J' \). Since both \( \hat{J}_k \) and \( \hat{J}'_k \) contain \( \pi_2(p) \), one has \( \hat{J}'_k \supseteq \hat{J}_k \).

We claim that \( \pi_2(z_{R'}) \notin \hat{J}_k \), hence \( z_{R'} \notin T_R \). Indeed, \( \pi_2(z_{R'}) \in \hat{J}'_k \), and \( \hat{J}_k \) is either contained in \( \hat{J}'_k \) or its dyadic sibling. But it is impossible for \( \hat{J}_k \) to be contained in \( \hat{J}'_k \) as it would imply \( \pi_2(p) \in \hat{J}_k \), which is an obvious contradiction. Therefore, \( \pi_2(z_{R'}) \notin \hat{J}_k \) and the proof for the case \( |I| > |I'| \) is complete.

The case \( |I| < |I'| \) can be treated symmetrically, as one has \( |J| < |J'| \) and can show in the same way as above that \( \pi_1(z_{R'}) \notin \hat{I}_\ell \).

**Remark 4.3.1.** Using the same method, one can easily show that there exists a bi-parameter dyadic shift of any given complexity which cannot have a \((1,1)\) sparse bound. We omit the details.

### 4.3.2 Failure of sparse bound in higher dimensions

For \( n \geq 1 \), define the \( n \)-parameter strong maximal function

\[
    \mathcal{M}_n(f)(x) := \sup_{R \ni x} \frac{1}{|R|} \int_R |f(y)| \, dy, \quad f : \mathbb{R}^n \to \mathbb{C},
\]

where \( R \) is any \( n \)-dimensional dyadic rectangle. Clearly, \( \mathcal{M}_S \) equals \( \mathcal{M}_2 \). We have the following theorem.

**Theorem 4.5.** Let \((D_n)\) denote the following statement: there exists \( C > 0 \) such that for all compactly supported integrable functions \( f, g \) on \( \mathbb{R}^n \), there exists a sparse collection of dyadic
rectangles $\mathcal{S}$ such that
\begin{equation*}
|\langle \mathcal{M}_n(f), g \rangle| \leq C \sum_{R \in \mathcal{S}} \langle |f| \rangle_R \langle |f| \rangle_R |R|.
\end{equation*}

Then for all $n \geq 2$, there holds
\begin{equation*}
(D_n) \implies (D_{n-1}).
\end{equation*}

Proof. The desired result follows from a tensor product argument. For the sake of simplicity, we will only prove $(D_3) \implies (D_2)$. Our goal is to show $(D_2)$: for any given functions $f, g$ on $\mathbb{R}^2$, we would like to find a sparse dominating form for $\langle \mathcal{M}_2(f), g \rangle$. Define $\tilde{f} = f \otimes \chi_{[0,1)}$ and $\tilde{g} = g \otimes \chi_{[0,1)}$, both are compactly supported integrable functions on $\mathbb{R}^3$. By the assumption $(D_3)$, there exists a sparse collection $\tilde{\mathcal{S}}$ of rectangles in $\mathbb{R}^3$ such that
\begin{equation*}
|\langle \mathcal{M}_3(\tilde{f}), \tilde{g} \rangle| \leq C \sum_{\tilde{R} = R \times J \in \tilde{\mathcal{S}}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}|,
\end{equation*}
where $R \in \mathbb{R}^2, J \in \mathbb{R}$.

By definition, for all $\tilde{x} = (x, x_3) \in \mathbb{R}^3$ with $x_3 \in [0,1),$
\begin{equation*}
M_3(\tilde{f})(\tilde{x}) = \sup_{R \times J \ni \tilde{x}} \frac{1}{|R| \cdot |J|} \int_{R \times J} |\tilde{f}(\tilde{y})| d\tilde{y} = M_2(f)(x)M(1_{[0,1)})(x_3) = M_2(f)(x).
\end{equation*}

Therefore,
\begin{equation*}
\langle M_3(\tilde{f}), \tilde{g} \rangle_{\mathbb{R}^3} = \langle M_2(f), g \rangle_{\mathbb{R}^2}.
\end{equation*}

It thus suffices to show that the sparse form in $\mathbb{R}^3$ can be dominated by a sparse form in $\mathbb{R}^2$.

To see this, rewrite
\begin{equation*}
\sum_{\tilde{R} = R \times J \in \tilde{\mathcal{S}}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}| = \sum_{\tilde{R} = R \times J \in \mathcal{R}^3} \alpha_{\tilde{R}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}|,
\end{equation*}
where $\alpha_{\tilde{R}} = 1$ if $\tilde{R} \in \mathcal{R}$, and 0 otherwise. The sparsity of $\mathcal{R}$ thus means for some constant $\Lambda$
there holds
\[ \sum_{\tilde{R} \subset \tilde{\Omega}} \alpha_{\tilde{R}} |\tilde{R}| \leq \Lambda |\tilde{\Omega}|, \quad \forall \text{ open set } \tilde{\Omega} \subset \mathbb{R}^3. \]

One can further simplify the expression
\[
\sum_{\tilde{R} = R \times J \subset \mathbb{R}^3} \alpha_{\tilde{R}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}|
\]
\[
= \sum_{R \subset \mathbb{R}^2} \langle |f| \rangle_{R} \langle |g| \rangle_{R} |R| \sum_{J \subset \mathbb{R}^2} \alpha_{R \times J}|J| \cdot \left( \frac{|[0,1) \cap J|}{|J|} \right)^2
\]
\[
=:\sum_{R \subset \mathbb{R}^2} \langle |f| \rangle_{R} \langle |g| \rangle_{R} |R| |\beta_R|.
\]

It suffices to show that \( \{ \beta_R \}_{R \subset \mathbb{R}^2} \) is a Carleson sequence.

Indeed, decomposing
\[
\beta_R = \sum_{j=0}^{\infty} 2^{-2j} \sum_{J \subset \mathbb{R}^2 : |J \cap [0,1)| \sim 2^{-j}|J|} \alpha_{R \times J}|J|
\]
one has for all open set \( \Omega \subset \mathbb{R}^2 \) that
\[
\sum_{R \subset \Omega} \beta_R |R| = \sum_{j=0}^{\infty} 2^{-2j} \sum_{R \subset \Omega} \sum_{J \subset \mathbb{R}^2 : |J \cap [0,1)| \sim 2^{-j}|J|} \alpha_{R \times J}|R \times J|
\]
\[
\leq \sum_{j=0}^{\infty} 2^{-2j} \sum_{R \times J \subset \tilde{\Omega}} \alpha_{R \times J}|R \times J|,
\]
where \( \tilde{\Omega} := \Omega \times \{ x_3 \in \mathbb{R} : M(\chi_{[0,1)})(x_3) > C2^{-j} \} \). By the weak \( L^1 \) boundedness of \( M \), one has
\[
|\tilde{\Omega}| \lesssim 2^j |\Omega|,
\]
hence
\[
\sum_{R \subset \Omega} \beta_R |R| \lesssim \Lambda \sum_{j=0}^{\infty} 2^{-j} |\Omega| \lesssim \Lambda |\Omega|.
\]
The proof is complete.

Similar results hold true for multi-parameter martingale transforms as well. One can define \( \tilde{f} = f \otimes h_{[0,1)} \) and \( \tilde{g} = g \otimes h_{[0,1)} \). Then one observes that \( \langle |T_\sigma(h_{[0,1)}), h_{[0,1)}| \rangle \geq C \) for some martingale transform \( T_\sigma \) in the third variable and therefore can argue similarly as above. The
4.3.3 Other sparse forms

As mentioned in the introduction, it is also interesting to consider the possibility of domination of the bi-parameter operators by bigger sparse forms, for instance those of type \((\phi, 1)\):

\[
\sum_{R \in S} \langle |f| \rangle_{R, \phi} \langle |g| \rangle_{R} |R|,
\]

where

\[
\langle f \rangle_{R, \phi} = \inf \left\{ \lambda > 0 : \frac{1}{|R|} \int_{R} \phi \left( \frac{f(x)}{\lambda} \right) \leq 1 \right\} dx.
\]

Interesting norms include the cases \(\phi(x) = x[\log(e + x)]^\alpha, \alpha > 0, \) and \(\phi(x) = x^p, p > 1.\) We can extend our main theorem to the following:

**Theorem 4.6.** Let \(\phi(x) = x[\log(e + x)]^\alpha, 0 < \alpha < \frac{1}{2}.\) Then there cannot be any \((\phi, 1)\)-sparse domination for the strong maximal function or the martingale transform \(T_\sigma.\)

**Proof.** Fix \(k \ll m.\) We need to first slightly modify the measure \(\mu\) to get a function that lies in the correct normed space. Let \(\nu\) be the same as before and define

\[
f = \frac{1}{\# \mathcal{P}} \sum_{p \in \mathcal{P}} \frac{1_{Q_p}}{|Q_p|},
\]

where for each \(p \in \mathcal{P}, Q_p \ni p\) is the small cube of side-length \(\sim 2^{-2m}\) that is constructed in the proof of Theorem 4.2. By the exact same argument in Section 2, it is easy to check that the same lower bound for the bilinear forms associated to \(M_S\) and the special martingale transform \(T_\sigma\) still hold true, i.e.

\[
\langle M_S(f), \nu \rangle \gtrsim 2^k, \quad \langle T_\sigma(f), \nu \rangle \gtrsim 2^k.
\]

We shall now focus on showing the upper bound. For every \(\Lambda\)-Carleson collection \(S\) we will show
that
\[
\sum_{R \in S} \langle f \rangle_{R, \phi} \nu_R |R| \lesssim \Lambda k m^\alpha \left( 1 + \frac{2^k}{m} \right). \tag{4.9}
\]
If (4.9) holds true, then by taking \( m = 2^{2k} \) and by noting that \( \alpha < \frac{1}{2} \), one immediately sees that \( \Lambda \) blows up to infinity as \( k \to \infty \), which completes the proof.

We now prove (4.9). Compared to the estimate of \( \langle \mu \rangle_R \) in the proof of Theorem 4.2, it suffices to show that there is at most an extra factor of \( m^\alpha \) in the upper estimate of the new average form \( \langle f \rangle_{R, \phi} \). Specifically, in the case \( |R| = |R \cap [0,1)^2| \geq 2^{-2m-1} \), we would like to show that
\[
\langle f \rangle_{R, \phi} \lesssim \frac{|R|}{|R|} \cdot \max \left( m^\alpha, \left[ \log \left( \frac{|R|}{|R|} \right) \right]^\alpha \right). \tag{4.10}
\]
Note that if this estimate holds true, then one can proceed as in the proof of Theorem 4.2 to conclude that contribution to \( \sum_{R \in S} \langle f \rangle_{R, \phi} \nu_R |R| \) from rectangles of this type is controlled by \( \Lambda k (1 + m^\alpha) \). Similarly, in the case \( |R| < 2^{-2m-1} \), we will show that
\[
\langle f \rangle_{R, \phi} \lesssim \frac{m^\alpha}{2^{2m} |R|}, \tag{4.11}
\]
which, combined with the estimates for large rectangles, completes the proof of (4.9).

To see why (4.10) holds, let \( \lambda > 0 \) and one has
\[
\frac{1}{|R|} \int_R \phi \left( \frac{f(x)}{\lambda} \right) dx = \frac{1}{|R|} \sum_{p \in \mathcal{P} \cap R} \int_{R \cap Q_p} \phi \left( \frac{\sum_{q \in \mathcal{P}} \frac{1}{|Q_q|} \cdot \# \mathcal{P} \cdot |Q_p|}{\lambda \cdot \# \mathcal{P}} \right) dx
\]
\[
\leq \frac{1}{|R|} \cdot \frac{\# \mathcal{P} \cap \overline{R}}{\# \mathcal{P}} \cdot |R \cap Q_p| \cdot \phi \left( \frac{1}{\lambda \cdot \# \mathcal{P} \cdot |Q_p|} \right)
\]
\[
\leq \frac{|R|}{|R|} \cdot 2^{-2m} \cdot \phi \left( \frac{2^m}{\lambda} \right). \tag{4.10}
\]
Take \( \lambda = \frac{|R|}{|R|} \cdot \max \left( m^\alpha, \left[ \log \left( \frac{|R|}{|R|} \right) \right]^\alpha \right) \), it suffices to show that the expression above with this \( \lambda \)
value is $\lesssim 1$. Indeed, note that $e \ll \frac{2^{2m}}{\lambda}$, so
\[
\frac{|R|}{|\overline{R}|} \cdot 2^{-2m} \cdot \phi \left( \frac{2^{2m}}{\lambda} \right) \lesssim \frac{2m + \log \left( \frac{|R|}{|\overline{R}|} \right) - \alpha \log \left( \max \left( m, \log \left( |R|/|\overline{R}| \right) \right) \right) \alpha}{\max \left( m^\alpha, \left[ \log \left( \frac{|R|}{|\overline{R}|} \right) \right]^\alpha \right) \lesssim 1.}
\]

To see (4.11), again let $\lambda > 0$ and one can calculate similarly as above to obtain
\[
\frac{1}{|R|} \int_{\mathbb{R}} \phi \left( \frac{f(x)}{\lambda} \right) dx \leq 2^{-2m} \phi \left( \frac{2^{2m}}{\lambda} \right) = \frac{1}{\lambda} \left[ \log (e + \frac{2^{2m}}{\lambda}) \right]^\alpha.
\]

Take $\lambda = \frac{m^\alpha}{2^{2m} |R|}$, then it suffices to show that the expression above with this $\lambda$ value is $\lesssim 1$. Indeed, the lower bound $|R| \gtrsim 2^{-2m-k}$ guarantees that $\frac{2^{2m}}{\lambda} \gg e$, hence
\[
\frac{1}{\lambda} \left[ \log (e + \frac{2^{2m}}{\lambda}) \right]^\alpha \lesssim \frac{2^{2m} |R|}{m^\alpha} [4m + \log |R| - \alpha \log m]^\alpha \lesssim 2^{2m} |R| \lesssim 1.
\]

The argument in the theorem above is not strong enough to produce a contradiction when $\alpha \geq \frac{1}{2}$. In other words, it could be possible that the bi-parameter operators that are considered have a $(\phi, 1)$-type sparse bound with some $\alpha \geq \frac{1}{2}$. Moreover, it is easy to check that a very similar argument as in Theorem 4.6 barely fails for concluding a contradiction when $\phi(x) = x^p$, $p > 1$. We think that positive or negative results for those types of bounds would be very interesting.

**Appendix: Endpoint Properties of the Sparse Form**

We will explain why the sparse-form bound considered in this chapter would not obviously contradict well-known mapping properties of $M_S$ near $L^1$.

Recall that if $T$ is an operator such that for all bounded, compactly supported $f, g$ there is
a sparse collection of cubes $\mathcal{S}$ such that

$$|\langle Tf, g \rangle| \leq C \sum_{Q \in \mathcal{S}} \langle f \rangle_Q \langle g \rangle_Q |Q|$$

then $T$ must map $L^1$ into $L^{1,\infty}$. The argument proceeds by using the standard duality result for the quasi-Banach space $L^{1,\infty}$. Indeed, in order to show that $Tf \in L^{1,\infty}$ it suffices to show that that for every set $G \subset \mathbb{R}^n$ with $0 < |G| < \infty$ there exists $G' \subset G$ with $|G| \leq 2|G'|$ such that if $g = 1_{G'}$ then

$$|\langle Tf, g \rangle| \lesssim 1$$

(see for example Section 2.4 in [60]). In particular one applies this estimate to the level sets

$$G = \{Tf > \mu\}, \quad \tilde{G} = \{-Tf > \mu\},$$

so $g$ above is usually taken to be the indicator function of a significant subset of some level set of $Tf$. One can use the sparse bound to prove (4.12) after picking $G' \subset G$ to be a certain subset of $G$ where $Mf \leq c|G|^{-1}$ (here $M$ is Hardy-Littlewood maximal function); see the appendix in [19] for details.

We briefly indicate why the argument outlined above does not extend to the bi-parameter setting. We let $f = 1_{Q_0}$ with $Q_0$ the unit cube, and note that if $G = \{M_S f > N^{-1}\}$ then $|G| \sim N \log N$ (indeed the level set is essentially a ‘dyadic staircase’ which contains $\sim \log N$ rectangles containing $Q_0$, each of area $N$). But similar reasoning shows that

$$|\{M_S f > c|G|^{-1}\}| \sim (N \log N) \log(N \log N) \gg |G|,$$

and so it will not be possible to extract a subset $G'$ with $|G'| \sim |G|$ such that $M_S f \leq c|G|^{-1}$ in $G'$. This step is a crucial part of the proof in [19] that a (one-parameter) sparse bound implies a weak $(1,1)$ mapping, and this argument will not extend to the bi-parameter setting.
Part II

Semiperiodic Strichartz Estimates
Chapter 5

Strichartz Estimates on $\mathbb{R}^n$ and $\mathbb{T}^d$

In this part of the thesis we discuss some recent results related to Strichartz-type estimates for solutions to the linear Schrödinger equation on product manifolds of the form $\mathbb{R}^n \times \mathbb{T}^d$, where $\mathbb{T}^d$ is a $d$-dimensional torus (the ‘semiperiodic’ Schrödinger equation). We also include a few applications at the end of the next chapter.

We begin with a brief overview of some important background material, in particular related to the linear Schrödinger equation in the Euclidean and periodic cases. This summary of Strichartz estimates in the Euclidean and periodic setting is far from complete, though we hope that it gives some motivation for the main results and arguments in the following chapter.

5.0.1 Strichartz Estimates on Euclidean Space

If $u$ is a solution to the linear Schrödinger equation on $\mathbb{R}^n$

\[
\begin{cases}
- i \partial_t u + \Delta u = 0, \\
 u(x, t) = u_0
\end{cases}
\]  

(5.1)

\[\text{125}\]
then by taking the Fourier transform and solving the resulting ODE, we see that solutions may be represented by the integral equation

\[ u(x, t) = \int_{\mathbb{R}^n} \hat{u}_0(\xi) e^{2\pi i (x \cdot \xi + t |\xi|^2)} d\xi, \]

at least if \( u_0 \) is sufficiently smooth. We may therefore interpret solutions to (5.1) as the space-time Fourier extension of the measure

\[ d\mu = \hat{u}_0(\xi) \delta(\tau - |\xi|^2) d\xi d\tau, \]

which is supported on the paraboloid \( P^n = \{(\xi, \tau) \in \mathbb{R}^n \times \mathbb{R} : \tau = |\xi|^2\} \) in \( \mathbb{R}^{n+1} \). In particular we have

\[ u(t, x) = \int_{P^n} e^{2\pi i (x \cdot \xi + t \tau)} d\mu(\xi, \tau). \]

This allows us to apply some deep results from Fourier restriction theory to analyze solutions to (5.1) (and also solutions to nonlinear versions of this equation). A fundamental result in this direction is the Stein-Tomas theorem, which in particular implies that if

\[ S = \{(\xi, \tau) \in P^n : |\xi| \leq 1\} \]

then

\[ \| \int_S e^{2\pi i (x \cdot \xi + t \tau)} d\mu(\xi, \tau) \|_{L^p(\mathbb{R}^n)} \leq C \| u_0 \|_{L^2(\mathbb{R}^n \times \mathbb{R})} \] (5.2)

for \( p \geq \frac{2(n+2)}{n} \) (see, for example, [65]). Moreover, in the endpoint case \( p = \frac{2(n+2)}{n} \) a parabolic rescaling \((x, t) \rightarrow (\lambda x, \lambda^2 t)\) shows that (5.2) remains true with the same constant if we replace \( S \) by any compact subset of \( P^n \). Then by a simple limiting argument we see that our solution \( u \) satisfies the following space-time estimate:

\[ \| u \|_{L^{\frac{2(n+2)}{n}}(\mathbb{R}^n \times \mathbb{R})} \leq C \| u_0 \|_{L^2(\mathbb{R}^n)}. \]
This is the simplest example of a Strichartz estimate. It turns out to be one of many possible space-time estimates one can prove for $u$:

**Proposition 5.1** (Euclidean Strichartz estimates). Let $u = e^{it\Delta}u_0$ be a solution to (5.1). Suppose $\frac{2}{q} + \frac{n}{p} = \frac{n}{2}$. Then

$$\|e^{it\Delta}u_0\|_{L_t^q L_x^p} \lesssim \|u_0\|_{L^2(\mathbb{R}^n)}. \tag{5.3}$$

The mixed norm is taken on the whole space $\mathbb{R}_t \times \mathbb{R}^n_x$, and so is in particular global in time. The original proof of the case $q = p$ due to Strichartz can be found in [82], and a proof of the more general case can be found in [84]. By considering rescaled functions $u_0(x) \to u_0(\lambda x)$ it is easy to check that the relation $\frac{2}{q} + \frac{n}{p} = \frac{n}{2}$ is necessary. A crucial element of the proof of (5.3) is the fact that solutions disperse in time like $t^{-\frac{n}{2}}$, which can be seen, for example, via the identity

$$u(x, t) = ct^{-\frac{n}{2}} \int_{\mathbb{R}^n} u_0(y) e^{-\frac{i|x-y|^2}{4t}} dy$$

for $u_0$ sufficiently nice (or from a stationary phase argument). Indeed, it is because of this exact decay in $t$ that one has $q = \frac{4p}{m(p-2)}$ in (5.3). Strichartz estimates are also important tools in the analysis of nonlinear Schrödinger equations of the form

$$iu_t + \Delta u = \pm |u|^{r-1}u$$

(see [84], for example).

### 5.0.2 Strichartz Estimates on the Torus

Over the last few decades there has been a wide range of research concerning Strichartz estimates of the type (5.3) for solutions to Schrödinger equations on manifolds other than $\mathbb{R}^n$. A particular example of interest is the periodic equation on the torus $\mathbb{T}^d$.

In the case where $\mathbb{T}^d$ is the standard torus one checks that solutions to the analogue of (5.1)
have the representation
\[ u(x, t) = \sum_{m \in \mathbb{Z}^d} a_m e^{2\pi i (x \cdot m + t |m|^2)}, \quad (5.4) \]
where
\[ u_0(x) = \sum_{m \in \mathbb{Z}^d} a_m e^{2\pi i m \cdot x}. \]

Since the solution is periodic in \( t \) it is impossible to prove analogues of (5.3) which are global in time, although one can hope to prove local variants of the form

\[ \|u\|_{L^p_{L^x}(T^d \times I)} \leq C_I \|u_0\|_{L^2(T^d)}, \]

where \( I \) is a bounded time interval. The study of these type of estimates dates back to work of Bourgain [70], though it is only very recently that the full range of (essentially sharp) local \( L^p_{L^x} \) estimates have been proved as a consequence of Bourgain and Demeter’s \( \ell^2 \) decoupling theorem [71]. Their result is the following:

**Proposition 5.2** ([71]). Let \( u(x, t) = e^{i t \Delta_{T^d}} u_0 \) be a solution to the Schrödinger equation on a (possibly irrational) torus \( T^d \). Then if \( p \geq \frac{2(d+2)}{d} \) one has

\[ \|e^{it \Delta_{T^d}} u_0\|_{L^p(T^d \times I)} \lesssim |I|^\frac{1}{p} \|u_0\|_{H^s(T^d)} \quad (5.5) \]

for any \( s > \frac{d}{2} - \frac{d+2}{p}. \)

See also the work of Killip and Visan [81], which sharpens this estimate to \( s = \frac{d}{2} - \frac{d+2}{p} \) in the non-endpoint case \( p > \frac{2(d+2)}{d} \). This dependence on \( s \) is essentially sharp, as can be seen by considering a Knapp-type example (see Proposition 6.5 in the next chapter).

We note that functions of the form (5.4) have Fourier support on lattice points of the form \((m, |m|^2) \subset P_d\), although the Stein-Tomas theorem is no longer enough to prove all cases of (5.5) (it is possible to prove partial results with a larger loss in \( s \) by using the Stein-Tomas theorem, however). Bourgain was also able to obtain certain cases of (5.5) in [70] by using...
number-theoretic arguments, inspired in part by the proof of the Stein-Tomas theorem. There were also several other partial results by other authors in the subsequent years (see the references in [71], [81]). The only known proof of (5.5) in the full range $p \geq \frac{2(d+2)}{d}$ requires the deep $\ell^2$ decoupling theorem of Bourgain and Demeter, which we will discuss further in the next chapter.

We end the section by pointing out a key difference from the Euclidean setting which will be relevant in the next chapter. If $P_N$ is a smooth Littlewood-Paley frequency cut-off to scale $N \geq 1$, then (5.5) says that for any $\epsilon > 0$ one has

$$\|e^{it\Delta_T} P_N u_0\|_{L^2\left(\mathbb{T}^d \times [0,1]\right)} \leq C \epsilon N^\epsilon \|P_N u_0\|_{L^2(T^d)}.$$ 

This contrasts the Euclidean setting, where there is no loss of $N^\epsilon$ at the Stein-Tomas endpoint.

It turns out that in the periodic case this loss is necessary, as shown by Bourgain. Indeed, in the case $d = 1, p = 6$ Bourgain showed that if

$$\|e^{it\Delta_T} P_N f\|_{L^6(T \times [0,1])} \leq A \|P_N f\|_{L^2(T)}$$

then $A \geq c (\log N) \frac{1}{6}$. In particular this lower bound holds when $f = \sum_{m=1}^{N} e^{2\pi i mx}$, as can be shown by restricting the $L^6$ norm to ‘major arcs’ where $x, t$ are well-approximated by certain rational numbers. See [70] for details. Bourgain’s argument also extends to higher dimensions (see [83] for details in the case $d = 2, p = 4$).
Chapter 6

The Semiperiodic Case

We now turn to the setting of product manifolds of the form $\mathbb{R}^n \times \mathbb{T}^d$, where $\mathbb{T}^d$ is a (rational or irrational) $d$-dimensional torus. There has been recent interest in the behavior of solutions to the linear and nonlinear Schrödinger equation on these manifolds (see for example [72],[75],[76],[78],[79],[85]). In particular, one can exploit dispersive effects coming from the Euclidean component of the manifold to obtain stronger asymptotic results than in the setting of $\mathbb{T}^d$. Indeed, as a starting point one can hope to prove global-in-time Strichartz-type estimates for solutions to the linear equation on $\mathbb{R}^n \times \mathbb{T}^d$ ([75]). This contrasts the situation on $\mathbb{T}^d$, where no global $L^p_{t,x}$ estimates are possible as we saw above. One can also hope to prove stronger results than in the more general setting of $\mathbb{R}^n \times M^d$, where $M^d$ is a $d$-dimensional Riemannian manifold, since the presence of the torus facilitates Fourier-analytic and number-theoretic methods in the vein of [70], [71].

The main object of study for the rest of the chapter is the linear semiperiodic Schrödinger equation

$$\begin{cases}
- i \partial_t u + \Delta_{\mathbb{R}^n \times \mathbb{T}^d} u = 0, \\
u(x, y, 0) = f(x, y) \quad x \in \mathbb{R}^n, y \in \mathbb{T}^d.
\end{cases}$$

(6.1)

We assume that our initial data $f$ is smooth and rapidly decaying, which we can do with no loss of generality as long as our estimates do not depend on the smoothness of $f$. For most
of our arguments in the following sections we will also assume that $\mathbb{T}^d$ is the standard torus $\mathbb{T}^d \sim [0,1]^d$. In this case solutions to (6.1) can be represented by 

$$e^{it\Delta_{\mathbb{R}^n \times \mathbb{T}^d}} f(x,y) = \sum_{m \in \mathbb{Z}^d} \int_{\mathbb{R}^n} \hat{f}_m(\xi) e^{2\pi i (x \cdot \xi + y \cdot m + t(|\xi|^2 + |m|^2))} d\xi,$$

where

$$f(x,y) = \sum_{m \in \mathbb{Z}^d} f_m(x) e^{2\pi i y \cdot m}.$$

In the more general setting of an irrational torus $\mathbb{T}^d \sim \prod_{i=1}^d [0, \beta_i]$ we have a similar representation formula obtained by replacing $|m|^2$ above by $\sum_i \beta_i^2 m_i^2$. We will explain when necessary how to adapt our arguments to the case $\mathbb{T}^d \sim \prod_{i=1}^d [0, \beta_i]$.

We will begin by surveying some earlier results in lower dimensions.

### 6.1 Local Estimates on $\mathbb{R} \times \mathbb{T}$

Since solutions to (6.1) are periodic in at least one direction, it is not immediately clear how one can generalize the classical Euclidean Strichartz estimates in the semiperiodic case. A starting point is to consider local variants of these estimates, as in the case of the torus. The first result in this direction was proved by Takaoka and Tzvetkov [83], who studied (6.1) and the cubic NLS on $\mathbb{R} \times \mathbb{T}$. This case is in some sense ‘intermediate’ between $\mathbb{R}^2$ and $\mathbb{T}^2$, and it turns out that one can prove the direct analogue of (5.5) with no extra loss of derivatives. The main geometric insight in their proof is the following lemma. Note that the corresponding result on $\mathbb{Z}^2$ (which is dual to $\mathbb{T}^2$) fails, since classical lattice-point-counting arguments show that there must be a loss of an extra logarithmic factor (recall the logarithmic loss in Bourgain’s example from the end of the last chapter).

**Lemma 6.1.** Suppose $A \geq 1$ and $\omega > 0$. We have

$$\sum_{m \in \mathbb{Z}} |\{(\xi, m) \in \mathbb{R} \times \mathbb{Z} : \xi^2 + m^2 \in [A, A+\omega]\}| \lesssim \omega,$$
and likewise
\[
\sum_{m \in \mathbb{Z}} |\{ (\xi, m) \in \mathbb{R} \times \mathbb{Z} : \xi^2 + (m + 1/2)^2 \in [A, A + \omega] \}| \lesssim \omega.
\]

Proof. Let \( L_m = |\{ (\xi, m) : \xi^2 + m^2 \in [A, A + \omega] \}|. \) Note that if \( \xi^2 + m^2 \in [A, A + \omega] \) then \( |m| \leq \sqrt{A + \omega} \), hence we need to estimate
\[
\sum_{|m| \leq \sqrt{A + \omega}} L_m.
\]

By symmetry it suffices to consider the positive \( m \) values.

If \( m \leq \sqrt{A} \) is fixed then
\[
L_m = 2(\sqrt{A + \omega} - m^2 - \sqrt{A - m^2}) = \frac{2\omega}{\sqrt{A + \omega} - m^2 + \sqrt{A - m^2}}.
\]

In this case it follows that
\[
\sum_{m=1}^{\lfloor \sqrt{A} \rfloor} L_m \lesssim \int_0^{\sqrt{A}} \frac{\omega}{\sqrt{A - x^2}} dx \lesssim \omega \int_0^1 \frac{1}{\sqrt{1 - z^2}} dz \lesssim \omega.
\]

Now if \( \sqrt{A} < m \leq \sqrt{A + \omega} \) we instead have
\[
L_m = 2\sqrt{A + \omega} - m^2
\]

Therefore by bounding the sum by an integral and applying a simple change of variables we get
\[
\sum_{m=\lfloor \sqrt{A} \rfloor + 1}^{\lfloor \sqrt{A + \omega} \rfloor} L_m \lesssim (A + \omega) \int_{\sqrt{A + \omega}}^1 \sqrt{1 + x^2} dx \lesssim \frac{(A + \omega)\omega}{(A + \omega) + \sqrt{A + \omega\sqrt{A}}}.\]

The last term is smaller than \( \omega \), so this completes the proof of the first part of the lemma. The
proof of the second part is essentially the same, since \( m \) is just shifted by a small constant factor. \( \square \)

An immediate consequence of this lemma is the following.

**Corollary 6.1.** Suppose \( c \in \mathbb{R} \) and \( c' \in \{0, 1/2\} \), and let \( R > 0 \) be arbitrary. Then

\[
\sum_m \int_\eta e^{-\pi((\eta - c)^2 + (m - c')^2 - R^2)} d\eta \lesssim 1,
\]

with implicit constant independent of \( c, c', R \).

**Proof.** Decompose the summation and integration into dyadic shells where

\[
|(\eta - c)^2 + (m - c')^2 - R^2| \sim 2^k, \quad k \in \mathbb{N}.
\]

In the case \( R \geq 1 \) we can directly apply Lemma 6.1 and then use exponential decay to sum over \( k \). If \( 0 < R < 1 \) and \(|m| \geq 2\) then the exponential is just smaller than the previous case, so we can apply the same argument. Finally if \( m = -1, 0, 1 \) we use the simple fact that the integral in \( \eta \) is uniformly bounded with respect to \( m \). \( \square \)

We now turn to the first example of a Strichartz estimate on \( \mathbb{R} \times T \).

**Theorem 6.1 (Takaoka-Tzvetkov [83]).** One has

\[
\|e^{it\Delta_{\mathbb{R}\times T}} \hat{f}\|_{L^4(\mathbb{R}\times \mathbb{T}\times [0,1])} \leq C \|f\|_{L^2(\mathbb{R}\times \mathbb{Z})}.
\]

The proof below is not quite the original argument used by Takaoka and Tzvetkov, although it is similar. In particular the argument relies on the key geometric estimate found in Lemma 6.1, which was also used in [83]. We learned of this argument from discussions with B. Pausader and M. Christ.

**Proof.** After inserting a Gaussian factor \( e^{-\pi t^2} \) to the \( L^4_{x,y,t}(\mathbb{R} \times \mathbb{T} \times [0,1]) \) norm of \( e^{it\Delta_{\mathbb{R}\times T}} \hat{f} \)
and expanding \(|e^{it\Delta_{\mathbb{R} \times T}}f|^4\) in terms of \(f\) we can estimate \(\|e^{it\Delta}\hat{f}\|_{L^4(\mathbb{R} \times T \times [0,1])}^4\) by the following integral:

\[
\int_{\mathbb{R} \times T \times \mathbb{R}} e^{-\pi t^2} \left[ \sum_{k,m,k',m'} \int_{\xi,\xi',\eta,\eta'} f(\xi,k) f(\xi',k') f(\eta,m) f(\eta',m') e^{i(x(\xi-\eta+\xi'-\eta')+y(k-m+k'-m'))} e^{it(\xi'^2-\eta'^2-(\eta')^2+k^2-m^2+(k')^2-(m')^2)} d\xi d\eta \right] dx dy dt
\]

where \(\xi = (\xi,\xi')\) and \(\eta = (\eta,\eta')\). Interchanging the order of integration and computing the Fourier transforms in the \(x, y,\) and \(t\) variables, this integral simplifies to

\[
\sum_{k,k',m,m'} \int_{\xi,\xi',\eta,\eta'} f(\xi,k) f(\xi',k') f(\eta,m) f(\eta',m') \delta(k-m+k'-m') \delta(\xi-\eta+\xi'-\eta') e^{-\pi Q^2} d\xi d\eta,
\]

where

\[
Q = \xi^2 - \eta^2 + (\xi')^2 - (\eta')^2 + k^2 - m^2 + (k')^2 - (m')^2
\]

and we interpret the continuous-frequency Dirac mass as

\[
\delta(\xi-\eta+\xi'-\eta') = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \mathbb{1}_{|\xi-\eta+\xi'-\eta'|<\epsilon}.
\]

We generalize a little and note that it will be enough to show that the quadrilinear form

\[
\sum_{k,k',m,m'} \int_{\xi,\xi',\eta,\eta'} f_1(\xi,k) f_1'(\xi',k') f_2(\eta,m) f_2'(\eta',m') \delta(k-m+k'-m') \delta(\xi-\eta+\xi'-\eta') e^{-\pi Q^2} d\xi d\eta (6.2)
\]

is bounded by \(\|f_1\|_{L^2} \|f_1'\|_{L^2} \|f_2\|_{L^2} \|f_2'\|_{L^2}\).

Assume first that the variables \(\xi, k, \xi', k'\) are held fixed. Because of the relations given by the Dirac masses we may write \(\eta' = \xi - \eta + \xi' + O(\epsilon)\) and \(m' = k - m + k'\). Taking the limit as
\( \epsilon \to 0 \) shows that we can factor

\[
-\xi^2 + \eta^2 - (\xi')^2 + (\eta')^2 = \eta^2 - \xi^2 - (\xi')^2 + (\xi - \eta + \xi')^2 = 2[(\eta - c)^2 - r^2],
\]

where \( c = (\xi + \xi')/2 \) and \( r = (\xi - \xi')/2 \) are independent of \( \eta, \eta' \). We can similarly factor \( k^2 - m^2 + (k')^2 - (m')^2 \). Using these reductions we see that (6.2) is equal to

\[
\sum_{k,k'} \int_{\xi,\xi'} f_1(\xi,k) f'_1(\xi',k') I(\xi,\xi',k,k') d\xi,'
\]

where

\[
I = \sum_{m} \int_{\eta} f_2(\eta,m) f'_2(\xi - \eta + \xi', k - m + k') e^{-\pi((\eta-c)^2 + (m-c')^2 - R^2)^2} d\eta
\]

with \( c, c' \) and \( R \) independent of \( m \) and \( \eta \). In particular

\[
I \leq \|f_2\|_\infty \|f'_2\|_\infty \sup_{\xi,\xi'} \sum_{k,k'} \int_{\eta} e^{-\pi((\eta-c)^2 + (m-c')^2 - R^2)^2} d\eta.
\]

The supremum is uniformly bounded by Corollary 6.1, and therefore

\[
(6.2) \leq \|f_1\|_{L^1} \|f'_1\|_{L^1} \|f_2\|_{L^\infty} \|f'_2\|_{L^\infty}.
\]

But this argument was symmetric in the choice of fixed variables, and so the same reasoning also shows that

\[
(6.2) \leq \|f_2\|_{L^1} \|f'_2\|_{L^1} \|f_1\|_{L^\infty} \|f'_1\|_{L^\infty}.
\]

Interpolating between these two bounds yields the desired \( L^2 \times L^2 \times L^2 \times L^2 \) estimate.

\( \square \)
Note that this argument proves the slightly more general bound
\[
\|e^{it\Delta_{\mathbb{R} \times T}}\hat{f}\|_{L^4(\mathbb{R} \times T \times [0,1])} \lesssim \|f\|_{L^p(\mathbb{R} \times T)}^{1/2} \|f\|_{L^{p'}(\mathbb{R} \times T)}^{1/2}
\]
for \(1 \leq p \leq \infty\), although it is not clear if this is useful.

### 6.2 Global Estimates on \(\mathbb{R}^n \times \mathbb{T}^d\)

Since a solution \(u(x, y, t)\) to (6.1) behaves like a solution to the Schrödinger equation on \(\mathbb{R}^n\) for each fixed \(y\), we expect dispersion on the order of what we saw in the purely Euclidean case as \(t \to \infty\). We therefore expect to be able to find some global-in-time (and possibly mixed) Lebesgue space that \(u\) lies in if the initial data is sufficiently nice. Indeed, by applying the triangle inequality to eliminate the periodic component it is easy to check that
\[
\|e^{it\Delta_{\mathbb{R} \times T}}f\|_{L^6(\mathbb{R} \times T \times \mathbb{R})} \lesssim \|f\|_{H^1(\mathbb{R} \times T)},
\]
where \(p = q = 6\) are the admissible exponents in dimension \(n = 1\) from (5.3). Although this loss in derivatives is too large to be useful in practice, the estimate at least shows that if the initial data \(f = u(x, y, 0)\) is in some Sobolev space then one has a global space-time estimate for \(u\); the question then becomes (i) what is the lowest regularity we can assume for \(u(x, y, 0)\) to obtain an estimate like (6.3), and (ii) for which types of (possibly mixed) Lebesgue spaces we can prove an estimate like (6.3). Of course we should be able to justify our answers to (i) and (ii) with some applications that parallel the Euclidean case.

It is possible to improve the \(H^s\) exponent in (6.3) by using duality and Sobolev embedding, although this approach only works for a limited range of \(L^p\) exponents. The following result is due to Tzvetkov and Visciglia.
Proposition 6.2 ([85]). One has

$$\|e^{i\Delta_{\mathbb{R}^n \times \mathbb{T}^d}} f\|_{L^6_{t,x} L^2_y} \lesssim \|f\|_{L^2(\mathbb{R} \times \mathbb{T})},$$

and hence by Sobolev embedding

$$\|e^{i\Delta_{\mathbb{R}^n \times \mathbb{T}^d}} f\|_{L^6(\mathbb{R} \times \mathbb{T} \times \mathbb{R})} \lesssim \|f\|_{H^{1/3}}.$$ 

The proof proceeds by considering the dual form of the first estimate, which enables us to exploit the fact that $e^{i\Delta_{\mathbb{R}^n \times \mathbb{T}^d}} f = e^{i\Delta_{\mathbb{R}^n}} (e^{i\Delta_{\mathbb{T}}} f)$ and then use the (dual) Euclidean $L^6_{t,x}$ estimate. Note that one cannot do better than $H^{1/3}$, as we can see by arguing as in Proposition 6.5 below.

However, since $\mathbb{R}^n \times \mathbb{T}^d$ locally looks like $\mathbb{R}^{n+d}$ (or $\mathbb{T}^{n+d}$) one hopes to be able to prove estimates near the Stein-Tomas endpoint $p^* = \frac{2(n+d+2)}{n+d}$. On $\mathbb{R} \times \mathbb{T}$ we have $p^* = 4$, and we already saw that a sharp local-in-time estimate does hold for this value of $p$. On the other hand, since periodic solutions to the Schrödinger equation do not disperse in time like in the Euclidean case, we do not expect the periodic component of the manifold to help with global estimates. We therefore expect the admissible $L^q_t$ exponents to be dictated by the relation $q = \frac{4p}{n(p-2)}$, as in the Euclidean case.

Unfortunately there does not appear to be any easy way to prove $L^q_t L^p_{x,y}$ estimates on $\mathbb{R}^n \times \mathbb{T}^d$ for $p < \frac{2(n+2)}{n}$, due to significant obstructions resulting from the periodic component of the solution.

The Hani-Pausader Estimate

Hani and Pausader approached the problem outlined and the end of the last section by changing the norm to include both a local and global component [75]. Their main Strichartz-type estimate is the following, which should be compared with Propositions 5.3 and 5.5:

$$\|e^{i\Delta_{\mathbb{R}^n \times \mathbb{T}^2}} P_N u_0\|_{L^p_x L^q_{x,y}}(\mathbb{R} \times \mathbb{T}^2 \times [\gamma, \gamma+1]) \lesssim N^{\frac{3}{2} - \frac{p}{2}} \|u_0\|_{L^2(\mathbb{R} \times \mathbb{T}^2)} \quad (6.4)$$
whenever
\[ p > 4 \quad \text{and} \quad \frac{2}{q} + \frac{1}{p} = \frac{1}{2}. \]

This estimate was then used as a starting point for their analysis of the defocusing quintic nonlinear Schrödinger equation on \( \mathbb{R} \times \mathbb{T}^2 \).

Notice that the value of \( q \) in (6.4) is exactly the Strichartz-admissible time endpoint for \( L^q_x L^p_t \) estimates for the Schrödinger equation on \( \mathbb{R} \), while the loss in \( N \) is the same as on \( \mathbb{T}^3 \) or \( \mathbb{R}^3 \) (and is in fact the best one can hope for, see Proposition 6.5 below). However, from the theory on \( \mathbb{R}^3 \) or \( \mathbb{T}^3 \) we expect to be able to push the exponent \( p \) down to values larger than \( \frac{10}{3} \) (or equal to \( \frac{10}{3} \) with a possible loss of \( N^\epsilon \)). The rest of the chapter is devoted to the proof that (6.4) is indeed true for \( p > \frac{10}{3} \), and also true for \( p = \frac{10}{3} \) with an arbitrarily small loss in the power of \( N \). We also extend this result to higher (and lower) dimensions and prove the scale-invariant analogue of (6.4) on \( \mathbb{R}^n \times \mathbb{T}^d \) for \( p \) away from the Stein-Tomas endpoint \( \frac{2(n+d+2)}{n+d} \).

The Main Theorem

Our main theorem is the following.

**Theorem 6.2.** Let \( \mathbb{T}^d \) be a \( d \)-dimensional rational or irrational torus and \( \Delta_{\mathbb{R}^n \times \mathbb{T}^d} \) the Laplacian on \( \mathbb{R}^n \times \mathbb{T}^d \). Let \( p^* = \frac{2(n+d+2)}{n+d} \) and fix \( p > p^* \). Then if \( q = q(p) := \frac{4p}{n(p-2)} \) and \( q > 2 \),

\[
\left( \sum_{\gamma \in \mathbb{Z}} \| e^{it \Delta_{\mathbb{R}^n \times \mathbb{T}^d}} u_0 \|_{L^q_x(\mathbb{R}^n \times \mathbb{T}^d \times [\gamma-1,\gamma+1])}^q \right)^{1/q} \lesssim \| u_0 \|_{H^s(\mathbb{R}^n \times \mathbb{T}^d)},
\]

(6.5)

where \( s = \frac{n+d}{2} - \frac{n+d+2}{p} \). Moreover, if \( p = p^* \) then the result holds with \( q = q(p^*) \) for any \( s > 0 \) (with a constant that blows up as \( s \to 0 \)).

Note that when \( p = p^* \) we have \( q(p) = \frac{2(n+d+2)}{n} \). Setting \( d = 0 \) or \( n = 0 \), we recover the usual \( L^p_{t,x} \) Strichartz estimates on \( \mathbb{R}^n \) and \( \mathbb{T}^d \), respectively (modulo a loss of \( N^\epsilon \) in the Euclidean case). Also note that \( q(p) \) is exactly the admissible \( q \) value corresponding to the \( L^q_x L^p_t \) Strichartz estimates on \( \mathbb{R}^n \), which is expected since we heuristically have \( n \) directions contributing to dispersion.
Indeed, one cannot prove an estimate of type (6.5) for $q < q(p)$ (see Proposition 6.4 below for a proof).

The proof of Theorem 6.2 is in Section 6.4, following a brief review of some preliminary material in Section 6.3. Our argument combines the approach of Hani and Pausader with the decoupling method of Bourgain and Demeter (see the next section for a precise statement of their decoupling theorem). We will initially prove (6.5) in the case $u_0 = P_{\leq N} u_0$ with an extra loss of $N^\epsilon$, but we show in Section 6.5 that this loss can be removed away from the Stein-Tomas endpoint $p^*$. In Section 6.6 we study some applications of Theorem 6.2 to the nonlinear theory of Schrödinger equations on $\mathbb{R}^n \times \mathbb{T}^d$. Our two main results in this direction are contained in the following theorem. Recall that the quintic NLS is the equation

$$-i \partial_t u + \Delta_{\mathbb{R}^n \times \mathbb{T}^d} u = \pm |u|^4 u, \quad u(x, y, 0) = u_0(x, y)$$

and the cubic NLS is the equation

$$-i \partial_t u + \Delta_{\mathbb{R}^n \times \mathbb{T}^d} u = \pm |u|^2 u, \quad u(x, y, 0) = u_0(x, y),$$

and that these equations are $H^{\frac{1}{2}}$ critical on $\mathbb{R} \times \mathbb{T}$ and $\mathbb{R}^2 \times \mathbb{T}$, respectively. Below $X^{\frac{1}{2}}$ is a Banach space of functions $u : \mathbb{R} \to H^{\frac{1}{2}}$ defined in Section 6.6, with the property that $X^{\frac{1}{2}} \hookrightarrow L^\infty(\mathbb{R}, H^{\frac{1}{2}})$.

**Theorem 6.3.** The quintic NLS on $\mathbb{R} \times \mathbb{T}$ with initial data $u_0 \in H^{\frac{1}{2}}(\mathbb{R} \times \mathbb{T})$ is locally well-posed. Moreover, there exists $\delta > 0$ such that if $\|u_0\|_{H^{\frac{1}{2}}} < \delta$ then the solution $u \in X^{\frac{1}{2}}$ is unique, exists globally in time, and scatters as $t \to \pm \infty$ in the sense that there are $v_\pm \in H^{\frac{1}{2}}$ such that

$$\lim_{t \to \pm \infty} \|u - e^{i t \Delta_{\mathbb{R} \times \mathbb{T}}} v_\pm\|_{H^{\frac{1}{2}}} = 0.$$

The cubic NLS on $\mathbb{R}^2 \times \mathbb{T}$ with initial data $u_0 \in H^{\frac{1}{2}}(\mathbb{R}^2 \times \mathbb{T})$ is also locally well-posed. Moreover, there exists $\delta > 0$ such that if $\|u_0\|_{H^{\frac{1}{2}}} < \delta$ then the solution $u \in X^{\frac{1}{2}}$ is unique, exists globally in time, and scatters as $t \to \pm \infty$. 

140
Once we have proved Theorem 6.2 we can prove Theorem 6.3 by using some standard machinery which we review at the beginning of Section 6.6. In this setting it is essential to have a global estimate of type (6.5) without any extra loss of $\mathcal{N}$. In particular the result in Proposition 6.6 below, which has a shorter proof than Theorem 6.2, is not sufficient to establish Theorem 6.3. We note that the local-in-time results in Theorem 6.3 do not require global Strichartz estimates; we will only use the full strength of Theorem 6.2 to establish the small-data global existence and scattering results.

Finally, in Section 6.7 we collect some additional remarks and related open problems.

**Notation and Basic Assumptions**

As is standard, we write $A \lesssim B$ if there is some constant $c > 0$ depending only on the dimension and various Lebesgue exponents such that $A \leq cB$. If $A \lesssim B$ and $B \lesssim A$ we will write $A \sim B$. Moreover, if $A \leq c(\alpha)B$ where the constant $c(\alpha)$ depends on some parameter $\alpha$ we will write $A \lesssim_{\alpha} B$. We will also often write $A \lesssim_{\epsilon} \mathcal{N}^{\epsilon}B$ as short-hand for the expression ‘for all $\epsilon > 0$ there is $c_{\epsilon}$ such that $A \leq c_{\epsilon} \mathcal{N}^{\epsilon}B$,’ to avoid having to write expressions involving constant multiples of an arbitrarily small parameter $\epsilon > 0$.

Let $S$ be a rectangle and let $S^{-1}$ denote the dual rectangle centered at the origin obtained by inverting the side lengths. We let $w_S$ be a weight adapted to $S$ in the following sense: $w_S(x)$ decays rapidly for $x \notin S$, and $\hat{w}_S(\xi)$ is supported in a fixed dilate of $S^{-1}$. We similarly define $w_S$ if $S$ is a ball, and if $\Omega = \bigcup_S S$ then we let $w_{\Omega} = \sum_S w_S$. Note that we can construct $w_S$ by taking a bump function $w$ adapted to the unit ball such that

$$|w(x)| \lesssim \frac{1}{(1 + |x|)^{1000(n+d)}}$$

and then applying a suitable affine transformation. If $c > 0$ and $S$ is a ball or rectangle, we also let $cS$ denote the centered dilate by $c$.

For a dyadic integer $N \geq 1$ we let $P_{\leq N} f = f \ast \varphi_{N}$, where $\varphi_N$ is a smooth function such that $\hat{\varphi}_N$ is supported in $B_{2N}(0)$. We also let $P_N$ denote a smooth Littlewood-Paley cut-off to the
annulus $A_N = \{ \xi : N/2 \leq |\xi| < N \}$.

Finally, we assume that all functions are smooth and rapidly decaying. We can do so with no loss of generality as long as our estimates are independent of the smoothness and decay parameters.

### 6.3 Preliminary Results

The $\ell^2$ decoupling theorem of Bourgain and Demeter will be an important tool in our argument.

**Theorem 6.4 ([71]).** Let $N \geq 1$ and suppose $f$ is a smooth function on $\mathbb{R}^{n+1}$ such that $\hat{f}$ is supported in an $O(N^{-2})$ neighborhood of the truncated paraboloid

$$P^n = \{ (\xi, \tau) \in \mathbb{R}^n \times \mathbb{R} : \tau = |\xi|^2, \ |\xi| \leq 1 \}.$$ 

Let $\{\theta\}$ be a finitely-overlapping collection of $N^{-1}$-caps covering the support of $\hat{f}$, let $\{\varphi_\theta\}$ be a partition of unity subordinate to this cover, and define $f_\theta = f * \varphi_\theta$. Then for any ball $B_{N^2}$ of radius $N^2$ in $\mathbb{R}^{n+1}$ one has

$$\|f\|_{L^p(w_{B_{N^2}})} \lesssim_N N^{c + \alpha_p} \left( \sum_\theta \|f_\theta\|_{L^p(w_{B_{N^2}})}^2 \right)^{\frac{1}{2}}$$

for $p \geq 2$, where $\alpha_p = 0$ if $2 \leq p \leq \frac{2(n+2)}{n}$ and $\alpha_p = \frac{n}{2} - \frac{n+2}{p}$ otherwise.

Decoupling inequalities date back to work of Wolff [86], who proved an analogue of the above theorem for the cone (in a limited $L^p$ range) and used this estimate to make progress on the local smoothing conjecture for the wave equation. There was partial progress on Theorem 6.4 in certain ranges before the proof of Bourgain and Demeter, and we send the reader to [71] for references. Theorem 6.4, which also implies Wolff’s result in the full range, is proved via a complicated induction-on-scales argument that incorporates some major results in contemporary harmonic analysis (in particular the multilinear restriction and Kakeya theorems play a key role).

We will also use the following discrete analogue of the Hardy-Littlewood-Sobolev inequality.
Proposition 6.3 (Discrete Hardy-Littlewood-Sobolev). Suppose $1 < p, q < \infty$ and $0 < \mu < 1$ such that

$$\frac{1}{p} + \frac{1}{q} + \mu = 2.$$

Then

$$\sum_{j \neq k} \frac{f(j)g(k)}{|j - k|^\mu} \lesssim \|f\|_{\ell^p} \|g\|_{\ell^q}.$$

Proof. We learned of the following proof from unpublished lecture notes by E.M. Stein. The argument is a modification of one of the possible proofs of the continuous Hardy-Littlewood-Sobolev inequality.

We let $\lambda = 1 - \mu$. By duality it suffices to show that the operator

$$I_\lambda(f)(n) = \sum_{m \neq 0 \atop m \in \mathbb{Z}} f(n - m)|m|^{\lambda - 1}$$

satisfies a bound

$$\|I_\lambda f\|_{\ell^q} \leq C\|f\|_{\ell^p}$$

for $\frac{1}{q} = \frac{1}{p} - \lambda$. This will follow by Hölder’s inequality if we can prove the following pointwise bound:

$$I_\lambda(f)(n) \lesssim M^d(f)(n)^{\frac{p}{q}}\|f\|_{\ell^p}^{1 - \frac{p}{q}}, \quad (6.6)$$

where $M^d f$ is the discrete Hardy-Littlewood maximal operator\(^1\)

$$M^d f(n) = \sup_{r > 0} \frac{1}{N(r)} \sum_{|m| \leq r} |f(n - m)|, \quad N(r) = \{n \in \mathbb{Z} : |n| < r\}.$$

\(^1\)Note that $M^d$ is bounded from $\ell^p(\mathbb{Z})$ to $\ell^p(\mathbb{Z})$ for $1 < p \leq \infty$ and also weak-type $(1,1)$; the weak-type bound can be proved by imitating the standard covering-lemma argument from the continuous case, and then the $\ell^p$ boundedness follows by interpolating with the trivial $\ell^\infty \to \ell^\infty$ bound.
To prove (6.6) we write

\[ I_\lambda f(n) = \sum_{0 < |m| < R} f(n - m)|m|^{\lambda - 1} + \sum_{|m| \geq R} f(n - m)|m|^{\lambda - 1}. \]

For the first sum we have

\[ \sum_{0 < |m| < R} f(n - m)|m|^{\lambda - 1} \lesssim \sum_{j=0}^{\infty} \sum_{|m| \sim 2^{-j} R} f(n - m)|m|^{\lambda - 1} \]
\[ \lesssim \sum_{j=0}^{\infty} (2^{-j} R)^{\lambda - 1} (2^{-j} R)M^d f(n) \]
\[ \lesssim R^{\lambda} M^d f(n). \]

On the other hand, by Hölder's inequality

\[ \sum_{|m| \geq R} f(n - m)|m|^{\lambda - 1} \leq \|f\|_{\ell^p} \left( \sum_{|m| \geq R} |m|^{(\lambda-1)p'} \right)^{\frac{1}{p'}} \]
\[ = \|f\|_{\ell^p} \left( \sum_{|m| \geq R} |m|^{-\left(\frac{q'}{q}+1\right)} \right)^{\frac{1}{p'}} \]
\[ \lesssim \|f\|_{\ell^p} R^{-\frac{1}{q}}, \]

using the fact that \( \frac{1}{q} = \frac{1}{p} - \lambda \). Now choose \( R \) such that

\[ R = \frac{\|f\|_{\ell^p}^p}{M^d f(n)^p}. \]

Using the fact that \( \lambda + \frac{1}{q} = \frac{1}{p} \) we conclude that for this choice

\[ I_\lambda f(n) \lesssim R^{\lambda} M^d f(n) + \|f\|_{\ell^p} R^{-\frac{1}{q}} \]
\[ \lesssim (M^d f(n))^p \|f\|_{\ell^p}^{1-\frac{q}{p}} \]

as desired.
On the Sharpness of Theorem 6.2

In this subsection we include two arguments demonstrating that one cannot improve on the \( \ell^q \) exponent in (6.5), and in the non-endpoint case one cannot improve on the value of \( s \).

**Proposition 6.4.** *The estimate (6.5) fails for \( q < \frac{4p}{n(p-2)} \).*

*Proof.* This is essentially a consequence of the sharpness of the \( L^q_t L^p_x \) Strichartz estimates on \( \mathbb{R}^n \). Indeed, suppose our initial data \( f \) is a function on \( \mathbb{R}^n \). Then \( e^{it\Delta_{\mathbb{R}^n}} f(x) \) and

\[
\|e^{it\Delta_{\mathbb{R}^n}} f\|_{L^q_t L^p_x(\mathbb{R}^n \times [\gamma,\gamma+1])} = \|e^{it\Delta_{\mathbb{R}^n}} f\|_{L^p(\mathbb{R}^n \times [\gamma,\gamma+1])}.
\]

Now an application of Minkowski’s inequality shows that if \( q \geq p \) then for any function \( F(x,t) \) one has

\[
\|F\|_{L^q_t L^p_x(\mathbb{R}^n \times \mathbb{R}^n)} \leq \left( \sum_{\gamma \in \mathbb{Z}} \left( \int_{\mathbb{R}^n} \left( \int_{\gamma}^{\gamma+1} |F(x,t)|^q \, dt \right)^{\frac{p}{q}} \, dx \right)^{\frac{1}{q}} \right)^{\frac{1}{p}}.
\]

Hence if there is some function \( G_\gamma(x) \) such that \( |F(x,t)| \sim |G_\gamma(x)| \) for \( t \in [\gamma,\gamma+1] \) then

\[
\|F\|_{L^q_t L^p_x(\mathbb{R}^n \times \mathbb{R}^n)} \lesssim \|F\|_{\ell^q_t L^p(\mathbb{R}^n \times [\gamma,\gamma+1])}. \tag{6.7}
\]

We choose \( f \) so that \( \hat{f} \) is a smooth function supported in a ball of radius one centered at the origin. A stationary phase argument (exploiting standard uncertainty principle heuristics) shows that \( |e^{it_1\Delta} f(x)| \sim |e^{it_2\Delta} f(x)| \) uniformly for \( t_1, t_2 \in [0,1] \). This property extends to \( t_1, t_2 \in [\gamma,\gamma+1] \) since \( e^{i(\gamma+t)\Delta} f = e^{it\Delta}(e^{i\gamma\Delta} f) \) and \( e^{i\gamma\Delta} f \) has the same Fourier support as \( f \). But then by (6.7) we have

\[
\|e^{it\Delta_{\mathbb{R}^n}} f\|_{L^q_t L^p_x(\mathbb{R}^n \times \mathbb{R}^n)} \lesssim \|e^{it\Delta_{\mathbb{R}^n}} f\|_{\ell^q_t L^p(\mathbb{R}^n \times [\gamma,\gamma+1])},
\]
and therefore if Theorem 6.2 held for this choice of $p,q$ we would obtain

$$\| e^{it\Delta_{\mathbb{R}^n}} f \|_{L_t^q L_x^p(\mathbb{R} \times \mathbb{R}^n)} \lesssim \| f \|_{L^2(\mathbb{R}^n)}.$$  

This estimate remains true for the rescaled functions $f_\lambda(x) = f(\lambda x)$ when $\lambda \leq 1$, since the Fourier support of $f_\lambda$ remains inside the unit ball. But then a simple change of variables implies that

$$\| e^{it\Delta_{\mathbb{R}^n}} f \|_{L_t^q L_x^p(\mathbb{R} \times \mathbb{R}^n)} \lesssim \lambda^{\frac{n}{2} + \frac{n}{p} - \frac{n}{2}} \| f \|_{L^2(\mathbb{R}^n)},$$

and in particular we must have $\frac{2}{q} + \frac{n}{p} \leq \frac{n}{2}$ and hence $q \geq \frac{4p}{n(p-2)}$.

\[ \square \]

**Proposition 6.5.** Suppose $p \geq \frac{2(n+d+2)}{n+d}$. Then (6.5) is false for $s < \frac{n+d}{2} - \frac{n+d+2}{p}$.

**Proof.** The example demonstrating the sharpness of $s$ is essentially the Knapp counterexample, although there are some minor extra technicalities involved. The basic idea is that if our initial data is supported in a very small ball centered at the origin, then for a short time interval there is no difference between the Euclidean and semi-periodic cases; but since there are extremal examples for the equation on $\mathbb{R}^{n+d}$ that are highly localized in space (variants of the Knapp example), one cannot do any better than in the Euclidean setting.

It suffices to prove the claim for $q = \infty$. We identify $\mathbb{T}^d \sim [-\frac{1}{2}, \frac{1}{2}]^d$ and choose a smooth and positive function $\phi_N(x,y) = \phi(Nx, Ny)$ supported in $[-\frac{c}{N}, \frac{c}{N}]^{n+d}$ such that $\hat{\phi}_N$ decays rapidly outside the ball of radius $N$ in $\mathbb{R}^{n+d}$ centered at the origin. Note that we can also view $\phi_N(x,y)$ as a function on $\mathbb{R}^n \times \mathbb{T}^d$. Now let $\varphi_N(y) = \varphi(Ny)$ be a smooth function that rapidly decays outside $[-\frac{c}{10N}, \frac{c}{10N}]^d$, with $\hat{\varphi}_N$ supported in a ball of radius $\sim N$ centered at the origin in $\mathbb{R}^d$. Also let $\chi_{N^2}(t) = \chi(N^2 t)$ be a smooth function decaying outside $[\frac{c}{20N^2}, \frac{c}{10N^2}]$ with Fourier support in an interval of length $\sim N^2$. We have

$$|e^{it\Delta_{\mathbb{R}^n \times \mathbb{T}^d}} \phi_N(x,y)| \geq c |e^{it\Delta_{\mathbb{R}^n \times \mathbb{T}^d}} \phi_N(x,y) \cdot \varphi_N(y) \chi_{N^2}(t)|,$$
and by computing Fourier transforms we see that

\[
e^{it\Delta_{\mathbb{R}^n \times T}} \phi_N(x, y) \cdot \varphi(y) = \int_{\mathbb{R}^{n+d+1}} \sum_m \hat{\varphi}_N(\eta - m) \hat{\phi}_N(\xi, m) \chi_{\mathbb{R}^{n+1}}(\tau - |\xi|^2 - |m|^2) e^{2\pi i (x \cdot \xi + y \cdot \eta + t \tau)} d\eta d\xi d\tau.
\]

(6.8)

After rescaling \((\xi, \eta, \tau) \rightarrow (N^{-1} \xi, N^{-1} \eta, N^{-2} \tau)\), we see that (6.8) is equal to

\[
\int_{\mathbb{R}^{n+d+1}} \sum_m \varphi(\eta - N^{-1} m) \hat{\phi}(\xi, N^{-1} m) \tilde{\chi}(\tau - |\xi|^2 - N^{-2}|m|^2) e^{2\pi i (N x \cdot \xi + N y \cdot \eta + (N^2 t) \tau)} d\eta d\xi d\tau
\]

:= \mathcal{E}(N x, N y, N^2 t).

The key point is that \(\mathcal{E}(x, y, t)\) has space-time Fourier support in a ball of constant radius, independent of \(N\). We have therefore shown that

\[
\int_0^1 \int_0^1 \int_{B_1^{n+d}} |\mathcal{E}(x, y, t)|^p \, dx \, dy \, dt \geq N^{-(n+d+2)} \int_0^1 \int_{B_1^{n+d}} |\mathcal{E}(x, y, t)|^p \, dx \, dy \, dt.
\]

Now if the estimate (6.5) holds with constant \(A\), then since \(\|\varphi\|_{L^p} \sim N^{-p(n+d)/2}\) we must have

\[
A^p \geq N^{-(n+d/2)} \int_0^1 \int_{B_1^{n+d}} |\mathcal{E}(x, y, t)|^p \, dx \, dy \, dt.
\]

Finally, since \(\mathcal{E}\) has space-time Fourier support in a ball of bounded radius, one easily checks that \(|\mathcal{E}| \geq c\) when \((t, x, y) \in [0, 1] \times B_1^{n+d}\). Therefore \(A \geq N^{-(n+d)/2} \frac{n+d+2}{p}\) as claimed.

\[\square\]

### 6.4 The Decoupling Argument

We begin by proving Theorem 6.2 with a loss of \(N^\epsilon\). 

\[\text{147}\]
Proposition 6.6. For all \( p \geq p^* \) and \( q \geq \frac{3p}{n(p-2)} \) with \( q > 2 \) we have

\[
\left( \sum_{\gamma \in \mathbb{Z}} \| e^{it\Delta_{\mathbb{R}^n \times \mathbb{T}^d}} P_{\leq N} f \|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times [\gamma-1, \gamma+1])}^q \right)^{1/q} \lesssim \epsilon N^{\epsilon + \frac{n+d}{2} - \frac{n+d+2}{p}} \| f \|_{L^2(\mathbb{R}^n \times \mathbb{T}^d)}.
\]

We show in Section 6.5 that this loss of \( N^\epsilon \) can be removed for \( p \) away from the endpoint.

Since our arguments use the \( \ell^2 \) decoupling theorem to handle the torus component of the manifold, the proposition is true regardless of whether \( \mathbb{T}^d \) is the standard torus or irrational torus. We include some more details in comments below.

6.4.1 Main Lemmas

In this subsection we collect some preliminary results. For the rest of the section we will write \( \Delta = \Delta_{\mathbb{R}^n \times \mathbb{T}^d} \).

Lemma 6.7. Suppose \( g, g_l \) are Schwartz functions on \( \mathbb{R}^n \times \mathbb{T}^d \) with \( g = P_{\leq N} g \), such that

\[
g(x, y) = \sum_{l \in \mathbb{Z}^d} \int_{[-N,N]^n} \hat{g}_l(\xi) e^{2\pi i (x \cdot \xi + y \cdot l)} d\xi.
\]

Cover \( [-N,N]^n \) by finitely-overlapping cubes \( Q_k \) of side-length \( \sim 1 \), let \( \{ \phi_k \} \) be a partition of unity adapted to the \( \{ Q_k \} \), and define \( g_{\theta_{m,k}} = e^{2\pi iy \cdot m} F_x^{-1}(\hat{g}_m \phi_k) \). Also let \( p^* = \frac{2(n+d+2)}{n+d} \) and \( p \geq p^* \). Then for any time interval \( I \) of length \( \sim 1 \) we have

\[
\| e^{it\Delta} g \|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times I)} \lesssim \epsilon N^{\epsilon + \frac{n+d}{2} - \frac{n+d+2}{p}} \left( \sum_{m,k} \| e^{it\Delta} g_{\theta_{m,k}} w_I \|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times \mathbb{R})}^2 \right)^{1/2},
\]

where \( w_I \) is a bump function adapted to \( I \).

The proof of Lemma 6.7 is similar to the proof of the discrete restriction theorem in [71]. One approximates functions on \( \mathbb{R}^n \times \mathbb{T}^d \) by functions on \( \mathbb{R}^{n+d} \), applies the \( \ell^2 \) decoupling theorem, and then takes limits. The details are in Section 6.4.3 below.
Lemma 6.8. Let \( h(x, y, t) = \sum_l h_l(x, t) e^{2\pi i y l} \) be a Schwartz function on \( \mathbb{R}^n \times T^d \times [-1, 1] \) such that \( h_l(x, t) \neq 0 \) for at most \( R \) values of \( l \). Let \( \psi \) be a bump function on \( \mathbb{R}^n \) supported in \( B_1 \) and define

\[
K_{\gamma}(x, y, t) = \sum_{l \in \mathbb{Z}^d |l| \leq cN} \int_{\mathbb{R}^n} \psi(N^{-1}\xi) e^{2\pi i (x \cdot \xi + y \cdot l + (t + \gamma)(|\xi|^2 + |l|^2))} d\xi
\]

for \( \gamma \in \mathbb{Z} \) with \( |\gamma| \geq 1 \). Then for all \( p > 2 \)

\[
\| K_{\gamma} * h \|_{L^p(\mathbb{R}^n \times T^d \times [-1, 1])} \lesssim R^{\frac{p-2}{p}} \| \gamma \|^{-\mu} \| h \|_{L^p(\mathbb{R}^n \times T^d \times [-1, 1])}, \tag{6.9}
\]

where \( \mu = \frac{n(p-2)}{2p} \). In particular, if \( p = p^* \) we have \( \mu = \frac{n}{n+d+2} \).

Proof. Let \( S \subset \mathbb{Z}^d \) be the set of \( l \) such that \( h_l \neq 0 \). By examining the Fourier coefficients we see that \( K_{\gamma} * h = K_{\gamma,S} * h \), where

\[
K_{\gamma,S} = \sum_{l \in S} \int_{\mathbb{R}^n} \psi(N^{-1}\xi) e^{2\pi i (x \cdot \xi + y \cdot l + (t + \gamma)(|\xi|^2 + |l|^2))} d\xi.
\]

By a stationary phase argument and the triangle inequality we have \( \| K_{\gamma,S} \|_{L^\infty} \lesssim R |\gamma|^{-n/2} \), and therefore \( \| K_{\gamma} * h \|_{L^\infty} \lesssim R |\gamma|^{-n/2} \| h \|_{L^1} \). On the other hand, since \( t \) is in a bounded interval we can use Plancherel’s theorem to get \( \| K_{\gamma,S} * h \|_{L^2} \lesssim \| h \|_{L^2} \). Interpolating these estimates yields \( \| K_{\gamma,S} * h \|_{L^p} \lesssim R^{\frac{(p-2)}{p}} |\gamma|^{-\mu} \| h \|_{L^{p'}} \), and this completes the proof since \( K_{\gamma} * h = K_{\gamma,S} * h \).

Finally, we record the following local Strichartz estimate.

Proposition 6.9. For any bounded time interval \( I \) and \( p \geq p^* \) one has

\[
\| e^{it\Delta_{\mathbb{R}^n \times T^d}} P \leq N f \|_{L^p(\mathbb{R}^n \times T^d \times I)} \lesssim I e^{N^{n + \frac{n+d}{2}} \frac{n+d+2}{p} \| f \|_{L^2(\mathbb{R}^n \times T^d)}}
\]

and

\[
\| \int_I e^{it\Delta} P \leq N h(x, y, t) dt \|_{L^2_{x,y,t}(\mathbb{R}^n \times T^d \times I)} \lesssim I e^{N^{n + \frac{n+d}{2}} \frac{n+d+2}{p} \| h \|_{L^{p'}(\mathbb{R}^n \times T^d \times I)}}.
\]
Proof. We prove the first estimate since the second follows easily by duality.

Suppose \( f = P_{\leq N} f \). By interpolating with \( p = \infty \) (via Bernstein’s inequality) it suffices to prove the endpoint case \( p = p^* \). By Lemma 6.7

\[
\| e^{it\Delta} f \|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times I)} \lesssim_{\epsilon, I} N^\epsilon \left( \sum_{m,k} \| e^{it\Delta} f_{\theta_{m,k}} \cdot w_I \|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times \mathbb{R})}^2 \right)^{1/2}. 
\]

By Plancharell’s theorem it suffices to prove the desired estimate when \( f = P_\theta f \) and \( \theta = \theta_{m,k} \).

In this case we apply Hölder’s inequality in time to get

\[
\| e^{it\Delta} f_{\theta} \cdot w_I \|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times \mathbb{R})} \lesssim \| e^{it\Delta} f_{\theta} \|_{L^q_t L^p_x},
\]

where \( q = \frac{2(n+d+2)}{n} \) is the admissible time exponent for the \( L^q_t L^p_x \) Strichartz estimate on \( \mathbb{R}^n \).

Applying this Strichartz bound completes the proof.

\[ \square \]

Remark 6.4.1. In the special case \( n = d = 1 \) we can instead apply the stronger local Strichartz estimate

\[
\| e^{it\Delta_R \times T} f \|_{L^q(\mathbb{R} \times \mathbb{T} \times I)} \leq C_I \| f \|_{L^2(\mathbb{R} \times \mathbb{T})}
\]
due to Takaoka and Tzvetkov [83]. This will simplify the \( \epsilon \)-removal argument in Section 4 when \( n = d = 1 \).

6.4.2 Main Argument

We prove Proposition 6.6 in the case \( p = p^* \). The the case \( p > p^* \) is essentially the same (or can be obtained by interpolation). Let \( q = \frac{4p^{*}}{n(p^{*}-2)} = \frac{2(n+d+2)}{n} \) and let \( w_\gamma \) be a bump function adapted to \([\gamma - 1, \gamma + 1] \). We wish to show that

\[
\left( \sum_{\gamma \in \mathbb{Z}} \| e^{it\Delta} P_{\leq N} f \cdot w_\gamma \|_{L^{p^*}(\mathbb{R}^n \times \mathbb{T}^d \times \mathbb{R})}^q \right)^{1/q} \lesssim_{\epsilon} N^\epsilon \| f \|_{L^2(\mathbb{R}^n \times \mathbb{T}^d)}. \tag{6.10}
\]
We will assume throughout the section that \( f = P_{\leq N} f \). Let \( M(f) \) denote the mixed norm on the left-hand side of (6.10). In order to show that \( M(f) \lesssim \varepsilon N \| f \|_{L^2} \), we decouple the frequencies to reduce to the case where \( f \) has small Fourier support. Let \( f_{\theta_{m,k}} \) be defined as in Lemma 6.7. Then by applying Lemma 6.7 for each \( \gamma \) and using Minkowski’s inequality in \( \ell^2 \) we obtain

\[
M(f) \lesssim \varepsilon N^\varepsilon \left( \sum_{m,k} \left( \sum_{\gamma} \left\| e^{i t \Delta} f_{\theta_{m,k}} \cdot w_\gamma \right\|_{L^p(\mathbb{R}^n)}^2 \right)^{q/2} \right)^{1/q} \\
\lesssim \varepsilon N^\varepsilon \left( \sum_{m,k} \left( \sum_{\gamma} \left\| e^{i t \Delta} f_{\theta_{m,k}} \cdot w_\gamma \right\|_{L^p(\mathbb{R}^n)}^q \right)^{2/q} \right)^{1/2} \\
= C \varepsilon N^\varepsilon \left( \sum_{m,k} M(f_{\theta_{m,k}})^2 \right)^{1/2}.
\] (6.11)

To complete the proof, we claim that

\[
M(f_{\theta_{m,k}}) \lesssim \| f_{\theta_{m,k}} \|_{L^2(\mathbb{R}^n \times T^d)}
\] (6.12)

for each \( m, k \). This will be enough, since by (6.11) and Plancherel’s theorem we then have

\[
M(f) \lesssim \varepsilon N^\varepsilon \left( \sum_{m,k} M(f_{\theta_{m,k}})^2 \right)^{1/2} \lesssim \varepsilon N^\varepsilon \left( \sum_{m,k} \| f_{\theta_{m,k}} \|_{L^2}^2 \right)^{1/2} \lesssim \varepsilon N^\varepsilon \| f \|_{L^2}.
\]

Now (6.12) shows that in order to prove (6.10) we can in fact assume that the Fourier transform of \( f \) is supported in a cube of side length \( \sim 1 \) in the frequency space \( \mathbb{R}^n \times \mathbb{Z}^d \).

Let \( P_\theta f \) denote a smooth frequency cut-off of \( f \) onto \( \theta \). Then by Hölder’s inequality in time one has

\[
\left( \sum_{\gamma \in \mathbb{Z}} \left\| e^{i t \Delta} P_\theta f \cdot w_\gamma \right\|_{L^p(\mathbb{R}^n \times T^d)}^q \right)^{1/q} \lesssim \left( \sum_{\gamma \in \mathbb{Z}} \left\| e^{i t \Delta} P_\theta f \cdot w_\gamma \right\|_{L^p(\mathbb{R}^n \times T^d)}^q \right)^{1/q}.
\]
where \( q = \frac{2(n+d+2)}{n} > p^* \). It is straightforward to check that \( \sum_\gamma w_\gamma \lesssim 1 \), and therefore

\[
\left( \sum_{\gamma \in \mathbb{Z}} \| e^{it\Delta} P_0 f \cdot w_\gamma \|_{L^q_t L^{p^*}([\mathbb{R}^n \times \mathbb{T}^d])}^q \right)^{1/q} \lesssim \| e^{it\Delta} P_0 f \|_{L^q_t L^{p^*}_x, y}.
\]

But \( P_0 f \) only has one non-zero Fourier coefficient, and therefore we can use the Euclidean \( L^q_t L^{p^*}_x \) Strichartz estimate to finally obtain

\[
\| e^{it\Delta} P_0 f \|_{L^q_t L^{p^*}_x, y} \lesssim \| P_0 f \|_{L^2([\mathbb{R}^n \times \mathbb{T}^d])},
\]

as desired.

### 6.4.3 Proof of Decoupling Lemma 6.7

Recall \( \alpha_p = \frac{n+d}{2} - \frac{n+d+2}{p} \). Let \( B_l \subset \mathbb{R}^n \) be a fixed ball of radius \( N \), and let \( w_l \) be a smooth weight adapted to \( B_l \times [-1, 1] \). To prove the lemma it will suffice to show that

\[
\| e^{it\Delta} g \|_{L^p(B_l \times \mathbb{T}^d \times [-1, 1])} \lesssim \varepsilon N^{\epsilon + \alpha_p} \left( \sum_{m,k} \| e^{it\Delta} g_{\theta_{m,k}} \|_{L^p(w)}^2 \right)^{1/2}. \quad (6.13)
\]

Then to prove the full estimate on \( \mathbb{R}^n \times \mathbb{T}^d \), we can choose a finitely-overlapping collection of balls \( B_l \) of radius \( N \) that cover \( \mathbb{R}^n \), and then apply (6.13) in each \( B_l \) and use Minkowski’s inequality to sum:

\[
\| e^{it\Delta} g \|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times [-1, 1])} \leq \sum_l \| e^{it\Delta} g \|_{L^p(B_l \times \mathbb{T}^d \times [-1, 1])}^{1/p} \lesssim \varepsilon N^{\epsilon + \alpha_p} \left( \sum_{m,k} \| e^{it\Delta} g_{\theta_{m,k}} \|_{L^p(w)}^2 \right)^{1/2} \lesssim \varepsilon N^{\epsilon + \alpha_p} \left( \sum_{m,k} \| e^{it\Delta} g_{\theta_{m,k}} \|_{L^p(w)}^2 \right)^{1/2},
\]

where the \( w_l \) are weights adapted to \( B_l \times [-1, 1] \) and \( w = \sum_l w_l \). Since \( w \leq C w_{[-1, 1]} \) this will complete the proof of Lemma 6.7.
Let \( u(x, y, t) = e^{t \Delta} g(x, y) \). We begin by rescaling \( u_0 \) to have frequency support in \([-1, 1]^{n+d} \).

Note that

\[
u(N^{-1}x, N^{-1}y, N^{-2}t) = \int_{B_1^n} \sum_{m \in \Lambda_N} \hat{g}(N\xi, Nm) e^{i(x\xi + y\cdot m + t(|\xi|^2 + |m|^2))} \, d\xi,
\]

where \( B_1^n \) and \( B_1^d \) are \([-1, 1]^n \) and \([-1, 1]^d \), respectively. We let \( f(\xi, m) = \hat{g}(N\xi, Nm) \), so if \( \Lambda_N = N^{-1} \mathbb{Z}^d \cap [-1, 1]^d \) then \( f \) is supported on \( B_1^n \times \Lambda_N \). Below we will exploit the fact that \( f \) can be viewed as a function on \( n \)-dimensional cubes of size \( \sim N^{-1} \) that are \( \sim N^{-1} \)-separated in \( B_1^{n+d} \) (this follows from the support property of \( \hat{g} \); we have one cube for each \( m \)). Let \( Ef \) denote the extension operator

\[
Ef = \sum_{m \in \Lambda_N} \int_{B^n_1} f(\xi, m) e^{i(x\xi + y\cdot m + t(|\xi|^2 + |m|^2))} \, d\xi.
\]

After applying a change of variables on the spatial side and using periodicity in the \( y \) variable, we see that

\[
\| u \|_{L^p(B_N \times \mathbb{T}^d \times [0, 1])} = N^n N^{-(n+d+2)/p} \| Ef \|_{L^p(B_{N^2} \times N\mathbb{T}^d \times [0, N^2])} = N^n (1 - \frac{1}{p}) \frac{d+2}{p} N^{-\frac{d}{p}} \| Ef \|_{L^p(B_{N^2} \times N^2\mathbb{T}^d \times [0, N^2])},
\]

In particular, for \( p^* = \frac{2(n+d+2)}{n+d} \) we have

\[
\| u \|_{L^{p^*}(B_N \times \mathbb{T}^d \times [0, 1])} = N^{\frac{n-d}{2} - \frac{d}{p^*}} \| Ef \|_{L^{p^*}(B_{N^2} \times N^2\mathbb{T}^d \times [0, N^2])},
\]

for any solution \( u \) with initial data \( u_0 \) and \( f = \hat{u}_0(N\xi, Nm) \).

We also introduce the operator \( \widetilde{E} \) on \( C^\infty([-1, 1]^{n+d}) \) defined by

\[
\widetilde{E} h(x, y, t) = \int_{B^n_1} \int_{B^d_1} h(\xi_1, \xi_2) e^{i(x\cdot \xi_1 + y\cdot \xi_2 + t(|\xi_1|^2 + |\xi_2|^2))} \, d\xi_1 d\xi_2
\]

(this is the usual extension operator associated to the paraboloid \( P^{n+d} \) in \( \mathbb{R}^{n+d+1} \)). Given a
function \( f \) on \([-1, 1]^n \times \Lambda_N\), let

\[
f^\delta(\xi_1, \xi_2) = \sum_{m \in \Lambda_N} c_d \delta^{-d} 1_{\{|\xi_2 - m| \leq \delta\}} f(\xi_1, m), \quad \delta < \frac{1}{N}
\]

where \( c_d \) is a dimensional constant chosen for normalization. Then by Lebesgue differentiation and Fatou’s lemma we have

\[
\| Ef \|_{L^p(B_{N^2} \times N^2 \times [0, N^2])} \leq \liminf_{\delta \to 0} \| \tilde{E} f^\delta \|_{L^p(B_{N^2} \times [0, N^2]^d \times [0, N^2])},
\]

where as usual we identify \( N^2 \mathbb{T}^d \) with \([0, N^2]^d\). We will begin by estimating \( \tilde{E} f \) for arbitrary \( f \) on \( B_1^{n+d} \) before specializing to \( f^\delta \) and passing to the limit later in the argument.

Fix a parameter \( R \lesssim N \) and choose a finitely-overlapping collection of \((n + d)\)-dimensional boxes \( \theta_{m,k} = I_k \times I_m \subset B_1^n \times B_1^d \) with the following properties:

(i) Each \( \theta_{m,k} \) has side lengths \( \sim R^{-1} \)

(ii) \( I_m \) is centered at \( m \in \Lambda_N \)

(iii) \( m \) varies over an \( \sim R^{-1} \)-separated subset of \( \Lambda_N \) such that the cubes \( I_m \) cover \([-1, 1]^d\).

Such a collection of \( \theta_{m,k} \) yields a finitely-overlapping tiling of \( B_1^{n+d} \) by boxes of side length \( \sim R^{-1} \), which corresponds to a covering of the paraboloid \( P^{n+d} \) by \( R^{-1} \)-caps. Let \( \{ \varphi_{m,k} \} \) be a smooth partition of unity subordinate to this cover, and let \( f_{m,k} = f \cdot \varphi_{m,k} \). By Theorem 6.4

\[
\| \tilde{E} f \|_{L^p(B_{N^2} \times [0, N^2]^d \times [0, N^2])} \lesssim \epsilon R^{d+\alpha} \left( \sum_{m,k} \| \tilde{E} f_{m,k} \|_{L^p(w_{N^2})}^2 \right)^{\frac{1}{2}},
\]

where \( w_{N^2} \) is a bump function adapted to \( B_{N^2} \times [0, N^2]^d \times [0, N^2] \). Setting \( R = N \) and \( p = p^* \) yields

\[
\| \tilde{E} f \|_{L^{p^*}(B_{N^2} \times [0, N^2]^d \times [0, N^2])} \lesssim \epsilon N^{d} \left( \sum_{m,k} \| \tilde{E} f_{m,k} \|_{L^{p^*}(w_{N^2})}^2 \right)^{\frac{1}{2}},
\]

(6.16)
with the $f_{m,k}$ supported on finitely-overlapping boxes of side lengths $\sim \frac{1}{N}$. Specializing to $f^\delta$ and taking a limit as in (6.15), we get

$$\|Ef\|_{L^{p^*}(B_N^2 \times 0, R^d \times [0, N^2])} \lesssim \epsilon \lim_{\delta \to 0} N^\epsilon \left( \sum_{m,k} \|\tilde{E}f^\delta_{m,k}\|_{L^{p^*}(w_{N^2})}^2 \right)^{\frac{1}{2}}.$$  

Then as a consequence of the scaling (6.14)

$$\|u\|_{L^{p^*}(B_N \times T^d \times [0, 1])} \lesssim \epsilon \lim_{\delta \to 0} N^{\frac{n-d}{p^*} - \frac{d}{p^*} + \epsilon} \left( \sum_{m,k} \|\tilde{E}f^\delta_{m,k}\|_{L^{p^*}(w_{N^2})}^2 \right)^{\frac{1}{2}},$$  

(6.17)

where as above $f(\xi, n) = \hat{u}_0(N\xi, Nn)$ for $n \in \Lambda_N$. Now by rescaling as before and writing out the definition of $f^\delta$ we get

$$N^{\frac{n-d}{p^*}} \|\tilde{E}f^\delta_{m,k}\|_{L^{p^*}(w_{N^2})} \approx N^{-\frac{d}{p^*}} \|u^\delta\|_{L^{p^*}(v)},$$

where

$$u^\delta = \sum_{l \in \mathbb{Z}^d \atop |l-m| \leq 1} \int_{R^d} \int_{R^n} c(N\delta)^{-d} e^{-f_{m,k}(N^{-1}\xi_1, N^{-1}l)} e^{2\pi i (x \cdot \xi_1 + y \cdot \xi_2 + t(\xi_1^2 + \xi_2^2))} d\xi_1 d\xi_2$$

and $v$ is a suitable bump function adapted to $B_N \times [0, N]^d \times [-1, 1]$. The desired result now follows from the pointwise estimate

$$|u^\delta| \leq \sum_{l \in \mathbb{Z}^d \atop |l-m| \leq 1} \left| \int_{R^n} f_{m,k}(N^{-1}\xi_1, N^{-1}l) e^{2\pi i (x \cdot \xi_1 + t\xi_1^2)} d\xi_1 \right|,$$

which is uniform in $\delta$.

The case $p > p^*$ follows by a similar argument (or interpolation).
6.5 The $\epsilon$-Removal Argument

In this section we show that the $N^\epsilon$ loss from our Strichartz estimates can be removed for $p > p^*$, where as before $p^*$ is the endpoint $p^* = \frac{2(n+d+2)}{n+d}$. The argument has a local and global component. In the local case the philosophy is the same as in [70] and [81]: one applies the Strichartz estimate with $\epsilon$ loss in the region where the operator is in some sense ‘small,’ and this leaves a ‘large’ portion of the operator which we can control with direct estimates for the associated kernel $K$, at least for $p$ away from the endpoint. To extend the $\epsilon$-removal to the global case we combine the local estimates with an interpolation argument to handle the global portion of a relevant bilinear form.

6.5.1 The Local-In-Time Case

We begin the argument by first showing that one can remove the $N^\epsilon$ loss from the local-in-time estimate from Proposition 6.9.

Proposition 6.10. Suppose $p > \frac{2(n+d+2)}{n+d}$. Then

$$\|e^{it\Delta_{\mathbb{R}^n \times \mathbb{T}^d}}P_{\leq N}f\|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times [0,1])} \lesssim N^{\frac{n+d}{2} - \frac{n+d+2}{p}}\|f\|_{L^2(\mathbb{R}^n \times \mathbb{T}^d)}.$$

To prove the proposition we reduce the problem to a situation where we can apply ideas due to Killip and Visan. The argument in this local setting is almost the same as their proof of $\epsilon$-removal in the case $n = 0$ from [81].

Normalize $\|f\|_{L^2} = 1$ and let $F = e^{it\Delta_{\mathbb{R}^n \times \mathbb{T}^d}}P_{\leq N}f$. Using the level set characterization of the $L^p$ norm and Bernstein’s inequality we get

$$\|F\|_{L^p(\mathbb{R}^n \times \mathbb{T}^d \times [0,1])}^p = p \int_0^{CN^{\frac{n+d}{2}}} \mu^{p-1} \mu^\delta \|\{(x, y, t) : |F(x, y, t)| > \mu\}\| d\mu.$$

Now if $\delta > 0$ then we can apply Chebyshev’s inequality and Proposition 6.9 to see that
\[
\int_0^{N \frac{n+d-\delta}{2}} \mu^{p-1}(|F > \mu|) \ d\mu \lesssim \epsilon \ N^{p^* (\frac{n+d}{2} - \frac{n+d+2}{p})} N^{(p-p^*)(\frac{n+d-\delta}{2})+\epsilon} \\
\leq C \epsilon N^{p^* (\frac{n+d}{2} - \frac{n+d+2}{p})} N^{\epsilon - \delta (p-p^*)} \\
\leq C \epsilon N^{p^* (\frac{n+d}{2} - \frac{n+d+2}{p})}
\]

provided \(\epsilon\) is chosen small enough. It remains to estimate the large portion of the integral

\[
A := \int_{N \frac{n+d-\delta}{2}}^{CN \frac{n+d}{2}} \mu^{p-1}(|F > \mu|) \ d\mu. \tag{6.18}
\]

We let \(\Omega = \{ F > \mu \} \) for \(\mu \geq N \frac{n+d-\delta}{2}\) fixed, and set \(\Omega_\omega = \{ \operatorname{Re}(e^{i\omega F}) > \mu/2 \}\). Note that there is some choice of \(\omega \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}\) such that \(|\Omega| \leq 4|\Omega_\omega|\), so we will estimate \(|\Omega_\omega|\) instead. In particular

\[
|\mu^2|\Omega_\omega|^2 \lesssim \langle 1_{\Omega_\omega}, K \ast 1_{\Omega_\omega} \rangle_{L^2_{x,y,t}}, \tag{6.19}
\]

where \(K\) is the kernel

\[
K(x,y,t) = \left( \sum_{l \in \mathbb{Z}^d} \psi(N^{-1} l) e^{2\pi i (y \cdot l + t |l|^2)} \right) \left( \int_{\mathbb{R}^n} \psi(N^{-1} \xi) e^{2\pi i (x \cdot \xi + t |\xi|^2)} d\xi \right)
\]

from above.

Since \(t \in [0,1]\) we cannot take advantage of the dispersion coming from the \(\mathbb{R}^n\) component of \(K\). Instead we discretize the kernel and proceed as if we were working on \(\mathbb{T}^{n+d}\).

**Lemma 6.11.** Let \(K_\mathbb{R}(x,t) = \int_{\mathbb{R}^n} \psi(N^{-1} \xi) e^{2\pi i (x \cdot \xi + t |\xi|^2)} d\xi\). Then if \(Q_0\) is the cube of side-length 1 centered at the origin we have

\[
|K_\mathbb{R}(x,t)| \lesssim \sup_{\alpha \in Q_0} \sum_{k \in \mathbb{Z}^n, |k| \leq cN} \psi(N^{-1} (\alpha + k)) e^{2\pi i ((x \cdot k + t |k|^2)}
\]

where

\[
|K_\mathbb{R}(x,t)| \lesssim \sup_{\alpha \in Q_0} \sum_{k \in \mathbb{Z}^n, |k| \leq cN} \psi(N^{-1} (\alpha + k)) e^{2\pi i ((x \cdot k + t |k|^2)}
\]
Proof. Let $Q_k = Q_0 + k$, where $k \in \mathbb{Z}^n$. We can write

$$K_R(x, t) = \sum_{k \in \mathbb{Z}^n} \int_{Q_k} \psi(N^{-1}\xi)e^{2\pi i(x \cdot \xi + t|\xi|^2)} d\xi.$$ 

Now making the change of variables $\xi = \alpha + k$ for each $k$, where $\alpha \in Q_0$, we get

$$K_R(x, t) = \int_{Q_0} e^{2\pi i(x \cdot \alpha + t|\alpha|^2)} \sum_{k \in \mathbb{Z}^n \atop |k| \leq cN} \psi(N^{-1}(\alpha + k))e^{2\pi i[(x + 2t\alpha) \cdot k + t|k|^2]} d\alpha.$$ 

The lemma follows by taking absolute values inside the integral and taking the supremum in $\alpha$. \hfill \Box

We will now use a pointwise estimate for $K$ originally due to Bourgain [70]. Given integers $1 \leq q < N$ and $1 \leq a < q$ with $(a, q) = 1$, let $S_{q, a} = \{t \in [0, 1] : |t - \frac{a}{q}| \leq \frac{1}{qN}\}$. We apply Lemma 6.11 and the proof of Lemma 3.18 in [70] for fixed $\alpha$ to obtain \footnote{Recall that Dirichlet’s approximation theorem implies that either $t \in S_{q, a}$ for some pair $(q, a)$, or $t$ is close to 0.}

$$|K(x, y, t)|_{S_{q, a}(t)} \lesssim \frac{N^{n+d}}{q^{\frac{n+d}{2}}(1 + N^{n+d}t - \frac{a}{q})^{\frac{n+d}{2}}}.$$ \hfill (6.20)

Following Killip and Visan [81], define the ‘large’ set

$$\mathcal{T} = \{t \in [0, 1] : qN^2|t - a/q| \leq N^{2\rho} \text{ for some } q \leq N^{2\rho}, \text{ and } (a, q) = 1\},$$

where $\rho > 0$ is a small parameter to be determined below. Also define

$$\tilde{K}(x, y, t) = K(x, y, t)1_{\mathcal{T}}(t),$$

and observe that from (6.20)

$$|K - \tilde{K}| \lesssim N^{n+d-(n+d)\rho}. \hfill (6.21)$$
This implies that
\[ |\langle 1_{\Omega}, (K - \widetilde{K}) \ast 1_{\Omega} \rangle_{L^2_{x,y,t}}| \lesssim N^{n+d-(n+d)\rho}|\Omega_{\omega}|^2, \]
and since \( \mu \geq N^{\frac{n+d}{2}-\delta} \) we can absorb the contribution of \( K - \widetilde{K} \) to the left-hand side of (6.19) provided we take \( \rho < \frac{2\delta}{n+d} \). To estimate the contribution from \( \widetilde{K} \) we use the following result due to Killip and Visan:

**Proposition 6.12 ([81]).** Suppose \( r > \frac{2(n+d+2)}{n+d} \). Then

\[ \| \widetilde{K} \ast F \|_{L^r(\mathbb{R}^n \times T^d \times [0,1])} \lesssim N^{2(n+d+2) - \frac{n+d+2}{r}} \| F \|_{L^{r'}(\mathbb{R}^n \times T^d \times [0,1])}, \]

provided \( \rho \) is chosen small enough (depending only on \( n,d,r \)).

**Proof.** The estimate follows by applying the argument from [81], Section 2, after bounding the kernel \( K \) pointwise via Lemma 6.11 and (6.20). \( \square \)

Applying this proposition to (6.19), we obtain

\[ \mu^2 |\Omega_{\omega}|^2 \lesssim |\langle 1_{\Omega_{\omega}}, \widetilde{K} \ast 1_{\Omega_{\omega}} \rangle_{L^2_{x,y,t}}| \lesssim |\Omega_{\omega}|^2 N^{2(n+d) - \frac{n+d+2}{r}}. \]

We can then conclude that

\[ |\Omega| \leq 4|\Omega_{\omega}| \lesssim N^{2(n+d) - \frac{2(n+d+2)}{r}} \mu^{-r} \tag{6.22} \]

for any \( r \in \left( \frac{2(n+d+2)}{n+d}, p \right) \). Plugging (6.22) into (6.18) gives

\[ A \lesssim N^{2 \left( n+d - \frac{2(n+d+2)}{r} \right)} \int_{N^{-\frac{n+d}{2} - \delta}}^{CN \frac{n+d}{2}} \mu^{p-r-1} d\mu \]

\[ \lesssim N^{p \left( \frac{n+d}{2} - \frac{n+d+2}{p} \right)}, \]

completing the proof of Proposition 6.10.
Remark 6.5.1. Proposition 6.12 remains true if $T_d$ is replaced by a $d$-dimensional irrational torus, after a suitable modification of the pointwise estimate (6.20) to take into account the irrational parameters. See [81] for more details.

6.5.2 The Global Case

We now extend Proposition 6.10 to the global setting and show that for all $p > p^*$ and $q = q(p) = \frac{4p}{n(p-2)}$,

$$\left( \sum_{\gamma \in \mathbb{Z}} \left\| e^{i\Delta_s P(\mathbb{R}^n \times T^d \times [\gamma-1, \gamma])} f \right\|_{L^2([0,T] \times \mathbb{R}^n \times T^d \times [\gamma, \gamma+1])} \right)^{1/q} \lesssim N^{\frac{n+d}{2} - \frac{n+d+2}{p}} \| f \|_{L^2(\mathbb{R}^n \times T^d)}.$$

This will complete the proof of Theorem 6.2.

The argument is an adaptation of the bilinear interpolation approach of Keel and Tao [80], and is also inspired by the $TT^*$ argument of Hani and Pausader [75]. We will use the fact that we already have both the local estimate without $\epsilon$-loss, and the global estimate for $q = q(p)$ with $\epsilon$ loss. We assume below that $p > p^* = \frac{2(n+d+2)}{n+d}$ is fixed. We are free to take $p$ as close to $p^*$ as needed, since the remaining cases can be handled by interpolation.

Initial Reductions

Let $U_{\alpha}(t) = e^{-it(t+\alpha)}\Delta_{\mathbb{R}^n \times T^d}$, and for functions $h(x,y,t)$ let $h_\alpha(x,y,t) = h(x,y,t + \alpha)$. By appealing to duality, we see that if $T(h,g)$ is the bilinear form

$$T(h,g) = \sum_{\alpha \in \mathbb{Z}} \sum_{\gamma \in \mathbb{Z}} \int_0^1 \int_0^1 \langle U_{\alpha}(s) P(\mathbb{R}^n \times T^d \times [\gamma, \gamma+1]) f(s,t) \rangle_{L^2(s, \mathbb{R}^n \times T^d)} ds dt \quad (6.23)$$

then it suffices to show that for $p > \frac{2(n+d+2)}{n+d}$ and $q = q(p) = \frac{4p}{n(p-2)}$ we have

$$|T(h,g)| \lesssim N^{2\alpha_p} \| h \|_{\ell^p_t L^{p'}(\mathbb{R}^n \times T^d \times [\gamma, \gamma+1])} \| g \|_{\ell^p_t L^{p'}(\mathbb{R}^n \times T^d \times [\gamma, \gamma+1])}. \quad (6.24)$$

Note that we can immediately prove (6.24) for the diagonal portion of $T$ where $|\gamma| \leq 10$ by using
Cauchy-Schwarz and Proposition 6.10 (in fact this argument gives a stronger estimate with an $\ell^2$ sum). Hence we can assume in (6.23) that $|\gamma| \geq 10$. We can also assume that $h = P_{\leq N}h$ and $g = P_{\leq N}g$.

We dyadically decompose $T$ and for $j \geq 3$ define

$$T_j(h, g) = \sum_{\alpha \in \mathbb{Z}} \sum_{\gamma \in \mathbb{Z}} \int_0^1 \int_0^1 \langle U_\alpha(s) h_\alpha(s), U_{\alpha + \gamma}(t) g_{\alpha + \gamma}(t) \rangle_{L^2_{x,y}} ds dt,$$

with the goal of showing that

$$\sum_{j \geq 3} |T_j(h, g)| \lesssim N^{2 \alpha_p} \|h\|_{\ell^{q'}_{\ell^p}} \|g\|_{\ell^{q'}_{\ell^p}}.$$

We claim that we can assume $h_\alpha(s)$ is zero for $\alpha$ outside an interval of length $2^j$. Indeed, let $I_l = \{ \alpha : l2^j \leq \alpha < (l+1)2^j \}$ and suppose that for some pair of exponents $(a, b)$ we have

$$\sup_{l \in \mathbb{Z}} |T_j(h1_{I_l}, g)| \leq A \|h\|_{\ell^a_{\ell^p}} \|g\|_{\ell^b_{\ell^p'}}.$$

Note that for fixed $l$ the terms in $T_j(h, g)$ are zero unless $\alpha + \gamma \in 2I_{l+1}$. Then

$$|T_j(h, g)| \leq A \sum_l \|h1_{I_l}\|_{\ell^a_{\ell^p'}} \|g1_{2I_{l+1}}\|_{\ell^b_{\ell^p'}}$$

$$\leq A \left( \sum_l \|h1_{I_l}\|_{\ell^a_{\ell^p'}} \right)^{\frac{1}{2}} \left( \sum_l \|g1_{2I_{l+1}}\|_{\ell^b_{\ell^p'}} \right)^{\frac{1}{2}}$$

$$\leq A \left( \sum_l \|h1_{I_l}\|_{\ell^a_{\ell^p'}} \right)^{\frac{1}{2}} \left( \sum_l \|g1_{2I_{l+1}}\|_{\ell^b_{\ell^p'}} \right)^{\frac{1}{2}}$$

$$\leq cA \|h\|_{\ell^a_{\ell^p'}} \|g\|_{\ell^b_{\ell^p'}},$$

using the fact that $q > q'$ in the second-to-last inequality. Hence it suffices to estimate $T_j(h, g)$ when $h_\alpha$ is supported with $\alpha \in I_l$ for some $l$, which additionally implies that we can assume $g$ is time-supported in an interval of $J$ length $\sim 2^j$. Note that with these assumptions $T_j(h, g) =$
The First Interpolation

The first step is to prove the following two-parameter family of estimates. The result is similar to Lemma 4.1 in Keel-Tao \cite{80}, though we have to interpolate in a smaller range to avoid too large of a loss in $N$.

**Lemma 6.13.** For all $(\frac{1}{a}, \frac{1}{b})$ in a neighborhood of $(\frac{1}{p}, \frac{1}{p})$ with $a, b > p^*$ we have

$$|T_j(h, g)| \lesssim \epsilon N^{c(a,b)+\epsilon} 2^{j\beta(a,b)} \|h\|_{L^a} \|g\|_{L^b},$$

where

$$c(a, b) = \begin{cases} (1 - \frac{p^*}{a})d + \frac{n+d+2}{a} - \frac{n+d+2}{b}, & \text{if } a \leq b \\ (1 - \frac{p^*}{b})d + \frac{n+d+2}{b} - \frac{n+d+2}{a}, & \text{if } a > b \end{cases}$$

and

$$\beta(a, b) = \frac{n}{2a} + \frac{n}{2b} - \frac{n}{p}.$$

**Proof.** We begin by proving the lemma in the case $(a, b) = (\infty, \infty)$ and then in the two symmetric cases where $a = r$ and $r < b < p$, and where $b = r$ and $r < a < p$. The full range of estimates is then obtained by interpolating between these cases. Recall from above that we can assume $h$ and $g$ have time support in an interval of length $\sim 2^j$.

First consider $(a, b) = (\infty, \infty)$. Since $|\gamma| \geq 10$ in the definition of $T_j$ we can use kernel estimates as in Lemma 6.8 to get

$$|T_j(h, g)| \lesssim N^d \sum_{\alpha, \gamma} |\gamma|^{-n/2} \|h_\alpha\|_{L^1} \|g_{\alpha+\gamma}\|_{L^1}.$$  

Letting $\mu = \frac{n(p-2)}{2p}$ and using the fact that $|\gamma| \sim 2^j$, this implies

$$|T_j(h, g)| \lesssim N^d 2^{-j\frac{n}{p}} \sum_{\alpha, \gamma} |\gamma|^{-\mu} \|h_\alpha\|_{L^1} \|g_{\alpha+\gamma}\|_{L^1},$$

162
and then by the discrete Hardy-Littlewood-Sobolev inequality

$$|T_j(h, g)| \lesssim N^{d_2 j \frac{n}{p}} \|h\|_{\ell^q L^1} \|g\|_{\ell^q L^1}. \quad (6.25)$$

This proves the lemma when $(a, b) = (\infty, \infty)$.

Next, suppose that $a = p^*$ and $p^* < b < p$. By bringing the sums and time integrals into the inner product in $T_j$ and then applying Cauchy-Schwarz we get

$$|T_j(h, g)| \lesssim \sup_{I, I'} \left\| \int_{\mathbb{R}} e^{-it\Delta} (h(t) 1_I) dt \right\|_{L^2_{x,y}} \left\| \int_{\mathbb{R}} e^{-it\Delta} (g(t) 1_{I'}) dt \right\|_{L^2_{x,y}},$$

where $I, I'$ are time intervals of length $\sim 2^j$. Applying the dual form of Theorem 6.2 and using Hölder’s inequality in time (taking into account the fact that both functions have bounded time support) yields the following:

$$|T_j(h, g)| \lesssim N^{\alpha_b + \epsilon} 2^j \left( \frac{1}{q(p^*)} - \frac{1}{q} \right) 2^j \left( \frac{1}{q(b')} - \frac{1}{q} \right) \|h\|_{\ell^q L^1(p^*)} \|g\|_{\ell^q L^1(b')},$$

Here we are using the fact that if $s < p$ then $q(s) > q(p)$, hence $q(s)' < q(p)'$. Now

$$\alpha_b = \frac{n + d}{2} - \frac{n + d + 2}{b} = \frac{n + d + 2}{p^*} - \frac{n + d + 2}{b},$$

so this gives the desired power of $N$. Moreover, we have

$$\left( \frac{1}{q(p^*)} - \frac{1}{q} \right) + \left( \frac{1}{q(b')} - \frac{1}{q} \right) = \frac{2}{q} - \frac{1}{q(p^*)} - \frac{1}{q(b)}.$$

It is easy to check using the definition of $q(s)$ that the last expression simplifies to $\beta(p^*, b) = \beta(a, b)$, giving the desired power of $2^j$ as well. The proof of the symmetric case $b = p^*$ and $p^* < a < p$ is essentially the same. 

\[\square\]
The Second Interpolation

The fact that we have a two-parameter family of estimates in a neighborhood of \((\frac{1}{p}, \frac{1}{p})\) in Lemma 6.13 will allow us to sum in \(j\), but we first have to suitably decompose the input functions \(h, g\).

In particular we will use the atomic decomposition due to Keel and Tao.

**Lemma 6.14** ([80]). Let \((X, \mu)\) be a measure space and \(0 < p < \infty\). Then any \(f \in L^p(X)\) can be written as \(f = \sum_{k \in \mathbb{Z}} c_k \chi_k\) where each \(\chi_k\) is a function bounded by \(O(2^{-k/p})\) and supported on a set of measure \(O(2^k)\), and the \(c_k\) are non-negative constants such that \(\|c_k\|_{\ell^p} \lesssim \|f\|_{L^p}\).

We apply the lemma on \(L^{p'}(\mathbb{R}^n \times \Gamma \times [\gamma, \gamma + 1])\) for each \(\gamma\). This allows us to write

\[
h(x, y, t) 1_{[\gamma, \gamma + 1]}(t) = \sum_{k \in \mathbb{Z}} h_k^\gamma \chi_k(x, y, t) 1_{[\gamma, \gamma + 1]}(t)
\]

with \(\chi_k^\gamma\) supported on a set of measure \(O(2^k)\) and \(|\chi_k^\gamma| \lesssim 2^{-k/p'}\) for each \(k\), such that

\[
(\sum_k |h_k^\gamma|^{p'})^{1/p'} \lesssim \|h\|_{L^{p'}(\mathbb{R}^n \times \Gamma \times [\gamma, \gamma + 1])}
\]

and hence

\[
\left\| (\sum_k |h_k^\gamma|^{p'})^{1/p'} \right\|_{\ell^q} \lesssim \|h\|_{\ell^q L^{p'}(\mathbb{R}^n \times \Gamma \times [\gamma, \gamma + 1])}.
\]  

(6.26)

Likewise we can decompose

\[
g(x, y, t) 1_{[\gamma, \gamma + 1]}(t) = \sum_{m \in \mathbb{Z}} g_m^\gamma \varphi_m(x, y, t) 1_{[\gamma, \gamma + 1]}(t)
\]

with \(\varphi_m^\gamma\) supported on a set of measure \(O(2^m)\) and \(|\varphi_m^\gamma| \lesssim 2^{-m/p'}\) for each \(m\), such that

\[
\left\| (\sum_m |g_m^\gamma|^{p'})^{1/p'} \right\|_{\ell^q} \lesssim \|g\|_{\ell^q L^{p'}(\mathbb{R}^n \times \Gamma \times [\gamma, \gamma + 1])}.
\]  

(6.27)

Define \(h_k \chi_k\) such that \(h_k \chi_k = h_k^\gamma \chi_k^\gamma\) when \(t \in [\gamma, \gamma + 1]\), and similarly define \(g_m \varphi_m\). Then
we have
\[
\sum_{j \geq 3} |T_j(h, g)| \leq \sum_{j \geq 3} \sum_{k \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} |T_j(h_k \chi_k, g_m \varphi_m)|. 
\] (6.28)

Fix one such pair \((k, m)\). Then for all \((\frac{1}{a}, \frac{1}{b})\) in a neighborhood of \((\frac{1}{p}, \frac{1}{p})\) Lemma 6.13 implies
\[
|T_j(h_k \chi_k, g_m \varphi_m)| \lesssim \epsilon^{2^j(a,b) N c(a,b) + \epsilon \|\chi_k h_k\|_{\ell^q(a)} \|\varphi_m g_m\|_{\ell^q(b)}} \lesssim \epsilon^{2^j(a,b) N c(a,b) + \epsilon \|\chi_k h_k\|_{\ell^q(a)} \|\varphi_m g_m\|_{\ell^q(b)}} \lesssim \epsilon^{c(a,b) + \epsilon} \lesssim \epsilon \lesssim \epsilon^{2^j(a,b) N c(a,b) + \epsilon \|\chi_k h_k\|_{\ell^q(a)} \|\varphi_m g_m\|_{\ell^q(b)}} 
\] (6.29)

(with implicit constants depending on \(\epsilon, a, b, p\)). We now optimize in \(a\) and \(b\) to get
\[
2^{j \frac{2}{n} k - \frac{n}{p}} = 2^{-\eta j \frac{2}{n} k} 
\] (6.30)
\[
2^{j \frac{2}{n} m - \frac{n}{p}} = 2^{-\eta j \frac{2}{n} m} 
\] (6.31)
for some uniform \(\eta > 0\). We explain how to control the power of \(N\) appearing in (6.29) in the following lemma.

**Lemma 6.15.** *If \(p\) is close enough to \(p^*\) and \((a, b)\) is in a small ball around \((p, p)\) then*
\[
c(a, b) < 2 \alpha_p. 
\] (6.32)

*Moreover, the size of the ball only depends on \(n, d\).*

**Proof.** Suppose without loss of generality that \(a \leq b\). Then \(c(a, b) < 2 \alpha_p\) is equivalent to the inequality
\[
(1 - \frac{p^*}{a})d + \frac{n + d + 2}{a} - \frac{n + d + 2}{b} + \frac{2(n + d + 2)}{p} < n + d. 
\] (6.33)

Note that \(\frac{1}{p} < \frac{1}{p^*}\), hence
\[
\frac{2(n + d + 2)}{p} < \frac{2(n + d + 2)}{p^*} = n + d. 
\]
The lemma will then follow if we can choose \( a, b \) such that

\[
(1 - \frac{p^*}{a})d + \frac{n + d + 2}{a} - \frac{n + d + 2}{b} < n + d - \frac{2(n + d + 2)}{p}.
\]

(6.34)

This is possible for any \( a, b \) close enough to \( p \), provided \( p \) has been chosen sufficiently close to \( p^* \) (which we are always free to assume). In particular, if we set \( a = b = p^* \) then the left-hand side of (6.34) is simply 0. But by continuity the strict inequality (6.34) is preserved if we vary \( a, b \) in a small neighborhood around \( p \), as long as \( p \) is close enough to \( p^* \). This proves (6.33) for \((a, b)\) in some small neighborhood of \((p, p)\), and the size of the neighborhood clearly only depends on \( p^* \) and hence only on \( n, d \).

By combining Lemma 6.15 with (6.30) and (6.31),

\[
|T_j(h_k\chi_k, g_m\varphi_m)| \lesssim N^{2\alpha_p} 2^{-\eta|j| - \frac{2}{n}|k|} 2^{-\eta|j| - \frac{2}{n}|m|} \|h_k\|_{\ell_q'} \|g_m\|_{\ell_q'}.
\]

(6.35)

Indeed, by Lemma 6.15 we can choose \( \epsilon \) small enough (depending only on \( p, n, d \) and the choice of \( a, b \)) such that \( c(a, b) + \epsilon \leq 2\alpha(p) \).

Summing in \( j \) then yields

\[
\sum_k \sum_m \sum_{j \geq 3} |T_j(h_k\chi_k, g_m\varphi_m)| \lesssim N^{2\alpha_p} \sum_k \sum_m \sum_{j \geq 3} 2^{-\eta|j| - \frac{2}{n}|k|} 2^{-\eta|j| - \frac{2}{n}|m|} \|h_k\|_{\ell_q'} \|g_m\|_{\ell_q'}
\]

\[
\lesssim N^{2\alpha_p} \sum_k \sum_m (1 + \frac{2}{n}|k - m|) 2^{-\frac{2}{n}\eta|k - m|} \|h_k\|_{\ell_q'} \|g_m\|_{\ell_q'}.
\]

The right-hand side of this bound is of the form

\[
\sum_{m, k} f(m - k)(\sum_{\gamma} |h_k^\gamma|_{q'}^{1/q'})^{1/q'} (\sum_{\gamma} |g_m^\gamma|_{q'}^{1/q'})^{1/q'} := \sum_{m, k} f(m - k)c_k d_m
\]

where \( f(m) \) is summable in \( m \). By Hölder’s and Young’s inequality we can control this by
Now $p' > q'$ so by Minkowski’s inequality

\[
\|c\|_{\ell^{q'}} = \left( \sum_{k} \left( \sum_{\gamma} |h_{k}^{\gamma}|^{q'} \right)^{\frac{q'}{p'}} \right)^{\frac{1}{q'}} \leq \left( \sum_{k} \left( \sum_{\gamma} |h_{k}^{\gamma}|^{p'} \right)^{\frac{p'}{q'}} \right)^{\frac{1}{q'}}
\]

which we can control by $\|h\|_{\ell^{q'}L^{p'}}$ using (6.26). Also note $\|d\|_{\ell^{q'}} \leq \|d\|_{\ell^{q'}}$ since $p > p'$, so the same argument (using (6.27)) works to give an acceptable bound for this term as well. We ultimately get

\[
\sum_{k} \sum_{m} \sum_{j \geq 3} |T_{j}(h_{k} \chi_{k}, g_{m} \varphi_{m})| \lesssim N^{2\alpha_{r}} \|h\|_{\ell^{q'}L^{p'}} \|g\|_{\ell^{q'}L^{p'}}
\]
as desired.

### 6.6 Some Applications

In this section we show how Theorem 6.2 can be used to prove Theorem 6.3. Below all mixed norms of the type $\ell^{q}L^{p}$ are defined as in Theorem 6.2.

#### 6.6.1 Function Spaces

We will employ the atomic and variational spaces that have frequently been used to study well-posedness problems for dispersive equations (for examples see [74], [77], [75], and [81]). We recall some basic definitions and properties and refer the reader to [74] for proofs.

Let $I \subset \mathbb{R}$ be a time interval. Given $1 \leq p < \infty$ and a Hilbert space $H$, a $U^{p}(I, H)$ atom $a$ is defined to be a function $a : I \rightarrow H$ such that

\[
a(t) = \sum_{k=1}^{K} 1_{(t_{k-1}, t_{k})} \phi_{k-1}, \quad \phi_{k} \in H
\]

for some partition $-\infty < t_{0} < t_{1} < \ldots < t_{K} \leq \infty$, with the additional property that

\[
\sum_{k=0}^{K-1} \|\phi_{k}\|_{H}^{p} = 1.
\]

Then $U^{p}(I, H)$ is the Banach space of functions $u : I \rightarrow H$ with a de-
composition of the form

\[ u = \sum_j \lambda_j a_j, \]

where \( \{\lambda_j\} \in \ell^1(\mathbb{C}) \) and \( a_j \) are \( U^p(I, H) \) atoms. The norm on \( U^p \) is taken to be

\[ \|u\|_{U^p} = \inf \{ \sum_j |\lambda_j| : u = \sum_j \lambda_j a_j \text{ with } a_j \text{ } U^p\text{-atoms} \}. \]

We also define the variational space \( V^p(I, H) \) to be the Banach space of functions \( v \) such that

\[ \|v\|_{V^p} := \sup_{\text{partitions } \{t_k\}} \left( \sum_{k=1}^K \|v(t_k) - v(t_{k-1})\|_H^p \right)^{\frac{1}{p}} < \infty. \]

We have the important duality relationship

\[ (U^p)^* = V^p'. \]

We introduce two further spaces \( X^* \) and \( Y^* \) which we will use to carry out the main iteration argument. First let \( U^p_\Delta \) denote the space of functions such that \( e^{-it\Delta}u \in U^p \), and similarly define \( V^p_\Delta \). Let \( \{C_z\}_{z \in \mathbb{Z}^{n+d}} \) be a tiling of frequency space by cubes of side-length \( \sim 1 \), and define \( X^*_0(\mathbb{R}) \) and \( Y^*(\mathbb{R}) \) to be the spaces of functions \( u \) and \( v \), respectively, such that the following norms are finite:

\[ \|u\|_{X^*_0(\mathbb{R})}^2 := \sum_{z \in \mathbb{Z}^{n+d}} \langle z \rangle^{2s} \|P_{C_z}u\|_{U^p_\Delta(\mathbb{R}, L^2(\mathbb{R}^n \times \mathbb{T}^d))}^2 \]

\[ \|v\|_{Y^*(\mathbb{R})}^2 := \sum_{z \in \mathbb{Z}^{n+d}} \langle z \rangle^{2s} \|P_{C_z}v\|_{V^p_\Delta(\mathbb{R}, L^2(\mathbb{R}^n \times \mathbb{T}^d))}^2. \]

One can likewise define the time-restriction norms \( X^*_0(I), Y^*(I) \) for \( I \subset \mathbb{R} \) (see [75]). Recall that we have the sequence of embeddings

\[ U^p_\Delta(\mathbb{R}, H^s) \hookrightarrow X^*_0(\mathbb{R}) \hookrightarrow Y^*(\mathbb{R}) \hookrightarrow V^p_\Delta(\mathbb{R}, H^s) \hookrightarrow U^p(\mathbb{R}, H^s) \hookrightarrow L^\infty(\mathbb{R}, H^s) \]

168
for any $p > 2$. Finally, since we wish to study scattering we introduce the modified space $X^s$ as in [75] defined to be

$$X^s(\mathbb{R}) := \{ u : \phi_{-\infty} := \lim_{t \to -\infty} e^{-it\Delta}u(t) \text{ exists in } H^s, \text{ and } u(t) - e^{it\Delta}\phi_{-\infty} \in X^s_0(\mathbb{R}) \},$$

with the norm

$$\|u\|_{X^s(\mathbb{R})}^2 := \|\phi_{-\infty}\|_{H^s(\mathbb{R}^n \times \mathbb{T}^d)}^2 + \|u - e^{it\Delta}\phi_{-\infty}\|_{X^s_0(\mathbb{R})}^2.$$

One can also define time-restriction spaces $X^s(I)$ similarly to $X^s_0(I)$. Note that it is not sufficient to work with $X^s_0(\mathbb{R})$ if we wish to study scattering as $t \to -\infty$ since if $u \in X^s_0$ then $\lim_{t \to -\infty} u(t) = 0$ in norm. \footnote{Indeed, given $\epsilon > 0$ one can use the atomic decomposition to decompose $u = u_1 + u_2$ where $u_1(t) = 0$ for $t < T$, and $\|u_2(t)\|_{H^s} \leq \epsilon$; see Proposition 2.2 in [74].}

### 6.6.2 The Quintic Equation on $\mathbb{R} \times \mathbb{T}$

As a first step we can transfer our Strichartz estimates to a result about $U^p_{\Delta}$ spaces.

**Lemma 6.16.** Let $q, p$ be as in Theorem 6.2. Let $C$ be a cube in frequency space $\mathbb{R}^n \times \mathbb{Z}^d$ with side-length $\sim N \geq 1$ and let $I \subset \mathbb{R}$ be a time interval. Then

$$\| 1_I \cdot P_C u \|_{L^q(I; L^p(\mathbb{R}^n \times \mathbb{T}^d \times [\gamma, \gamma + 1]))} \lesssim N^{\frac{n+4d}{2} - \frac{n+d+2}{p}} \| P_C u \|_{U^{\min(p,q)}(I; L^2(\mathbb{R}^n \times \mathbb{T}^d))}.$$

The lemma is a direct consequence of the atomic decomposition of $e^{-it\Delta}P_C u$.

We begin with the case of the quintic equation on $\mathbb{R} \times \mathbb{T}$ with initial data in $H^{\frac{1}{2}}$. (recall this equation is $H^{\frac{1}{2}}$-critical). Let $F(u) = |u|^4 u$. We will apply an iteration argument to the Duhamel operator

$$u_0 + \int_0^t e^{i(t-s)\Delta} F(u)(s) ds. \tag{6.36}$$

The presentation here is similar to [81], although we will be able to prove more in this semiperiodic setting since we have global-in-time estimates.
A first step is the following multilinear estimate, which we will see is a corollary of Theorem 6.2 and the basic properties of the function spaces outlined above.

**Lemma 6.17.** Let $I \subset \mathbb{R}$ be a time interval and suppose $u^{(j)} \in X^{\frac{1}{2}}(I)$ and $v \in Y^{-\frac{1}{2}}(I)$. Then

\[
\left| \int_I \int_{\mathbb{R} \times T} v \cdot \prod_{j=1}^5 u^{(j)} dx dy dt \right| \lesssim \|v\|_{Y^{-\frac{1}{2}}} \prod_{j=1}^5 \|u^{(j)}\|_{X^{\frac{1}{2}}(I)}. \tag{6.37}
\]

**Proof.** The proof of this lemma follows the same basic approach as in [77] (see also [81], [75], and [74]). Below all norms are taken with respect to the time interval $I$ (note we can have $I = \mathbb{R}$). Since we have embeddings $X^{\frac{1}{2}} \hookrightarrow Y^{-\frac{1}{2}}$ it suffices to prove (6.37) with the $X^{\frac{1}{2}}$ norms replaced by $Y^{-\frac{1}{2}}$ norms.

We will use the short-hand $u_N = P_{\leq N} u$. By performing Littlewood-Paley decompositions on all six functions and exploiting symmetry of the resulting expression, we reduce matters to showing that

\[
\sum_{N_0 \geq 1} \sum_{N_1 \geq N_2 \geq \ldots \geq N_5} \left| \int_I \int_{\mathbb{R} \times T} v_{N_0} \cdot \prod_{j=1}^5 u^{(j)}_{N_j} dx dy dt \right| \lesssim \|v\|_{Y^{-\frac{1}{2}}} \prod_{j=1}^5 \|u^{(j)}\|_{Y^{\frac{1}{2}}(I)}, \tag{6.38}
\]

where $N_j \geq 1$ are dyadic integers. Moreover, by Plancherel’s theorem we know that for fixed $N_0, \ldots, N_5$ the corresponding term in (6.38) vanishes unless the two largest frequencies are comparable. Hence we have two cases to consider, when $N_0 \sim N_1 \geq N_2 \geq \ldots \geq N_5$ (**Case I**), and when $N_0 \lesssim N_1 \sim N_2 \geq \ldots \geq N_5$ (**Case II**).

**Case I.** We fix $N_0, N_1$ with $N_0 \sim N_1$ and decompose the support of $P_{\leq N_0}$ and $P_{\leq N_1}$ into subcubes $\{C_j\}$ of side-lengths $N_2$. We say $C_j \sim C_k$ if the sumset $C_j + C_k$ overlaps the Fourier support of $P_{\leq 2N_2}$. After decomposing $v_{N_0}$ and $u^{(1)}_{N_1}$ with respect to the $C_j$'s, we see by Plancherel's theorem that it suffices to bound

\[
\sum_{C_j \sim C_k} \sum_{N_2 \geq N_3 \geq N_4 \geq N_5} \left| \int_I \int_{\mathbb{R} \times T} P_{C_j} v_{N_0} P_{C_k} u^{(1)}_{N_1} \cdot \prod_{j=2}^5 u^{(j)}_{N_j} dx dy dt \right| \tag{6.39}
\]

by a factor that will be summable over $N_0$ and $N_1$ (in particular, if $C_j$ is not comparable to $C_k$.
then the integral is 0). We turn to this task now.

We apply the mixed-norm Hölder’s inequality with $L^p$ exponents corresponding to

$$
\frac{4}{18} + \frac{4}{18} + \frac{4}{18} + \frac{4}{18} + \frac{1}{18} + \frac{1}{18} = 1
$$

and $\ell^q$ exponents corresponding to

$$
\frac{5}{36} + \frac{5}{36} + \frac{5}{36} + \frac{2}{9} + \frac{2}{9} = 1
$$

to obtain

$$
\left(6.39\right) \leq \sum_{C_i \sim C_k \text{ } N_2 \geq N_3 \geq N_4 \geq N_5} \frac{N_4^{\frac{5}{36}} N_5^{\frac{5}{18}}}{N_2^{\frac{5}{36}} N_3^{\frac{5}{18}}} \|P_{C_j} v_{N_0}\|_{\ell^\frac{36}{5} L^\infty} \|P_{C_k} u_{N_1}^{(1)}\|_{\ell^\frac{36}{5} L^\infty} \|u_{N_2}^{(2)}\|_{\ell^\frac{36}{5} L^\infty} \|u_{N_3}\|_{\ell^\frac{36}{5} L^\infty} \cdot \|u_{N_4}\|_{\ell^\frac{9}{7} L^{18}} \|u_{N_5}\|_{\ell^\frac{9}{7} L^{18}},
$$

where the norms are all localized in time to $I$. By applying Lemma 6.16 and the embedding $Y^0 \hookrightarrow U^r$ for $r > 2$ we ultimately obtain

$$
\left(6.39\right) \lesssim \sum_{C_i \sim C_k \text{ } N_2 \geq N_3 \geq N_4 \geq N_5} \frac{N_4^{\frac{5}{36}} N_5^{\frac{5}{18}}}{N_2^{\frac{5}{36}} N_3^{\frac{5}{18}}} \|P_{C_j} v_{N_0}\|_{Y^{-\frac{1}{2}}} \|P_{C_k} u_{N_1}^{(1)}\|_{Y^{\frac{1}{2}}} \|u_{N_2}^{(2)}\|_{Y^{\frac{1}{2}}} \|u_{N_3}\|_{Y^{\frac{1}{2}}} \cdot \|u_{N_4}\|_{Y^{\frac{1}{2}}} \|u_{N_5}\|_{Y^{\frac{1}{2}}}.
$$

Now by using Cauchy-Schwarz (or Schur’s test) to sum, we conclude that

$$
\left(6.39\right) \lesssim \sum_{C_i \sim C_k} \|P_{C_j} v_{N_0}\|_{Y^{-\frac{1}{2}}} \|P_{C_k} u_{N_1}^{(1)}\|_{Y^{\frac{1}{2}}} \|u_{N_2}^{(2)}\|_{Y^{\frac{1}{2}}} \|u_{N_3}\|_{Y^{\frac{1}{2}}} \|u_{N_4}\|_{Y^{\frac{1}{2}}} \|u_{N_5}\|_{Y^{\frac{1}{2}}} \cdot \|u_{N_4}\|_{Y^{\frac{1}{2}}} \|u_{N_5}\|_{Y^{\frac{1}{2}}}.
$$

\begin{align*}
&\lesssim \|v_{N_0}\|_{Y^{-\frac{1}{2}}} \|u_{N_1}^{(1)}\|_{Y^{\frac{1}{2}}} \|u_{N_2}^{(2)}\|_{Y^{\frac{1}{2}}} \|u_{N_3}\|_{Y^{\frac{1}{2}}} \|u_{N_4}\|_{Y^{\frac{1}{2}}} \|u_{N_5}\|_{Y^{\frac{1}{2}}} \cdot \left(\prod_{j=2}^{5} \|u_{N_j}\|_{Y^{\frac{1}{2}}}\right),
\end{align*}

171
and therefore

\[
(6.38) \lesssim \sum_{N_0 \sim N_1} \|v_{N_0}\|_{Y^{-\frac{1}{2}}} \|u_{N_1}^{(1)}\|_{Y^{\frac{1}{2}}} \prod_{j=2}^{5} \|u^{(j)}\|_{Y^{\frac{1}{2}}} \\
\lesssim \|v\|_{Y^{-\frac{1}{2}}} \prod_{j=1}^{5} \|u^{(j)}\|_{Y^{\frac{1}{2}}},
\]

as desired.

**Case II.** In this case we do not decompose into subcubes, and instead directly apply the mixed-norm Hölder inequality with the same exponents as in Case I and then Lemma 6.16. This leads to a bound of the form

\[
(6.38) \lesssim \sum_{N_1} \sum_{N_0 \sim N_1} \sum_{N_2 \geq N_3 \geq N_4 \geq N_5} \frac{N_0^{\frac{11}{18}} N_1^{\frac{5}{18}} N_2^{\frac{5}{18}}}{N_1^{\frac{5}{18}} N_2^{\frac{1}{18}} N_3^{\frac{1}{18}}} \|v_{N_0}\|_{Y^{-\frac{1}{2}}} \|u_{N_1}^{(1)}\|_{Y^{\frac{1}{2}}} \|u_{N_2}^{(2)}\|_{Y^{\frac{1}{2}}} \|u_{N_3}^{(3)}\|_{Y^{\frac{1}{2}}} \\
\quad \cdot \|u_{N_4}^{(4)}\|_{Y^{\frac{1}{2}}} \|u_{N_5}^{(5)}\|_{Y^{\frac{1}{2}}},
\]

Then repeatedly using Cauchy-Schwarz or Schur’s test as before to sum, we get

\[
(6.38) \lesssim \sum_{N_1 \sim N_2} \frac{N_1^{\frac{5}{18}}}{N_2^{\frac{1}{18}}} \|v\|_{Y^{-\frac{1}{2}}} \|u_{N_1}^{(1)}\|_{Y^{\frac{1}{2}}} \|u_{N_2}^{(2)}\|_{Y^{\frac{1}{2}}} \prod_{j=3}^{5} \|u^{(j)}\|_{Y^{\frac{1}{2}}} \\
\lesssim \|v\|_{Y^{-\frac{1}{2}}} \prod_{j=1}^{5} \|u^{(j)}\|_{Y^{\frac{1}{2}}},
\]

completing the proof.

Below we will write

\[
\mathcal{I}(u) = \int_0^t e^{i(t-s)\Delta} F(u(s))ds.
\]
Proposition 6.18. For any time interval $I \subset \mathbb{R}$ we have

$$
\|I(u)\|_{\dot{X}^{\frac{1}{2}}(I)} \lesssim \|u\|_{\dot{X}^{\frac{5}{2}}(I)}^{5}
$$

(6.40)

and

$$
\|I(u + w) - I(u)\|_{\dot{X}^{\frac{1}{2}}(I)} \lesssim \|w\|_{\dot{X}^{\frac{1}{2}}(I)} (\|u\|_{\dot{X}^{\frac{1}{2}}(I)} + \|w\|_{\dot{X}^{\frac{1}{2}}(I)})^{4}.
$$

(6.41)

Proof. By duality

$$
\|I(u)\|_{\dot{X}^{\frac{1}{2}}(I)} \leq \sup_{v \in \dot{Y}^{-\frac{1}{2}}(I), \|v\|_{\dot{Y}^{-\frac{1}{2}}(I)} = 1} \left| \int_I \int_{\mathbb{R} \times T} v \cdot F(u) dx dy dt \right|
$$

for any $u \in X^{\frac{1}{2}}$ (the proof is similar to the proof of Proposition 2.11 in [77], see also [75]). We prove the second part of the proposition, since the first is then a simple consequence.

It suffices to prove that

$$
\left| \int_I \int_{\mathbb{R} \times T} v \cdot (|u + w|^{4}(u + w) - |u|^{4}u) dx dy dt \right| \lesssim \|v\|_{\dot{Y}^{-\frac{1}{2}}} \|w\|_{\dot{X}^{\frac{1}{2}}} (\|u\|_{\dot{X}^{\frac{1}{2}}} + \|w\|_{\dot{X}^{\frac{1}{2}}})^{4}
$$

for a fixed $v$ with $\|v\|_{\dot{Y}^{-\frac{1}{2}}} = 1$. By expanding the expression inside the integral we see that this estimate is an immediate consequence of Lemma 6.17 with $u^{(j)} \in \{\pm u, \pm w, \pm \bar{u}, \pm \bar{w}\}$. This completes the proof.

\[\square\]

We can now prove Theorem 6.3 using a standard iteration argument. As above we fix a time interval $I \subset \mathbb{R}$ containing 0 (in particular we could have $I = \mathbb{R}$).

The Global Case for Small Data

Suppose one has small initial data $u_0 \in H^{\frac{1}{2}}(\mathbb{R} \times \mathbb{T})$ with

$$
\|u_0\|_{H^{\frac{1}{2}}} \leq \eta < \eta_0,
$$

173
with $\eta_0$ some fixed parameter to be determined. We apply a contraction mapping argument to the operator

$$\Phi(u)(t) := e^{it\Delta}u_0 \pm i\mathcal{I}(u)(t)$$

on the ball

$$B := \{ u \in X^{1/2}(I) \cap C_t H^{1/2}_{\mp y}(I, \mathbb{R} \times \mathbb{T}) : \| u \|_{X^{1/2}(I)} \leq 4\eta \}$$

with respect to the metric $d(u, v) = \| u - v \|_{X^{1/2}(I)}$. By Proposition 6.18 we have

$$\| \Phi(u) \|_{X^{1/2}(I)} \leq \| e^{it\Delta}u_0 \|_{X^{1/2}(I)} + \| \mathcal{I}(u) \|_{X^{1/2}(I)}$$

$$\leq 3\| u_0 \|_{H^{1/2}} + C\| u \|^5_{X^{1/2}(I)}$$

$$\leq 3\eta + C(4\eta)^5 \leq 4\eta$$

provided $\eta_0$ is chosen small enough. Therefore $\Phi$ maps $B$ into itself. Now the second part of Proposition 6.18 implies that

$$d(\Phi(u), \Phi(v)) \lesssim \| u - v \|_{X^{1/2}(I)} (\| u \|_{X^{1/2}(I)} + \| v \|_{X^{1/2}(I)})^4$$

$$\lesssim d(u, v)(8\eta)^4$$

$$\lesssim \frac{1}{2} d(u, v),$$

again provided $\eta_0$ is chosen sufficiently small. Now apply the contraction mapping theorem on $B$ to obtain a solution. By taking $I = \mathbb{R}$ we see in particular that for small initial data in $H^{1/2}(\mathbb{R} \times \mathbb{T})$ one has global-in-time solutions to the quintic equation. We can additionally show that such solutions scatter in $H^{1/2}(\mathbb{R} \times \mathbb{T})$ by appealing to the fact that if $G \in X^{1/2}(\mathbb{R})$ then $\lim_{t \to \pm \infty} G(t)$ exists in $H^{1/2}$ (for the case $t \to \infty$ see Proposition 2.2 in [74]). Indeed, to prove scattering it suffices to show that if $u$ is a solution to the quintic equation then the Duhamel integral $\mathcal{I}(u)(t)$ is conditionally convergent in $H^{1/2}$ as $t \to \pm \infty$.

We can also prove scattering directly from the argument above with a small amount of
additional work. We record some details since these estimates are also useful in the local theory for large data. To facilitate the argument we introduce the following time-divisible norm as in [75]:

\[ \|u\|_{Z(I)} = \sum_{p=9/2, 18} \left( \sum_{N \geq 1} N^{p(\frac{4}{p} - \frac{1}{2})} \|1_{I} P_{N} u\|_{\ell^{p}(p) L^{p}}^{\frac{1}{p}} \right), \]

where \( q(p) = \frac{4p}{p-2} \) as above. We also let

\[ \|u\|_{Z'(I)} = \|u\|_{Z(I)}^{\frac{3}{4}} \|u\|_{X^{\frac{1}{2}}(I)}^{\frac{1}{4}}. \]

Note that \( \|u\|_{Z(I)} \lesssim \|u\|_{X^{\frac{1}{2}}(I)} \) by Lemma 6.16. By repeating the argument given in the proofs of Lemma 6.17 and Proposition 6.18 we obtain the stronger results

\[ \|\mathcal{I}(u)\|_{X^{\frac{1}{2}}(I)} \lesssim \|u\|_{X^{\frac{1}{2}}(I)}^{\frac{3}{4}} \|u\|_{Z'(I)} \] (6.42)

and

\[ \|\mathcal{I}(u + w) - \mathcal{I}(u)\|_{X^{\frac{1}{2}}(I)} \lesssim \|w\|_{X^{\frac{1}{2}}(I)} \left( \|u\|_{X^{\frac{1}{2}}(I)} + \|w\|_{X^{\frac{1}{2}}(I)} \right) \left( \|u\|_{Z(I)} + \|w\|_{Z(I)} \right)^{3}. \]

Now if \( u \) is a solution in \( B \) with \( \|u\|_{H^{\frac{1}{2}}} \leq \eta_{0} \) then the norm \( \|u\|_{Z'(I)}^{5} \) is uniformly bounded independent of \( I \). Given \( \epsilon > 0 \), we can therefore find a time \( T_{\epsilon} \) such that if \( r, t > T_{\epsilon} \) then

\[ \|\mathcal{I}(1_{[r, t]} u)\|_{X^{\frac{1}{2}}(\mathbb{R})} < \epsilon. \]

But this implies that \( \mathcal{I}(u)(t) = \int_{0}^{t} e^{-is\Delta} F(u)(s) ds \) is Cauchy in \( H^{\frac{1}{2}} \), and hence converges in norm as \( t \to \infty. \)
The Local Case for Large Data

Suppose $\|u_0\|_{H^{\frac{1}{2}}} \leq A$ and fix $\delta \geq 0$ to be determined. We apply a contraction mapping argument to the operator $\Phi$ on

$$B' := \{ u \in X^{1/2}(I) \cap C_t H^{1/2}_{x,y} (I, \mathbb{R} \times \mathbb{T}) : \|u\|_{X^{1/2}(I)} \leq 2A \text{ and } \|u\|_{Z(I)} \leq 2\delta \}.$$  

Now if $\|e^{it\Delta} u_0\|_{Z(I)} \leq \delta$ then by (6.42) we see that $\Phi : B' \to B'$ if $A\delta$ is small enough, and likewise

$$\|\Phi(u) - \Phi(v)\|_{X^{1/2}(I)} \leq \frac{1}{2} \|u - v\|_{X^{1/2}(I)}$$

if $A\delta$ is small enough. But for any $u_0 \in H^{1/2}$ we can find a time interval $I$ containing 0 such that $\|e^{it\Delta} u_0\|_{Z(I)} \leq \delta$, and this completes the argument.

6.6.3 The Cubic Equation on $\mathbb{R}^2 \times \mathbb{T}$

The skeleton of the argument in this setting is essentially the same as what we saw in the last section, so we only provide a brief sketch of the details. The key point here is that the $L^q(p)$ exponent in dimension $(n, d) = (2, 1)$ is $q(p) = \frac{4p}{2(p-2)} = \frac{2p}{p-2}$. It follows that if $\sum_{i=0}^{3} \frac{1}{p_i} = 1$ then $\sum_{i=0}^{3} \frac{1}{q(p_i)} = 1$, and therefore one has the quadrilinear estimate

$$\left| \int_I \int_{\mathbb{R}^2 \times \mathbb{T}} v \cdot \prod_{i=1}^{3} u^{(i)} dx dy dt \right| \lesssim \|v\|_{L^q(p_0)} \prod_{i=1}^{3} \|u^{(i)}\|_{L^q(p_i)}.$$  \hspace{1cm} (6.43)

Note that (6.43) will not hold in general for the admissible exponents $r, q(r)$ on $\mathbb{R} \times \mathbb{T}^2$, since with these dimensions $q(r) = \frac{4r}{r-2}$.

The main nonlinear estimate is the following.

**Lemma 6.19.** Let $I \subset \mathbb{R}$ be a time interval and suppose $u^{(j)} \in X^{1/2}(I)$ and $v \in Y^{-1/2}(I)$. Then

$$\left| \int_I \int_{\mathbb{R}^2 \times \mathbb{T}} v \cdot \prod_{j=1}^{3} u^{(j)} dx dy dt \right| \lesssim \|v\|_{Y^{-1/2}} \prod_{j=1}^{3} \|u^{(j)}\|_{X^{1/2}(I)}.$$  \hspace{1cm} (6.44)
Proof. We use the same Littlewood-Paley decomposition argument as the proof of Lemma 6.17 with two cases corresponding to $N_0 \sim N_1 \geq N_2 \geq N_3$ (Case I), and $N_0 \lesssim N_1 \sim N_2 \geq N_3$ (Case II). In both cases we apply (6.43) with exponents $p_0, p_1, p_2 = \frac{7}{2}$ and $p_3 = 7$.

**Case I.** As in the proof of Lemma 6.17 we decompose the support of $P_{N_0}$ and $P_{N_1}$ into subcubes $C_j$ of side length $N_2$. It then suffices to estimate

$$\sum_{N_0 \sim N_1 \geq N_2 \geq N_3} \sum_{C_j \sim C_k} \left| \int_I \int_{\mathbb{R}^2 \times \mathbb{T}} P_{C_j} v_{N_0} P_{C_j} u_{N_1}^{(1)} \cdot u_{N_2}^{(2)} u_{N_3}^{(3)} \, dx \, dy \, dt \right|. \tag{6.45}$$

By (6.43) and then Lemma 6.16 we have

$$\tag{6.45} \lesssim \sum_{N_0 \sim N_1 \geq N_2 \geq N_3} \sum_{C_j \sim C_k} \left( \frac{N_3}{N_2} \right)^{\frac{5}{2}} \| P_{C_k} v_{N_0} \|_{L^2} \| P_{C_j} u_{N_1}^{(1)} \|_{L^4} \| P_{C_j} u_{N_1}^{(2)} \|_{L^4} \| u_{N_2}^{(2)} \|_{L^2} \| u_{N_3}^{(3)} \|_{L^1}$$

which is bounded above by the desired quantity by Cauchy-Schwarz or Schur’s test. Here all spacetimes norms are taken relative to $I \times \mathbb{R}^2 \times \mathbb{T}$.

**Case II.** In this case we estimate

$$\sum_{N_0 \lesssim N_1 \sim N_2 \geq N_3} \left| \int_I \int_{\mathbb{R}^2 \times \mathbb{T}} v_{N_0} u_{N_1}^{(1)} u_{N_2}^{(2)} u_{N_3}^{(3)} \, dx \, dy \, dt \right| \tag{6.46}$$

by applying (6.43) and Lemma 6.16 with the same exponents as in Case I. We get

$$\tag{6.46} \lesssim \sum_{N_0 \lesssim N_1 \sim N_2 \geq N_3} \left( \frac{N_3}{N_2} \right)^{\frac{5}{2}} \frac{N_1^{\frac{3}{2}}}{N_1^2 N_2^2} \| v_{N_0} \|_{L^2} \| u_{N_1}^{(1)} \|_{L^4} \| u_{N_2}^{(2)} \|_{L^4} \| u_{N_3}^{(3)} \|_{L^1}$$

which is bounded by the desired quantity (again by Cauchy-Schwarz or Schur’s test).

\[ \square \]

The rest of the cubic case of Theorem 6.3 can now be proved by routine modifications of arguments from the last section, using Lemma 6.19 in place of Lemma 6.17.

177
6.7 Additional Remarks

Remark 6.7.1. Theorem 6.2 remains true with a loss of $N^\epsilon$ if the operator $e^{it\Delta}$ is replaced by

$$e^{it\phi(D)}f := \sum_m \int \hat{f}_m(\xi)e^{2\pi i (x\xi + y\cdot m + t\phi(\xi, m))}d\xi,$$

where $\phi(\xi, \eta)$ is a $C^3$ function on $\mathbb{R}^{n+d}$ such that $D^2\phi$ is uniformly positive-definite on $\mathbb{R}^{n+d}$ and

$$|\partial_I \phi(\xi, \eta)| \lesssim \frac{1}{1 + (|\xi|^2 + |\eta|^2)^{\frac{I}{2}}}$$

for any triple index $I$. Let $\rho = (\xi, \eta)$ and define the rescaled function $\phi_1(\rho) = N^{-2}\phi(N\xi, N\eta)$. By Taylor expansion of the phase we see that if $\tau \subset B_1$ has radius $\sim N^{-\frac{3}{4}}$ then

$$\int_\tau \hat{g}(\rho)e^{2\pi i (z\cdot \rho + t\phi_1(\rho))}d\rho$$

has space-time Fourier support in an $O(N^{-2})$-neighborhood of the paraboloid in $\mathbb{R}^{n+d+1}$. One then obtains a decoupling result for $e^{it\phi(D)}$ by following the iteration scheme outlined in Section 7 of [71], and then completes the argument by following the same steps from Section 3.

It is not clear if one can extend the $\epsilon$-removal argument from Section 4 to operators with these types of phases. The main issue is with the local Strichartz estimate in this context. The argument of Killip and Visan relies on some subtle number-theoretic properties of the kernel associated to $e^{it\Delta}$, and it is not immediately clear that the same argument will work even if $e^{it\phi(D)}$ is a small perturbation of $e^{it\Delta}$.

Remark 6.7.2. We do not know if the loss of $N^\epsilon$ is necessary in the endpoint case $p = \frac{2(n+d+2)}{n+d}$ and $q = \frac{2(n+d+2)}{n}$ of Theorem 6.2 for the case when $n + d > 2$. In the case $d = 0$ the theorem is true without any loss (this is just the Stein-Tomas theorem), while in the case $n = 0$ some loss in $N$ is necessary (as shown by Bourgain [70]).

A simple argument using Sobolev embedding and the Strichartz estimate from $\mathbb{R}^n$ shows
that Theorem 6.2 holds for $p = \frac{2(n+2)}{n}$ with no loss of $N^\epsilon$ (note that in this case $q = p$). We also know that at least in the low-dimensional case $n = d = 1$ the local Strichartz estimate holds with no loss in $N$ for $p = 4$, as shown by Takaoka and Tzvetkov [83]. However, as far as we know there are no higher-dimensional analogues of the Takaoka-Tzvetkov result beyond what is present in this paper.

It is possible to prove the endpoint case of Theorem 6.2 when $n = d = 1$ with no loss of derivatives (that is, with $s = 0$) by inserting Gaussians adapted to the time intervals $[\gamma, \gamma + 1]$ and then expanding out the $L^4$ and $\ell^8$ norms and computing Fourier transforms [73]. One then obtains a uniform bound on the resulting kernel. This argument heavily relies on the fact that $p = 4$ and $q = 8$ are even exponents, which no longer holds as soon as $n + d > 2$. It is not clear if the endpoint case of the local or global estimate is true without any loss of derivatives in higher dimensions; we at least expect Theorem 6.2 to be true at the endpoint with $s = 0$ when $n > d$, although this may be hard to prove.
Bibliography


[50] Lacey, M.T. Personal Communication.


[73] Christ, M. Pausader, B. Personal communication.


[75] Hani, Z., Pausader, B. On scattering for the quintic defocusing nonlinear Schrödinger equation on $\mathbb{R} \times \mathbb{T}^2$. Com. on Pure and Applied Math. 67 (9), 1466-1542


