

## MAXIMAL MULTILINEAR OPERATORS

CIPRIAN DEMETER, TERENCE TAO, AND CHRISTOPH THIELE

ABSTRACT. We establish multilinear  $L^p$  bounds for a class of maximal multilinear averages of functions of one variable, reproving and generalizing the bilinear maximal function bounds of Lacey (2000). As an application we obtain almost everywhere convergence results for these averages, and in some cases we also obtain almost everywhere convergence for their ergodic counterparts on a dynamical system.

### 1. INTRODUCTION

Let  $n > 1$ ,  $m \geq 1$  and consider an  $(n-1) \times m$  real-valued matrix  $A = (a_{i,j})_{i=1}^{n-1}{}_{j=1}^m$ . This naturally gives rise to the multilinear averages:

$$(1) \quad T_{A,\mathbf{R},r}(f_1, \dots, f_{n-1})(x) := \frac{1}{(2r)^m} \int_{|t_1|, \dots, |t_m| \leq r} \prod_{i=1}^{n-1} f_i(x + \sum_{j=1}^m a_{i,j} t_j) dt \vec{t},$$

where  $r > 0$  and  $f_1, \dots, f_{n-1}$  are arbitrary measurable functions on  $\mathbf{R}$ . Part of the motivation for considering such averages comes from ergodic theory. Let  $\mathbf{X} = (X, \Sigma, m, S)$  be a dynamical system, i.e. a complete probability space  $(X, \Sigma, m)$  endowed with an invertible bimeasurable transformation  $S : X \rightarrow X$  such that  $mS^{-1} = m$ . We define the iterates  $S^n : X \rightarrow X$  for  $n \in \mathbf{Z}$  in the usual manner. In case the matrix  $A$  has integer entries, one can consider the following ergodic averages:

$$(2) \quad T_{A,\mathbf{X},L}(f_1, \dots, f_{n-1})(x) := \frac{1}{(2L+1)^m} \sum_{|t_1|, \dots, |t_m| \leq L} \prod_{i=1}^{n-1} f_i(S^{\sum_{j=1}^m a_{i,j} t_j} x).$$

We use  $L^p(\mathbf{R})$  to denote the usual Lebesgue spaces on  $\mathbf{R}$ , and  $L^p(\mathbf{X})$  to denote the Lebesgue spaces on the dynamical system  $\mathbf{X}$ .

In this paper we shall be primarily concerned with the problem of almost everywhere convergence of these averages as  $r \rightarrow 0$  or  $L \rightarrow \infty$  in the case that the  $f_i$  obey some  $L^{p_i}$  type integrability condition. As is well known, such problems are

---

Received by the editors November 30, 2005 and, in revised form, October 27, 2006.

2000 *Mathematics Subject Classification*. Primary 42B25; Secondary 37A45.

*Key words and phrases*. Maximal operators, multilinear averages.

The first author was supported by NSF Grant DMS-0556389.

The second author was supported by NSF Grant CCF-0649473 and a grant from the McArthur Foundation.

The third author was supported by NSF Grants DMS-0400879 and DMS-0701302.

related to the boundedness properties of the maximal operator

$$(3) \quad \begin{aligned} T_{A,\mathbf{R}}^*(f_1, \dots, f_{n-1})(x) &:= \sup_{r>0} |T_{A,\mathbf{R},r}(f_1, \dots, f_{n-1})(x)| \\ &= \sup_{r>0} \left| \frac{1}{(2r)^m} \int_{|t_1|, \dots, |t_m| \leq r} \prod_{i=1}^{n-1} f_i(x + \sum_{j=1}^m a_{i,j} t_j) d\vec{t} \right| \end{aligned}$$

or the closely related maximal operator

$$(4) \quad \begin{aligned} T_{A,\mathbf{X}}^*(f_1, \dots, f_{n-1})(x) &:= \sup_{L>0} |T_{A,\mathbf{X},L}(f_1, \dots, f_{n-1})(x)| \\ &= \sup_{L>0} \left| \frac{1}{(2L+1)^m} \sum_{|t_1|, \dots, |t_m| \leq L} \prod_{i=1}^{n-1} f_i(S^{\sum_{j=1}^m a_{i,j} t_j} x) \right|. \end{aligned}$$

It turns out that standard transference arguments allow one to convert any positive or negative boundedness result for  $T_{A,\mathbf{R}}^*$  to one for  $T_{A,\mathbf{X}}^*$  and vice versa; see Proposition 14.1. Thus we shall view the boundedness problems for these two maximal operators as being equivalent.

Since one can easily establish convergence for (1) in any reasonable topology when the  $f_1, \dots, f_{n-1}$  are smooth, compactly supported functions, a standard density argument then shows that as soon as the maximal operator  $T_{A,\mathbf{R}}^*$  maps  $L^{p_1}(\mathbf{R}) \times \dots \times L^{p_{n-1}}(\mathbf{R})$  to weak  $L^q(\mathbf{R})$  for some  $0 < q < \infty$ , then the averages (1) will converge pointwise almost everywhere when  $f_i \in L^{p_i}(\mathbf{R})$  for  $1 \leq i \leq n-1$ , at least in the case when all the  $p_1, \dots, p_{n-1}$  are finite.<sup>1</sup> In fact these averages will converge almost everywhere to the pointwise product  $f_1 \dots f_{n-1}$ . In the converse direction, Stein's maximal principle [21] shows that in many cases, almost everywhere convergence of (1) can only be established via such weak  $L^q$  bounds on the maximal operator  $T_{A,\mathbf{R}}^*$ .

For the ergodic averages (2), the situation is more difficult because there is no obvious counterpart of the class  $C_c^\infty(\mathbf{R})$  of smooth, compactly supported functions on which the convergence is easy to establish.<sup>2</sup> However, one can use the class  $L^\infty(\mathbf{X})$  as a substitute, in the sense that once almost everywhere convergence for  $T_{A,\mathbf{X},L}$  is established for  $f_1, \dots, f_{n-1} \in L^\infty(\mathbf{X})$ , one can extend this convergence result to the case when  $f_i \in L^{p_i}(\mathbf{X})$  provided that one knows that the maximal operator  $T_{A,\mathbf{R}}^*$  maps  $L^{p_1}(\mathbf{R}) \times \dots \times L^{p_{n-1}}(\mathbf{R})$  to weak  $L^q(\mathbf{R})$  for some  $0 < q < \infty$ , since transference arguments then give an analogous boundedness statement for  $T_{A,\mathbf{X}}^*$ . Thus the problem of almost everywhere convergence of  $T_{A,\mathbf{X},L}$  for functions  $f_i \in L^{p_i}(\mathbf{X})$  factors into two rather distinct problems, namely establishing convergence for  $L^\infty(\mathbf{X})$  functions (which is a problem in ergodic theory), and establishing a bound for  $T_A^*$  (which is a problem in multilinear harmonic analysis). In this paper we shall focus almost exclusively on the latter problem. The former problem is quite difficult, except when  $n = 2$ ; the  $n = 3$  case already requires a deep result of Bourgain [5], and convergence for higher  $n$  is only proven for very special averages

<sup>1</sup>When one or more of the exponents is  $\infty$  one can proceed by localization arguments, exploiting the fact that an  $L^\infty$  function is locally in  $L^p$  for any  $p < \infty$ . This costs us an epsilon in the exponents but in most of our results the range of exponents will be open and so this will not make any difference.

<sup>2</sup>An alternate approach would be to establish either a  $V^q$  variational estimate on  $T_{A,\mathbf{X},L}$  in  $L$  for some  $q < \infty$ , or an oscillation inequality, since any of these automatically implies convergence as  $L \rightarrow \infty$ , in the spirit of Doob's inequality or Lepingle's inequality. We will not pursue such an approach here, but see for instance [8], [10].

(see e.g. [1]) or with additional spectral assumptions on the shift  $S$  (see [2], [16], [17]), and we will not make progress on these issues here.

Let us now discuss some important special cases of the above general setup.

**1.1. Linear averages.** If  $n = 2$ ,  $m = 1$ , and  $A = (a)$  for some nonzero integer  $a$ , then we have

$$T_{A,\mathbf{R},r}f_1(x) = \frac{1}{2r} \int_{-r}^r f_1(x + at) \, dt$$

and

$$T_{A,\mathbf{X},L}f_1(x) = \frac{1}{2L+1} \sum_{l=-L}^L f_1(S^{al}x).$$

If  $f_1$  is in  $L^p(\mathbf{R})$  (resp.  $L^p(\mathbf{X})$ ) for some  $1 \leq p \leq \infty$ , then the Lebesgue differentiation theorem (resp. the Birkhoff ergodic theorem) shows that  $T_{A,\mathbf{R},r}f_1$  (resp.  $T_{A,\mathbf{X},L}$ ) are almost everywhere convergent. Both of these results require the *Hardy-Littlewood maximal inequality*, which asserts that the Hardy-Littlewood maximal operator

$$Mf_1(x) := \sup_{r>0} \frac{1}{r} \int_{-r}^r |f_1|(x+t) \, dt$$

maps  $L^1$  to weak  $L^1$ . The Lebesgue differentiation theorem follows immediately from the maximal inequality, whereas the Birkhoff ergodic theorem requires that one first establish almost everywhere convergence for a dense class such as  $L^\infty(\mathbf{X})$ .

**1.2. Bilinear averages.** Let  $n = 3$ ,  $m = 1$ , and  $A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$  for some distinct nonzero integers  $a_1, a_2$ ; thus

$$T_{A,\mathbf{R},r}(f_1, f_2)(x) = \frac{1}{2r} \int_{-r}^r f_1(x + a_1t)f_2(x + a_2t) \, dt$$

and

$$T_{A,\mathbf{X},L}(f_1, f_2)(x) = \frac{1}{2L+1} \sum_{l=-L}^L f_1(S^{a_1l}x)f_2(S^{a_2l}x).$$

As a consequence of a deep theorem of Bourgain [5] (relying on Fourier analysis on the torus), it is known that the averages  $T_{A,\mathbf{X},L}(f_1, f_2)$  converge almost everywhere whenever  $f_1, f_2 \in L^\infty(X)$ . Almost everywhere convergence in other classes then pivots on understanding the bilinear maximal operator

$$T_{A,\mathbf{R}}^*(f_1, f_2)(x) = \sup_{r>0} \left| \frac{1}{2r} \int_{-r}^r f_1(x + a_1t)f_2(x + a_2t) \, dt \right|.$$

This operator clearly maps  $L^\infty(\mathbf{R}) \times L^\infty(\mathbf{R}) \rightarrow L^\infty(\mathbf{R})$ , and from the Hardy-Littlewood maximal inequality it also maps  $L^\infty(\mathbf{R}) \times L^1(\mathbf{R})$  or  $L^1(\mathbf{R}) \times L^\infty(\mathbf{R})$  to weak  $L^1$ . This, combined with bilinear interpolation, is enough to establish almost everywhere convergence of the ergodic averages  $T_{A,\mathbf{X},L}$  for  $f_1 \in L^{p_1}(X)$ ,  $f_2 \in L^{p_2}(X)$  when  $1/p_1 + 1/p_2 < 1$  (one also obtains the edge  $1/p_1 + 1/p_2 = 1$  from this argument as long as  $p_1, p_2 < \infty$ ). It was shown by Lacey [15], using time-frequency analysis, that  $T_{A,\mathbf{R}}^*$  in fact maps  $L^{p_1}(\mathbf{R}) \times L^{p_2}(\mathbf{R})$  to  $L^q(\mathbf{R})$  whenever  $\frac{1}{q} = \frac{1}{p_1} + \frac{1}{p_2}$  and  $q > \frac{2}{3}$ . This allows one to extend the almost everywhere convergence result to the larger range  $1/p_1 + 1/p_2 < 3/2$ . It is an interesting question as to whether this is the true limit for these results. Certainly one has boundedness for a single-scale operator  $T_{A,\mathbf{R},r}$  or  $T_{A,\mathbf{X},L}$  all the way up to the range  $1/p_1 + 1/p_2 \leq 2$ . On the other

hand, the time-frequency approach is known to break down at  $1/p_1 + 1/p_2 = 3/2$  (see [15]).

**1.3. Furstenberg averages.** Let  $n \geq 2$ ,  $m = 1$ , and let  $A$  be the matrix

$$A := \begin{pmatrix} 1 \\ 2 \\ \vdots \\ n-1 \end{pmatrix}.$$

Then (1) becomes the multilinear average

$$T_{A,\mathbf{R},r}(f_1, \dots, f_{n-1})(x) = \frac{1}{2r} \int_{-r}^r \prod_{i=1}^{n-1} f_i(x + it) dt$$

and (2) becomes the Furstenberg average

$$T_{A,\mathbf{X},L}(f_1, \dots, f_{n-1})(x) = \frac{1}{2L+1} \sum_{l=-L}^L f_i(S^{il}x).$$

Note the cases  $n = 2, n = 3$  are special cases of the linear and bilinear averages considered earlier. These averages are related to the Furstenberg recurrence theorem [12] and to Szemerédi's theorem on arithmetic progressions [22], and are also connected to the recent result in [13] that the primes contain arbitrarily long progressions. For instance, the Furstenberg recurrence theorem is essentially the assertion that

$$\liminf_{L \rightarrow \infty} \int_X T_{A,\mathbf{X},L}(f, \dots, f) f dm > 0$$

whenever  $f$  is nonnegative and does not vanish almost everywhere. The question of norm convergence of  $T_{A,\mathbf{X},L}$  is more difficult and has only recently been treated in the independent works of Host and Kra [14] and Ziegler [23]. They showed that if  $f_1, \dots, f_{n-1} \in L^\infty(X)$ , then  $T_{A,\mathbf{X},L}(f_1, \dots, f_{n-1})$  converges in the  $L^2(X)$  norm (and hence in the  $L^p(X)$  norm for any  $1 \leq p < \infty$ ). Their approach relies on the reduction to convergence for functions in a sub- $\sigma$ -algebra  $\mathcal{Z}_{n-1}$  of  $\Sigma$ , known as a *characteristic factor*, on which  $T$  can be represented as an inverse limit of translations on nilmanifolds. The advantage of such a concrete representation is that this particular type of translation is quite well understood. In particular,  $\mathcal{Z}_0$  is the  $\sigma$ -algebra spanned by the invariant sets of powers of  $T$ , while the action of  $T$  on the Kronecker factor  $\mathcal{Z}_1$  is isomorphic with a rotation on some abelian group. The  $\sigma$ -algebras  $\mathcal{Z}_k$  with  $k \geq 2$  give rise to noncommutative factors which require a more delicate analysis. The work in this paper will however proceed in a different direction, focusing on the quantitative bounds of various operators associated with these averages rather than analyzing characteristic factors. It is of course possible to extend these norm convergence results to functions  $f_i$  in other spaces  $L^{p_i}(X)$  by exploiting boundedness properties  $T_{A,\mathbf{X},L}$  or  $T_{A,\mathbf{R},r}$ , but we will not pursue this issue here, though we will mention that some surprising subtleties in this problem in the case  $1/p_1 + \dots + 1/p_n > 1$  have been uncovered by Christ [6].

The problem of almost everywhere convergence, as opposed to norm convergence, for the Furstenberg averages remains open even for  $n = 4$ . One can obtain some bounds of the corresponding maximal operators in  $L^p$  spaces by leveraging the corresponding bounds in the bilinear setting. For instance one can extend Lacey's

bilinear estimates mentioned earlier to the multilinear setting by estimating all but two of the functions in  $L^\infty$ . This ultimately leads to a bound on  $T_{A,\mathbf{R}}^*$  from  $L^{p_1}(\mathbf{R}) \times \dots \times L^{p_n}(\mathbf{R})$  to  $L^q(\mathbf{R})$  whenever  $1 < p_1, \dots, p_n \leq \infty$  and  $1/q = 1/p_1 + \dots + 1/p_n < 3/2$ .

**1.4. Averages along cubes.** The work of Host and Kra [14] related the norm convergence of the above Furstenberg averages to the norm convergence of averages of cubes, which is a special case of (2) with  $n = 2^m$ . To define them, let  $V_m$  be the index set  $V_m := \{0, 1\}^m \setminus \{0\}^m$ . The averages on the  $m$ -dimensional cubes are

$$(5) \quad \frac{1}{(2L+1)^m} \sum_{\vec{i} \in \{-L, \dots, L\}^m} \prod_{\epsilon \in V_m} f_\epsilon(S^{\vec{i} \cdot \epsilon} x).$$

For example, when  $m = 1$  (so  $n = 2$ ) we just have a linear averaging operator. When  $m = 2$  (and so  $n = 4$ ), this averaging operator along squares is essentially the same as  $T_{A,\mathbf{X},L}$  with

$$A := \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$

while when  $m = 3$  (and  $n = 8$ ) the averaging operator along cubes is essentially  $T_{A,\mathbf{X},L}$  with

$$A := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

It is proved in [14] that the averages<sup>3</sup> in (5) have  $\mathcal{Z}_{m-1}$  as a characteristic factor for  $L^2$ -norm convergence, and as a consequence that these averages converge in  $L^2(\mathbf{X})$  whenever  $f_\epsilon \in L^\infty(\mathbf{X})$ . Using these characteristic factors, Assani [1] showed that these averages also converged pointwise almost everywhere when  $f_\epsilon \in L^\infty(\mathbf{X})$ . It is somehow peculiar that these techniques do not seem to be able to give an alternative (nonFourier analytical) proof to Bourgain’s pointwise result mentioned earlier.

To extend the latter  $L^\infty(\mathbf{X})$  convergence result to an  $L^p(\mathbf{X})$  convergence result requires control of a maximal function. For the sake of concreteness let us just focus on the case  $m = 2$ , where the relevant maximal function is

$$\sup_{r>0} \left| \frac{1}{(2r)^2} \int_{-r}^r \int_{-r}^r f_{10}(x+t_1) f_{01}(x+t_2) f_{11}(x+t_1+t_2) dt_1 dt_2 \right|.$$

One can deduce a certain number of bounds on this maximal function from the Hardy-Littlewood maximal inequality and multilinear interpolation. Indeed, the maximal inequality and Hölder’s inequality imply that this maximal function lies in weak  $L^{1/2}$  whenever two of  $f_{10}, f_{01}, f_{11}$  lie in  $L^1$  and the other one lies in  $L^\infty$ , while this maximal operator is trivially in  $L^\infty$  when all three of  $f_{10}, f_{01}, f_{11}$  lie in  $L^\infty$ . Interpolation then gives bounds (and hence almost everywhere convergence of

<sup>3</sup>Actually, a more general class of averages is shown in [14] to have  $\mathcal{Z}_{m-1}$  as a characteristic factor; we refer the reader to [14] for the details.

the associated averages along squares) when  $f_{01} \in L^{p_{01}}, f_{10} \in L^{p_{10}}, f_{11} \in L^{p_{11}}$  with  $1/p_{01} + 1/p_{10} + 1/p_{11} < 2$ , with an extension to the boundary  $1/p_{01} + 1/p_{10} + 1/p_{11} = 2$  when all of the exponents are finite. As a corollary of our main result (which is proven using time-frequency techniques) we shall be able to extend this range to  $1/p_{01} + 1/p_{10} + 1/p_{11} < 5/2$ , in analogy with the situation for bilinear averages discussed earlier (see Corollaries 1.6, 1.7 below).

**1.5. Main results.** We now study the maximal operator  $T_{A,\mathbf{R}}^*$  defined in (3) for a general  $(n - 1) \times m$  matrix  $A = (a_{i,j})$ ; we will allow the  $a_{i,j}$  here to be noninteger as one can still define  $T_{A,\mathbf{R}}^*$  in this case. To state the main result we need some notation. We introduce the extended matrix  $\mathbb{E}(A)$ , which is the  $n \times (m + 1)$  matrix

$$\mathbb{E}(A) := \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} & 1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} & 1 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n-1,1} & a_{n-1,2} & \dots & a_{n-1,m} & 1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Note that the range of this matrix consists of all  $n$ -tuples of the form

$$\left(x + \sum_{j=1}^m a_{1,j}t_j, \dots, x + \sum_{j=1}^m a_{n-1,j}t_j, x\right)$$

for  $x, t_1, \dots, t_m \in \mathbf{R}$ .

A set of row indices  $i$  is said to be a *set of linear independence* for a matrix  $B$  if the set of corresponding rows of  $B$  is linearly independent. Given a matrix  $A$ , let  $S_{A,\epsilon}$  for  $0 < \epsilon < 1/4$  be the set of all tuples  $(x_1, \dots, x_{n-1})$  where  $x_i \in \{0, 1/2 + \epsilon, 1 - \epsilon\}$  for all  $i$ , there is at most one index  $i$  with  $x_i = 1/2 + \epsilon$ , the indices  $i$  with  $x_i = 1 - \epsilon$  form a set of linear independence for  $A$ , and the indices  $i$  with  $x_i \in \{1/2 + \epsilon, 1 - \epsilon\}$  form a set of linear independence for  $\mathbb{E}(A)$ . Let  $H_{A,\epsilon}$  be the convex hull of  $S_{A,\epsilon}$  and let  $H_A$  be the union of all  $H_{A,\epsilon}$  with  $0 < \epsilon < 1/4$ .

The following is our main theorem:

**Theorem 1.1.** *Assume  $n \geq 3$  and let  $A$  be a matrix as above. Let  $(p_1, \dots, p_{n-1})$  be a tuple of real numbers with*

$$(6) \quad 1 < p_i \leq \infty$$

for  $1 \leq i \leq n - 1$  and set

$$(7) \quad \frac{1}{p_n'} = \sum_{i=1}^{n-1} \frac{1}{p_i}.$$

If

$$(1/p_1, \dots, 1/p_{n-1}) \in H_A,$$

then the operator  $T_{A,\mathbf{R}}^*$ ,

$$T_{A,\mathbf{R}}^* : L^{p_1} \times \dots \times L^{p_{n-1}} \rightarrow L^{p_n'}$$

is bounded.

*Remark 1.1.* The condition (7) is mandated by scaling considerations (i.e. dimensional analysis). As we shall see shortly, the theorem is trivial if one restricts the

tuples  $(1/p_i)$  to the convex hull of those points in  $S_{A,\epsilon}$  which do not have a component equal to  $1/2 + \epsilon$ . This happens in particular when  $n = 2$ . Thus, in a nutshell, we are gaining  $1/2 - \epsilon$  over the trivial estimates.

*Remark 1.2.* For some matrices  $A$  we can obtain a better range of exponents than stated in the theorem. Namely, when the matrix  $A$  is a diagonal block matrix, we may gain  $1/2$  for every block. More precisely, the argument works for  $A$  upper-block triangular and  $\mathbb{E}(A)$ , modulo the last column and restricted to the rows other than the last row, is block diagonal. The argument involves only separation of variables and Hölder’s inequality, so we shall not elaborate on this.

The following corollary is weaker than the theorem, but has the advantage of an easy description of the range of exponents and covers many of the cases of interest. Define the *nondegeneracy rank* of the matrix  $A$ , denoted by  $\text{rank}^*(A)$ , to be the largest integer  $r$  such that any  $r$  rows of  $A$  are linearly independent. It is an immediate observation that  $\text{rank}^*(A) + 1 \geq \text{rank}^*(\mathbb{E}(A)) \geq \text{rank}^*(A)$ .

**Corollary 1.2.** *Assume  $n \geq 2$  and let  $A$  be a matrix as above. Define the complexity parameter  $k = n - \text{rank}^*(\mathbb{E}(A))$ . Let  $(p_1, \dots, p_{n-1})$  be a tuple with*

$$1 < p_i \leq \infty$$

for  $1 \leq i \leq n - 1$  and set

$$\frac{1}{p_{n'}} = \sum_{i=1}^{n-1} \frac{1}{p_i}.$$

If

$$(8) \quad \frac{1}{p_1} + \dots + \frac{1}{p_{n-1}} < n - k - \frac{1}{2},$$

then the operator  $T_{A,\mathbf{R}}^*$ ,

$$T_{A,\mathbf{R}}^* : L^{p_1} \times \dots \times L^{p_{n-1}} \rightarrow L^{p_{n'}},$$

is bounded.

*Proof.* The closure of the region of tuples  $(1/p_i)$  in the corollary is the intersection of the cube  $[0, 1]^{n-1}$  with a half space. All extremal points of this set are on an edge of the cube and thus have all but at most one coordinate in  $\{0, 1\}$ . The only possible value for the exceptional coordinate is  $1/2$  as the right-hand side of (8) is equal to  $1/2$  modulo the integers. Thus the region in the corollary is the convex hull of all tuples  $(x_1, \dots, x_2)$  with at most  $n - k - 1$  components equal to  $1 - \epsilon$ , at most one component equal to  $1/2 + \epsilon$  and the remaining components equal to  $0$ . The corollary then follows from the rank conditions on  $A$  and  $\mathbb{E}(A)$  and the fact that  $\text{rank}^*(A) \geq n - k - 1$ .  $\square$

*Remark 1.3.* As discussed earlier, the boundedness results in Theorem 1.1 and Corollary 1.2 immediately imply almost everywhere convergence for  $T_{A,\mathbf{R},\epsilon}(f_1, \dots, f_{n-1})$  as  $\epsilon \rightarrow 0$  when  $f_i \in L^{p_i}(\mathbf{R})$  if all the  $p_i$  are finite, since this convergence is trivial for  $f_i$  in the dense class  $C_c^\infty(\mathbf{R})$ . The  $p_i = \infty$  cases can also be handled by a localization argument and exploiting some openness properties of  $H_A$ . The situation for the ergodic averages is however substantially more difficult.

*Remark 1.4.* If  $\text{rank}^*(\mathbb{E}(A)) = \text{rank}(\mathbb{E}(A))$ , then the regions described in Theorem 1.1 and Corollary 1.2 are equal.

*Remark 1.5.* It is worth noting that  $p'_n$  can be less than 1; indeed it is less than 1 in all nontrivial cases. In some cases one can get below 1 by using just Hölder's inequality and interpolation; see for instance the discussion in Section 1.4.

*Remark 1.6.* Theorem 1.1 is a direct analog of the singular integral version in [18, Theorem 1.1], which roughly speaking replaces  $T_{A,\mathbf{R}}^*$  with the related expression

$$p.v. \int_{\mathbf{R}^m} \prod_{i=1}^{n-1} f_i(x + \sum_{j=1}^m a_{i,j} t_j) K(\vec{t}) d\vec{t}$$

for some Calderón-Zygmund kernel  $K$ . As a consequence the methods of proof are quite similar. The parameter  $k$  in Corollary 1.2 plays the same role as the parameter  $k$  appearing in [18, Theorem 1.1], measuring the complexity of the averages under investigation. The case  $k = 0$  for the singular integral version can be solved with classical methods, namely Littlewood-Paley theory or wavelets, just as the case  $k = 0$  for the maximal version can be solved using the classical Hardy-Littlewood maximal inequality.

Readers familiar with [18] will observe that the range of exponents in Theorem 1.1 is somewhat more permissive than that in [18]. More precisely, the restriction  $k < \frac{n}{2}$  as well as several restrictions on the exponents  $p_i$  from [18] are not needed in Theorem 1.1. This is a consequence of the fact that there are trivial reductions in the maximal operator case if there are exponents  $p_i = \infty$ , while in the singular integral setting there are no such trivial reductions. This explains why for instance we can obtain nontrivial estimates for the trilinear maximal operator ( $n = 4, k = 2$ )

$$(9) \quad T_a^*(f_1, f_2, f_3) := \sup_{\epsilon > 0} \frac{1}{\epsilon} \int_{|t| \leq \epsilon} |f_1(x + a_1 t) f_2(x + a_2 t) f_3(x + a_3 t)| dt$$

with  $a_1, a_2, a_3, 0$  pairwise different (see Example 1.5 below), despite the fact that no  $L^p$  bounds of any sort are known for the trilinear Hilbert transform

$$p.v. \int_{\mathbf{R}} f_1(x + a_1 t) f_2(x + a_2 t) f_3(x + a_3 t) \frac{dt}{t}.$$

*Remark 1.7.* It should be emphasized that the nontrivial estimates from the  $k > 1$  cases are all obtained by such trivial reductions to the case  $k = 1$  and multilinear interpolation. In other words, there is no special theory developed yet to address the case  $k > 2$ . It is quite probable that more sophisticated techniques will extend the range of the exponents in this case. An interesting connection concerns the fact that averages corresponding to some  $k \geq 0$  appear to have  $\mathcal{Z}_k$  as a characteristic factor for  $L^2$ -norm convergence. In particular, it is an exercise based on the techniques from [1] and from [5] to show that  $\mathcal{Z}_k$  is the characteristic factor even for a.e. convergence, when  $k = 0, 1$ . This would support the evidence that, as in the case of norm convergence,  $k$  is the only parameter which dictates the complexity of the averages and of the techniques needed for the proof.

Similar difficulties are encountered when dealing with polynomial maximal operators such as  $P^*(f_1, f_2)(x) := \sup_{\epsilon > 0} \frac{1}{\epsilon} \int_{|t| \leq \epsilon} |f_1(x + t) f_2(x + t^2)| dt$ . In all these instances, the decomposition of the maximal operator, as explained in the third section below, gives rise to a summation over a larger family of multidimensional cubes, each of which is indexed by more than just one parameter. Curiously, the boundedness of the maximal operator associated with polynomial averages, unlike the multilinear averages studied here (see Proposition 14.1), does not in general transfer from



harmonic analysis to ergodic theory. It is really that the results in these two contexts have different meaning and most probably distinct ideas behind their proofs. An illuminating contrast comes from the fact that  $\sup_{\epsilon>0} \frac{1}{\epsilon} \int_{|t|\leq\epsilon} |f(x+t^2)|dt$  can be easily bounded by the Hardy-Littlewood maximal function, while Bourgain showed that the convergence of the ergodic averages along squares needs completely new ideas [4].

Let us illustrate Theorem 1.1 and Corollary 1.2 with some examples.

**Example 1.3.** Consider the bilinear averages from Section 1.2. Here the extended matrix is

$$\mathbb{E}(A) = \begin{pmatrix} a_1 & 1 \\ a_2 & 1 \\ 0 & 1 \end{pmatrix}.$$

One can check that  $\text{rank}(A) = 1$  and  $\text{rank}^*(\mathbb{E}(A)) = \text{rank}(\mathbb{E}(A)) = 2$ , and

$$S_{A,\epsilon} = \{(0, 0), (0, 1/2+\epsilon), (1/2+\epsilon, 0), (0, 1-\epsilon), (1-\epsilon, 0), (1/2+\epsilon, 1-\epsilon), (1-\epsilon, 1/2+\epsilon)\}$$

and hence

$$H_A = \{(a, b) : 0 \leq a, b < 1; a + b < 3/2\}.$$

In this case Theorem 1.1 and Corollary 1.2 give the same results, namely recovering the bilinear maximal function estimates of Lacey [15] described earlier. Indeed we give a reasonably self-contained<sup>4</sup> proof of the main results from [15] here, following Lacey’s approach.

**Example 1.4.** Consider the  $n = 4$  Furstenberg average from Section 1.3. Here the extended matrix is

$$\mathbb{E}(A) := \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 0 & 1 \end{pmatrix}.$$

One can check that  $\text{rank}(A) = 1$  and  $\text{rank}^*(\mathbb{E}(A)) = \text{rank}(\mathbb{E}(A)) = 2$ , and  $S_{A,\epsilon}$  consists of those triples  $(a, b, c)$  with  $a, b, c \in \{0, 1/2 + \epsilon, 1 - \epsilon\}$ , at most one of  $a, b, c$  equal to  $1/2 + \epsilon$ , and at most one of  $a, b, c$  equal to  $1 - \epsilon$ . This gives

$$H_A = \{(a, b, c) : 0 \leq a, b, c < 1; a + b < 3/2\}.$$

In this case, Theorem 1.1 and Corollary 1.2 recover the multilinear estimates mentioned at the end of Section 1.3 that can be trivially obtained from Lacey’s bilinear result. Similar considerations apply to higher values of  $n$ .

**Example 1.5.** Consider the  $m = 2$  average along squares from Section 1.4. Here the extended matrix is

$$\mathbb{E}(A) := \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

---

<sup>4</sup>We will require some results from other papers, notably the multilinear interpolation theory from [18], the weak Bessel inequality for forests (see e.g. [20]), a maximal Fourier inequality of Bourgain [4], and an interval selection lemma of Lacey [15].

One can check that  $\text{rank}(A) = 2$  and  $\text{rank}^*(\mathbb{E}(A)) = \text{rank}(\mathbb{E}(A)) = 3$ , and  $S_{A,\epsilon}$  consists of those triples  $(a, b, c)$  with  $a, b, c \in \{0, 1/2 + \epsilon, 1 - \epsilon\}$ , at most one of  $a, b, c$  equal to  $1/2 + \epsilon$ , and at most two of  $a, b, c$  equal to  $1 - \epsilon$ . This gives

$$H_A = \{(a, b, c) : 0 \leq a, b, c < 1; a + b < 5/2\}.$$

Combining the above example with Proposition 14.1 from Appendix 14, and the result of Assani [1], we obtain the following corollary.

**Corollary 1.6.** *Let  $1 < p_1, p_2, p_3 \leq \infty$  be such that  $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} < \frac{5}{2}$ . For every dynamical system  $\mathbf{X} = (X, \Sigma, m, S)$ , the averages on squares*

$$\frac{1}{N^2} \sum_{i=-N}^N \sum_{j=-N}^N f_1(S^i x) f_2(S^j x) f_3(S^{i+j} x)$$

converge a.e.  $x$ , for each  $f_i \in L^{p_i}(\mathbf{X})$ .

*Remark 1.8.* A version of Corollary 1.6 holds for all averages with  $k = 1$ . The convergence for  $L^\infty$  functions follows by using the aforementioned fact that these averages have characteristic factor  $\mathcal{Z}_1$  for pointwise convergence. We omit the details.

*Remark 1.9.* In [9] we use combinatorial methods involving sum set estimates to get nontrivial positive results in Corollary 1.6. This completely different approach gives the result only in a small range,  $p'_4 > \frac{1}{2+\epsilon}$  for some unspecified  $\epsilon$ , and does not seem to extend to the case when  $p'_4$  is smaller than or even close to  $\frac{2}{5}$ .

*Remark 1.10.* An interesting contrast to the results of Theorem 1.1 is provided by the constructions from [7], showing that some maximal operators fail to be bounded when the indices  $p_i, 1 \leq i \leq n - 1$  are sufficiently close to 1. As a consequence, both Furstenberg's averages with  $n \geq 4$  and the averages on cubes with  $m \geq 3$  are proved to diverge a.e. in some range of  $L^p$  spaces. The trilinear maximal operator from (9) has been proved in [6] to be unbounded for  $p_1 = p_2 = p_3 = p, 1 \leq p < \frac{3}{2}$ , for appropriate choices of  $\vec{a}$  depending on  $p$ . The main ingredient behind these negative results is the fact that the polynomials  $x + \sum_{j=1}^m a_{i,j} t_j, 1 \leq i \leq n - 1$  are linearly dependent in  $\mathbf{R}[x, t_1, \dots, t_m]$  and hence  $\text{rank}^*(\mathbb{E}(A)) \leq n - 2$  and  $k \geq 2$ . In other words, our tools provide negative results only when  $k \geq 2$ , and all positive results are trivially deduced from positive results when  $k = 0, 1$ . Further progress would require breaking this barrier in the complexity  $k$  either for positive or for negative results.

The following is the straightforward application of Corollary 1.2 to averages on cubes. In this case, while  $\text{rank}(A)$  is the dimension of the cube, we have  $\text{rank}^*(A) = 2$ , an obstruction for higher nondegeneracy rank being the linear dependence of the polynomials  $t_1, t_2$ , and  $t_1 + t_2$ . On the other hand,  $\text{rank}^*(\mathbb{E}(A)) = 3$ , and hence:

**Corollary 1.7.** *Let  $1 < p_\epsilon \leq \infty, \epsilon \in V_m$ , be such that*

$$(10) \quad \sum_{\epsilon \in V_m} \frac{1}{p_\epsilon} < \frac{5}{2}.$$

*For every dynamical system  $\mathbf{X} = (X, \Sigma, m, S)$ , the averages on  $m$ -dimensional cubes (5) converge a.e. for each  $f_\epsilon \in L^{p_\epsilon}(\mathbf{X})$ .*

This of course generalizes Corollary 1.6. It would be interesting to know whether one can improve upon the  $5/2$  on the right-hand side of (10). Certainly the methods of this paper do not yield such an improvement, and [7] provides an upper bound of  $28/5$  for the right-hand side of (10) for three-dimensional cubes.

Theorem 1.1 is proven using standard time-frequency strategies, and in particular follows the approach of Lacey [15], though it is more self-contained and employs some technical simplifications over that in [15]. In Section 2 we use the theory of multilinear interpolation to reduce Theorem 1.1 to a model case, Theorem 2.3, in which the matrix  $A$  is in a simplified normal form, the functions  $f_1, \dots, f_{n-1}$  have become  $L^2$ -normalized functions adapted to certain sets  $E_1, \dots, E_{n-1}$ , and the output is being measured in another set  $E'_n$  which excludes a certain exceptional set determined by the Hardy-Littlewood maximal function. In Sections 3, 4 we use the Fourier transform and wave packet decomposition to reduce matters to bounding a certain model sum (Theorem 4.4) involving the inner product of the functions  $f_1, \dots, f_n$  with various wave packets (and maximal wave packets) associated with a certain “rank one” collection of multitiles. To estimate this model sum, we organize the collection of multitiles into trees; after obtaining an upper bound for the contribution of a single tree (see Proposition 6.2 and Section 7) one quickly reduces (essentially by summing a geometric series; see Section 6) to that of proving estimates for a tree selection algorithm (Lemma 6.3), which in turn reduces to a certain maximal Bessel inequality concerning wave packets in a forest (Theorem 9.1, slightly improving and simplifying a similar result from [15]). This Bessel inequality will involve a certain logarithmic-type loss involving the size parameter  $2^m$ , but by some “good- $\lambda$ ” type reductions in Section 9 we can replace this factor with another logarithmic factor involving instead the multiplicity  $\|N_{\mathcal{F}}\|_{L^\infty}$  of the forest (Theorem 9.2). After some sparsification of the tile set, some elimination of exceptional tiles, and duality, one reduces to establishing a certain maximal Bessel inequality on two families of tiles (see (60) and (61)). These inequalities are proven by using the time localization properties of wave packets, a nonmaximal Bessel inequality (proven in Section 13), and the Rademacher-Menshov inequality. In the case of one of these inequalities (61), one also needs a maximal inequality of Bourgain [4]. Finally, in an Appendix (Section 14) we present a standard correspondence principle equating boundedness of maximal functions on  $\mathbf{R}$  with maximal functions on measure-preserving systems.

## 2. INTERPOLATION REDUCTIONS

The rest of the paper is devoted to the proof of Theorem 1.1. We shall use the methods of multilinear time-frequency analysis and work entirely on  $\mathbf{R}$ ; thus we will not make any further reference to the dynamical system  $\mathbf{X}$ .

In this section we use some multilinear interpolation techniques to reduce the operator  $T_{A,\mathbf{R}}^*$  and the exponents  $p_1, \dots, p_n$  to a standard form, and then to also reduce the input functions  $f_1, \dots, f_{n-1}$  (and an additional output function  $f_n$  arising from duality) to another standard form.

We first introduce some basic notation. If  $E$  is a measurable subset of  $\mathbf{R}$ , we use  $1_E$  to denote the indicator function of  $E$  and  $|E|$  to denote the Lebesgue measure. Also  $Mf(x) := \sup_{r>0} \frac{1}{2r} \int_{x-r}^{x+r} |f(y)| dy$  denotes the classical Hardy-Littlewood maximal function. The notation  $a \lesssim b$  or  $a = O(b)$  means that  $a \leq cb$  for some universal constant  $C$  (which will be allowed to depend on parameters such

as  $n$  and  $p_1, \dots, p_n$ ), and  $a \sim b$  means that  $a \lesssim b$  and  $b \lesssim a$ . In some cases we will subscript the  $\lesssim$  notation by a parameter to emphasize the fact that the constant  $C$  involved can depend on that parameter; thus for instance  $a \lesssim_\mu b$  means that  $C$  can depend on  $\mu$ . If  $x \in \mathbf{R}^n$  we use  $\|x\|$  to denote the Euclidean norm of  $x$ .

Now we can reduce the operator  $T_{A, \mathbf{R}}^*$  and the exponents  $p_1, \dots, p_n$  to a standard form.

**Theorem 2.1** (First reduction). *Let  $n \geq 3$ , let  $\Sigma$  be a hyperplane in  $\mathbf{R}^{n-1}$  containing the origin but not containing any of the  $n - 1$  coordinate vectors  $e_1, \dots, e_{n-1}$  or the vector  $(1, \dots, 1)$ . Then the  $(n - 1)$ -linear operator  $T^*$  defined by*

$$(11) \quad \begin{aligned} & T^*(f_1, \dots, f_{n-1})(x) \\ &= \sup_{r>0} \frac{1}{r^{n-2}} \int_{\vec{t} \in \Sigma: \|\vec{t}\| \leq r} |f_1(x + t_1) \cdots f_{n-1}(x + t_{n-1})| d\vec{t} \end{aligned}$$

is bounded from  $L^{p_1}(\mathbf{R}) \times \dots \times L^{p_{n-1}}(\mathbf{R})$  to  $L^{p'_n}(\mathbf{R})$  whenever  $1 < p_1, \dots, p_{n-1} < 2$ ,  $\frac{5}{2} - n < \frac{1}{p_n} < 3 - n$ , and

$$\frac{1}{p'_n} = \frac{1}{p_1} + \dots + \frac{1}{p_{n-1}}.$$

The bound of course depends on  $p_1, \dots, p_n$  and the  $\lambda_i$ .

*Remark 2.1.* Note that  $\text{rank}(\mathbb{E}(A)) = \text{rank}^*(\mathbb{E}(A)) = n - 1$ . Hence we are in the case  $k = 1$  of Corollary 1.2 and the corollary is equivalent to Theorem 1.1 in this case. The condition that  $\Sigma$  does not contain  $e_1, \dots, e_{n-1}$  or  $(1, \dots, 1)$  corresponds to the nondegeneracy condition in [18].

*Proof of Theorem 1.1 assuming Theorem 2.1.* By multilinear interpolation as in [18] it suffices to prove the estimate for tuples  $(1/p_i)$  in  $S_{A, \epsilon}$  for some  $0 < \epsilon < 1/2$ , so in particular  $1/p_i \in \{1/2 + \epsilon, 1 - \epsilon, 0\}$  for all  $i$ . We may of course assume the  $f_i$  are nonnegative. For each index  $i$  with  $p_i = \infty$  we can trivially estimate  $f_i$  by its supremum norm and remove it from the maximal operator:

$$\begin{aligned} & \sup_{r>0} \frac{1}{(2r)^m} \int_{|t_1|, \dots, |t_m| \leq r} \prod_{i=1}^{n-1} f_i(x + \sum_{j=1}^m a_{i,j} t_j) d\vec{t} \\ & \lesssim \|f_j\|_{L^\infty} \sup_{r>0} \frac{1}{(2r)^m} \int_{|t_1|, \dots, |t_m| \leq r} \prod_{i \neq j} f_i(x + \sum_{j=1}^m a_{i,j} t_j) d\vec{t}. \end{aligned}$$

Doing this to each such exponent, we may assume without loss of generality that  $1/p_i \in \{1/2 + \epsilon, 1 - \epsilon\}$  for all  $i$ .

If  $1/p_i = 1 - \epsilon$  for all  $i$ , then by definition of  $S_{A, \epsilon}$  the rows of the matrix  $A$  are linearly independent and we may do a change of variables so that  $a_{i,j}$  is the Kronecker delta for  $1 \leq i, j \leq n - 1$ . Of course the cube of integration in the parameter space  $\{(t_1, \dots, t_n)\}$  will be a parallelepiped in the new variables, but we may use the positivity of the  $f_i$  and estimate the characteristic function of the parallelepiped by that of a cube, conceding a bounded loss in the estimates. We may also assume that  $A$  is a square matrix of dimension  $m = n - 1$ , since in the case  $m > n - 1$  we may fix the variables  $t_j$  with  $j > n - 1$  and apply the result in the square matrix case to fixed translates of the function  $f_i$  obtaining an  $L^p(\mathbf{R})$  bound independently of the translation. Then we perform a dummy average in the

variable  $t_j$  with  $j > n - 1$  to obtain the desired estimate. In the square matrix case we estimate

$$\begin{aligned} & \sup_{r>0} \frac{1}{(2r)^m} \int_{|t_1|, \dots, |t_{n-1}| \leq r} \prod_{i=1}^{n-1} |f_i(x + t_i)| d\vec{t} \\ & \leq \prod_{i=1}^{n-1} \sup_{\epsilon>0} \frac{1}{\epsilon} \int_{|t| \leq \epsilon} |f_i(x + t)| dt \end{aligned}$$

and then apply the Hardy-Littlewood maximal theorem for  $L^{1+\epsilon}$  and Hölder’s inequality to obtain the desired estimate.

It remains to consider the case when  $1/p_j = 1/2 + \epsilon$  for one index  $j$  and  $1/p_i = 1 - \epsilon$  for  $i \neq j$ ; note that this places these exponents in the situation of Theorem 2.1. We may assume that  $n \geq 3$  since the  $n = 2$  case follows from the Hardy-Littlewood maximal inequality. By symmetry we may assume that  $j = n - 1$ . The first  $n - 2$  rows of  $A$  are linearly independent, and we may assume that  $(a_{i,j})_{1 \leq i, j, \leq n-2}$  is the Kronecker delta. We may assume that the last row of  $A$  is a linear combination of the other rows, or otherwise we can apply the reasoning of the previous paragraph. By a reasoning as in the previous paragraph we may also assume that  $m \leq n - 2$ . Thus after a change of variables if necessary (and covering the resulting parallelepiped by a ball) the operator  $T_{A, \mathbf{R}}^*$  takes the form (11) for some hyperplane  $\Sigma$ . If  $\Sigma$  contains  $e_i$ , then we perform the  $t_i$  average first, estimate the average using the Hardy-Littlewood maximal function of  $f_i$ , and use Hölder’s inequality to reduce matters to the case with one fewer function. We may thus assume that  $\Sigma$  does not contain any of the  $e_i$ . Finally, the hypothesis that the first  $n - 1$  rows of  $\mathbb{E}(A)$  are linearly independent implies that  $\Sigma$  does not contain  $(1, \dots, 1)$ , and the claim now follows from Theorem 2.1.  $\square$

*Remark 2.2.* If the hypothesis  $\frac{5}{2} - n < \frac{1}{p_n} < 3 - n$  is replaced by  $1 < p_n < \infty$ , then Theorem 2.1 is easy to prove. Indeed, in this case we can use Hölder’s inequality to obtain the pointwise estimate

$$T^*(f_1, \dots, f_{n-1})(x) \lesssim \left( \prod_{i=1}^{n-1} M|f_i|^{p_i/p'_n} \right)^{p'_n/p_i}(x),$$

at which point the claim follows from the Hardy-Littlewood maximal inequality.

To prove Theorem 2.1, it suffices to prove the following “restricted weak-type” analogue. For any measurable  $E \subset \mathbf{R}$ , let  $X(E)$  denote the space of functions supported on  $E$  which are bounded in magnitude by 1.

**Theorem 2.2** (Second reduction). *Let  $n \geq 3$ , and let  $\Sigma$  and  $T^*$  be as in Theorem 2.1. Let  $E_1, \dots, E_n$  be subsets of  $\mathbf{R}$  of positive finite measure. Let  $p_1, \dots, p_n$  be such that  $1 < p_1, \dots, p_{n-1} < 2$ ,  $\frac{5}{2} - n < 1/p_n < 3 - n$ , and*

$$\frac{1}{p'_n} = \frac{1}{p_1} + \dots + \frac{1}{p_{n-1}}.$$

*Then there exists a subset  $E'_n$  of  $E_n$  with  $|E'_n| \geq \frac{1}{2}|E_n|$  such that one has*

$$\left| \int T^*(g_1, \dots, g_{n-1})g_n \right| \lesssim |E_1|^{1/p_1} \dots |E_n|^{1/p_n}$$

*for all  $g_1 \in X(E_1), \dots, g_{n-1} \in X(E_{n-1}), g_n \in X(E'_n)$ . Here the implied constant is allowed to depend on  $n, p_1, \dots, p_n$  and  $\Sigma$ .*

In the notation of [18], Theorem 2.2 asserts that the  $n$ -sublinear form  $\int T^*(f_1, \dots, f_{n-1})f_n$  is of restricted type  $(1/p_1, \dots, 1/p_n)$  with  $n$  as the bad index. The deduction of Theorem 2.1 from Theorem 2.2 follows from a variant of the Marcinkiewicz interpolation theorem and is a minor modification of the argument in [18, Lemma 3.11]; the details will be omitted here. The point of Theorem 2.2 is that the functions  $g_1, \dots, g_n$  have been normalized; indeed  $g_j$  can be thought of as essentially the indicator function of  $E_j$  (or  $E'_n$  when  $j = n$ ).

By a limiting argument we may take  $E_1, \dots, E_n$  to be finite unions of intervals, and  $g_1, \dots, g_n$  to be smooth; this allows us to justify a number of formal computations in the sequel without difficulty, and we shall do so without any further comment.

To prove Theorem 2.2, we may apply a rescaling argument to normalize  $|E_n| = 1$ . From the Hardy-Littlewood maximal inequality we may then set

$$(12) \quad E'_n := E_n \setminus \Omega$$

where  $\Omega$  is the exceptional set

$$(13) \quad \Omega := \bigcup_{i=1}^n \{M1_{E_i} \geq C|E_i|\}$$

for a sufficiently large absolute constant  $C$ , so that  $|E'_n| \sim 1$ . It is convenient to renormalize for each  $i < n$ ,  $\alpha_i := 1/p_i - 1/2$  and  $f_i := g_i/|E_i|^{1/2}$ ; thus  $f_i$  lives in the  $L^2$ -normalized space  $X_2(E_i)$  of functions supported on  $E_i$  and bounded in magnitude by  $1/|E_i|^{1/2}$ . We also set  $\alpha_n := \frac{n-2}{2} - \alpha_1 - \dots - \alpha_{n-1}$ ; thus  $0 < \alpha_n < 1/2$ . Theorem 2.2 now reduces to

**Theorem 2.3** (Third reduction). *Let  $n \geq 3$ , and let  $\Sigma$  and  $T^*$  be as in Theorem 2.1. Let  $E_1, \dots, E_n$  be finite unions of intervals with  $|E_n| = 1$ , and let  $E'_n$  be defined by (12), (13), so that  $|E'_n| \sim 1$ . Then one has*

$$\left| \int T^*(f_1, \dots, f_{n-1})f_n \right| \lesssim |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}$$

for all smooth  $f_1 \in X_2(E_1), \dots, f_{n-1} \in X_2(E_{n-1}), f_n \in X_2(E'_n)$  and any  $0 < \alpha_1, \dots, \alpha_n < 1/2$  with  $\alpha_1 + \dots + \alpha_n = \frac{n-2}{2}$ . The implied constant can depend on  $n, \alpha_1, \dots, \alpha_n, \Sigma$ .

This reduction is slightly more convenient to work in as the  $L^2$  normalization of  $f_1, \dots, f_n$  will be useful for a certain “(maximal) Bessel inequality” which is crucial to a later stage of the argument.

### 3. FOURIER REPRESENTATION

Our task is now to prove Theorem 2.3. As in [15], we begin by replacing the rather rough truncation in (11) by a smoother one which has a more tractable Fourier representation. As is customary, for any  $f \in L^1(\mathbf{R})$ , we define the Fourier transform

$$\hat{f}(\xi) := \int_{\mathbf{R}} e^{-2\pi\sqrt{-1}x\xi} f(x) dx$$

and the inverse Fourier transform

$$\check{f}(x) := \int_{\mathbf{R}} e^{2\pi\sqrt{-1}x\xi} f(\xi) d\xi;$$

we use  $\sqrt{-1}$  here instead of  $i$  in order to free up the letter  $i$  for use as an integer-valued index.

Let us fix the hyperplane  $\Sigma$ . We view the hyperplane  $\Sigma$  as an  $(n-2)$ -dimensional Euclidean space with Lebesgue measure  $d\vec{t}$ , and thus endowed with its own Fourier transform; thus if  $\theta$  is a Schwartz function on  $\Sigma$  we have the inverse Fourier transform

$$\check{\theta}(\vec{t}) := \int_{\Sigma} e^{2\pi\sqrt{-1}\vec{t}\cdot\vec{\xi}}\theta(\vec{\xi}) \, d\xi.$$

We now introduce the multilinear operator

$$T_{\theta}(f_1, \dots, f_{n-1})(x) := \int_{\Sigma} \left(\prod_{j=1}^{n-1} f_j(x + t_j)\right)\check{\theta}(\vec{t})d\vec{t};$$

this operator can also be written in Fourier space as

$$T_{\theta}(f_1, \dots, f_{n-1}) = C_{\Sigma} \int_{\mathbf{R}^{n-1}} \left(\prod_{j=1}^{n-1} \hat{f}_j(\xi_j)\right)\theta(\pi(\xi))e^{2\pi\sqrt{-1}x(\xi_1+\dots+\xi_{n-1})}d\xi^{\vec{}}$$

where  $\pi : \mathbf{R}^{n-1} \rightarrow \Sigma$  is the orthogonal projection onto  $\Sigma$  and  $C_{\Sigma} > 0$  is a normalization constant depending only on  $\Sigma$ . For any integer  $k$ , write  $\theta_k(\xi) := \theta(2^k\xi)$ . We define the associated maximal function  $T_{\theta}^*$  as

$$T_{\theta}^*(f_1, \dots, f_{n-1})(x) := \sup_{k \in \mathbf{Z}} |T_{\theta_k}(f_1, \dots, f_{n-1})(x)|.$$

We shall deduce Theorem 2.3 from

**Theorem 3.1** (Fourth reduction). *Let  $n \geq 3$ , and let  $\Sigma$  be as in Theorem 2.1. Let  $0 < \alpha_1, \dots, \alpha_n < 1/2$  with  $\alpha_1 + \dots + \alpha_n = \frac{n-2}{2}$ . Let  $\theta$  be a smooth function supported on a ball  $\{\xi \in \Sigma : \|\xi\| \leq 4\}$  which is constant on a ball  $\{\xi \in \Sigma : \|\xi\| \leq 1/4\}$ , and obeys the estimate*

$$(14) \quad |\check{\theta}(t)| \lesssim \frac{1}{(1 + \|t\|)^{N^3}} \text{ for all } t \in \Sigma$$

for some large integer  $N$  depending on  $\alpha_1, \dots, \alpha_n$ . Let  $E_1, \dots, E_n$  be finite unions of intervals with  $|E_n| = 1$ , and let  $E'_n$  be defined by (12), (13). Then one has

$$\left| \int T_{\theta}^*(f_1, \dots, f_{n-1})f_n \right| \lesssim |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}$$

for all smooth  $f_1 \in X_2(E_1), \dots, f_{n-1} \in X_2(E_{n-1}), f_n \in X_2(E'_n)$ . The implied constant can depend on  $\Sigma, \alpha_1, \dots, \alpha_{n-1}, N$  and on the implicit constant in (14).

*Proof of Theorem 2.3 assuming Theorem 3.1.* We may take  $f_1, \dots, f_{n-1}$  nonnegative.

Let  $\eta$  be a fixed real-valued symmetric Schwartz function<sup>5</sup> on  $\Sigma$  supported on the ball  $\{\|\xi\| \leq 1\}$  whose Fourier transform is nonnegative and  $\check{\eta}(0) = 1$ . Observe that

$$T_{\eta_k}(f_1, \dots, f_{n-1})(x) := \frac{1}{2^{k(n-2)}} \int_{\Sigma} \left(\prod_{j=1}^{n-1} f_j(x + t_j)\right)\check{\eta}(\vec{t}/2^k)d\vec{t}.$$

From this, the positivity of the  $f_i$  and  $\check{\eta}$ , it is easy to establish the pointwise estimate

$$T^*(f_1, \dots, f_{n-1})(x) \lesssim T_{\eta}^*(f_1, \dots, f_{n-1})(x)$$

<sup>5</sup>Such a function can be constructed by starting with a real-valued symmetric function on the ball  $\{\|\xi\| \leq 1/2\}$ , then convolving it with itself and normalizing it.

(where the implied constant depends on  $\eta$ ), so it suffices to show that

$$|\int T_\eta^*(f_1, \dots, f_{n-1})f_n| \lesssim |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}.$$

We cannot yet apply Theorem 3.1, because  $\eta$  is not constant near the origin. Indeed the requirement that  $\tilde{\eta}$  be nonnegative forces  $\eta$  to have a negative Laplacian at the origin. Fortunately, we can rectify this by a further dyadic decomposition. More precisely, we split

$$\eta(\xi) = \eta_2(\xi) + \sum_{l=-\infty}^0 \phi_l(\xi)$$

with  $\eta_2$  smooth, symmetric, supported in  $\|\xi\| \leq 11/10$  and equal to 1 on  $\|\xi\| \leq 1$ , while  $\phi_l(\xi) := (\eta - \eta_2)(\xi)(\eta_2(\xi/2^l) - \eta_2(\xi/2^{l-1}))$ . One can easily verify that the function  $\eta_2$  is already of the form required for Theorem 3.1 and so  $T_{\eta_2}^*$  gives an acceptable contribution to  $T_\eta^*$ . As for the tail terms  $\phi_l$ , we observe the Fourier estimates

$$|\frac{1}{2^l} \check{\phi}_l(\frac{\xi}{2^l})| \lesssim 2^{-|l|} \frac{1}{(1 + \|\xi\|)^{N^3}}$$

uniformly in  $l$ . Also,  $\phi_l$  is constant on  $\|\xi\| \leq 2^l/4$  and zero when  $\|\xi\| \geq 4 \times 2^l$ . A simple rescaling argument using Theorem 3.1 (noting that  $T_\phi^*$  is unchanged if one replaces  $\phi$  by  $\phi(2^l \cdot)$ ) then shows that

$$|\int T_{\phi_l}^*(f_1, \dots, f_{n-1})f_n| \lesssim 2^{-|l|} |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}.$$

The claim now follows from the triangle inequality. □

#### 4. DISCRETIZATION

It remains to prove Theorem 3.1. We now perform the usual dyadic decompositions to reduce matters to estimating a certain sum over dyadic objects, namely a collection of “multitiles”, after first doing some additional refinements to ensure that these multitiles obey some good geometrical properties (specifically, a rank one condition).

We introduce two large constants  $1 \ll C_0 \ll C_1$  (depending on  $\Sigma$ , and  $C_1$  assumed to be large compared to  $C_0$ ) that will be used to sparsify the time-frequency geometry. We will take some care to specify how the implied constants in the  $\lesssim$  notation depend on  $C_0$  and  $C_1$ ; however we will allow these constants to depend freely on  $n, \alpha_1, \dots, \alpha_n, N, \Sigma$ .

It will be convenient to dilate  $\theta$  by  $C_0$ , so that  $\theta$  is now supported on  $\{\xi \in \Sigma : \|\xi\| \leq 4C_0\}$ , which is constant on a ball  $\{\xi \in \Sigma : \|\xi\| \leq C_0/4\}$ ; this affects our final bounds by some factor depending on  $C_0$ , but as we shall eventually choose  $C_0$  to be a quantity depending only on existing parameters such as  $n, \alpha_1, \dots, \alpha_n, N, \Sigma$ , this shall be of no consequence. We perform the dyadic decomposition

$$\theta_k(\xi) = \sum_{i \geq k} \varphi_i(\xi)$$

where  $\varphi(\xi) := \theta(\xi) - \theta(2\xi)$  is a smooth function supported on an annulus  $\|\xi\| \sim C_0$ , and  $\varphi_i(\xi) := \varphi(2^i \xi)$ . Thus

$$T_{\theta_k} = \sum_{i \geq k} T_{\varphi_i},$$



and hence for any  $f_1, \dots, f_n$  we have

$$\int T_\theta^*(f_1, \dots, f_{n-1})f_n = \sum_i \int T_{\varphi_i}(f_1, \dots, f_{n-1})(x)f_n(x)1_{i \geq k(x)} dx$$

for some integer-valued measurable function  $k : \mathbf{R} \rightarrow \mathbf{Z}$ . Thus it suffices to establish the multilinearized estimate

$$(15) \quad \left| \sum_i \int T_{\varphi_i}(f_1, \dots, f_{n-1})(x)f_n(x)1_{i \geq k(x)} dx \right| \lesssim_{C_0, C_1} |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}$$

for each such function  $k : \mathbf{R} \rightarrow \mathbf{Z}$ , which we now fix. Note that we can write the left-hand side as

$$(16) \quad \int_{\mathbf{R}} \int_{\Sigma} \left( \prod_{j=1}^n f_{j,i}(x + t_i) \right) \check{\varphi}_i(\vec{t}) d\vec{t} dx$$

where  $f_{j,i} := f_j$  for  $1 \leq j \leq n - 1$  and  $f_{n,i}(x) := f_n(x)1_{i \geq k(x)}$ , and we adopt the convention that  $t_n = 0$ . One should think of  $i$  as a scale parameter, corresponding to the terms with frequency uncertainty  $\sim 2^{-i}$  and time uncertainty  $\sim 2^i$ . Note that the annulus that  $\varphi_i$  is supported in has thickness  $\sim C_0 2^{-i}$  and can thus tolerate the frequency uncertainty associated with the scale  $i$ .

The next (standard) step is wave packet decomposition. We shall adopt the usual trick of covering the time domain  $\mathbf{R}$  by three overlapping dyadic grids to eliminate some artificial boundary effects caused by dyadicity.

For each  $1 \leq j \leq n$ , let us pick a Schwartz function  $\psi_j$  such that  $\hat{\psi}_j$  is supported in  $[0.1, 0.9]$ , and that  $\psi_j$  is rapidly decreasing; in particular we have the bounds

$$(17) \quad |\psi_j(x)| \lesssim (1 + |x|)^{-10N} \text{ for all } x \in \mathbf{R}$$

and we have the following property for every  $\xi \in \mathbf{R}$ :

$$\sum_{l \in \mathbf{Z}} \left| \hat{\psi}_j \left( \xi - \frac{l}{3} \right) \right|^2 = 1.$$

This is possible because the translates of  $[0.1, 0.9]$  by integer multiples of  $\frac{1}{3}$  cover the real line  $\mathbf{R}$  with some room to spare for smooth cutoffs. For each scale  $i \in \mathbf{Z}$  we can then decompose

$$f_{i,j} = \sum_{m,l \in \mathbf{Z}} \langle f_{i,j}, \psi_{j,i,m,\frac{l}{3}} \rangle \psi_{j,i,m,\frac{l}{3}},$$

where

$$\psi_{j,i,m,l}(x) := 2^{-\frac{i}{2}} \psi_j(2^{-i}x - m) e^{2\pi \sqrt{-1} 2^{-i}xl}$$

and  $\langle f, g \rangle := \int f \bar{g}$  is the usual inner product. Inserting this decomposition into (15), (16) and using the triangle inequality, we reduce to showing that

$$(18) \quad \sum_{i \in \mathbf{Z}} \sum_{\vec{m}, \vec{l} \in \mathbf{Z}^n} C_{\vec{m}, \vec{l}, i} 2^{i(1-\frac{\alpha}{2})} \prod_{j=1}^n |\langle f_{j,i}, \psi_{j,i,m_j,l_j} \rangle| \lesssim_{C_0, C_1} |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}$$

where  $\vec{m} = (m_1, \dots, m_n)$ ,  $\vec{l} = (l_1, \dots, l_n)$ , and  $C_{\vec{m}, \vec{l}, i}$  are the operator coefficients

$$(19) \quad C_{\vec{m}, \vec{l}, i} := \frac{1}{2^{i(1-\frac{\alpha}{2})}} \left| \int_{\mathbf{R}} \int_{\Sigma} \prod_{j=1}^n \psi_{j,i,m_j,l_j}(x + t_j) \check{\theta}_i(\vec{t}) d\vec{t} dx \right|.$$

One should think of  $\vec{m}$  as containing the time location, and  $\vec{l}$  as containing the frequency location information; roughly speaking, the summand in (18) is the contribution when  $f_j$  is localized in space to  $2^i m_j + O(2^i)$  and localized in frequency to  $2^{-i} l_j + O(2^{-i})$ .

We now use the geometry of the hyperplane  $\Sigma$  to obtain localization estimates on the coefficients  $C_{\vec{m}, \vec{l}, i}$ . We let  $\Gamma \subset \mathbf{R}^n$  denote the hyperplane  $\Gamma := \{(\xi_1, \dots, \xi_n) : \xi_1 + \dots + \xi_n = 0\}$ .

**Lemma 4.1.** *We have the estimate*

$$(20) \quad C_{\vec{m}, \vec{l}, i} \lesssim_{C_0, C_1} (1 + \text{diam}\{m_1, \dots, m_n\})^{-N^2}.$$

Furthermore, if  $C_{\vec{m}, \vec{l}, i}$  is nonzero, then

$$(21) \quad l_1 + \dots + l_n = O(1)$$

and

$$(22) \quad \|\pi(l_1, \dots, l_{n-1})\| \sim C_0.$$

*Remark 4.1.* In the notation of [18], these conditions are essentially asserting that the tuples  $(\vec{m}, \vec{l}, i)$  with a sizeable coefficient  $C_{\vec{m}, \vec{l}, i}$  form a collection of multitiles of rank one (which is also the situation with the bilinear Hilbert transform). See also Definition 4.3 below.

*Proof.* We first observe that by rescaling by  $2^i$  that  $C_{\vec{m}, \vec{l}, i}$  is actually independent of  $i$ . Thus we may assume  $i = 0$  throughout the proof.

To prove (20), we then use the physical space representation (19) of  $C_{\vec{m}, \vec{l}, 0}$ , followed by the triangle inequality, to obtain

$$C_{\vec{m}, \vec{l}, 0} \lesssim \int_{\mathbf{R}} \int_{\Sigma} \left| \left( \prod_{j=1}^n \psi_j(x + t_j - m_j) \right) \check{\theta}_0(\vec{t}) \right| d\vec{t} dx.$$

Now as  $\psi$  is rapidly decreasing, we conclude from (14) that

$$C_{\vec{m}, \vec{l}, 0} \lesssim_{C_0, C_1} \int_{\mathbf{R}} \int_{\Sigma} (1 + \|t\|)^{-2N^2} \prod_{j=1}^n (1 + |x + t_j - m_j|)^{-2N^2} d\vec{t} dx$$

(say), and the claim (20) follows from the pointwise estimate

$$\prod_{j=1}^n (1 + |x + t_j - m_j|)^{-N^2} \lesssim (1 + \text{diam}\{m_1, \dots, m_n\})^{-N^2} (1 + \|t\|)^{N^2}.$$

Now suppose that  $C_{\vec{m}, \vec{l}, i}$  is nonzero. To exploit this we use the Fourier representation, converting (19) to

$$C_{\vec{m}, \vec{l}, 0} \sim \left| \int_{\Gamma} \left( \prod_{j=1}^n \hat{\psi}(\xi_j - l_j) \right) \theta(\pi(\xi_1, \dots, \xi_{n-1})) d\vec{\xi} \right|.$$

Thus there exists  $\vec{\xi} \in \Gamma$  such that  $\xi_j - l_j$  lies in the support of  $\hat{\psi}$  for all  $1 \leq j \leq n$  and  $\pi(\xi_1, \dots, \xi_{n-1})$  lies in the support of  $\theta$ . From the former property we have  $l_j = \xi_j + O(1)$ , and (21) follows from the definition of  $\Gamma$ . From the latter property we have  $\|\pi(\xi_1, \dots, \xi_{n-1})\| \sim C_0$ , and the claim follows by using the approximation  $l_j = \xi_j + O(1)$  and the homogeneity of  $\pi$ .  $\square$

In view of the above lemma, it now suffices to show that

$$(23) \quad \sum_{(\vec{m}, \vec{l}, i) \in \Omega} (1 + \text{diam}\{m_1, \dots, m_n\})^{-100n^2} 2^{i(1-\frac{\alpha}{2})} \prod_{j=1}^n |\langle f_{j,i}, \psi_{j,i,m_j,l_j} \rangle| \\ \lesssim_{C_0, C_1} |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}$$

where  $\Omega$  is a collection of triples  $(\vec{m}, \vec{l}, i) \in \mathbf{Z}^n \times \mathbf{Z}^n \times \mathbf{Z}$  obeying (21) and (22).

We now perform a number of refinements to improve the nesting properties of the set  $\Omega$ . First we observe that for each  $(\vec{m}, \vec{l}, i) \in \Omega$  and  $1 \leq j \leq n$ , the Fourier transform of  $\psi_{j,i,m_j,l_j}$  is contained in the interval  $[2^{-i} \frac{l_j}{3}, 2^{-i}(\frac{l_j}{3} + 1)]$  (in fact they are contained in the slightly smaller interval  $[2^{-i}(\frac{l_j}{3} + 0.1), 2^{-i}(\frac{l_j}{3} + 0.9)]$ ). These intervals are almost dyadic, but for the denominator of 3. However this factor of 3 can be eliminated in the following standard manner. Let  $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$  be the dyadic grids

$$(24) \quad \begin{aligned} \mathcal{D}_0 &:= \{[2^{-i}l, 2^{-i}(l + 1)] : i, l \in \mathbf{Z}\}, \\ \mathcal{D}_1 &:= \{[2^{-i}(l + (-1)^i/3), 2^{-i}(l + 1 + (-1)^i/3)] : i, l \in \mathbf{Z}\}, \\ \mathcal{D}_2 &:= \{[2^{-i}(l - (-1)^i/3), 2^{-i}(l + 1 - (-1)^i/3)] : i, l \in \mathbf{Z}\}. \end{aligned}$$

Thus  $\mathcal{D}_0$  is the standard dyadic grid, and the other two grids are essentially similar (one can view the latter two grids as translates of the first by the nonterminating 2-adic  $\pm 1/3$ ). In particular, within a single grid we have the nesting property that if two intervals intersect, then the shorter one is contained by the longer one. Observe that every interval  $[2^{-i} \frac{l_j}{3}, 2^{-i}(\frac{l_j}{3} + 1)]$  belongs to one of these three grids. By pigeonholing once for each  $j$  (conceding a factor of  $3^n$  in the estimates), we can assume that for fixed  $j$ , the intervals  $[2^{-i} \frac{l_j}{3}, 2^{-i}(\frac{l_j}{3} + 1)]$  belong to a single dyadic grid. For ease of exposition we shall assume that these intervals always lie in the standard dyadic grid  $\mathcal{D}_0$ ; thus the intervals  $[2^{-i} \frac{l_j}{3}, 2^{-i}(\frac{l_j}{3} + 1)]$  are genuine dyadic intervals. The other cases are handled similarly but with some minor changes in notation.

Morally speaking, the localizing factor  $(1 + \text{diam}\{m_1, \dots, m_n\})^{-N^2}$  in (23) implies that the diagonal contribution  $m_1 = \dots = m_n$  is the dominant contribution. Again to simplify the exposition, we shall focus entirely on this diagonal case  $m_1 = \dots = m_n$ . We now briefly sketch how to pass from the diagonal case to the general case. Write  $m_j = m_1 + r_j$ . For each fixed  $(n-1)$ -tuple of integers  $r_2, \dots, r_n$ , one can convert the case  $m_j = m_1 + r_j$  to the diagonal case  $m_j = m_1$  by shifting the function  $\psi_j$  by  $r_j$ . This affects the bounds (17) but only by  $(1 + |r_j|)^{10N}$  at worst. This gives a total loss of  $\prod_{j=1}^n (1 + |r_j|)^{10nN}$  for this contribution, but one is also gaining a factor of  $(1 + \text{diam}(0, r_2, \dots, r_n))^{-N^2}$ , and the product is then summable in  $r$  if  $N$  is large enough. Thus it suffices to treat the diagonal case.

Another application of the pigeonhole principle (giving up a constant factor of  $C_1$  in the estimates) allows one to refine the scale parameter  $i$  to not take values in the integers, but to instead take values in a residue class  $\{i = c \pmod{C_1}\}$  for some residue  $c$ . This “sparsification” of the scales will be useful in obtaining a certain rank separation condition in the frequencies below.

Finally, we analyze the conditions (21) and (22). Observe that if we instead had the exact constraints  $l_1 + \dots + l_n = 0$  and  $\pi(l_1, \dots, l_{n-1}) = 0$ , then  $(l_1, \dots, l_n)$  would be restricted to a one-dimensional subspace of  $\mathbf{R}^n$ . Since  $\Sigma$  did not contain

$e_1, \dots, e_{n-1}$  or  $(1, \dots, 1)$ , it is easy to see that the nonzero vectors in this one-dimensional subspace have no zero coordinates; thus we have  $l_j = c_{j,j'} l_{j'}$  for all  $1 \leq j, j' \leq n$  and some explicit nonzero finite constants  $c_{j,j'}$  depending only on  $\Sigma$ ; furthermore we have  $c_{j,j} = 1$ ,  $c_{1,j'} + \dots + c_{n,j'} = 0$  and  $\pi(c_{1,j'}, \dots, c_{n-1,j'}) = 0$ . Returning now to the inexact constraints (21), (22), we conclude that

$$l_j = c_{j,j'} l_{j'} + O(C_0)$$

for all  $1 \leq j, j' \leq n$ . By pigeonholing (and conceding a factor of  $C_0^{m^2}$  at worst) we may thus assume that

$$(25) \quad l_j = \lfloor c_{j,j'} l_{j'} \rfloor + a_{j,j'}$$

on  $\Omega$  for all  $1 \leq j, j' \leq n$  and some fixed integers  $a_{j,j'} = O(C_0)$ ; note that  $a_{j,j}$  is necessarily zero. Thus each frequency  $l_j$  is now uniquely determined by any of the other frequencies  $l_{j'}$ . Furthermore, from (21), (22) we have

$$a_{1,j'} + \dots + a_{n,j'} = O(1) \text{ and } \|\pi(a_{1,j'}, \dots, a_{n-1,j'})\| \sim C_0.$$

If  $C_0$  is large enough, this implies the following basic fact:

**Lemma 4.2.** *For each  $j'$ , there exist at least two  $j$  distinct from  $j'$  such that  $|a_{j,j'}| \sim C_0$ .*

The upshot of this lemma is that whenever we fix one of the frequencies of  $f_1, \dots, f_n$ , at least two other frequencies depend in a ‘‘lacunary’’ manner on the scale parameter  $i$ . This fact will be crucial in controlling the geometry of certain ‘‘trees’’ which will appear later.

The estimate (23) has now been reduced to

$$(26) \quad \sum_{(\vec{m}, \vec{l}, i) \in \Omega} 2^{i(1-\frac{n}{2})} \prod_{j=1}^n |\langle f_{j,i}, \psi_{j,i,m_j,l_j} \rangle| \lesssim |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}.$$

We now convert (26) into the more traditional language of multitiles and wave packets.

**Definition 4.2 (Tiles).** A *tile*  $P$  is a rectangle  $P = I_P \times \omega_P$  with both  $I_P$  and  $\omega_P$  dyadic intervals, obeying the Heisenberg relation  $|I_P| \cdot |\omega_P| = 1$ ; we refer to  $I_P$  as the *time interval* of  $P$  and  $\omega_P$  as the *frequency interval*. A *multitile*  $s$  is an  $n$ -tuple  $s = (s_1, \dots, s_n)$  of tiles with the same time interval  $I_s := I_{s_1} = \dots = I_{s_n}$ . If  $I$  is an interval and  $C > 0$  is a number, we let  $CI$  denote the interval with the same center as  $I$  but  $C$  times the length (note that this interval will most likely not be dyadic). Let us say that a function  $\psi_P$  is a *wave packet adapted to a tile*  $P$  if  $\hat{\psi}_P$  is supported in  $0.8\omega_P$  and we have the pointwise estimate

$$(27) \quad |\psi_P(x)| \lesssim |I_P|^{-1/2} \chi_{I_P}^{10N}(x) \text{ for all } x \in \mathbf{R}$$

where for any interval  $I$ ,  $\chi_I$  is the weight function

$$\chi_I(x) := \left(1 + \frac{(x - c(I))^2}{|I|^2}\right)^{-1/2}$$

and  $c(I)$  is the center of  $I$ ; in particular observe that  $\psi_P$  is normalized to have an  $L^2$  norm of  $O(1)$ .

Note that because of all the reductions we have already achieved, every triple  $(\vec{m}, \vec{l}, i)$  in  $\Omega$  gives rise to a multitile  $s$  with  $s_j := [2^i m_j, 2^i(m_j + 1)] \times \omega_{s_j} := [2^{-i} \frac{l_j}{3}, 2^{-i}(\frac{l_j}{3} + 1)]$ . In particular we have  $|I_s| = 2^i$ . Let  $\mathbf{S}_{\max}$  denote the collection of all multitiles obtained this way. For each multitile  $s \in \mathbf{S}_{\max}$  arising from a triple  $(\vec{m}, \vec{l}, i)$ , define the functions  $\psi_{s,j}$  for  $1 \leq j \leq n$  by setting

$$\psi_{s,j}(x) := \psi_{j,i,m_j,l_j}(x).$$

Observe that for each  $1 \leq j \leq n$ ,  $\psi_{s,j}$  is a wave packet adapted to  $s_j$ . We also observe the following important consequence of Lemma 4.2.

**Definition 4.3** (Rank one). A collection  $\mathbf{S}$  of multitiles is said to have *rank one* if for every  $j \in \{1, \dots, n\}$  there exist distinct  $j_1(j), j_2(j) \in \{1, \dots, n\} \setminus \{j\}$  and signs  $\epsilon_1(j), \epsilon_2(j) \in \{-1, +1\}$  (not necessarily distinct) with the following properties.

- (Scale separation) If  $s, s' \in \mathbf{S}$  are such that  $|\omega_{s_j}| > |\omega_{s'_j}|$ , then  $|\omega_{s_j}| \geq 2^{C_1} |\omega_{s'_j}|$ .
- (One independent frequency parameter) If  $s, s' \in \mathbf{S}$  are such that  $\omega_{s_j} = \omega_{s'_j}$ , then  $\omega_{s_{j'}} = \omega_{s'_{j'}}$  for all  $1 \leq j' \leq n$ .
- (Nearby  $j$ -frequencies implies nearby  $j'$ -frequencies) If  $s, s' \in \mathbf{S}$  are such that  $10\omega_{s_j} \cap 10\omega_{s'_{j'}} \neq \emptyset$  and  $|I_s| \geq |I_{s'}|$ , then  $\text{dist}(\omega_{s_{j'}}, \omega_{s'_{j'}}) \lesssim C_0 |I_{s'}|^{-1}$  for all  $1 \leq j' \leq n$ .
- (Lacunarity property) If  $s, s' \in \mathbf{S}$  are such that  $10\omega_{s_j} \cap 10\omega_{s'_{j_t}} \neq \emptyset$  and  $|I_s| > |I_{s'}|$ , then  $\text{dist}(\omega_{s_{j_t}}, \omega_{s'_{j_t}}) \sim C_0 |I_{s'}|^{-1}$  for  $t = 1, 2$ . In particular  $10\omega_{s_{j_t}}$  and  $10\omega_{s'_{j_t}}$  are disjoint. Furthermore we require  $\epsilon_t(j)(\xi' - \xi) \geq 0$  for all  $\xi \in 10\omega_{s_{j_t}}$  and  $\xi' \in 10\omega_{s'_{j_t}}$ .

*Remark 4.4.* For the definition of higher order rank (which we will not need here), see [18]. Actually our definition of rank one is slightly stronger than that in [18] in that we require the indices  $j_1, j_2, \epsilon_1(j), \epsilon_2(j)$  to depend only on  $j$ , and not be dependent on  $s, s'$ , but this is only a minor technical change.

**Lemma 4.3** (Rank one property).  $\mathbf{S}_{\max}$  has rank one.

This lemma shows, among other things, that the multitiles in  $\mathbf{S}_{\max}$  have essentially one independent frequency parameter. Note that if  $\mathbf{S}_{\max}$  has rank one, then so does any subset  $\mathbf{S}$  of  $\mathbf{S}_{\max}$ .

*Proof.* The scale separation property follows since for each multitile  $s \in \mathbf{S}_{\max}$ , we have  $|\omega_{s_j}| = |I_s|^{-1} = 2^{-i}$  for all  $1 \leq j \leq n$  and some integer  $i = C \pmod{C_1}$ . The remaining properties follow from (25) and Lemma 4.2, setting  $j_1, j_2$  to be the indices distinct from  $j$  such that  $|a_{j_1,j}|, |a_{j_2,j}| \sim C_0$ , and  $\epsilon_t(j)$  to be the sign of  $a_{j_t,j}$ .  $\square$

We also define the modified wave packets  $\phi_{s,j}$  by setting

$$(28) \quad \phi_{s,j} := \psi_{s,j} \text{ for } 1 \leq j \leq n - 1$$

and

$$(29) \quad \phi_{s,n}(x) := \psi_{s,n}(x) 1_{|I_s| > 2^{k(x)}}.$$

The estimate (26) can now be rewritten as

$$\sum_{s \in \mathbf{S}_{\max}} |I_s|^{(1-\frac{\alpha}{2})} \prod_{j=1}^n |\langle f_j, \phi_{s,j} \rangle| \lesssim |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}.$$

By the monotone convergence theorem we can replace  $\mathbf{S}_{\max}$  by a finite subset  $\mathbf{S}$  of  $\mathbf{S}_{\max}$ , so long as our estimates are uniform in  $\mathbf{S}$ . Note that the properties in Lemma 4.3 will be preserved if we pass from  $\mathbf{S}_{\max}$  to  $\mathbf{S}$ . We can now deduce (26) (and hence Theorem 1.1) from the following more abstract result.

**Theorem 4.4** (Fifth reduction). *Let  $n \geq 3$ , let  $0 < \alpha_1, \dots, \alpha_n < 1/2$  with  $\alpha_1 + \dots + \alpha_n = \frac{n-2}{2}$ , and let  $N$  be a sufficiently large integer depending on  $\alpha_1, \dots, \alpha_n$ . Let  $\mathbf{S}$  be a finite collection of multitiles which is rank one. For each  $s \in \mathbf{S}$  and  $1 \leq j \leq n$ , let  $\psi_{s,j}$  be a wave packet adapted to  $s_j$ . Let  $k : \mathbf{R} \rightarrow \mathbf{Z}$  be an arbitrary measurable function, and let  $\phi_{s,j}$  be defined by (28), (29). Let  $E_1, \dots, E_n$  be finite unions of intervals with  $|E_n| = 1$ , and let  $E'_n$  be defined by (12), (13). Then one has*

$$\sum_{s \in \mathbf{S}} |I_s|^{(1-\frac{n}{2})} \prod_{j=1}^n |\langle f_j, \phi_{s,j} \rangle| \lesssim |E_1|^{\alpha_1} \dots |E_n|^{\alpha_n}$$

for all smooth  $f_1 \in X_2(E_1), \dots, f_{n-1} \in X_2(E_{n-1}), f_n \in X_2(E'_n)$ . The implied constant can depend on  $\alpha_1, \dots, \alpha_{n-1}, N$  and on the bounds in the rank one condition and (27) but is uniform in  $\mathbf{S}$ .

*Remark 4.5.* If the  $\phi_{s,j}$  were replaced by  $\psi_{s,j}$  (i.e. if the cutoff  $|I_s| > 2^{k(x)}$  were not present), then this result would follow from the results in [18]. Thus the novelty (which is also present in [15]) is the cutoff  $|I_s| > 2^{k(x)}$ , which ultimately arises from the maximal function nature of  $T_{A,\mathbf{R}}^*$ .

### 5. TREES

It remains to prove Theorem 4.4. To do this we use the standard strategy of organizing the multitiles into trees, estimating the contribution of each tree separately, controlling the total number of trees of a certain “size”, and then summing up.

Henceforth we fix the tile collection  $\mathbf{S}$  and the functions  $f_1, \dots, f_n$  and sets  $E_1, \dots, E_n$ , as well as the exponents  $\alpha_1, \dots, \alpha_n$  and  $N$ , the function  $k(x)$  and the wave packet functions  $\psi_{s,j}$  (which of course determine  $\phi_{s,j}$ ). We now recall a standard notion of tile order.

**Definition 5.1** (Tile order). For any two tiles  $P$  and  $P'$ , we write  $P < P'$  if  $I_P \subsetneq I_{P'}$  and  $3\omega_P \supseteq 3\omega_{P'}$ , and  $P \leq P'$  if  $P < P'$  or  $P = P'$ .

Note that this is a partial order on tiles. The factor of 3 is convenient for technical reasons to provide a little more frequency separation; the presence of the large constants  $C_0$  and  $C_1$  in the rank condition will allow us to have this additional factor.

**Definition 5.2** (Trees). A multitile tree, or tree for short, is a triplet  $(\mathbf{T}, T, i)$  where  $1 \leq i \leq n$  is the index of the tree,  $T \in \mathbf{S}$  is a multitile, and  $\mathbf{T} \subset \mathbf{S}$  is a collection of multitiles such that  $s_i \leq T_i$  for all  $s \in \mathbf{T}$ . We shall often abuse notation and abbreviate a tree  $(\mathbf{T}, T, i)$  as  $\mathbf{T}$ . We refer to  $I_{\mathbf{T}} := I_T$  as the time interval of the tree. If  $1 \leq j \leq n$  and  $\epsilon \in \{-1, +1\}$ , we say that a tree  $(\mathbf{T}, T, i)$  is  $(j, \epsilon)$ -separated if  $j = j_t(i)$  and  $\epsilon = \epsilon_t(i)$  for some  $t \in \{1, 2\}$ . We say that a tree is  $j$ -separated if it is  $(j, \epsilon)$ -separated for some  $\epsilon \in \{-1, +1\}$ .

**Example 5.1.** For any tile  $T$  and  $1 \leq i \leq n$ , the singleton tree  $(\{T\}, T, i)$  is a multitile tree.

*Remark 5.3.* We use the rather clumsy terminology *multitile tree* to distinguish from the notion of a *lacunary tree*, which consists of tiles rather than multitiles, that we will introduce in Section 9. Note that we do not require that the tree  $\mathbf{T}$  contains its top  $T$ , although this is often the case; also note that if  $(\mathbf{T}, T, i)$  is a tree, then so is  $(\mathbf{T} \cup \{T\}, T, i)$  and  $(\mathbf{T} \setminus \{T\}, T, i)$  (so one can always add or remove the top from a tree). This additional flexibility in our definition of a tree (not present in some other literature) is convenient because it makes the notion of a tree more stable with respect to the passage to subsets. In particular, if  $(\mathbf{T}, T, i)$  is a  $(j, \epsilon)$ -separated tree and  $\mathbf{T}' \subset \mathbf{T}$ , then  $(\mathbf{T}', T, i)$  is also a  $j$ -separated tree. Furthermore, if  $\mathbf{T}'$  takes the form  $\mathbf{T}' := \{s \in \mathbf{T} : s_i \leq T'_i\}$  for some multitile  $T'$ , then  $(\mathbf{T}', T', i)$  is also a  $j$ -separated tree.

The rank one condition implies certain geometric facts about trees, which we collect below for the reader's convenience.

**Lemma 5.2.** *Let  $(\mathbf{T}, T, i)$  and  $(\mathbf{T}', T', i)$  be  $(j, \epsilon)$ -separated multitile trees.*

- (i) *The frequency intervals of a multitile in  $\mathbf{T}$  are determined entirely by the size of the spatial interval. In other words, if  $s, s' \in \mathbf{T}$  and  $|I_s| = |I_{s'}|$ , then  $\omega_{s_k} = \omega_{s'_k}$  for all  $1 \leq k \leq n$ .*
- (ii) *Each multitile in  $\mathbf{T}$  has a distinct time interval: if  $s, s' \in \mathbf{T}$  and  $s \neq s'$ , then  $I_s \neq I_{s'}$ .*
- (iii) *If  $s \in \mathbf{T}$  and  $s \neq T$ , then  $\text{dist}(10\omega_{s_j}, 10\omega_{T_j}) \sim C_0|I_s|^{-1}$ ; in particular,  $10\omega_{s_j}$  and  $10\omega_{T_j}$  are disjoint.*
- (iv) *Suppose that  $s \in \mathbf{T}$  and  $s' \in \mathbf{T}'$  are such that  $\omega_{s_j} \subsetneq \omega_{s'_j}$  and  $I_{s'} \cap I_T \neq \emptyset$ . Then  $s'_j < T_j$ , and furthermore we have  $\epsilon(\xi - \xi') > 0$  whenever  $\xi \in \omega_{T_i}$  and  $\xi' \in \omega_{T'_i}$ .*

*Proof.* If  $|I_s| = |I_{s'}|$ , then  $|\omega_{s_j}| = |\omega_{s'_j}|$ ; since these intervals intersect, we must have  $\omega_{s_j} = \omega_{s'_j}$  and then (i) follows from the rank one condition. Property (ii) follows immediately from (i). Now we show (iii). From (ii) we see that  $I_s$  is strictly smaller than  $I_{\mathbf{T}}$ , and so  $\omega_{s_i}$  strictly contains  $\omega_{T_i}$ . The claim then follows from the lacunarity property of the rank condition. Finally, we show (iv). We have  $|\omega_{s_j}| < |\omega_{s'_j}|$  and  $|I_s| \leq |I_T|$  and hence  $|I_{s'}| < |I_T|$ . By dyadic nesting this means that  $I_{s'} \subsetneq I_T$ , and to show that  $s'_j < T_j$  it will suffice to show that  $3\omega_{s'_j}$  intersects  $3\omega_{T_j}$ . But  $\omega_{T_j}$  lies within  $\lesssim C_0|\omega_{s_j}|^{-1}$  of  $\omega_{s_j}$ , which is contained inside  $\omega_{s'_j}$ . Since  $|\omega_{s'_j}| \geq 2^{C_1}|\omega_{s_j}|$  by scale separation, the claim  $s'_j < T_j$  follows if  $C_1$  is sufficiently large depending on  $C_0$ . To show the remaining claim in (iv), we observe from the rank separation condition that  $\text{dist}(\omega_{s'_j}, \omega_{T'_j}) \sim C_0|\omega_{s'_j}|$ , with  $\omega_{T'_j}$  lying below  $\omega_{s'_j}$  if  $\epsilon = +1$  and above if  $\epsilon = -1$ . The claim follows.  $\square$

We can now introduce the concept of size. There will be one size for each of the functions  $f_1, \dots, f_n$ .

**Definition 5.4 (Size).** For a set of multitiles  $\mathbf{S}' \subset \mathbf{S}$  and  $1 \leq j \leq n$  define its  $j$ -size as

$$\text{size}_j(\mathbf{S}') := \sup_{\mathbf{T}} \left( \frac{1}{|I_{\mathbf{T}}|} \sum_{s \in \mathbf{T}} |\langle f_j, \phi_{s,j} \rangle|^2 \right)^{\frac{1}{2}}$$

where the supremum is taken over all the  $j$ -separated trees  $(\mathbf{T}, T, i)$  with  $\mathbf{T} \subset \mathbf{S}'$ .

*Remark 5.5.* In the above definition the trees  $\mathbf{T}$  are not required to contain their top  $T$ . However it is easy to see that a tree without a top can be partitioned into trees with tops that have disjoint time intervals, and because of this one could replace the supremum in the definition of size by a supremum over trees that contain their tops without affecting the size. However we will not need to do this in this paper.

6. HIGH-LEVEL OVERVIEW OF PROOF

Following the usual time-frequency approach, we can now reduce the task of proving Theorem 4.4 to that of verifying a number of lemmas concerning trees.

The first lemma is easy to state and prove:

**Lemma 6.1** (Contribution of a single tree). *If  $(\mathbf{T}, T, i_0)$  is a tree, then*

$$\sum_{s \in \mathbf{T}} |I_s|^{1-\frac{n}{2}} \prod_{i=1}^n |\langle f_i, \phi_{s,i} \rangle| \leq |I_{\mathbf{T}}| \prod_{i=1}^n \text{size}_i(\mathbf{T}).$$

*Proof.* By definition of size we have

$$\left( \sum_{s \in \mathbf{T}} |\langle f_i, \phi_{s,i} \rangle|^2 \right)^{1/2} \leq |I_{\mathbf{T}}|^{1/2} \text{size}_i(\mathbf{T})$$

when  $i = j_1(i_0)$  or  $i = j_2(i_0)$ . Also, since a singleton multitile is always a tree, we also have

$$|\langle f_i, \phi_{s,i} \rangle| \leq |I_s|^{1/2} \text{size}_i(\mathbf{T})$$

for the other  $n - 2$  values of  $i$ . The claim then follows from Hölder’s inequality.  $\square$

In light of this lemma, the task is now to subdivide the collection  $\mathbf{S}$  into distinct trees  $\mathbf{T}$  for which one has the bound

$$(30) \quad \sum_{\mathbf{T}} |I_{\mathbf{T}}| \prod_{i=1}^n \text{size}_i(\mathbf{T}) \lesssim |E_1|^{\alpha_1} \dots |E_{n-1}|^{\alpha_{n-1}}.$$

This will be accomplished via a number of propositions. First we need a basic upper bound on the size of a tree, which we prove in Section 7.

**Proposition 6.2** (Size estimate). *Let  $1 \leq j \leq n$ , let  $\mathbf{S}'$  be a collection of multitiles in  $\mathbf{S}$ , and let*

$$\mathcal{P}_{\mathbf{S}'} := \{I \text{ dyadic} : I_s \subseteq I \subseteq I_{s'} \text{ for some } s, s' \in \mathbf{S}'\}$$

*be the time convexification of  $\mathbf{S}'$ . Then*

$$\text{size}_j(\mathbf{S}') \lesssim |E_j|^{-1/2} \sup_{I \in \mathcal{P}_{\mathbf{S}'}} \frac{1}{|I|} \int_{E_j} \chi_I^N.$$

Note that this bound is consistent with the hypothesis  $f_j \in X_2(E_j)$  and the intuition that the  $j$ -size is something like a BMO average of  $f_j$ .

To decompose the collection of multitiles  $\mathbf{S}$  into trees, we need the following result.

**Lemma 6.3** (Splitting lemma). *Let  $\mathbf{S}'$  be a finite collection of multitiles,  $1 \leq j \leq n$  and suppose that  $\text{size}_j(\mathbf{S}') \leq 2^{m+1}$ . Let  $\mu > 0$  and suppose that  $N$  is sufficiently large depending on  $\mu$ . Then  $\mathbf{S}'$  can be written as a disjoint union*

$$(31) \quad \mathbf{S}' = \left( \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T} \right) \cup \mathbf{S}_2$$



where  $\mathcal{F}$  is a collection of trees such that

$$(32) \quad \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| \lesssim_{\mu} 2^{-2m} \left( \frac{1}{|E_j|^{1/2} 2^m} \right)^{\frac{2}{\mu}},$$

while

$$(33) \quad \text{size}_j(\mathbf{S}_2) \leq 2^m.$$

This lemma is quite difficult and will be proven in Sections 8-13. Assuming the lemma for the moment, we may iterate it in the standard way (see e.g. [20]) and conclude

**Corollary 6.4** (Tree selection algorithm). *Let  $\mathbf{S}'$  be a finite collection of multitiles and  $1 \leq j \leq n$ . Let  $\mu > 0$  and suppose that  $N$  is sufficiently large depending on  $\mu$ . Then, after discarding tiles  $s$  of  $j$ -size zero (in the sense that  $\langle f_j, \phi_{s,j} \rangle = 0$ ), there exists a partition*

$$\mathbf{S}' = \bigcup_{m: 2^m \leq \text{size}_j(\mathbf{S}')} \bigcup_{\mathbf{T} \in \mathcal{F}^{m,j}} \mathbf{T}$$

where for each  $m$ ,  $\mathcal{F}^{m,j}$  is a collection of trees such that  $\text{size}_j(\mathbf{T}) \leq 2^{m+1}$  and

$$(34) \quad \sum_{\mathbf{T} \in \mathcal{F}^{m,j}} |I_{\mathbf{T}}| \lesssim_{\mu} 2^{-2m} \left( \frac{1}{|E_j|^{1/2} 2^m} \right)^{\frac{2}{\mu}}.$$

Now we prove (30). It will suffice for each  $l \geq 0$  to prove the stronger estimate

$$(35) \quad \sum_{\mathbf{T}} |I_{\mathbf{T}}| \prod_{i=1}^n \text{size}_i(\mathbf{T}) \lesssim 2^{-l} |E_1|^{\alpha_1} \dots |E_n|^{\alpha_n}$$

under the additional assumption that

$$(36) \quad 2^l \leq 1 + \frac{\text{dist}(I_s, \mathbf{R} \setminus \Omega)}{|I_s|} < 2^{l+1}$$

for all tiles  $s \in \mathbf{S}$ , since the original claim (30) then follows by dyadic decomposition of  $\mathbf{S}$ .

From (36) and Proposition 6.2 we have

$$(37) \quad \text{size}_i(\mathbf{S}) \lesssim |E_i|^{\frac{1}{2}} 2^l \text{ for } 1 \leq i < n$$

and

$$(38) \quad \text{size}_n(\mathbf{S}) \lesssim 2^{(1-N)l}.$$

Now use the selection algorithm in Theorem 6.4 for  $\mathbf{S}$  to get for each  $i$  the collections of trees  $\mathcal{F}^{m,i}$ ; the tiles of  $i$ -size zero can be safely discarded (viewing them as singleton trees) as they make no contribution to (35). We can then partition

$$\mathbf{S} = \bigcup_{m_1, \dots, m_n} \mathbf{S}^{m_1,1} \cap \dots \cap \mathbf{S}^{m_n,n}$$

where  $\mathbf{S}^{m,i} := \bigcup_{\mathbf{T} \in \mathcal{F}^{m,i}} \mathbf{T}$  and we implicitly assume that

$$(39) \quad 2^{m_i} \leq \text{size}_i(\mathbf{S}).$$

By pigeonholing we can restrict to the case when  $m_j = \max(m_1, \dots, m_n)$  for some fixed  $1 \leq j \leq n$ . We then have the partition

$$\mathbf{S} = \bigcup_{m_1, \dots, m_n: m_j = \max(m_1, \dots, m_n)} \bigcup_{\mathbf{T} \in \mathcal{F}^{m_j, j}} (\mathbf{T} \cap \mathbf{S}^{m_1, 1} \cap \dots \cap \mathbf{S}^{m_n, n}).$$

Note that  $\mathbf{T} \cap \mathbf{S}^{m_1, 1} \cap \dots \cap \mathbf{S}^{m_n, n}$  is a tree with the same top as  $\mathbf{T}$ , and with  $j$ -size at most  $2^{m_j+1}$ ; this tree need not contain its top, but this is of no consequence for us. To verify (35) it thus suffices to show that

$$(40) \quad \sum_{m_1, \dots, m_n: m_j = \max(m_1, \dots, m_n)} \sum_{\mathbf{T} \in \mathcal{F}^{m_j, j}} |I_{\mathbf{T}}| 2^{m_1} \dots 2^{m_n} \lesssim 2^{-l} |E_1|^{\alpha_1} \dots |E_n|^{\alpha_n}.$$

Meanwhile, from (34) we have

$$\sum_{\mathbf{T} \in \mathcal{F}^{m_j, j}} |I_{\mathbf{T}}| \lesssim_{\mu} 2^{-2m_j} \left( \frac{1}{|E_j|^{1/2} 2^{m_j}} \right)^{\frac{2}{\mu}}$$

where  $\mu$  is a large parameter to be chosen later. Also, from (39) we have

$$2^{m_1} \dots 2^{m_n} \leq 2^{m_j} \prod_{i \neq j} \text{size}_i(\mathbf{S})^{2\alpha_i} 2^{(1-2\alpha_i)m_i}.$$

From these bounds and summing the geometric series in all the  $m_i$  for  $i \neq j$ , we have

$$\begin{aligned} & \sum_{m_1, \dots, m_n: m_j = \max(m_1, \dots, m_n)} \sum_{\mathbf{T} \in \mathcal{F}^{m_j, j}} |I_{\mathbf{T}}| 2^{m_1} \dots 2^{m_n} \\ & \lesssim_{\mu} \prod_{i \neq j} \text{size}_i(\mathbf{S})^{2\alpha_i} \sum_{m_j} 2^{m_j} \left( \prod_{i \neq j} 2^{(1-2\alpha_i)m_j} \right) 2^{-2m_j} \left( \frac{1}{|E_j|^{1/2} 2^{m_j}} \right)^{\frac{2}{\mu}}. \end{aligned}$$

Since  $\alpha_1 + \dots + \alpha_n = (n - 2)/2$ , we can rewrite the right-hand side as

$$\prod_{i \neq j} \text{size}_i(\mathbf{S})^{2\alpha_i} \sum_{m_j} 2^{2\alpha_j m_j} \left( \frac{1}{|E_j|^{1/2} 2^{m_j}} \right)^{\frac{2}{\mu}}.$$

Summing the geometric series, we can bound this (for  $\mu$  sufficiently large) by

$$\lesssim_{\mu} \prod_i \text{size}_i(\mathbf{S})^{2\alpha_i} \left( \frac{1}{|E_j|^{1/2} \text{size}_j(\mathbf{S})} \right)^{\frac{2}{\mu}}.$$

Applying (37), (38) we obtain (40) as desired if  $N$  and  $\mu$  are chosen sufficiently large. This proves Theorem 4.4 and hence Theorem 1.1.

It remains to prove Theorem 6.2 and Lemma 6.3. This will occupy the remainder of the paper.

### 7. SINGLE TREE SIZE ESTIMATE

In this section we prove Theorem 6.2. This estimate is well known in the case  $j \leq n - 1$ , when the cutoff  $|I_s| > 2^{k(x)}$  has no effect; see [20, Lemma 6.8]. Thus we shall focus instead on the more difficult case  $j = n$ .

Our task is to show that

$$\sum_{s \in \mathbf{T}} |\langle f, \phi_{s,n} \rangle|^2 \lesssim |I_{\mathbf{T}}| \left( \sup_{I \in \mathcal{P}_{\mathbf{T}}} \frac{1}{|I|} \int_E \chi_I^N \right)^2$$

for each  $n$ -separated multitile tree  $(\mathbf{T}, T, i)$ , and each  $f \in X(E)$ .

Fix  $(\mathbf{T}, T, i)$ . By frequency translation invariance we may assume that  $0 \in \omega_{T,n}$ . If  $f$  is supported outside  $2I_{\mathbf{T}}$ , then from the decay of  $\phi_{s,n}$  we get

$$|\langle f, \phi_{s,n} \rangle| \lesssim \left( \frac{|I_s|}{|I_{\mathbf{T}}|} \right)^N |I_{\mathbf{T}}|^{-\frac{1}{2}} \int_E \chi_{I_s}^N$$

for all  $s \in \mathbf{T}$ , which proves the result in this case. Thus we may assume without loss of generality that  $f$  is supported on  $2I_{\mathbf{T}}$ .

Using duality it hence suffices to prove that

$$\frac{1}{|I_{\mathbf{T}}|^{\frac{1}{2}}} \left| \int f \sum_{s \in \mathbf{T}} a_s \phi_{s,n} \right| \lesssim \sup_{I \in \mathcal{P}_{\mathbf{T}}} \frac{1}{|I|} \int_E \chi_I^N$$

for all  $(a_s)_{s \in \mathbf{T}}$  with  $\|(a_s)\|_{l^2(\mathbf{T})} \leq 1$ .

Fix the  $a_s$ . We can estimate

$$\left| \sum_{s \in \mathbf{T}} a_s \phi_{s,n}(x) \right| \leq \sup_k \left| \sum_{\substack{s \in \mathbf{T} \\ |I_s| > 2^k}} a_s \psi_{s,n}(x) \right|.$$

Since  $T$  is  $n$ -separated, we see from Lemma 5.2 that the tiles  $s \in \mathbf{T}$  with  $|I_s| > 2^k$  have a disjoint frequency support from the tiles  $s \in \mathbf{T}$  with  $|I_s| \leq 2^k$ . Indeed we can write  $\sum_{\substack{s \in \mathbf{T} \\ |I_s| > 2^k}} a_s \psi_{s,n}(x)$  as a Fourier multiplier applied to the function  $F := \sum_{s \in \mathbf{T}} a_s \psi_{s,n}(x)$ , where the symbol of the multiplier is a cutoff smoothly adapted to an interval of length  $\sim C_0 2^{-k}$ . From this and standard kernel estimates, we conclude that

$$\sup_k \left| \sum_{\substack{s \in \mathbf{T} \\ |I_s| > 2^k}} a_s \psi_{s,n}(x) \right| \lesssim MF,$$

and so it will suffice to show that

$$(41) \quad \frac{1}{|I_{\mathbf{T}}|^{\frac{1}{2}}} \int fM(F) \lesssim \sup_{I \in \mathcal{P}_{\mathbf{T}}} \frac{1}{|I|} \int_E \chi_I^N.$$

For a dyadic interval  $J$  denote by  $J_1, J_2, J_3$  the three dyadic intervals of the same length with  $J$ , sitting at the left of  $J$ , with  $J_3$  being adjacent to  $J$ . Similarly let  $J_5, J_6, J_7$  be the three dyadic intervals of the same length with  $J$ , sitting at the right of  $J$ , with  $J_5$  being adjacent to  $J$ . Also define  $J_4 := J$ . Let  $\mathcal{J}$  be the set of all dyadic intervals  $J$  with the following properties:

- (a)  $J \cap 2I_{\mathbf{T}} \neq \emptyset$ ,
- (b)  $\nexists I \in \mathcal{P}_{\mathbf{T}} : |I| < |J|$  and  $I \subset 3J$ ,
- (c)  $J_i \in \mathcal{P}_{\mathbf{T}}$  for some  $1 \leq i \leq 7$ .

We claim that  $2I_{\mathbf{T}} \subset \bigcup_{J \in \mathcal{J}} J$ . Indeed, assume by contradiction that there exists some  $x \in 2I_{\mathbf{T}} \setminus \bigcup_{J \in \mathcal{J}} J$ . Let  $J^{(0)} \subset J^{(1)} \subset J^{(2)} \subset \dots$  be the sequence of dyadic intervals of consecutive lengths containing  $x$ , with  $|J^{(0)}| = \min_{I \in \mathcal{P}_{\mathbf{T}}} |I|$ . Since  $J^{(0)} \notin \mathcal{J}$  and since (a) and (b) are certainly satisfied for  $J^{(0)}$ , it follows that  $J_i^{(0)} \notin \mathcal{P}_{\mathbf{T}}$  for each  $1 \leq i \leq 7$ . Moreover, note that for each  $1 \leq i \leq 7$  there is no  $I \in \mathcal{P}_{\mathbf{T}}$  with  $I \subset J_i^{(0)}$ . We proceed now by induction. Assume that for some

$j \geq 0$  we proved that for each  $1 \leq i \leq 7$  we have  $J_i^{(j)} \notin \mathcal{P}_{\mathbf{T}}$  and also that there is no  $I \in \mathcal{P}_{\mathbf{T}}$  with  $I \subset J_i^{(j)}$ . Note that this implies the same for  $j + 1$ . Indeed, since  $3J^{(j+1)} \subset 7J^{(j)}$  and by the induction hypothesis, it follows that (b) is satisfied for  $J^{(j+1)}$ . Hence  $J_i^{(j+1)} \notin \mathcal{P}_{\mathbf{T}}$  for each  $1 \leq i \leq 7$ . We verify now the second statement of the induction. Note that if there was an  $I \in \mathcal{P}_{\mathbf{T}}$  with  $I \subset J_i^{(j+1)}$ , then the hypothesis of the induction and the fact that  $3J^{(j+1)} \subset 7J^{(j)}$  would imply that  $i \in \{1, 2, 6, 7\}$ . Hence  $I \subset J_i^{(j+1)} \subset I_T$ , and by convexity of  $\mathcal{P}_{\mathbf{T}}$  it would follow that  $J_i^{(j+1)} \in \mathcal{P}_{\mathbf{T}}$ , which is impossible. This closes the induction. To see how the claim follows from here, observe that  $I_T = J_i^{(j)}$  for some  $i, j$ , which certainly contradicts the fact that  $I_T \in \mathcal{P}_{\mathbf{T}}$ .

The next thing we prove is that on each interval  $2J$  with  $J \in \mathcal{J}$ , the oscillation of  $F$  is well controlled. More exactly we will show that for each  $x, y \in 2J$ ,  $|F(x) - F(y)| \lesssim \frac{1}{|J|^{\frac{1}{2}}}$ . We have

$$\begin{aligned} |F(x) - F(y)| &\lesssim |J| \sum_{s \in \mathbf{T}} \sup_{z \in 2J} |\psi'_{s,n}(z)| |a_s| \\ &\lesssim |J| \sum_{s \in \mathbf{T}} \frac{1}{|I_s|^{\frac{3}{2}}} \chi_{I_s}^M(c(J)) |a_s|, \end{aligned}$$

since by definition there exists no  $I_s \subset 3J$ . Now

$$\begin{aligned} |J| \sum_{\substack{s \in \mathbf{T} \\ |I_s| \geq |J|}} \frac{1}{|I_s|^{\frac{3}{2}}} \chi_{I_s}^M(c(J)) |a_s| &\lesssim |J| \left( \sum_{\substack{s \in \mathbf{T} \\ |I_s| \geq |J|}} |a_s|^2 \right)^{\frac{1}{2}} \left( \sum_{2^k \geq |J|} \sum_{i \geq 1} \frac{1}{i^M 2^{3k}} \right)^{\frac{1}{2}} \\ &\lesssim \frac{1}{|J|^{\frac{1}{2}}} \end{aligned}$$

and also

$$\begin{aligned} |J| \sum_{\substack{s \in \mathbf{T} \\ |I_s| < |J|}} \frac{1}{|I_s|^{\frac{3}{2}}} \chi_{I_s}^M(c(J)) |a_s| &\lesssim \left( \sum_{\substack{s \in \mathbf{T} \\ |I_s| < |J|}} |a_s|^2 \right)^{\frac{1}{2}} \left( \sum_{2^k < |J|} \sum_{i \geq \frac{|J|}{2^k}} \frac{1}{i^M 2^{3k}} \right)^{\frac{1}{2}} \\ &\lesssim \frac{1}{|J|^{\frac{1}{2}}}, \end{aligned}$$

due to the fact that there exists no  $I_s \subset 3J$  with  $|I_s| < |J|$ .

Define now the measure space  $X = \bigcup_{J \in \mathcal{J}} J$  and its  $\sigma$ -algebra  $\Upsilon$  generated by the maximal intervals  $J \in \mathcal{J}$ . Recall that  $2I_T \subset \bigcup_{J \in \mathcal{J}} J = X \subset 10I_T$ . We will see that for each  $x \in J$ ,

$$(42) \quad \mathbb{M}(F)(x) \lesssim \frac{1}{|J|} \int_J \mathbb{M}(F)(z) dz + \frac{1}{|J|^{\frac{1}{2}}}.$$

Indeed, if  $r > \frac{1}{2}|J|$ ,

$$\begin{aligned} \frac{1}{2r} \int_{x-r}^{x+r} |F|(z) dz &\lesssim \inf_{y \in J} \mathbb{M}(F)(y) \\ &\lesssim \frac{1}{|J|} \int_J \mathbb{M}(F)(z) dz. \end{aligned}$$

On the other hand, if  $r \leq \frac{1}{2}|J|$ ,

$$\begin{aligned} \frac{1}{2r} \int_{x-r}^{x+r} |F|(z) dz &\lesssim \sup_{y \in 2J} |F|(y) \\ &\lesssim \inf_{y \in J} |F|(y) + \frac{1}{|J|^{\frac{1}{2}}} \\ &\lesssim \frac{1}{|J|} \int_J M(F)(z) dz + \frac{1}{|J|^{\frac{1}{2}}}. \end{aligned}$$

From (42) we can write

$$\begin{aligned} \frac{1}{|I_T|^{\frac{1}{2}}} \int f M(F) &\lesssim \frac{1}{|I_T|^{\frac{1}{2}}} \int_X f \mathbb{E}(M(F)|\Upsilon) + \sup_{J \in \mathcal{J}} \frac{1}{|J|} \int_J f \\ &= \frac{1}{|I_T|^{\frac{1}{2}}} \int_X \mathbb{E}(f|\Upsilon) \mathbb{E}(M(F)|\Upsilon) + \sup_{J \in \mathcal{J}} \frac{1}{|J|} \int_J f \\ &\leq \frac{1}{|I_T|^{\frac{1}{2}}} \|\mathbb{E}(f|\Upsilon)\|_{L^\infty} \int_X \mathbb{E}(M(F)|\Upsilon) + \sup_{J \in \mathcal{J}} \frac{1}{|J|} \int_E \chi_J^N \\ &\lesssim \left( 1 + \left[ \int_X \mathbb{E}(M(F)|\Upsilon)^2 \right]^{\frac{1}{2}} \right) \sup_{J \in \mathcal{J}} \frac{1}{|J|} \int_E \chi_J^N \\ &\lesssim \sup_{J \in \mathcal{J}} \frac{1}{|J|} \int_E \chi_J^N, \end{aligned}$$

where  $\mathbb{E}(\cdot|\Upsilon)$  denotes the conditional expectation relative to  $\Upsilon$ . Finally, note that since for each  $J \in \mathcal{J}$ ,  $J_i \in \mathcal{P}_{\mathbf{T}}$  for some  $i$ , we have that

$$\sup_{J \in \mathcal{J}} \frac{1}{|J|} \int_E \chi_J^N \lesssim \sup_{I \in \mathcal{P}_{\mathbf{T}}} \frac{1}{|I|} \int_E \chi_I^N,$$

which yields (41). This concludes the proof of Theorem 6.2. □

### 8. REDUCTION TO BESSEL INEQUALITY

We still have to prove Lemma 6.3. This will be achieved by means of a certain maximal Bessel inequality and a stopping time argument. We first recall a definition.

**Definition 8.1** ([18]). Let  $j \in \{1, 2, \dots, n\}$ . Two  $j$ -separated multitile trees  $(\mathbf{T}, T, i)$  and  $(\mathbf{T}', T', i)$  with the same index are said to be *strongly  $j$ -disjoint* if  $\mathbf{T} \cap \mathbf{T}' = \emptyset$ , and furthermore whenever  $s \in \mathbf{T}, s' \in \mathbf{T}'$  are such that  $\omega_{s_j} \subsetneq \omega_{s'_j}$ , then one has  $I_T \cap I_{s'} = \emptyset$ , and similarly with  $\mathbf{T}$  and  $\mathbf{T}'$  reversed. A collection of  $j$ -separated multitile trees is called *mutually strongly  $j$ -disjoint* if each two multitile trees in the collection are strongly  $j$ -disjoint.

*Remark 8.2.* If two  $j$ -separated multitile trees  $(\mathbf{T}, T, i)$  and  $(\mathbf{T}', T', i)$  are strongly  $j$ -disjoint, then one has  $s_j \cap s'_j = \emptyset$  for each  $s \in \mathbf{T}, s' \in \mathbf{T}'$ . This is because if  $s_j$  and  $s'_j$  intersect, then since  $\mathbf{T} \cap \mathbf{T}' = \emptyset$ , we must have either  $\omega_{s_j} \subsetneq \omega_{s'_j}$  or  $\omega_{s'_j} \subsetneq \omega_{s_j}$ , and the claim then follows from the definition of strong  $j$ -disjointness. This may help explain the terminology “strong  $j$ -disjointness”.

The next estimate controls the extent to which disjoint trees can each absorb a lot of  $L^2$  energy. It is the main technical estimate used in the proof, and the core

of Lacey’s original argument in [15]. The proof is rather difficult and will occupy the remainder of this paper.

**Theorem 8.1** (Maximal Bessel inequality, multitile version). *Let  $\mathcal{F}$  be a finite collection of strongly  $j$ -disjoint,  $j$ -separated multitile trees. Let  $\mu > 0$  and suppose that  $N$  is sufficiently large depending on  $\mu$  (recall the definition of  $N$  from (14)). Assume also that*

$$(43) \quad 2^m \leq \left( \frac{1}{|I_{\mathbf{T}}|} \sum_{s \in \mathbf{T}} |\langle f_j, \phi_{s,j} \rangle|^2 \right)^{\frac{1}{2}} \leq 2^{m+1}$$

and

$$(44) \quad \left( \frac{1}{|I_{T'}|} \sum_{\substack{s \in \mathbf{T} \\ I_s \subset I_{T'}}} |\langle f_j, \phi_{s,j} \rangle|^2 \right)^{\frac{1}{2}} \leq 2^{m+1}$$

for each  $T' \in \mathbf{T} \in \mathcal{F}$ . Then if  $\mu > 0$  we have

$$(45) \quad \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| \lesssim_{\mu} 2^{-2m} \left( \frac{1}{|E_j|^{1/2} 2^m} \right)^{\frac{2}{\mu}}.$$

*Remark 8.3.* The factor  $\left( \frac{1}{|E_j|^{1/2} 2^m} \right)^{\frac{2}{\mu}}$  in (45) is technical and should be ignored. Intuitively, the condition (43) asserts that the function  $f_j$ , when “restricted” to a tree  $\mathbf{T}$  in  $\mathcal{F}$ , has  $L^2$  norm roughly comparable to  $2^m |I_{\mathbf{T}}|^{1/2}$ . The strong disjointness of the trees is an assertion that these restrictions are in some sense “almost orthogonal”. Since  $f_j$  has an  $L^2$  norm of  $O(1)$ , we see that (45) is indeed a kind of Bessel inequality. This estimate is standard (and fairly straightforward) when  $j \neq n$ , but when  $j = n$  the presence of the cutoff  $1_{|I_s| > 2^{k(x)}}$  in the modified wave packet  $\phi_{s,n}$  presents some significant difficulties (already encountered in [15]).

Let us now show how Lemma 6.3 follows from Theorem 8.1. This will be a standard stopping time argument of the type which has been commonly used in time-frequency analysis, see for instance [18], but for the sake of completeness we present the argument here.

We perform the following algorithm to construct  $\mathbf{S}_2$  and  $\mathcal{F}$ .

- Step 0. Initialize  $\mathcal{F}$  to be empty, and  $\mathbf{S}_2$  to equal  $\mathbf{S}'$ .
- Step 1. If  $\text{size}_j(\mathbf{S}_2) \leq 2^m$ , then we terminate the algorithm. Otherwise, we have

$$2^m < \text{size}_j(\mathbf{S}_2) \leq \text{size}_j(\mathbf{S}') \leq 2^{m+1}.$$

By definition of size, we can find a  $j$ -separated multitile tree  $\mathbf{T} = (\mathbf{T}, T, i)$  in  $\mathbf{S}_2$  obeying (44).

- Step 2. The multitile tree  $(\mathbf{T}, T, i)$  mentioned above is a  $(j, \epsilon)$ -separated tree for some  $\epsilon = \pm 1$ . For fixed  $i$  and  $\epsilon$ , we may assume that this tree maximizes the quantity  $\epsilon \xi_{T_i}$ , where  $\xi_{T_i}$  is the center of the frequency tile  $T_i$ , subject to the constraints (44) and  $\mathbf{T} \subseteq \mathbf{S}_2$ .
- Step 3. Clearly the multitile tree  $\mathbf{T}$  is nonempty, since it has positive size. Add the multitile tree  $\mathbf{T}$  to the collection  $\mathcal{F}$ , and delete the multitiles in  $\mathbf{T}$  from  $\mathbf{S}_2$ . Note that this removes at least one multitile from  $\mathbf{S}_2$ .

- Step 4. Next, define the (possibly empty) companion tree  $(\tilde{\mathbf{T}}, T, j)$  where  $\tilde{\mathbf{T}} := \{s \in \mathbf{S}_2 : s_j \leq T_j\}$ , add this tree  $\tilde{\mathbf{T}}$  to  $\mathcal{F}$  also, and delete the multitiles in  $\tilde{\mathbf{T}}$  from  $\mathbf{S}_2$ . Then return to Step 1.

This algorithm terminates in finite time since  $\mathbf{S}_2$  was initially finite, and every iteration of the algorithm removes at least one multitime from  $\mathbf{S}_2$ . It is also clear that this algorithm will obtain a decomposition (31) obeying (33). The only remaining task is to verify the bound (32). It suffices to do this for each fixed  $1 \leq i \leq n$ , thus restricting the summation to those trees  $\mathbf{T} = (\mathbf{T}, T, i)$  in  $\mathcal{F}$  with index  $i$ . This in turn fixes the quantity  $\epsilon$  appearing in Step 2 above, so if one indexes the trees  $\mathbf{T}$  in the order that they are added to  $\mathcal{F}$ , then  $\epsilon \xi_{T_j}$  will be nonincreasing. We also only need to focus on those trees selected using Step 3 rather than Step 4, since the trees in Step 4 have the same time interval as those in Step 3 and so we are only giving up a factor of 2 by doing this.

To prove (32), it suffices by Theorem 8.1 to show that the trees in  $\mathcal{F}$  with fixed  $i$  and  $\epsilon$  arising from Step 3 are mutually strongly  $j$ -disjoint. Suppose for contradiction that there were two trees  $\mathbf{T}, \mathbf{T}'$  in  $\mathcal{F}$  of this type which were not strongly  $j$ -disjoint. Since these trees have distinct multitiles by construction, the only way that strong  $j$ -disjointness can fail (up to swapping  $\mathbf{T}$  and  $\mathbf{T}'$ ) is if there exist  $s \in \mathbf{T}, s' \in \mathbf{T}'$  with  $\omega_{s,j} \subsetneq \omega_{s',j}$  and  $I_T \cap I_{s'} \neq \emptyset$ . From Lemma 5.2 we conclude that  $s'_j \leq T_j$  and  $\epsilon(\xi_{T_i} - \xi_{T'_i}) \geq 0$ . The latter condition, combined with the nonincreasing nature of the  $\epsilon \xi_{T_j}$ , ensures that  $T$  was selected earlier in the algorithm than  $T'$ . But then  $s'$  would have been selected in the companion tree  $\tilde{\mathbf{T}}$  and could not have remained in  $\mathbf{S}_2$  by the time  $T'$  was selected, a contradiction. This ensures the strong  $j$ -disjointness and concludes the deduction of Lemma 6.3 from Theorem 8.1.

It remains to prove Theorem 8.1. This will occupy the remainder of the paper.

### 9. GOOD- $\lambda$ REDUCTION

The only remaining task in the proof of Theorem 1.1 is to show the maximal Bessel inequality in Theorem 8.1. This will be accomplished in stages. In this section we rephrase the inequality as an inequality concerning tiles rather than multitiles, and use some “BMO theory” for tiles to replace the  $\left(\frac{|E_j|^{1/2}}{2^m}\right)^\mu$  factor in (45) by a factor which depends instead on the counting function  $N_{\mathcal{F}}$ . This BMO theory is quite elementary and may have some independent interest.

We will focus on the hardest case  $j = n$ , in which one must deal with the presence of the cutoff  $1_{|I_s| > 2^{k(x)}}$  in the modified wave packet  $\phi_{s,n}$ . The cases  $j \neq n$  are significantly simpler (see for instance [18]) and in any event can be handled by the argument here (e.g. by the artificial expedient of setting  $k(x)$  to be so low that the cutoff  $1_{|I_s| > 2^{k(x)}}$  disappears).

The Bessel inequality is now really only a statement about the  $n$ -tiles of the multitiles in  $\mathbf{S}$ , and so we shall introduce new notation to focus only on these tiles rather than on the multitiles.

**Definition 9.1** (Lacunary tree). A *lacunary tree*  $\mathbf{T} = (\mathbf{T}, I_{\mathbf{T}}, \xi_{\mathbf{T}})$  is a collection  $\mathbf{T}$  of tiles, together with a dyadic time interval  $I_{\mathbf{T}} \in \mathbf{D}_0$  and a center frequency  $\xi_{\mathbf{T}} \in \mathbf{R}$ , such that for all  $P \in \mathbf{T}$  we have  $I_P \subseteq I_{\mathbf{T}}$  and  $\text{dist}(\omega_P, \xi_{\mathbf{T}}) \sim C_0 |\omega_P|$ , and such that the frequency interval  $\omega_P$  of a tile is determined entirely by the length of the time interval; thus if  $P, P' \in \mathbf{T}$  and  $|I_P| = |I_{P'}|$ , then  $\omega_P = \omega_{P'}$ . (In particular,

this means that distinct tiles in  $\mathbf{T}$  have distinct time intervals.) We say that one lacunary tree  $(\mathbf{T}', I_{\mathbf{T}'}, \xi_{\mathbf{T}'})$  is a *subtree* of another  $(\mathbf{T}, I_{\mathbf{T}}, \xi_{\mathbf{T}})$  if  $\mathbf{T}' \subseteq \mathbf{T}$  (thus we allow subtrees to have a different time interval and center than the supertree). We say that two lacunary trees  $(\mathbf{T}, I_{\mathbf{T}}, \xi_{\mathbf{T}})$ ,  $(\mathbf{T}', I_{\mathbf{T}'}, \xi_{\mathbf{T}'})$  are *strongly disjoint* if  $\mathbf{T} \cap \mathbf{T}' = \emptyset$ , and whenever  $P \in \mathbf{T}$ ,  $P' \in \mathbf{T}'$  are such that  $\omega_P \subsetneq \omega_{P'}$ , then one has  $I_{\mathbf{T}} \cap I_{P'} = \emptyset$ , and similarly with  $\mathbf{T}$  and  $\mathbf{T}'$  reversed. We define a *forest* to be any collection  $\mathcal{F}$  of lacunary trees such that any two distinct trees  $\mathbf{T}, \mathbf{T}'$  in  $\mathcal{F}$  are strongly disjoint.

Observe from Lemma 5.2 that if  $(\mathbf{T}, T, i)$  is an  $n$ -separated multitile tree, then  $(\mathbf{T}_n, I_{T_n}, \xi_{T_n, n})$  is a lacunary tree, where  $\mathbf{T}_n := \{s_n : s \in \mathbf{T}\}$  is the set of  $n$ -tiles of the multitile tree  $\mathbf{T}$ , and  $\xi_{T_n, n}$  is a frequency such that  $\text{dist}(\omega_{T_n, n}, \xi_{T_n, n}) \sim C_0 |\omega_{T_n, n}|$ . Furthermore, if  $(\mathbf{T}, T, i)$  and  $(\mathbf{T}', T', i)$  are strongly  $n$ -disjoint, then  $(\mathbf{T}_n, I_{T_n}, \xi_{T_n, n})$  and  $(\mathbf{T}'_n, I_{T'_n}, \xi_{T'_n, n})$  are strongly disjoint. Thus, we can deduce Theorem 8.1 from

**Theorem 9.1** (Maximal Bessel inequality, first reduction). *Let  $\mathcal{F}$  be a forest. Let  $\mu > 0$  and suppose that  $N$  is sufficiently large depending on  $\mu$ . For each tile  $P$  in  $\bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}$ , let  $\psi_P$  be a wave packet adapted to  $P$ , and let  $\phi_P$  be the function*

$$\phi_P(x) := 1_{|I_s| > 2^{k(x)}} \psi_P(x).$$

Let  $E$  be a finite union of intervals, and let  $f \in X_2(E)$  be such that

$$2^m \leq \left( \frac{1}{|I_{\mathbf{T}}|} \sum_{P \in \mathbf{T}} |\langle f, \phi_P \rangle|^2 \right)^{\frac{1}{2}} \leq 2^{m+1}$$

and

$$\left( \frac{1}{|I_{T'}|} \sum_{\substack{P \in \mathbf{T}' \\ I_P \subset I_{T'}}} |\langle f, \phi_P \rangle|^2 \right)^{\frac{1}{2}} \leq 2^{m+1}$$

for each  $T' \in \mathbf{T} \in \mathcal{F}$ . Then we have

$$\sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| \lesssim_{\mu} 2^{-2m} \left( \frac{1}{|E|^{1/2} 2^m} \right)^{\frac{2}{\mu}},$$

for all  $\mu > 0$ .

We will now eliminate the role of the set  $E$ , replacing it with a certain counting function multiplicity, and also eliminate the role of the size parameter  $2^m$ . More precisely, in this section we shall deduce Theorem 9.1 from

**Theorem 9.2** (Maximal Bessel inequality, second reduction). *Let  $\mathcal{F}$  be a forest. Let  $\mu > 0$  and suppose that  $N$  is sufficiently large depending on  $\mu$ . Let  $\psi_P, \phi_P$  be as in Theorem 9.1. Let  $f \in L^2(\mathbf{R})$  be such that*

$$(46) \quad 1 \leq \left( \frac{1}{|I_{\mathbf{T}}|} \sum_{P \in \mathbf{T}} |\langle f, \phi_P \rangle|^2 \right)^{\frac{1}{2}} \leq 2$$



for each  $\mathbf{T} \in \mathcal{F}$ , and

$$(47) \quad \left( \frac{1}{|I_{T'}|} \sum_{\substack{P \in \mathbf{T} \\ I_P \subset I_{T'}}} |\langle f, \phi_P \rangle|^2 \right)^{\frac{1}{2}} \leq 2$$

for each  $T' \in \mathbf{T}$ . Let  $N_{\mathcal{F}}$  be the counting function

$$(48) \quad N_{\mathcal{F}} := \sum_{\mathbf{T} \in \mathcal{F}} 1_{I_{\mathbf{T}}},$$

and let  $I_0$  be any interval which contains the support of  $N_{\mathcal{F}}$ . Then we have the Bessel-type inequality

$$(49) \quad \sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}} |\langle f, \phi_P \rangle|^2 \lesssim_{\mu} \|N_{\mathcal{F}}\|_{L^{\infty}}^{\frac{1}{2}} \int |f|^2 \chi_{I_0}^{10}.$$

We shall prove Theorem 9.2 in later sections. For now, we show how it implies Theorem 8.1. The argument is similar to the “good- $\lambda$ ” type estimates used to prove John-Nirenberg BMO inequalities, and to emphasize this connection (and because this theory may be of some independent interest) we shall proceed in a somewhat abstract manner.

The following observation is trivial, but is still worth recording.

**Lemma 9.3** (Forest refinement). *Let  $\mathcal{F}$  be a forest. For each tree  $\mathbf{T}$  in  $\mathcal{F}$ , let  $\mathcal{F}_{\mathbf{T}}$  be a collection of subtrees of  $\mathbf{T}$  with disjoint time intervals. Then  $\bigcup_{\mathbf{T} \in \mathcal{F}} \mathcal{F}_{\mathbf{T}}$  is also a forest.*

Let  $\mathcal{F}$  be any forest. The quantity  $\|N_{\mathcal{F}}\|_{L^{\infty}}$  measures the maximum possible overlap of the time intervals  $I_{\mathbf{T}}$  of the trees  $\mathbf{T}$  in  $\mathcal{F}$ . We shall introduce a closely related quantity  $\|\mathcal{F}\|_{\text{BMO}}$ , defined as

$$\|\mathcal{F}\|_{\text{BMO}} := \sup_I \frac{1}{|I|} \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subset I} |I_{\mathbf{T}}|,$$

where the supremum is taken over all the dyadic intervals  $I$ .

*Remark 9.2.* One can relate this BMO-type norm to the genuine (dyadic, vector-valued) BMO norm by the formula  $\|\mathcal{F}\|_{\text{BMO}} = \|\vec{N}_{\mathcal{F}}\|_{\text{BMO}}^2$ , where  $\vec{N}_{\mathcal{F}} := \sum_{\mathbf{T} \in \mathcal{F}} 1_{I_{\mathbf{T}}} e_{\mathbf{T}}$  is a vector-valued counting function, with the  $e_{\mathbf{T}}$  being orthonormal vectors in an abstract Hilbert space. However, we will not adopt this approach since the theory of vector-valued BMO is not as familiar as that of ordinary BMO, preferring instead a more direct and elementary approach.

It is clear that  $\|\mathcal{F}\|_{\text{BMO}} \leq \|N_{\mathcal{F}}\|_{L^{\infty}}$ . Indeed,

$$\begin{aligned} \frac{1}{|I|} \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subset I} |I_{\mathbf{T}}| &= \frac{1}{|I|} \int_I \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subset I} 1_{I_{\mathbf{T}}} \\ &\leq \left\| \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subset I} 1_{I_{\mathbf{T}}} \right\|_{L^{\infty}} \\ &\leq \|N_{\mathcal{F}}\|_{L^{\infty}}. \end{aligned}$$

While the converse is not quite true, we do expect the  $L^{\infty}$  norm and BMO norm to be very close.

Now we obtain some good-lambda inequalities for the BMO norm. We first observe that to control the BMO norm of a collection  $\mathcal{F}$  of trees, it suffices to control the BMO norm of subcollections of trees which are already controlled in  $L^\infty$ .

**Lemma 9.4.** *Let  $\mathcal{F}$  be a forest such that*

$$\|\mathcal{F}'\|_{\text{BMO}} \leq B$$

*whenever  $\mathcal{F}' \subset \mathcal{F}$  is such that  $\|N_{\mathcal{F}'}\|_{L^\infty} \leq 2B$ . Then we have*

$$\|\mathcal{F}\|_{\text{BMO}} \leq 2B.$$

*Proof.* Let  $I_0$  be a dyadic interval. Call a dyadic interval  $J \subseteq I_0$  *heavy* if  $|\{\mathbf{T} \in \mathcal{F} : J \subseteq I_{\mathbf{T}} \subseteq I_0\}| > 2B$ , and let  $\mathcal{F}'$  be the collection of those trees  $\mathbf{T} \in \mathcal{F}$  such that  $I_{\mathbf{T}} \subseteq I_0$  and such that  $I_{\mathbf{T}}$  is not heavy. Then by construction we have  $\|\mathcal{F}'\|_{L^\infty} \leq 2B$ , and hence by hypothesis  $\|\mathcal{F}'\|_{\text{BMO}} \leq B$ . In particular

$$\sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subseteq I_0; I_{\mathbf{T}} \text{ not heavy}} |I_{\mathbf{T}}| \leq B|I_0|.$$

Now we deal with the heavy intervals. If we let  $\mathbf{J}$  denote the set of maximal dyadic heavy intervals, then we have

$$\begin{aligned} \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subseteq I_0; I_{\mathbf{T}} \text{ heavy}} |I_{\mathbf{T}}| &\leq \sum_{J \in \mathbf{J}} \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subseteq J} |I_{\mathbf{T}}| \\ &\leq \sum_{J \in \mathbf{J}} \|\mathcal{F}\|_{\text{BMO}} |J| \\ &= \frac{\|\mathcal{F}\|_{\text{BMO}}}{2B} \int_{\bigcup_{J \in \mathbf{J}} J} 2B \\ &\leq \frac{\|\mathcal{F}\|_{\text{BMO}}}{2B} \int_{\bigcup_{J \in \mathbf{J}} J} \sum_{\mathbf{T} \in \mathcal{F}'} 1_{I_{\mathbf{T}}} \\ &\leq \frac{\|\mathcal{F}\|_{\text{BMO}}}{2B} \int_{I_0} \sum_{\mathbf{T} \in \mathcal{F}'} 1_{I_{\mathbf{T}}} \\ &\leq \frac{\|\mathcal{F}\|_{\text{BMO}}}{2B} |I_0| \|\mathcal{F}'\|_{\text{BMO}} \\ &\leq \frac{\|\mathcal{F}\|_{\text{BMO}}}{2} |I_0|. \end{aligned}$$

Summing these two estimates, we obtain

$$\frac{1}{|I_0|} \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \subseteq I_0} |I_{\mathbf{T}}| \leq B + \frac{\|\mathcal{F}\|_{\text{BMO}}}{2},$$

and then taking the supremum over  $I_0$  we obtain

$$\|\mathcal{F}\|_{\text{BMO}} \leq B + \frac{\|\mathcal{F}\|_{\text{BMO}}}{2}.$$

The claim follows. □

Similarly, to control the  $L^1$  norm of  $N_{\mathcal{F}}$ , it suffices to control the  $L^1$  norm of subcollections  $\mathcal{F}'$  which are controlled in  $L^\infty$  by the BMO norm of  $\mathcal{F}$ :

**Lemma 9.5.** *Let  $\mathcal{F}$  be a forest such that*

$$\|N_{\mathcal{F}'}\|_{L^1} \leq A$$

*whenever  $\mathcal{F}' \subset \mathcal{F}$  is such that  $\|N_{\mathcal{F}'}\|_{L^\infty} \leq \|\mathcal{F}\|_{\text{BMO}}$ . Then we have*

$$\|N_{\mathcal{F}}\|_1 \leq 2A.$$

*Proof.* Set  $B := \|\mathcal{F}\|_{\text{BMO}}$ . As before, we call a dyadic interval  $J \subseteq I_0$  *heavy* if  $|\{\mathbf{T} \in \mathcal{F} : J \subseteq I_{\mathbf{T}} \subseteq I_0\}| > B$ , and let  $\mathcal{F}'$  be the collection of those trees  $\mathbf{T} \in \mathcal{F}$  such that  $I_{\mathbf{T}} \subseteq I_0$  and that  $I_{\mathbf{T}}$  is not heavy. Then by construction we have  $\|N_{\mathcal{F}'}\|_{L^\infty} \leq B$  and hence by hypothesis  $\|N_{\mathcal{F}'}\|_1 \leq A$ . Now if we let  $\mathbf{J}$  be the collection of maximal heavy intervals, then we have

$$\begin{aligned} \|N_{\mathcal{F} \setminus \mathcal{F}'}\|_1 &= \sum_{J \in \mathbf{J}} \sum_{\mathbf{T} \in \mathcal{F} : I_{\mathbf{T}} \subseteq J} |I_{\mathbf{T}}| \\ &\leq \sum_{J \in \mathbf{J}} \|\mathcal{F}\|_{\text{BMO}} |J| \\ &= \int_{\bigcup_{J \in \mathbf{J}} J} B \\ &\leq \int_{\bigcup_{J \in \mathbf{J}} J} \sum_{\mathbf{T} \in \mathcal{F}'} 1_{I_{\mathbf{T}}} \\ &\leq \|N_{\mathcal{F}'}\|_1 \\ &\leq A \end{aligned}$$

and the claim follows. □

We can of course combine these two lemmas to obtain

**Corollary 9.6.** *Let  $\mathcal{F}$  be a forest such that*

$$\|N_{\mathcal{F}'}\|_1 \leq A \text{ and } \|\mathcal{F}'\|_{\text{BMO}} \leq B$$

*whenever  $\mathcal{F}' \subset \mathcal{F}$  is such that  $\|N_{\mathcal{F}'}\|_{L^\infty} \leq 2B$ . Then we have*

$$\|N_{\mathcal{F}}\|_1 \leq 2A \text{ and } \|\mathcal{F}\|_{\text{BMO}} \leq 2B.$$

A specific case of this is

**Corollary 9.7.** *Let  $\mathcal{F}$  be a forest such that for some  $\mu > 1$ ,*

$$\|N_{\mathcal{F}'}\|_1 \leq A \|N_{\mathcal{F}'}\|_{L^\infty}^{\frac{1}{\mu}} \text{ and } \|\mathcal{F}'\|_{\text{BMO}} \leq B \|N_{\mathcal{F}'}\|_{L^\infty}^{\frac{1}{\mu}}$$

*for all  $\mathcal{F}' \subseteq \mathcal{F}$ . Then we have*

$$\|N_{\mathcal{F}}\|_1 \lesssim_{\mu} AB^{\frac{1}{\mu-1}} \text{ and } \|\mathcal{F}\|_{\text{BMO}} \lesssim_{\mu} BB^{\frac{1}{\mu-1}}.$$

Thus to prove a counting function estimate on  $\|N_{\mathcal{F}}\|_1$ , we are permitted to lose a small power of the  $\|N_{\mathcal{F}}\|_{L^\infty}$  as long as the argument also works for all subtrees and localizes to a BMO version as well (with a different constant  $B$ ).

Now we can finally prove Theorem 9.1.

*Proof of Theorem 9.1.* Let  $\mathcal{F}' \subset \mathcal{F}$  be arbitrary. From Theorem 9.2 with  $f$  replaced by  $f/2^m$ , and  $I_0$  chosen to be so large as to contain all the time intervals arising from  $\mathcal{F}'$ , we have

$$\begin{aligned} \|N_{\mathcal{F}'}\|_1 &= \sum_{\mathbf{T} \in \mathcal{F}'} |I_{\mathbf{T}}| \\ &\lesssim \sum_{P \in \cup_{\mathbf{T} \in \mathcal{F}'} \mathbf{T}} |\langle f/2^m, \phi_P \rangle|^2 \\ &\lesssim_{\mu} \|N_{\mathcal{F}'}\|_{L^\infty}^{\frac{1}{\mu}} \int |f/2^m|^2 \\ &\lesssim 2^{-2m} \|N_{\mathcal{F}'}\|_{L^\infty}^{\frac{1}{\mu}} \end{aligned}$$

thanks to the  $L^2$  normalization of  $f \in X_2(E)$ . If we let  $I_0$  be an arbitrary dyadic interval, then by replacing  $\mathcal{F}'$  by  $\{\mathbf{T} \in \mathcal{F}' : I_{\mathbf{T}} \subseteq I_0\}$  in the above argument we see that

$$\begin{aligned} \frac{1}{|I_0|} \sum_{\mathbf{T} \in \mathcal{F}' : I_{\mathbf{T}} \subseteq I_0} |I_{\mathbf{T}}| &\lesssim_{\mu} \frac{1}{|I_0|} \|N_{\mathcal{F}'}\|_{L^\infty}^{\frac{1}{\mu}} \int |f/2^m|^2 \chi_{I_0}^{10} \\ &\lesssim \|N_{\mathcal{F}'}\|_{L^\infty}^{\frac{1}{\mu}} 2^{-2m} |E|^{-1} \end{aligned}$$

thanks to the uniform bound of  $|E|^{-1/2}$  on  $f \in X_2(E)$ . Taking suprema over  $I_0$  we conclude that  $\|\mathcal{F}'\|_{\text{BMO}} \lesssim_{\mu} \|N_{\mathcal{F}'}\|_{L^\infty}^{\frac{1}{\mu}} 2^{-2m} |E|^{-1}$ . Applying Corollary 9.7 we conclude that

$$\sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| = \|N_{\mathcal{F}}\|_1 \lesssim_{\mu} 2^{-2m} (2^{-2m} |E|^{-1})^{\frac{1}{\mu-1}}.$$

Replacing  $\mu$  by  $\mu + 1$  we obtain Theorem 9.1. □

### 10. TILESET REFINEMENTS

It remains to prove Theorem 9.2. In this section we perform some additional elementary reductions. First we eliminate the localizing weight  $\chi_{I_0}^{10}$  and we permit the deletion of those tiles which lie inside a small exceptional set. Then we sparsify the tile set and remove some logarithmic pileups of time interval multiplicity.

We begin with the first reduction. We assert that to prove Theorem 9.2 it suffices to prove the same assertion with the weight  $\chi_{I_0}^{10}$  not present in (49). The reason for this is that  $\chi_{I_0}^{-10}$  is a polynomial, and because of this (and the hypothesis that all the tiles have time interval contained in  $I_0$ )  $\chi_{I_0}^{-10} \psi_P$  is a wave packet adapted to  $P$ , except for the trivial change that the exponent of  $10N$  in (27) must be reduced slightly to  $10(N-1)$ . But this clearly makes no essential difference to the argument since we are free to take  $N$  as large as we wish. Since  $\langle f, \phi_P \rangle = \langle f \chi_{I_0}^{10}, \chi_{I_0}^{-10} \phi_P \rangle$ , we thus see that Theorem 9.2 with the localizing weight  $\chi_{I_0}^{10}$  follows more or less automatically from Theorem 9.2 without the localizing weight.

The next step is to eliminate the hypotheses (46), (47) and also give the ability to delete a small exceptional collection of tiles.

**Theorem 10.1** (Maximal Bessel inequality, third reduction). *Let  $\mathcal{F}$  be a forest. Let  $\mu > 0$  and suppose that  $N$  is sufficiently large depending on  $\mu$ . Let  $\psi_P, \phi_P$  be*

as in Theorem 9.1, and let  $N_{\mathcal{F}}$  be the counting function (48). Then there exists an exceptional set  $\mathbf{P}_* \subset \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}$  of tiles with

$$(50) \quad \left| \bigcup_{P \in \mathbf{P}_*} I_P \right| \leq \frac{1}{10} \frac{\|N_{\mathcal{F}}\|_{L^1}}{\|N_{\mathcal{F}}\|_{L^\infty}}$$

such that we have the Bessel-type inequality

$$(51) \quad \sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T} \setminus \mathbf{P}_*} |\langle f, \phi_P \rangle|^2 \lesssim_\mu \|N_{\mathcal{F}}\|_{L^\infty}^{\frac{1}{2}} \|f\|_{L^2}^2$$

for all  $f \in L^2(\mathbf{R})$ .

*Proof of Theorem 9.2 assuming Theorem 10.1.* Write  $\Omega := \bigcup_{P \in \mathbf{P}_*} I_P$ . Then

$$\sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T} : I_P \not\subseteq \Omega} |\langle f, \phi_P \rangle|^2 \leq \sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T} \setminus \mathbf{P}_*} |\langle f, \phi_P \rangle|^2.$$

To prove (49), it thus suffices in view of (51) to show that

$$\sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T} : I_P \subseteq \Omega} |\langle f, \phi_P \rangle|^2 \leq \frac{1}{2} \sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}} |\langle f, \phi_P \rangle|^2.$$

From (46), it thus suffices to show that

$$\sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T} : I_P \subseteq \Omega} |\langle f, \phi_P \rangle|^2 \leq \frac{1}{2} \|N_{\mathcal{F}}\|_{L^1}.$$

For each tree  $\mathbf{T} \in \mathcal{F}$ , consider the tile set  $\{P \in \mathbf{T} : I_P \subseteq \Omega\}$ . If  $Q$  is any tile in this set with  $I_Q$  maximal with respect to set inclusion, then  $I_Q \subseteq \Omega$  and from (47) we have

$$\sum_{P \in \mathbf{T} : I_P \subseteq I_Q \subseteq \Omega} |\langle f, \phi_P \rangle|^2 \leq 4|I_Q|.$$

Summing this over all such  $Q$  (noting that the  $I_Q$  are disjoint by dyadicity and maximality) we conclude

$$\sum_{P \in \mathbf{T} : I_P \subseteq \Omega} |\langle f, \phi_P \rangle|^2 \leq 4|I_{\mathbf{T}} \cap \Omega| = 4 \int_{\Omega} 1_{I_{\mathbf{T}}}.$$

Summing this over all  $\mathbf{T} \in \mathcal{F}$  we obtain

$$\sum_{P \in \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T} : I_P \subseteq \Omega} |\langle f, \phi_P \rangle|^2 \leq 4 \int_{\Omega} N_{\mathcal{F}} \leq 4|\Omega| \|N_{\mathcal{F}}\|_{L^\infty}$$

and the claim follows from (50). □

We still have to prove Theorem 10.1. The next step will be to sparsify the collection of tiles. Recall the three dyadic grids  $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$  from (24). One can easily verify that for every interval  $J$  (not necessarily dyadic) there exists a  $d \in \{0, 1, 2\}$  and a shifted dyadic interval  $J' \in \mathcal{D}_d$  such that  $J \subseteq J' \subseteq 3J$ ; we will say that  $J$  is *d-regular*.

Let  $A \geq 1$ , and let  $d \in \{0, 1, 2\}$ . We shall say that a collection of  $\mathcal{I} \subset \mathcal{D}_0$  of time intervals is *(A, d)-sparse* if we have the following properties:

- (i) If  $I, I' \in \mathcal{I}$  are such that  $|I| > |I'|$ , then  $|I| \geq 2^{100A}|I'|$ .
- (ii) If  $I, I' \in \mathcal{I}$  are such that  $|I| = |I'|$  and  $I \neq I'$ , then  $\text{dist}(I, I') \geq 100A|I'|$ .

(iii) If  $I \in \mathcal{I}$ , then  $AI$  is  $d$ -regular; thus there exists an interval  $I_A \in \mathcal{D}_d$  such that  $AI \subseteq I_A \subseteq 3AI$ . We refer to  $I_A$  as the  $A$ -enlargement of  $I$ .

If  $\mathcal{I}$  is an  $(A, d)$ -sparse set of time intervals and  $P$  is a tile whose time interval  $I_P$  lies in  $\mathcal{I}$ , we write  $I_{P,A}$  for the  $A$ -enlargement of  $I_P$ . Similarly if  $\mathbf{T}$  is a tree whose time interval  $I_{\mathbf{T}}$  lies in  $\mathcal{I}$ , we write  $I_{\mathbf{T},A}$  for the  $A$ -enlargement of  $I_{\mathbf{T}}$ .

To prove Theorem 10.1, it suffices to prove a variant for  $(A, d)$ -sparse sets of tiles. More precisely, we can reduce to

**Theorem 10.2** (Maximal Bessel inequality, fourth reduction). *Let  $A, D, \nu > 1$ , and suppose that  $N$  is sufficiently large depending on  $\nu$ . Let  $\mathcal{F}$  be a forest with  $\|N_{\mathcal{F}}\|_{L^\infty} \leq D$ . Let  $\mathbf{P} := \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}$ , and suppose that the time intervals*

$$\{I_P : P \in \mathbf{P}\} \cup \{I_{\mathbf{T}} : \mathbf{T} \in \mathcal{F}\}$$

*are  $(A, d)$ -sparse. Let  $\psi_P, \phi_P$  be as in Theorem 9.1. Then there exists an exceptional set  $\mathbf{P}_* \subset \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}$  of tiles with*

$$(52) \quad \left| \bigcup_{P \in \mathbf{P}_*} I_P \right| \lesssim_\nu (A^{-\nu} + D^{-\nu}) \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}|$$

*such that we have the Bessel-type inequality*

$$\sum_{P \in \mathbf{P} \setminus \mathbf{P}_*} |\langle f, \phi_P \rangle|^2 \lesssim_\nu ((\log(2 + AD))^{10} + A^{10-\nu} D^{10}) \|f\|_{L^2}^2$$

*for all  $f \in L^2(\mathbf{R})$ .*

*Proof of Theorem 10.1 assuming Theorem 10.2.* Let  $A, \nu$  be chosen later, and set  $D := \|N_{\mathcal{F}}\|_{L^\infty}$ . We need the following lemma:

**Lemma 10.3** (Sparsification). *Let  $\mathcal{I}$  be a collection of time intervals. Then we can split  $\mathcal{I} = \mathcal{I}_1 \cup \dots \cup \mathcal{I}_L$  with  $L = O(A^2)$  such that each  $\mathcal{I}_l$  for  $1 \leq l \leq L$  is  $(A, d)$ -sparse for some  $d = 0, 1, 2$ .*

*Proof.* By pigeonholing the scale parameter into cosets of  $100AZ$ , we can partition  $\mathcal{I}$  into  $100A$  subcollections, such that on each subcollection we have the scale separation property (i) from the definition of  $(A, d)$ -sparseness. Similarly if we partition the position parameter at each fixed scale into cosets of  $100A$ , we can partition further into  $(100A)^2$  subcollections on which we also have the position separation property (ii). Finally, we make the elementary observation that for each dyadic  $I \in \mathbf{D}_0$  there exists  $d = 0, 1, 2$  such that there exists  $I_A \in \mathcal{D}_d$  with  $AI \subseteq I_A \subseteq 3AI$ . A final pigeonholing based on the  $d$  parameter concludes the claim.  $\square$

We apply this lemma to the set  $\mathcal{I} := \{I_P : P \in \mathbf{P}\}$  to split  $\mathcal{I}$  into  $\mathcal{I}_1, \dots, \mathcal{I}_L$  for some  $L = O(A^2)$ . Then we have  $\mathbf{P} = \mathbf{P}_1 \cup \dots \cup \mathbf{P}_L$ , where  $\mathbf{P}_l := \{P \in \mathbf{P} : I_P \in \mathcal{I}_l\}$ . Observe that if  $\mathbf{T}$  is a lacunary tree in  $\mathcal{F}$ , then  $\mathbf{T} \cap \mathbf{P}_l$  is also a lacunary tree. The time interval  $I_{\mathbf{T}}$  of this tree need not lie in  $\mathcal{I}_l$ ; however, one can partition  $\mathbf{T} \cap \mathbf{P}_l$  into subtrees with this property. More precisely, if we let  $I$  be any interval in  $\{I_P : P \in \mathbf{T} \cap \mathbf{P}_l\}$  which is maximal with respect to set inclusion, then  $(\{P \in \mathbf{T} \cap \mathbf{P}_l : I_P \subseteq I\}, I, \xi_{\mathbf{T}})$  is a lacunary tree whose time interval  $I$  also lies in  $\mathcal{I}_l$ . Let  $\mathcal{F}_l$  be the collection of all the trees obtained in this manner for fixed  $l$ , as  $\mathbf{T}$  varies over  $\mathcal{F}$  and  $I$  varies over the maximal intervals in  $\{I_P : P \in \mathbf{T} \cap \mathbf{P}_l\}$ ; thus  $\mathbf{P}_l = \bigcup_{\mathbf{T} \in \mathcal{F}_l} \mathbf{T}$ . Since for each fixed  $\mathbf{T}$  the intervals  $I$  are disjoint, one easily verifies the pointwise estimate  $N_{\mathcal{F}_l} \leq N_{\mathcal{F}}$ , and hence  $\|N_{\mathcal{F}_l}\|_{L^\infty} \leq D$ . Applying

Theorem 10.2 (if  $N$  is large depending on  $\nu$ ), one can then obtain an exceptional set  $\mathbf{P}_{l,*} \subset \mathbf{P}_l$  obeying (52) such that

$$\sum_{P \in \mathbf{P}_l \setminus \mathbf{P}_{l,*}} |\langle f, \phi_P \rangle|^2 \lesssim_\nu ((\log(1 + AD))^{10} + A^{10-\nu} D^{10}) \|f\|_{L^2}^2.$$

Setting  $\mathbf{P}_* := \bigcup_{1 \leq l \leq L} \mathbf{P}_{l,*}$  we thus conclude

$$\left| \bigcup_{P \in \mathbf{P}_*} I_P \right| \lesssim_\nu A^2 (A^{-\nu} + D^{-\nu}) \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}|$$

and

$$\sum_{P \in \mathbf{P} \setminus \mathbf{P}_*} |\langle f, \phi_P \rangle|^2 \lesssim_\nu A^2 ((\log(2 + AD))^{10} + A^{10-\nu} D^{10}) \|f\|_{L^2}^2.$$

If we then set  $\nu := 100 + 400\mu$  and  $A := C_\mu D^{1/4\mu}$  for a large constant  $C_\mu$  we obtain the claim.

The hypothesis in Theorem 10.2 is currently assuming some control on the quantity  $\|N_{\mathcal{F}}\|_{L^\infty} = \|\sum_{\mathbf{T} \in \mathcal{F}} 1_{I_{\mathbf{T}}}\|_{L^\infty}$ . In the arguments which follow, it is more convenient to assume control on the larger quantity  $\|\sum_{\mathbf{T} \in \mathcal{F}} M1_{I_{\mathbf{T}}}\|_{L^\infty}$ , where of course  $M$  is the Hardy-Littlewood maximal function. It is not necessarily the case that control of the former implies control of the latter, due to “logarithmic pile-ups” such as those where the intervals  $I_{\mathbf{T}}$  are lacunary around a fixed origin; this is also related to the failure of the Fefferman-Stein vector-valued maximal inequality [11] at this endpoint. Nevertheless, by removing all the tiles in a small set it is possible to control the latter from the former. More precisely, we have

**Lemma 10.4.** *Let  $\mathcal{I}$  be a finite set of intervals in  $\mathcal{D}_d$  for some  $d = 0, 1, 2$  such that  $\|\sum_{I \in \mathcal{I}} 1_I\|_{L^\infty} \leq D$  for some  $D$ . Then  $\mathcal{I}$  can be split into two collections  $\mathcal{I} = \mathcal{I}^\sharp \cup \mathcal{I}^\flat$  such that*

$$\left\| \sum_{I \in \mathcal{I}^\sharp} (M1_I)^2 \right\|_{L^\infty} \lesssim_\nu D^3$$

and

$$(53) \quad \left| \bigcup_{I \in \mathcal{I}^\flat} I \right| \lesssim_\nu D^{-\nu} \sum_{I \in \mathcal{I}} |I|.$$

*Proof.* See [15, Lemma 3.14]. □

As a consequence, we can reduce Theorem 10.2 to

**Theorem 10.5** (Maximal Bessel inequality, fifth reduction). *Let  $A, M, \nu > 1$ , and suppose that  $N$  is sufficiently large depending on  $\nu$ . Let  $\mathcal{F}$  be a forest with*

$$(54) \quad \left\| \sum_{\mathbf{T} \in \mathcal{F}} M1_{I_{\mathbf{T}}} \right\|_{L^\infty} \leq M.$$

Let  $\mathbf{P} := \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}$ , and suppose that the time intervals

$$\{I_P : P \in \mathbf{P}\} \cup \{I_{\mathbf{T}} : \mathbf{T} \in \mathcal{F}\}$$

are  $(A, d)$ -sparse. Suppose also that we have the technical condition

$$(55) \quad \sup_{x \in I_P} \text{dist}(x, \partial I_{\mathbf{T}}) \geq A^{-\nu} |I_{\mathbf{T}}|$$

for all  $P \in \mathbf{P}$  and  $\mathbf{T} \in \mathcal{F}$  (this ensures that tiles do not cluster near the edges of trees). Let  $\psi_P, \phi_P$  be as in Theorem 9.1. Then we have the Bessel-type inequality

$$(56) \quad \sum_{P \in \mathbf{P}} |\langle f, \phi_P \rangle|^2 \lesssim_\nu ((\log(2 + AM))^{10} + A^{1-\nu} M^2) \|f\|_{L^2}^2$$

for all  $f \in L^2(\mathbf{R})$ .

*Proof of Theorem 10.2 assuming Theorem 10.5.* Apply Lemma 10.4 to the collection  $\mathcal{I} := \{I_{\mathbf{T}} : \mathbf{T} \in \mathcal{F}\}$  to create the partition  $\mathcal{I} = \mathcal{I}^\sharp \cup \mathcal{I}^\flat$  with the desired properties. Set

$$\mathbf{P}_* := \bigcup_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \in \mathcal{I}^\flat} \mathbf{T} \cup \bigcup_{\mathbf{T} \in \mathcal{F}} \{P \in \mathbf{P} : \sup_{x \in I_P} \text{dist}(x, \partial I_{\mathbf{T}}) < A^{-\nu} |I_{\mathbf{T}}|\}.$$

Observe that

$$\bigcup_{P \in \mathbf{P}_*} I_P \subseteq \bigcup_{I \in \mathcal{I}^\flat} I \cup \bigcup_{\mathbf{T} \in \mathcal{F}} \{x \in I_P : \text{dist}(x, \partial I_{\mathbf{T}}) < A^{-\nu} |I_{\mathbf{T}}|\}$$

and hence by (53),

$$\bigcup_{P \in \mathbf{P}_*} I_P \lesssim_\nu (D^{-\nu} + A^{-\nu}) \sum_{\mathbf{T} \in \mathcal{I}^\sharp} |I_{\mathbf{T}}|.$$

Now since the intervals  $I_{\mathbf{T}}$  have multiplicity at most  $D$ , we have

$$\left\| \sum_{\mathbf{T} \in \mathcal{F}: I_{\mathbf{T}} \notin \mathcal{I}^\flat} M1_{I_{\mathbf{T}}} \right\|_{L^\infty} \leq D \left\| \sum_{I \in \mathcal{I}^\sharp} M1_I \right\|_{L^\infty} \lesssim_\nu D^4.$$

Applying Theorem 10.2 with  $M \sim_\nu D^4$  (and all the trees with spatial interval in  $\mathcal{I}^\flat$  have been completely removed from  $\mathbf{P} \setminus \mathbf{P}_*$ ) we obtain

$$\sum_{P \in \mathbf{P} \setminus \mathbf{P}_*} |\langle f, \phi_P \rangle|^2 \lesssim_\nu ((\log(2 + AD^4))^{10} + A^{1-\nu} D^8) \|f\|_{L^2}^2$$

and the claim follows. □

It remains to prove Theorem 10.5. We may dualize (56), observing that it is equivalent to the estimate

$$\left\| \sum_{P \in \mathbf{P}} a_P \phi_P \right\|_{L^2}^2 \lesssim_\nu ((\log(2 + AM))^{10} + A^{1-\nu} M^2) \|a\|_{l^2}$$

for any sequence  $a = (a_P)_{P \in \mathbf{P}}$  of complex numbers. By definition of  $\phi_P$ , it thus suffices to show the maximal Bessel-type inequality

$$(57) \quad \left\| \sup_k \left| \sum_{P \in \mathbf{P}: |I_P| > 2^k} a_P \psi_P \right\| \right\|_{L^2} \lesssim_\nu ((\log(2 + AM))^{10} + A^{1-\nu} M^2) \|a\|_{l^2}.$$

At this point we shall pause to sketch the general strategy we shall employ to prove (57), following [15]. First we shall split the tile set  $\mathbf{P}$  into layers  $\mathbf{P} = \mathbf{P}_1 \cup \dots \cup \mathbf{P}_J$ . Roughly speaking, the idea is to arrange these layers so that the time intervals of  $\mathbf{P}_{j'}$  tend to be (locally) wider than those of  $\mathbf{P}_j$  when  $j' < j$ . Since  $\psi_P$  is essentially concentrated in  $I_P$  (or more accurately  $I_{P,A}$ ), this heuristically gives rise to an estimate of the form

$$\sup_k \left| \sum_{P \in \mathbf{P}: |I_P| > 2^k} a_P \psi_P \right| \leq \sup_j \left| \sum_{P \in \bigcup_{j' < j} \mathbf{P}_{j'}} a_P \psi_P \right| + \sup_k \left| \sum_{P \in \mathbf{P}_j: |I_P| > 2^k} a_P \psi_P \right|.$$



To deal with the former expression we shall use the Rademacher-Menshov inequality and a nonmaximal Bessel inequality (which is essentially (57) without the supremum in  $k$ , and with somewhat fewer logarithmic losses on the right-hand side). To deal with the second term we replace the supremum in  $j$  by a square function, and reduce to controlling the contribution of a localized expression over a single generation  $\mathbf{P}_j$  (which will ultimately reduce to a certain maximal inequality of Bourgain [4]).

For technical reasons it turns out that one needs to treat the “boundary” of the layers  $\mathbf{P}_j$  separately from the rest of the  $\mathbf{P}_j$ , in order to improve the separation properties between layers. As such we will have to execute the above strategy twice, once for the boundary tiles and once for the interior tiles.

We now turn to the details, beginning with the selection of the layers. Introduce the sets  $\mathcal{I} \subset \mathcal{D}_0$  and  $\mathcal{I}_A \subset \mathcal{D}_d$  by

$$\mathcal{I} := \{I_{\mathbf{T}} : \mathbf{T} \in \mathcal{F}\}; \quad \mathcal{I}_A := \{I_{\mathbf{T},A} : \mathbf{T} \in \mathcal{F}\}.$$

Observe that the  $(A, d)$ -sparseness of  $\mathcal{I}$  ensures that the map  $I \mapsto I_A$  is a bijection from  $\mathcal{I}$  to  $\mathcal{I}_A$  which preserves the set inclusion relation. Since  $AI_{\mathbf{T}} \subseteq I_{\mathbf{T},A} \subseteq 3AI_{\mathbf{T}}$  we see that

$$1_{\mathbf{T},A} \subseteq 10A1_{\mathbf{T}},$$

and hence by (54) we have the multiplicity bound

$$(58) \quad \left\| \sum_{I \in \mathcal{I}_A} 1_I \right\|_{L^\infty} \leq 10AM.$$

We then partition

$$\mathcal{I}_A = \mathcal{I}_A^{(1)} \cup \mathcal{I}_A^{(2)} \cup \dots \cup \mathcal{I}_A^{(10AM)}$$

recursively by defining  $\mathcal{I}_A^{(j)}$  to be those intervals in  $\mathcal{I}_A \setminus \bigcup_{i < j} \mathcal{I}_A^{(i)}$  which are maximal with respect to set inclusion; thus  $\mathcal{I}_A^{(j)}$  is a collection of disjoint intervals in the dyadic grid  $\mathcal{D}_d$ . Observe that for  $1 < j \leq 10AM$ , each interval in  $\mathcal{I}_A^{(j)}$  is contained in exactly one interval in  $\mathcal{I}_A^{(j-1)}$ ; since  $\left\| \sum_{I \in \mathcal{I}} 1_I \right\|_{L^\infty} \leq 10AM$ , we conclude that  $\mathcal{I}_A^{(1)}, \dots, \mathcal{I}_A^{(10AM)}$  do indeed partition  $\mathcal{I}_A$ . Using the bijection between  $\mathcal{I}$  and  $\mathcal{I}_A$ , we thus induce a partition  $\mathcal{I} = \mathcal{I}^{(1)} \cup \dots \cup \mathcal{I}^{(10AM)}$  of  $\mathcal{I}$ .

Let  $1 \leq j \leq 10AM$ . For each  $I \in \mathcal{I}^{(j)}$ , let  $\mathbf{P}_I$  denote the tiles with time interval  $I$ :

$$\mathbf{P}_I := \{P \in \mathbf{P} : I_P = I\}.$$

Observe that each tree  $\mathbf{T}$  in  $\mathcal{F}$  contributes at most one tile to  $\mathbf{P}_I$ , by definition of a lacunary tree, and if  $\mathbf{T}$  does contribute a tile, then  $1_I \leq 1_{\mathbf{T}}$ . By (54) we thus have

$$(59) \quad \#\mathbf{P}_I \leq M \text{ for all } I \in \mathcal{I}.$$

We also introduce the tileset  $\mathbf{P}_{<I}$  for  $I \in \mathcal{I}^{(j)}$  by

$$\mathbf{P}_{<I} := \{P \in \mathbf{P} : I_P \subsetneq I; I_P \not\subseteq J \text{ for all } J \in \bigcup_{i>j} \mathcal{I}^{(i)}\};$$

thus  $\mathbf{P}_{<I}$  is the collection of tiles whose time interval is strictly contained in the interval  $I \in \mathcal{I}^{(j)}$ , but is not contained in any interval from a later layer of  $\mathcal{I}$ . Since every tile  $P \in \mathbf{P}$  has its time interval  $I_P$  contained in some interval in  $\mathcal{I}$  (because

$P$  is contained in some tree  $\mathbf{T} \in \mathcal{F}$ , and hence  $I_P \subseteq I_{\mathbf{T}} \in \mathcal{I}$ ) we see that we have the partition

$$\mathbf{P} = \bigcup_{1 \leq j \leq 10AM} \bigcup_{I \in \mathcal{I}^{(j)}} (\mathbf{P}_I \cup \mathbf{P}_{<I}).$$

To prove (57), it thus suffices by the triangle inequality to prove the estimates

$$(60) \quad \left\| \sup_k \left| \sum_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_I: |I_P| > 2^k} a_P \psi_P \right\|_{L^2} \right\| \lesssim_\nu ((\log(2 + AM))^{10} + A^{1-\nu} M^2) \|a\|_{l^2}$$

and

$$(61) \quad \left\| \sup_k \left| \sum_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}: |I_P| > 2^k} a_P \psi_P \right\|_{L^2} \right\| \lesssim_\nu ((\log(2 + AM))^{10} + A^{1-\nu} M^2) \|a\|_{l^2}.$$

The estimate (60) is easier and is proven in Section 11. The estimate (61) is more difficult, relying in particular on a certain inequality of Bourgain, and is proven in Section 12. To conclude this section, we present two tools which will be used to prove both (60) and (61). The first is a nonmaximal Bessel inequality, and more precisely the bound

$$(62) \quad \left\| \sum_{P \in \mathbf{P}} a_P \psi_P \right\|_{L^2}^2 \lesssim \log(1 + M) \|a\|_{l^2}.$$

This inequality may be of some independent interest and is proven in Section 13. Secondly, we will rely on the following form of the standard Rademacher-Menshov inequality, whose proof we include for the sake of completeness. We observe first the trivial bound

$$(63) \quad \left\| \sup_i |f_i| \right\|_{L^2} \leq \left\| \left( \sum_i |f_i|^2 \right)^{1/2} \right\|_{L^2} = \left( \sum_i \|f_i\|_{L^2}^2 \right)^{1/2}$$

valid for any finite collection of  $L^2$  functions  $f_i$ . This bound is usually too crude for applications, as the summation in  $i$  usually creates an undesirable polynomial loss in the estimates; however, one can refine this polynomial loss to a logarithmic loss in the following way.

**Theorem 10.6** (Rademacher-Menshov). *Let  $(f_l)_{l=1}^L$  be a sequence of functions in  $L^2(\mathbf{R})$  which are almost orthogonal in the sense that there exists a constant  $B$  such that for each finite sequence  $\epsilon_1, \dots, \epsilon_L \in \{-1, +1\}$  of signs we have*

$$\left\| \sum_{l=1}^L \epsilon_l f_l \right\|_{L^2} \leq B.$$

*Then we have the maximal inequality*

$$\left\| \sup_{L' \leq L} \left| \sum_{l=1}^{L'} f_l \right| \right\|_{L^2} \lesssim B \log(2 + L).$$

*Proof.* We may take the  $f_l$  to be real-valued. By adding dummy  $f_l$  if necessary, we may assume that  $L = 2^m$  for some integer  $m \geq 1$ . For each set  $I \subseteq \{1, \dots, L\}$  let  $f_I := \sum_{l \in I} f_l$ . For each  $0 \leq m' \leq m$  let  $\mathcal{I}_{m'}$  denote the collection of sets of the

form  $\{2^{m'}j + 1, \dots, 2^{m'}j + 2^{m'}\}$  for  $j = 0, \dots, 2^{m-m'} - 1$ . For each fixed  $m'$ , the sets in  $\mathcal{I}_{m'}$  partition  $\{1, \dots, L\}$ , and thus by hypothesis we have

$$\left\| \sum_{I \in \mathcal{I}_{m'}} \epsilon_I f_I \right\|_{L^2} \leq B$$

for all signs  $\epsilon_I = \pm 1$ . If we square this inequality we obtain

$$\sum_{I \in \mathcal{I}_{m'}} \|f_I\|_{L^2}^2 + \sum_{I, J \in \mathcal{I}_{m'}: I \neq J} \epsilon_I \epsilon_J \langle f_I, f_J \rangle \leq B^2.$$

If we then set  $\epsilon_I$  to be independent random signs and take expectations, we conclude

$$\sum_{I \in \mathcal{I}_{m'}} \|f_I\|_{L^2}^2 \leq B^2.$$

By (63) this implies that

$$\left\| \sup_{I \in \mathcal{I}_{m'}} |f_I| \right\|_{L^2} \leq B.$$

By representing  $L'$  in binary and using the triangle inequality we have the pointwise estimate

$$\left| \sum_{l=1}^{L'} f_l \right| \leq \sum_{0 \leq m' \leq m} \sup_{I \in \mathcal{I}_{m'}} |f_I|$$

for all  $L' \leq L$ . Taking suprema over all  $L'$ , taking  $L^2$  norms, and applying the triangle inequality, the claim follows.  $\square$

### 11. PROOF OF (60)

We first prove the estimate (60), which is relatively easy and serves as a model for the more complicated estimate (61).

Intuitively, the contribution of the wave packets  $\psi_P$  for  $P \in \mathbf{P}_I$  should be localized to the time interval  $I_A$ . To exploit this we introduce the tail error

$$E(x) := \sum_j \sum_{I \in \mathcal{I}^{(j)}: x \notin I_A} \sum_{P \in \mathbf{P}_I} |a_P| |\psi_P(x)|.$$

This error is small:

**Lemma 11.1** (Tail estimate). *We have*

$$\|E\|_{L^2} \lesssim_\nu A^{-\nu} M^{1/2} \|a\|_{l^2}.$$

*Proof.* From (27) one easily verifies the pointwise estimates

$$|\psi_P| \lesssim |I|^{-1/2} M 1_I$$

and the  $L^1$  bound

$$\| |\psi_P| (1 - 1_{I_A}) \|_{L^1} \lesssim A^{-10N+10} |I|^{1/2}$$

whenever  $P \in \mathbf{P}_I$ . The former bound and (54) imply the estimate

$$\left\| \sum_j \sum_{I \in \mathcal{I}^{(j)}: x \notin I_A} \sum_{P \in \mathbf{P}_I} |I|^{1/2} |a_P| |\psi_P(x)| \right\|_{L^\infty} \lesssim M \|a\|_{l^\infty}$$

while the latter bound and the triangle inequality imply the bound

$$\left\| \sum_j \sum_{I \in \mathcal{I}^{(j)}: x \notin I_A} \sum_{P \in \mathbf{P}_I} |I|^{-1/2} |a_P| |\psi_P(x)| \right\|_{L^1} \lesssim A^{-10N+10} \|a\|_{l^1}.$$

The claim then follows from interpolation (or from Cauchy-Schwarz), since we assume  $N$  sufficiently large depending on  $\nu$ . □

To exploit this tail estimate we use the following pointwise inequality:

**Lemma 11.2.** *For almost every  $x$  we have*

$$\sup_k \left| \sum_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_I: |I_P| > 2^k} a_P \psi_P(x) \right| \lesssim \sup_{j_0} \left| \sum_{j \leq j_0} \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_I} a_P \psi_P(x) \right| + E(x).$$

*Proof.* We may assume that  $x$  is not the endpoint of any dyadic interval. It suffices to show that for every  $k$  and  $x$  there exists a  $j_0$  such that

$$\left| \sum_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_I: |I_P| > 2^k} a_P \psi_P(x) \right| \lesssim \left| \sum_{j \leq j_0} \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_I} a_P \psi_P(x) \right| + O(E(x)).$$

Since  $I_P = I$ , we can write the left-hand side as

$$\left| \sum_j \sum_{I \in \mathcal{I}^{(j)}: |I| > 2^k} \sum_{P \in \mathbf{P}_I} a_P \psi_P(x) \right|.$$

By definition of  $E$  and the triangle inequality, we can bound this by

$$\left| \sum_j \sum_{I \in \mathcal{I}^{(j)}: |I| > 2^k; x \in I_A} \sum_{P \in \mathbf{P}_I} a_P \psi_P(x) \right| + O(E(x)).$$

For each  $1 \leq j \leq 10AM$ , we know that there is at most one interval  $I_j \in \mathcal{I}^{(j)}$  whose dilate  $I_{j,A}$  contains  $x$ , and furthermore these intervals are decreasing in  $j$  (adopting the convention that  $I_j = \emptyset$  if no interval in  $\mathcal{I}^{(j)}$  contains  $x$ ). Thus if we let  $j_0$  be the largest  $j$  for which  $|I_{j_0}| > 2^{k(x)}$  (with  $j_0 = 0$  if no such  $j$  exists), then we see that if  $1 \leq j \leq 10AM$  and  $I \in \mathcal{I}^{(j)}$  are such that  $x \in I_A$ , then  $|I| > 2^{k(x)}$  if and only if  $j \leq j(x)$ . Thus we can bound the preceding expression by

$$\left| \sum_{j \leq j_0} \sum_{I \in \mathcal{I}^{(j)}: x \in I_A} \sum_{P \in \mathbf{P}_I} a_P \psi_P(x) \right| + O(E(x)).$$

One can then remove the constraint  $x \in I_A$  by definition of  $E(x)$  and the triangle inequality. □

In light of the above two lemmas, we see that to prove (60) it suffices to show that

$$(64) \quad \left\| \sup_{j_0} \left| \sum_{j \leq j_0} \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_I} a_P \psi_P \right| \right\|_{L^2} \lesssim (\log(2 + AM))^{10} \|a\|_{l^2}.$$

Applying the Rademacher-Menshov inequality (Theorem 10.6), it suffices to show that

$$\left\| \sum_j \epsilon_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_I} a_P \psi_P \right\|_{L^2} \lesssim (\log(2 + AM))^9 \|a\|_{l^2}$$

for all choices  $\epsilon_1, \dots, \epsilon_{10AM} \in \{-1, +1\}$  of signs. But this follows from the non-maximal Bessel inequality (62) (with some room to spare), since the  $\mathbf{P}_I$  are disjoint in  $\mathbf{P}$ . This concludes the proof of (60).

12. PROOF OF (61)

Now we prove (61). We shall argue as in the proof of (60), although the details shall be more technical, and we shall also rely crucially on a maximal inequality of Bourgain.

In the previous section we localized the contribution of  $\mathbf{P}_I$  to the interval  $I_A$ . It turns out (because of the  $(A, d)$ -sparseness hypothesis) that the contribution of  $\mathbf{P}_{<I}$  can be localized even further, to the interval  $I$  itself. To formalize this we again introduce a tail error

$$\tilde{E}(x) := \sum_{I \in \mathcal{I}^{(j)}: x \notin I} \sum_{P \in \mathbf{P}_{<I}} |a_P| |\psi_P(x)|.$$

**Lemma 12.1** (Tail estimate). *We have*

$$\|\tilde{E}\|_{L^2} \lesssim_\nu A^{1-\nu} M^2 \|a\|_{l^2}.$$

*Proof.* Since there are only  $10AM$  values of  $j$ , it suffices by the triangle inequality to show that

$$(65) \quad \left\| \sum_{I \in \mathcal{I}^{(j)}: x \notin I} \sum_{P \in \mathbf{P}_{<I}} |a_P| |\psi_P(x)| \right\|_{L^2} \lesssim_\nu A^{-\nu} M \|a\|_{l^2}$$

for each  $j$ , which we now fix. Suppose for the moment that we could show the pointwise estimate

$$(66) \quad \sum_{P \in \mathbf{P}_{<I}} |a_P| |\psi_P(x)| \lesssim_\nu A^{-\nu} M c_I |I|^{-1/2} M 1_I(x)^2$$

for each  $I$  and  $x \notin I$ , where  $c_I := (\sum_{P \in \mathbf{P}_{<I}} |a_P|^2)^{1/2}$ . Then the left-hand side of (65) is bounded by

$$\lesssim_\nu A^{-\nu} M \left\| \sum_{I \in \mathcal{I}^{(j)}} c_I |I|^{-1/2} M 1_I(x)^2 \right\|_{L^2}.$$

Applying the Fefferman-Stein maximal inequality [11], which among other things asserts that

$$\left\| \sum_i M f_i^2 \right\|_{L^2} = \left\| \left( \sum_i M f_i^2 \right)^{1/2} \right\|_{L^4}^2 \lesssim \left\| \left( \sum_i |f_i|^2 \right)^{1/2} \right\|_{L^4} = \left\| \sum_i |f_i|^2 \right\|_{L^2},$$

we can bound the left-hand side of (65) by

$$\lesssim_\nu A^{-\nu} M \left\| \sum_{I \in \mathcal{I}^{(j)}} c_I |I|^{-1/2} 1_I(x)^2 \right\|_{L^2}.$$

Since the intervals in  $\mathcal{I}^{(j)}$  are disjoint, this expression is bounded by

$$A^{-\nu} M \left( \sum_{I \in \mathcal{I}^{(j)}} |c_I|^2 \right)^{1/2} \leq A^{-\nu} M \|a\|_{l^2}$$

as desired.

It remains to prove (66). By Cauchy-Schwarz it suffices to verify the estimates

$$(67) \quad \sum_{P \in \mathbf{P}_{<I}} |I_P| |\psi_P(x)| \lesssim M^2 |I|^{1/2}$$

and

$$(68) \quad \sum_{P \in \mathbf{P}_{<I}} |a_P|^2 |I_P|^{-1} |\psi_P(x)| \lesssim_\nu A^{-2\nu} c_I^2 |I|^{-3/2} M 1_I(x)^4.$$

To prove (67), we break  $\mathbf{P}_{<I}$  up into  $\mathbf{P}_{<I} \cup \mathbf{T}$ , where  $\mathbf{T}$  ranges over the forest  $\mathcal{F}$ . Observe that  $\mathbf{P}_{<I} \cup \mathbf{T}$  is empty unless  $I \subseteq I_{\mathbf{T}}$ ; by (54) we thus see that there are at most  $M$  trees  $\mathbf{T}$  for which  $\mathbf{P}_{<I} \cup \mathbf{T}$  is nonempty. Thus it suffices to show that

$$\sum_{P \in \mathbf{P}_{<I} \cap \mathbf{T}} |I_P| |\psi_P(x)| \lesssim |I|^{1/2}.$$

But from (27) we have  $|I_P| |\psi_P(x)| \lesssim |I_P|^{1/2} (M1_{I_P}(x))^{100}$  (say). Since the  $I_P$  are dyadic subintervals of  $I$  and each interval can occur at most once in  $\mathbf{T}$ , the claim follows.

It remains to prove (68). From the definition of  $c_I$  and the triangle inequality it suffices to prove that

$$|I_P|^{-1} |\psi_P(x)| \lesssim_\nu A^{-2\nu} |I|^{-3/2} M1_I(x)^4$$

for each  $P \in \mathbf{P}_{<I}$  and  $x \notin I$ . But from the  $(A, d)$ -sparseness hypothesis we see that  $I_P \subsetneq I$  and  $|I_P| \leq 2^{-100A} |I|$ , while from (55) (recalling that  $I$  is the time interval of some tree  $\mathbf{T}$ ) we have  $\sup_{x \in I_P} \text{dist}(x, \partial I) \geq A^{-\nu} |I|$ . The claim now follows from (27), the exponential gain of  $|I|/|I_P| \geq 2^{100A}$  being more than sufficient to compensate for any polynomial losses in  $A$  or in  $|I|/|I_P|$ .  $\square$

The analog of Lemma 11.2 is

**Lemma 12.2.** *For almost every  $x$  we have*

$$\begin{aligned} \sup_k \left| \sum_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}: |I_P| > 2^k} a_P \psi_P(x) \right| &\lesssim \sup_{j_0} \left| \sum_{j < j_0} \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}} a_P \psi_P(x) \right| \\ &+ \sup_{I \in \mathcal{I}} \sup_k \left| \sum_{P \in \mathbf{P}_{<I}: |I_P| > 2^k} a_P \psi_P(x) \right| \\ &+ \tilde{E}(x). \end{aligned}$$

*Proof.* We again may assume that  $x$  is not the endpoint of a dyadic interval. We fix  $k$ ; it would suffice to find a  $j_0$  and an  $I_0 \in \mathcal{I}$  such that

$$\begin{aligned} \left| \sum_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}: |I_P| > 2^k} a_P \psi_P(x) \right| &\leq \left| \sum_{j < j_0} \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}} a_P \psi_P(x) \right| \\ (69) \qquad \qquad \qquad &+ \left| \sum_{P \in \mathbf{P}_{<I_0}: |I_P| > 2^k} a_P \psi_P(x) \right| \\ &+ O(\tilde{E}(x)). \end{aligned}$$

By definition of  $\tilde{E}(x)$ , we have

$$\begin{aligned} \left| \sum_j \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}: |I_P| > 2^k} a_P \psi_P(x) \right| &\leq \left| \sum_j \sum_{I \in \mathcal{I}^{(j)}; x \in I} \sum_{P \in \mathbf{P}_{<I}: |I_P| > 2^k} a_P \psi_P(x) \right| + O(\tilde{E}(x)). \end{aligned}$$

Let  $j_0$  be the largest  $j$  for which there exists an interval in  $\mathcal{I}^{(j)}$  which contains  $x$  and has length greater than  $2^k$ . There is only one such interval; call it  $I_0$ . We can

thus estimate the contribution of the  $j = j_0$  term by (69) and reduce to showing that

$$\left| \sum_{j < j_0} \sum_{I \in \mathcal{I}^{(j)}; x \in I} \sum_{P \in \mathbf{P}_{<I}; |I_P| > 2^k} a_P \psi_P(x) \right| \lesssim \left| \sum_{j < j_0} \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}} a_P \psi_P(x) \right| + O(\tilde{E}(x)).$$

But if  $I \in \mathcal{I}^{(j)}$  and  $x \in I$ , then  $I$  and  $I_0$  overlap. Since  $I_0$  belongs to a later layer  $\mathcal{I}^{(j_0)}$  than  $I$  we must have  $I_0 \subsetneq I$ , and thus  $|I| > |I_0| > 2^k$ . Hence the constraint  $|I_P| > 2^k$  is redundant and can be removed. The claim now follows from the triangle inequality.  $\square$

In light of the above two lemmas, to prove (61) it would suffice to show that

$$\left\| \sup_{j_0} \left| \sum_{j < j_0} \sum_{I \in \mathcal{I}^{(j)}} \sum_{P \in \mathbf{P}_{<I}} a_P \psi_P(x) \right| \right\|_{L^2} \lesssim (\log(2 + AM))^{10} \|a\|_{l^2}$$

and

$$\left\| \sup_{I \in \mathcal{I}} \sup_k \left| \sum_{P \in \mathbf{P}_{<I}; |I_P| > 2^k} a_P \psi_P(x) \right| \right\|_{L^2} \lesssim (\log(2 + AM))^{10} \|a\|_{l^2}.$$

The first inequality is proven in exactly the same way as (64) and is omitted, so we now turn to the second inequality. By (63) it would suffice to show that

$$\left( \sum_{I \in \mathcal{I}} \left\| \sup_k \left| \sum_{P \in \mathbf{P}_{<I}; |I_P| > 2^k} a_P \psi_P(x) \right| \right\|_{L^2}^2 \right)^{1/2} \lesssim (\log(2 + AM))^{10} \|a\|_{l^2},$$

which in turn would follow from the estimate

$$\left\| \sup_k \left| \sum_{P \in \mathbf{P}_{<I}; |I_P| > 2^k} a_P \psi_P(x) \right| \right\|_{L^2} \lesssim (\log(2 + AM))^{10} \left( \sum_{P \in \mathbf{P}_{<I}} |a_P|^2 \right)^{1/2}$$

for each fixed  $I$ .

Let  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_J$  be all the trees in  $\mathcal{F}$  which intersect  $\mathbf{P}_{<I}$ ; the time interval of such trees must contain  $I$ , and so from (54) we have  $J \leq M$ . We can then write

$$\sum_{P \in \mathbf{P}_{<I}; |I_P| > 2^k} a_P \psi_P(x) = \sum_{j=1}^J \sum_{P \in \mathbf{P}_{<I} \cap \mathbf{T}_j; |I_P| > 2^k} a_P \psi_P(x).$$

Let  $\xi_1, \dots, \xi_J$  be the base frequencies of  $\mathbf{T}_1, \dots, \mathbf{T}_J$ . Since  $\mathbf{T}_j$  is a lacunary tree, we see that if  $P \in \mathbf{P}_{<I} \cap \mathbf{T}_j$ , then  $\psi_P$  has Fourier support in an interval of width  $|I_P|^{-1}$  and distance  $\sim C_0 |I_P|^{-1}$  from  $\xi_j$ . By the strong disjointness of the  $\mathbf{T}_j$  we see that these intervals must be disjoint. This implies that

$$\sum_{j=1}^J \sum_{P \in \mathbf{P}_{<I} \cap \mathbf{T}_j; |I_P| > 2^k} a_P \psi_P(x) = \mathbf{P}i_k \sum_{j=1}^J \sum_{P \in \mathbf{P}_{<I} \cap \mathbf{T}_j} a_P \psi_P(x)$$

where  $\mathbf{P}i_k$  is a Fourier projection to the union of  $J$  intervals centered at  $\xi_1, \dots, \xi_J$ , each of radius  $\sim C_0 2^{-k}$ . We now invoke a deep maximal inequality of Bourgain [4, Lemma 4.11], which asserts in our notation that

$$\left\| \sup_k |\mathbf{P}i_k f| \right\|_{L^2} \lesssim \log(2 + J)^2 \|f\|_{L^2}.$$

Using this, we reduce to showing that

$$\left\| \sum_{j=1}^J \sum_{P \in \mathbf{P}_{<I} \cap \mathbf{T}_j} a_P \psi_P \right\|_{L^2} \lesssim (\log(2 + AM))^8 \left( \sum_{P \in \mathbf{P}_{<I}} |a_P|^2 \right)^{1/2}.$$

But this follows from the nonmaximal Bessel inequality (62). This concludes the proof of (61).

### 13. PROOF OF (62)

We now prove (62). We shall in fact prove the slightly more general statement, which may have some independent interest:

**Proposition 13.1** (Nonmaximal Bessel inequality). *Let  $\mathcal{F}$  be a forest, let  $\mathbf{P} := \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}$ , and for each tile  $P \in \mathbf{P}$  let  $\psi_P$  be a wave packet adapted to  $P$ . Suppose also that  $\|\sum_{\mathbf{T} \in \mathcal{F}} \mathbf{1}_{\mathbf{T}}\|_{L^\infty} \leq M$ . Then we have*

$$\left\| \sum_{P \in \mathbf{P}} a_P \psi_P \right\|_{L^2} \lesssim \log(2+M) \|a\|_{l^2}$$

for any sequence  $a = (a_P)_{P \in \mathbf{P}}$  of complex numbers.

*Remark 13.1.* By duality and the  $TT^*$  method, this inequality is also equivalent to the assertion that

$$\left( \sum_{P \in \mathbf{P}} |\langle f, \psi_P \rangle|^2 \right)^{1/2} \lesssim \log(2+M) \|f\|_{L^2}$$

or that

$$\left\| \sum_{P \in \mathbf{P}} \langle f, \psi_P \rangle \psi_P \right\|_{L^2} \lesssim \log(2+M)^2 \|f\|_{L^2}$$

for all  $f \in L^2$ . The logarithmic loss can probably be lowered to  $\log(2+M)^{1/2}$  but cannot be removed entirely; see [3].

We prove Proposition 13.1 in stages. The most important step is to establish a restricted version of the proposition without the logarithmic loss in  $M$ .

**Proposition 13.2** (Restricted Bessel inequality). *Let  $\mathcal{F}, \mathbf{P}, \psi_P, M$  be as in Proposition 13.1. Suppose that  $a = (a_P)_{P \in \mathbf{P}}$  obeys the Carleson condition  $\sum_{P \in \mathbf{T}'} |a_P|^2 \lesssim 2^{2m} |I_{\mathbf{T}'}|$  for all  $\mathbf{T} \in \mathcal{F}$  and all subtrees  $\mathbf{T}'$  of  $\mathbf{T}$ , where  $m$  is a fixed integer. Then we have*

$$\left\| \sum_{P \in \mathbf{P}} a_P \psi_P \right\|_{L^2} \lesssim 2^m \left( \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| \right)^{1/2}.$$

*Proof.* See [20, Lemma 6.6]. The main idea is to square both sides, use standard estimates on the inner products  $|\langle \psi_P, \psi_Q \rangle|$ , and exploit the strong disjointness of the trees  $\mathbf{T}$  in the forest  $\mathcal{F}$ .  $\square$

Next, we establish restricted  $L^p$ -type estimates with a polynomial loss in  $M$ .

**Proposition 13.3** (Crude Bessel inequality). *Let  $\mathcal{F}, \mathbf{P}, \psi_P, M, m, a$  be as in Proposition 13.2. Then for any  $1 < p < \infty$  we have*

$$\left\| \sum_{P \in \mathbf{P}} a_P \psi_P \right\|_{L^p} \lesssim_p 2^m M \left( \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| \right)^{1/p}.$$

*Remark 13.2.* One can improve the factor of  $M$  here by interpolation with Proposition 13.2, and at the endpoint  $p = 1$  one can remove the loss in  $M$  entirely. However for our purposes any polynomial factor in  $M$  will suffice.



*Proof.* First observe that we can partition the forest  $\mathcal{F}$  into forests  $\mathcal{F}_1 \cup \dots \cup \mathcal{F}_M$ , with each  $\mathcal{F}_j$  having multiplicity one in the sense that  $\|\sum_{\mathbf{T} \in \mathcal{F}_j} 1_{\mathbf{T}}\|_{L^\infty} \leq 1$ . Indeed one could set  $\mathcal{F}_M$  to be a maximal collection of trees in  $\mathcal{F}$  whose time intervals are distinct and are maximal with respect to set inclusion, remove  $\mathcal{F}_M$  from  $\mathcal{F}$  (dropping the multiplicity by 1), and induct; we leave the details to the reader. From the triangle inequality we see that it thus suffices to verify the claim when  $M = 1$ . We may also normalize  $m = 0$ .

Let  $\mathbf{T}$  be a tree in  $\mathcal{F}$ . We can partition the dyadic interval  $I_{\mathbf{T}}$  into four equally sized dyadic subintervals  $I_{\mathbf{T},1}, I_{\mathbf{T},2}, I_{\mathbf{T},3}, I_{\mathbf{T},4}$ , from left to right. Let  $\mathbf{T}_l, \mathbf{T}_r \subset \mathbf{T}$  be the trees  $\mathbf{T}_l := \{P \in \mathbf{T} : I_P \subseteq I_{\mathbf{T},1}\}$  and  $\mathbf{T}_r := \{P \in \mathbf{T} : I_P \subseteq I_{\mathbf{T},4}\}$  with spatial intervals  $I_{\mathbf{T},1}$  and  $I_{\mathbf{T},4}$  respectively, and let  $\mathcal{F}'$  be the forest formed by these trees  $\mathbf{T}_l$  and  $\mathbf{T}_r$ , and  $\mathbf{P}' := \bigcup_{\mathbf{T} \in \mathcal{F}'} \mathbf{T}$ . Observe that this forest also has multiplicity one and that  $\sum_{\mathbf{T} \in \mathcal{F}'} |I_{\mathbf{T}}| = \frac{1}{2} \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}|$ . It thus suffices by the obvious recursion argument to prove the Bessel inequality with  $\mathbf{P}$  replaced by  $\mathbf{P} \setminus \mathbf{P}'$  (conceding a factor of  $\frac{1}{1-2^{-p}} \sim p$  in the implicit constant). The practical upshot of this reduction is that for any tree  $\mathbf{T}$  in the forest  $\mathcal{F}$ , we may assume without loss of generality that none of the tiles in  $\mathbf{T}$  have time interval contained in the left quarter  $I_{\mathbf{T},l}$  or right quarter  $I_{\mathbf{T},r}$  of the tree.

From the Carleson condition we have the crude bound  $|a_P| \lesssim |I_P|^{1/2}$  for all  $P \in \mathbf{P}$ . From this, (27), and the above reduction on the trees  $\mathbf{T}$  one easily verifies the pointwise estimate

$$(1 - 1_{I_{\mathbf{T}}}(x)) \left| \sum_{P \in \mathbf{T}} a_P \psi_P(x) \right| \lesssim M 1_{I_{\mathbf{T}}}(x)^{10}$$

(say) for all  $\mathbf{T} \in \mathcal{F}$  and  $x \in \mathbf{R}$ . From the Fefferman-Stein maximal inequality [11] and the multiplicity one nature of  $\mathcal{F}$  we thus have

$$\left\| \sum_{\mathbf{T} \in \mathcal{F}} (1 - 1_{I_{\mathbf{T}}}(x)) \left| \sum_{P \in \mathbf{T}} a_P \psi_P(x) \right| \right\|_{L^p} \lesssim \left\| \sum_{\mathbf{T} \in \mathcal{F}} M 1_{I_{\mathbf{T}}}(x)^{10} \right\|_{L^p} \lesssim_p \left( \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| \right)^{1/p},$$

and hence by the triangle inequality it will suffice to show that

$$\left\| \sum_{\mathbf{T} \in \mathcal{F}} 1_{I_{\mathbf{T}}} \left| \sum_{P \in \mathbf{T}} a_P \psi_P \right| \right\|_{L^p} \lesssim_p \left( \sum_{\mathbf{T} \in \mathcal{F}} |I_{\mathbf{T}}| \right)^{1/p}.$$

From the disjointness of the intervals  $I_{\mathbf{T}}$  it thus suffices to show that

$$\left\| \sum_{P \in \mathbf{T}} a_P \psi_P \right\|_{L^p(I_{\mathbf{T}})} \lesssim_p |I_{\mathbf{T}}|^{1/p}$$

for each tree  $\mathbf{T}$ . By shifting the frequency dyadic grid if necessary we may assume that  $\xi_T = 0$ ; this essentially turns the wave packets  $\psi_P$  into wavelets. The Carleson condition on the  $a_P$  and standard almost orthogonality estimates then give the  $L^2$  estimate

$$\left\| \sum_{P \in \mathbf{T}} a_P \psi_P \right\|_{L^2} \lesssim |I_{\mathbf{T}}|^{1/2}$$

and the BMO estimate

$$\left\| \sum_{P \in \mathbf{T}} a_P \psi_P \right\|_{\text{BMO}} \lesssim 1$$

from which the claim follows by the John-Nirenberg inequality. □

The idea is now to combine the above two propositions via some sort of real interpolation method to obtain Proposition 13.1. It may well be possible to use one of the existing real interpolation theorems in the literature to obtain this conclusion, but we will use a more explicit argument, based on the following decomposition of an arbitrary  $l^2$  sequence  $a$  into Carleson sequences.

**Lemma 13.4** (Stopping time algorithm). *Let  $\mathcal{F}, \mathbf{P}, \psi_P, M$  be as in Proposition 13.1. Suppose that  $a = (a_P)_{P \in \mathbf{P}}$  obeys the Carleson condition  $\sum_{P \in \mathbf{T}'} |a_P|^2 \leq 2^{2m} |I_{\mathbf{T}'}|$  for all  $\mathbf{T} \in \mathcal{F}$  and all subtrees  $\mathbf{T}'$  of  $\mathbf{T}$ , where  $m$  is a fixed integer. Then we can partition  $\mathbf{P} = \bigcup_{\mathbf{T} \in \mathcal{F}_1} \mathbf{T} \cup \bigcup_{\mathbf{T} \in \mathcal{F}_2} \mathbf{T}$ , where  $\mathcal{F}_1$  is a collection of subtrees of trees in  $\mathcal{F}$  such that*

$$(70) \quad 2^{2m} \sum_{\mathbf{T} \in \mathcal{F}_1} |I_{\mathbf{T}}| \sim \sum_{\mathbf{T} \in \mathcal{F}_1} \sum_{P \in \mathbf{T}} |a_P|^2$$

and  $\mathcal{F}_2$  is a collection of subtrees of trees in  $\mathcal{F}$  such that  $\sum_{P \in \mathbf{T}'} |a_P|^2 \leq 2^{2(m-1)} |I_{\mathbf{T}'}|$  for all  $\mathbf{T} \in \mathcal{F}_2$  and all subtrees  $\mathbf{T}'$  of  $\mathbf{T}$ . Furthermore we have  $\|\sum_{\mathbf{T} \in \mathcal{F}_j} 1_{I_{\mathbf{T}}}\|_{L^\infty} \leq M$  for  $j = 1, 2$ .

*Proof.* It suffices to establish this lemma in the case when the forest  $\mathcal{F}$  consists of a single tree,  $\mathcal{F} = \{\mathbf{T}\}$ , with  $M = 1$ , since the general case then follows by applying the lemma to each tree separately and taking unions, using Lemma 9.3, as well as the observation that the contribution to  $\sum_{\mathbf{T} \in \mathcal{F}_j} 1_{I_{\mathbf{T}}}$  arising from a single tree  $\mathbf{T}_0$  in  $\mathcal{F}$  will be bounded pointwise by  $1_{\mathbf{T}_0}$ .

Let  $\mathcal{I}$  be the set of all dyadic intervals  $I$  in  $I_{\mathbf{T}}$  such that  $\sum_{P \in \mathbf{T}: I_P \subseteq I} |a_P|^2 > 2^{2(m-1)} |I|$ , and such that  $I$  is maximal with respect to set inclusion among all such intervals with the property; thus the intervals in  $\mathcal{I}$  are disjoint and lie inside  $I_{\mathbf{T}}$ . We then let  $\mathcal{F}_1$  be the forest consisting of trees  $\mathbf{T}_I = (\mathbf{T}_I, I, \xi_{\mathbf{T}})$  of the form  $\mathbf{T}_I := \{P \in \mathbf{T} : I_P \subseteq I\}$ , where  $I$  ranges over  $\mathcal{I}$ . By construction it is clear that  $\mathcal{F}_1$  is indeed a forest, and that  $\sum_{P \in \mathbf{T}'} |a_P|^2 \sim 2^{2m} |I_{\mathbf{T}'}|$  for all  $\mathbf{T}' \in \mathcal{F}_1$ ; summing over all  $\mathbf{T}'$  we obtain (70). If we let  $\mathcal{F}_2$  consist of the single tree  $\mathbf{T}_2 = (\mathbf{T}_2, I_{\mathbf{T}}, \xi_{\mathbf{T}})$  consisting of all the tiles not covered by  $\mathcal{F}_1$ , thus  $\mathbf{T}_2 := \mathbf{T} \setminus \bigcup_{\mathbf{T}' \in \mathcal{F}_1} \mathbf{T}'$ , then we see from construction that  $\sum_{P \in \mathbf{T}'} |a_P|^2 \leq 2^{2(m-1)} |I_{\mathbf{T}'}|$  for all subtrees  $\mathbf{T}'$  of  $\mathbf{T}_2$ . The claim follows.  $\square$

Iterating this lemma in the usual manner, starting with  $m$  extremely large and exploiting the fact that the forest  $\mathcal{F}$  contains only finitely many tiles, we obtain

**Corollary 13.5** (Iterated stopping time algorithm). *Let  $\mathcal{F}, \mathbf{P}, \psi_P, M$  be as in Proposition 13.1. Then there exist forests  $\mathcal{F}_m$  for each integer  $m$ , together with a tile set  $\mathbf{P}_{-\infty}$ , such that we have the partition*

$$\mathbf{P} = \bigcup_m \bigcup_{\mathbf{T} \in \mathcal{F}_m} \mathbf{T} \cup \mathbf{P}_{-\infty},$$

such that we have the Carleson condition  $\sum_{P \in \mathbf{T}'} |a_P|^2 \leq 2^{2m} |I_{\mathbf{T}'}|$  for all  $m$ , all  $\mathbf{T} \in \mathcal{F}_m$  and all subtrees  $\mathbf{T}'$  of  $\mathbf{T}$ , we have the bound

$$(71) \quad \sum_m 2^{2m} \sum_{\mathbf{T} \in \mathcal{F}_m} |I_{\mathbf{T}}| \sim \sum_{P \in \mathbf{P}} |a_P|^2$$

and such that  $a_P = 0$  for all  $P \in \mathbf{P}_{-\infty}$ . Finally we have  $\|\sum_{\mathbf{T} \in \mathcal{F}_m} 1_{I_{\mathbf{T}}}\|_{L^\infty} \leq M$  for all  $m$ .

Of course, all but finitely many of the  $\mathcal{F}_m$  will be empty.

We can now prove Proposition 13.1. We apply Corollary 13.5. The tiles in  $\mathbf{P}_{-\infty}$  yield no contribution and can be discarded. We are reduced to establishing that

$$\left\| \sum_m F_m \right\|_{L^2} \lesssim \log(2 + M) \|a\|_{l^2}$$

where  $F_m := \sum_{\mathbf{T} \in \mathcal{F}_m} \sum_{P \in T} a_P \psi_P$ . If we let  $L$  be the first integer greater than  $100 \log(2 + M)$ , it suffices by the triangle inequality to show that

$$\left\| \sum_{m: m \equiv l \pmod L} F_m \right\|_{L^2} \lesssim \|a\|_{l^2}$$

for all residue classes  $l \pmod L$ . Squaring this and using symmetry it suffices to show that

$$\sum_{m: m \equiv l \pmod L} \|F_m\|_{L^2}^2 + \sum_{m, m': m, m' \equiv l \pmod L; m' > m} |\langle F_m, F_{m'} \rangle| \lesssim \|a\|_{l^2}^2.$$

Note that if  $m, m' \equiv l \pmod L$  and  $m' > m$ , then  $m' \geq m + L$ . Introduce the quantities  $A_m := 2^{-2m} \sum_{\mathbf{T} \in \mathcal{F}_m} |I_{\mathbf{T}}|$ ; from (71) it suffices to show that

$$\sum_m \|F_m\|_{L^2}^2 + \sum_{m, m': m' > m+L} |\langle F_m, F_{m'} \rangle| \lesssim \sum_m A_m.$$

From (13.3) we have  $\|F_m\|_{L^2} \lesssim A_m^{1/2}$ , and so we are reduced to showing that

$$\sum_{m, m': m' > m+L} |\langle F_m, F_{m'} \rangle| \lesssim \sum_m A_m.$$

We now use Proposition 13.3 to obtain

$$\|F_m\|_{L^4} \lesssim 2^m M \left( \sum_{\mathbf{T} \in \mathcal{F}_m} |I_{\mathbf{T}}| \right)^{1/4} = 2^{m/2} M A_m^{1/4}$$

and

$$\|F_{m'}\|_{L^{4/3}} \lesssim 2^{m'} M \left( \sum_{\mathbf{T} \in \mathcal{F}_{m'}} |I_{\mathbf{T}}| \right)^{3/4} = 2^{-m'/2} M A_{m'}^{3/4}$$

and hence by Hölder's inequality,

$$|\langle F_m, F_{m'} \rangle| \lesssim 2^{-(m'-m)/2} M^2 A_m^{1/4} A_{m'}^{3/4} \lesssim 2^{-(m'-m)/2} M^2 (A_m + A_{m'}).$$

Summing this and using the geometric series formula we conclude

$$\sum_{m, m': m' > m+L} |\langle F_m, F_{m'} \rangle| \lesssim 2^{-L/2} M^2 \sum_m A_m,$$

and the claim follows from the definition of  $L$ . This concludes the proof of Proposition 13.1, and (62) follows. The proof of Theorem 1.1 (and hence Corollary 1.2) is now complete.  $\square$

14. APPENDIX: A CORRESPONDENCE PRINCIPLE

The purpose of this appendix is to prove the following correspondence principle.

**Proposition 14.1.** *Let  $A$  be an  $(n-1) \times m$  matrix with integer entries. In addition to the operators  $T_{A,\mathbf{R}}^*$  and  $T_{A,\mathbf{X}}^*$  defined in (3) and (4), we introduce the operator  $T_{A,\mathbf{Z}}^*$  defined on functions  $\phi_i : \mathbf{Z} \rightarrow \mathbf{R}$  of compact support:*

$$T_{A,\mathbf{Z}}^*(\phi_1, \dots, \phi_{n-1})(l) := \sup_{N>0} \frac{1}{(2N+1)^m} \sum_{|n_1|, \dots, |n_m| \leq N} \prod_{i=1}^{n-1} |\phi_i(l + \sum_{j=1}^m a_{i,j}n_j)|.$$

Let  $1 < p_1, \dots, p_{n-1} \leq \infty$  and  $p'_n$  be such that  $1/p_1 + \dots + 1/p_{n-1} = 1/p'_n$ . Then the following claims are equivalent.

- (i)  $T_{A,\mathbf{R}}^*$  maps  $L^{p_1}(\mathbf{R}) \times \dots \times L^{p_{n-1}}(\mathbf{R})$  to  $L^{p'_n}(\mathbf{R})$ .
- (ii)  $T_{A,\mathbf{Z}}^*$  maps  $l^{p_1}(\mathbf{Z}) \times \dots \times l^{p_{n-1}}(\mathbf{Z})$  to  $l^{p'_n}(\mathbf{Z})$ .
- (iii) For every dynamical system  $\mathbf{X}$ ,  $T_{A,\mathbf{X}}^*$  maps  $L^{p_1}(\mathbf{X}) \times \dots \times L^{p_{n-1}}(\mathbf{X})$  to  $L^{p'_n}(\mathbf{X})$ , with a bound uniform in  $\mathbf{X}$ .

*Proof.* We first show that (i) implies (ii). Let  $\phi_1, \dots, \phi_{n-1} : \mathbf{Z} \rightarrow \mathbf{R}$  have finite support. For each such  $\phi_i$  define  $f_i : \mathbf{R} \rightarrow \mathbf{R}$  in such a way that  $f_i(x) = \phi_i(l)$  if  $x \in [l - \frac{1}{3}, l + \frac{1}{3}]$  for some  $l \in \mathbf{Z}$ , and 0 otherwise. Note that for each  $x \in [l - \frac{1}{6}, l + \frac{1}{6}]$  and  $N \geq 1$ ,

$$\begin{aligned} & \frac{1}{(2N+1)^m} \sum_{|n_1|, \dots, |n_m| \leq N} \prod_{i=1}^{n-1} |\phi_i(l + \sum_{j=1}^m a_{i,j}n_j)| \\ & \lesssim \frac{1}{(2N+1)^m} \int_{|t_1|, \dots, |t_m| \leq N+1} \prod_{i=1}^{n-1} |f_i(x + \sum_{j=1}^m a_{i,j}t_j)| dt \\ & \lesssim T_A^*(f_1, \dots, f_{n-1})(x). \end{aligned}$$

From the hypothesis (i) we thus conclude (ii).

Now we show that (ii) implies (i). Without loss of generality we may take  $f_1, \dots, f_{n-1}$  to be smooth, positive and compactly supported. Approximating an integral by the Riemann sum, we obtain

$$\begin{aligned} & \|T_{A,\mathbf{R}}(f_1, \dots, f_{n-1})\|_{L^{p'_n}(\mathbf{R})} \\ & = \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1/p'_n} \left\| \sup_{N>0} \frac{1}{(2N+1)^m} \sum_{|n_1|, \dots, |n_{m-1}| \leq N} \prod_{i=1}^{n-1} f_i(\varepsilon(l + \sum_{j=1}^m a_{i,j}n_j)) \right\|_{l^{p'_n}(\mathbf{Z})}. \end{aligned}$$

Applying the hypothesis (ii) we obtain

$$\begin{aligned} & \|T_{A,\mathbf{R}}(f_1, \dots, f_{n-1})\|_{L^{p'_n}(\mathbf{R})} \\ & \lesssim \limsup_{\varepsilon \rightarrow 0} \varepsilon^{-1/p'_n} \prod_{i=1}^{n-1} \|f_i(\varepsilon \cdot)\|_{l^{p_i}(\mathbf{Z})}. \end{aligned}$$

Approximating integrals by Riemann sums again and using the scaling hypothesis  $1/p_1 + \dots + 1/p_{n-1} = 1/p'_n$  we obtain (i) as desired.

Now we show that (ii) implies (iii). Define  $M := \max\{\sum_{j=1}^m |a_{i,j}| : 1 \leq i \leq n-1\}$ . Let  $f_i \in L^{p_i}(\mathbf{X})$ , let  $L \geq 1$  be an arbitrary number and let  $x \in X$

also be arbitrary. By applying the hypothesis (ii) to the functions  $\phi_i$  defined by  $\phi_i(l) = f_i(S^l x)$  if  $|l| \leq (M+1)L$  and  $\phi_i(l) = 0$  otherwise, we get that

$$\sum_{|l| \leq L} (T_{A, \mathbf{X}, L}^*(f_1, \dots, f_{n-1})(S^l x))^{p'_n} \lesssim \prod_{i=1}^{n-1} \left( \sum_{|l| \leq L} |f_i|^{p_i}(S^l x) \right)^{\frac{p'_n}{p_i}},$$

with an implicit constant independent of  $x$  and  $L$ . The quantity  $T_{A, \mathbf{X}, L}^*(f_1, \dots, f_{n-1})(x)$  denotes the maximal operator over averages with  $N \leq L$ . Integration with respect to  $x$  and Hölder's inequality imply that

$$\|T_{A, \mathbf{X}, L}^*(f_1, \dots, f_{n-1})\|_{L^{p'_n}(\mathbf{X})} \lesssim \prod_{i=1}^{n-1} \|f_i\|_{L^{p_i}(\mathbf{X})}.$$

By letting  $L \rightarrow \infty$  we obtain (iii).

To show that (iii) implies (ii), we specialize (iii) to the finitary dynamical system  $X = \mathbf{Z}/N\mathbf{Z}$  with the standard shift  $Sx := x + 1$  and the uniform probability measure. Letting  $N \rightarrow \infty$  (taking advantage of the uniformity of the bounds in (iii) in  $N$ ) and renormalizing the probability measure to be counting measure (taking advantage of the scaling condition) we obtain (ii); we omit the details.  $\square$

#### REFERENCES

- [1] I. Assani, *Pointwise convergence of ergodic averages along cubes*, preprint.
- [2] I. Assani, *Multiple recurrence and almost sure convergence for weakly mixing dynamical systems*, Israel J. Math. **103** (1998), 111–124. MR1613556 (99f:28021)
- [3] J. Barrionuevo and M. Lacey, *A weak-type orthogonality principle*, Proc. Amer. Math. Soc. **131** (2003), no. 6, 1763–1769. MR1955263 (2004f:42025)
- [4] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, Publ. Math. IHES **69** (1989), 5–45. MR1019960 (90k:28030)
- [5] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161. MR1037434 (91d:28029)
- [6] M. Christ, *On certain elementary trilinear operators*, Math. Res. Lett. **8** (2001), no. 1-2, 43–56. MR1825259 (2002e:47077)
- [7] C. Demeter, *Divergence of combinatorial averages*, preprint.
- [8] C. Demeter, *Pointwise convergence of the ergodic bilinear Hilbert transform*, accepted for publication in the Illinois Journal of Mathematics.
- [9] C. Demeter, T. Tao and C. Thiele, *A trilinear maximal function via arithmetic combinatorics*, work in progress.
- [10] C. Demeter, M. Lacey, T. Tao and C. Thiele, *Breaking the duality in the Return Times Theorem*, preprint.
- [11] C. Fefferman and E. M. Stein, *Some maximal inequalities*, Amer. J. Math. **93** (1971), 107–115. MR0284802 (44:2026)
- [12] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerdi on arithmetic progressions*, J. Analyze Math. **31** (1977), 204–256. MR0498471 (58:16583)
- [13] B. J. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, preprint.
- [14] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488. MR2150389 (2007b:37004)
- [15] M. Lacey, *The bilinear maximal functions map into  $L^p$  for  $2/3 < p \leq 1$* , Ann. of Math. (2) **151** (2000), no. 1, 35–57. MR1745019 (2001b:42015)
- [16] E. Lesigne, *Sur la convergence ponctuelle de certaines moyennes ergodiques*, C. R. Acad. Sci. Paris Sér. I Math. **298** (1984), no. 17, 425–428. MR765266 (86d:28019)
- [17] J.-M. Derrien and E. Lesigne, *Un théorème ergodique polynomial ponctuel pour les endomorphismes exacts et les  $K$ -systèmes*, Ann. Inst. H. Poincaré Probab. Statist. **32** (1996), no. 6, 765–778. MR1422310 (98k:28023)

- [18] C. Muscalu, T. Tao and C. Thiele, *Multilinear operators given by singular multipliers*, J. Amer. Math. Soc. **15** (2002), no. 2, 469–496. MR1887641 (2003b:42017)
- [19] C. Muscalu, T. Tao and C. Thiele,  *$L^p$  estimates for the “Biest” I. The Walsh case*, Math. Ann. **329** (2004), no. 3, 401–426. MR2127984 (2005k:42053)
- [20] C. Muscalu, T. Tao and C. Thiele,  *$L^p$  estimates for the “Biest” II. The Fourier model*, Math. Ann. **329** (2004), no. 3, 427–461. MR2127985 (2005k:42054)
- [21] E. M. Stein, *On limits of sequences of operators*, Ann. of Math. **74** (1961): 140–170. MR0125392 (23:A2695)
- [22] E. Szemerédi, *On sets of integers containing no  $k$  elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245. MR0369312 (51:5547)
- [23] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), 53–97. MR2257397

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT LOS ANGELES, LOS ANGELES,  
CALIFORNIA 90095-1555

*E-mail address:* `demeter@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT LOS ANGELES, LOS ANGELES,  
CALIFORNIA 90095-1555

*E-mail address:* `tao@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT LOS ANGELES, LOS ANGELES,  
CALIFORNIA 90095-1555

*E-mail address:* `thiele@math.ucla.edu`