TIME-FREQUENCY ANALYSIS

TUOMAS HYTÖNEN

Lecture notes of a course at the University of Helsinki, Winter 2012.

1. Fourier series and Walsh series

1.1. **Fourier series.** The functions $e_k(x) := e^{i2\pi kx}$, $k \in \mathbb{Z}$, form an orthonormal basis of $L^2(0,1)$. Thus, by basic Hilbert space functional analysis, we have the following identity:

$$f = \sum_{k=-\infty}^{\infty} \langle f, e_k \rangle e_k = \lim_{n\to\infty} S_n f, \qquad S_n f = \sum_{k=-n}^{n} \langle f, e_k \rangle e_k,$$

with convergence in the norm of $L^2(0,1)$; thus $\|f - S_n f\|_{L^2} \to 0$ as $n \to \infty$. The scalar products

$$\langle f, e_k \rangle := \int_0^1 f \cdot \overline{e_k} = \int_0^1 f(x) e^{-i2\pi kx}\,\mathrm{d}x$$

are called the *Fourier coefficients* of f, and denoted by $\hat{f}(k) := \langle f, e_k \rangle$. A much deeper theorem is the following *pointwise* convergence result:

Theorem 1.1 (L. Carleson 1966)**.** *Let $f \in L^2(0,1)$. Then we have*

$$f(x) = \lim_{n\to\infty} S_n f(x)$$

for almost every $x \in [0,1)$.

The essence of this theorem is captured in an estimate for the *maximal partial sum operator*

$$S^* f(x) := \sup_{n\in\mathbb{N}} |S_n f(x)|.$$

Namely, this nonlinear operator satisfies the estimate:

Theorem 1.2. *For some constant $C < \infty$, we have for all $f \in L^2(0,1)$ the bound*

$$\|S^* f\|_{L^{2,\infty}} := \sup_{\lambda>0} \lambda |\{x \in [0,1) : |S^* f(x)| > \lambda\}|^{1/2} \leq C \|f\|_{L^2}.$$

Once we know such maximal control (which is a very deep result!), the proof of the convergence is relatively easy. It suffices to observe that the convergence holds for a dense class of functions in $L^2(0,1)$. We take for granted that the *trigonometric polynomials*

$$g(x) = \sum_{k=-m}^{m} a_k e_k(x)$$

form such a dense subspace. But for such a g, we have $S_n g(x) = g(x)$ as soon as $n \geq m$, so the convergence is trivial. For a general $f \in L^2(0,1)$ and $\epsilon > 0$, we can find a trigonometric polynomial (by density) g so that $\|f - g\|_{L^2} < \epsilon$. Then we argue as follows:

$$\begin{aligned}\{x : S_n f(x) \not\to f(x)\} &= \{x : \limsup_{n\to\infty} |S_n f(x) - f(x)| > 0\} \\ &= \bigcup_{j=1}^{\infty} \{x : \limsup_{n\to\infty} |S_n f(x) - f(x)| > \frac{1}{j}\},\end{aligned}$$

Version: March 5, 2012.

and

$$\begin{aligned}|S_nf(x)-f(x)| &\le |S_n(f-g)(x)|+|S_ng(x)-g(x)|+|g(x)-f(x)| \\ &\le S^*(f-g)(x)+|S_ng(x)-g(x)|+|g(x)-f(x)|,\end{aligned}$$

so that

$$\limsup_{n\to\infty}|S_nf(x)-f(x)|\le S^*(f-g)(x)+|g(x)-f(x)|,$$

observing that $\limsup_{n\to\infty}|S_ng(x)-g(x)|=0$. Thus

$$\begin{aligned}|\{x:\limsup_{n\to\infty}|S_nf(x)-f(x)|>\frac{1}{j}\}| &\le |\{x:|S^*(f-g)(x)|>\frac{1}{2j}\}|+|\{x:|g(x)-f(x)|>\frac{1}{2j}\}| \\ &\le \Big(\frac{C}{1/2j}\|f-g\|_{L^2}\Big)^2+\Big(\frac{1}{1/2j}\|f-g\|_{L^2}\Big)^2 \\ &= 4j^2(C^2+1)\|f-g\|_{L^2}^2<4j^2(C^2+1)\epsilon^2.\end{aligned}$$

This holds for every $\epsilon>0$, so taking $\epsilon\to 0$ we deduce that the left side is 0, and hence also

$$|\{x:S_nf(x)\not\to f(x)\}|=0.$$

Thus $S_nf(x)\to f(x)$ for almost every $x\in[0,1)$.

The previous argument is quite standard, and applies to different pointwise convergence problems. The difficult part is controlling the maximal operator. Instead of attempting this now, we look instead at a simpler model case.

1.2. **Walsh series.** The Walsh model consist of replacing the smooth sine wave $\sin(2\pi x)$ (the imaginary part of $e^{i2\pi x}$) by the block wave

$$w_1(x):=r_0(x):=\sum_{m\in\mathbb{Z}}(1_{[m,m+1/2)}-1_{[m+1/2,m+1)})(x),$$

which is in fact the sign $\operatorname{sgn}\sin(2\pi x)$ (expect at the zeros of sine, where we define the values as above). We give this function two names, it is both the first *Walsh function* w_1 and the zeroth *Rademacher function* r_0. Note that $1=2^0$.

More generally, we let

$$w_{2^k}(x):=r_k(x):=r_0(2^kx)=\sum_{m\in\mathbb{Z}}(1_{2^{-k}[m,m+1/2)}-1_{2^{-k}[m+1/2,m+1)})(x),\qquad k\in\mathbb{N}:=\{0,1,2,\ldots\}.$$

For other values $n\neq 2^k$, we make a more complicated definition of w_n. This is motivated by the following observation about the functions e_n: Let

$$n=\sum_{i=0}^{\infty}n_i2^i,\qquad n_i\in\{0,1\}$$

be the *binary expansion* of $n\in\mathbb{N}$. Of course, only finitely many n_i are different from zero. Then

$$e_n(x)=\exp(i2\pi nx)=\exp(i2\pi\sum_{j=0}^{\infty}2^in_ix)=\prod_{j=0}^{\infty}\exp(i2\pi 2^in_ix)=\prod_{j=0}^{\infty}(e_{2^i}(x))^{n_i}.$$

For the trigonometric function e_n, this is a lemma, a *consequence* of their definition. For the Walsh functions, we make it the definition:

$$w_n(x):=\prod_{i=0}^{\infty}(r_i(x))^{n_i}=\prod_{i\in\mathbb{N}:n_i=1}r_i(x),\qquad n=\sum_{i=0}^{\infty}n_i2^i.$$

Let us record a useful observation:

Lemma 1.1. *If the binary expansions of $n,m\in\mathbb{N}$ satisfy $n_im_i=0$ for all $i\in\mathbb{N}$, then $w_{n+m}=w_nw_m$.*

Proof. That $n_i m_i = 0$ means that $\{i : n_i = 1\}$ and $\{i : m_i = 1\}$ are disjoint, and then $\{i : (n+m)_i = 1\} = \{i : n_i = 1\} \cup \{i : m_i = 1\}$. Thus

$$w_{n+m} = \prod_{i:(n+m)_i=1} r_i = \prod_{i:n_i=1} r_i \times \prod_{i:m_i=1} r_i = w_n \times w_m.$$

□

Lemma 1.2. *If $0 \leq b < 2^k$, then $w_{2^k a+b} = w_{2^k a} w_b$.*

Proof. We have $(2^k a)_i = 0$ for $i < k$, whereas $b_i = 0$ for $i \geq k$. The claim follows from the previous lemma. □

Lemma 1.3. $w_{2^k a}(x) = w_a(2^k x)$.

Proof. Easy from the definition and $r_{i+k}(x) = r_0(2^{i+k}x) = r_i(2^k x)$. □

We now develop for the Walsh function w_n a theory analogous to the trigonometric functions e_n.

Proposition 1.1. *$(w_n)_{n=0}^{\infty}$ is an orthonormal basis of $L^2(0,1)$.*

Proof. Checking the identity

$$\langle w_n, w_m \rangle = \int_0^1 w_n \cdot w_m = \delta_{mn} = \begin{cases} 1 & m = n, \\ 0 & m \neq n, \end{cases}$$

is left as an exercise. It remains to check that $\operatorname{span}(w_n)_{n=0}^{\infty}$ is dense in $L^2(0,1)$. This in turn follows from the known fact that step functions (with arbitrarily small step) are dense in $L^2(0,1)$, and the following lemma: □

Lemma 1.4.

$$\Big\{ \sum_{j=0}^{2^k-1} a_j 1_{2^{-k}[j,j+1)} : a_j \in \mathbb{C} \Big\} = \operatorname{span}(w_n)_{n=0}^{2^k-1}.$$

Proof. First observe that both spaces have equal dimension 2^k. Indeed, the step functions $1_{2^{-k}[j,j+1)}$ are clearly linearly independent, and so are the w_n, since they are orthogonal. Thus it suffices to prove "$\supseteq$". For $n < 2^k$, we have $n_i \neq 0$ only for $i < k$, hence

$$w_n = \prod_{i=0}^{k-1} r_i.$$

But $r_i(x) = r_0(2^i x)$ is constant on intervals of length $2^{-i-1} \geq 2^{-k}$, and so hence is w_n as the product. This proves the inclusion $\supseteq$. □

Let us redefine the notation S_N to denote the partial Walsh sums now:

$$S_N f(x) := \sum_{n=0}^{N-1} \langle f, w_n \rangle w_n(x).$$

By elementary Hilbert space theory, we have that $\|S_N f - f\|_{L^2} \to 0$ as $N \to \infty$. The analogue of Carleson's theorem is this:

Theorem 1.3 (Billard 1966)**.** *For every $f \in L^2(0,1)$, we have*

$$S_N f(x) = \sum_{n=0}^{N-1} \langle f, w_n \rangle w_n(x) \to f(x)$$

for almost every $x \in [0,1)$.

Note that by the previous lemma, the convergence is trivial for the step functions, since then $S_n f(x) = f(x)$ for all sufficiently large n. By a similar argument as before, the general convergence then follows from:

Theorem 1.4. *There is a constant $C < \infty$ such that for every $f \in L^2(0,1)$,*

$$S^* f(x) := \sup_{N\in\mathbb{N}} |S_N f(x)|$$

satisfies

$$\|S^* f\|_{L^{2,\infty}} \leq C\|f\|_{L^2}.$$

We will prove this by using the modern *time-frequency analysis* of M. Lacey and C. Thiele.

1.3. **Time-frequency analysis for the Walsh model.** Let us work with the (horizontal) *time axis* $\mathbb{R}_+ = [0,\infty)$, and the (vertical) *frequency axis*, also equal to $\mathbb{R}_+ = [0,\infty)$. In both time and frequency, we consider the *dyadic intervals*

$$\mathscr{D} := \{2^{-k}[m, m+1) : k \in \mathbb{Z}, m \in \mathbb{N}\}.$$

A key feature of $\mathscr{D}$ is the following *nestedness*:

$$\forall I, J \in \mathscr{D} : \quad I \cap J \in \{\varnothing, I, J\};$$

any two intervals are either disjoint, or one contains the other.

We write $I^{(1)}$ for the *parent* of I: the smallest dyadic interval which strictly contains I. The two intervals I_0, I_1, which have I as their parent, are called the *children* of I, and they are called *siblings* of each other. We usually denote time intervals by I and frequency intervals by ω.

The product $\mathbb{R}_+ \times \mathbb{R}_+$ of time and frequency axes is called the *phase plane*. The sets

$$P = I \times \omega, \qquad I, \omega \in \mathscr{D}$$

are called *dyadic rectangles*. A *tile* is a dyadic rectangle of area $|P| = 1$ and a *bitile* is a dyadic rectangle of area $|P| = 2$. A bitile $P = I \times \omega$ naturally splits into its up-tile $P_u = I \times \omega_u$ and down-tile $P_d = I \times \omega_d$, where ω_d and ω_u are two siblings whose parent is ω. We often write I_P for the time interval and ω_P for the frequency interval of a tile or bitile P.

Since a tile has area 1 and both time and frequency intervals are dyadic, it can also be written as

$$P = I \times \frac{1}{|I|}[n, n+1), \qquad I \in \mathscr{D}, n \in \mathbb{N}.$$

Then we define the *Walsh wave packet* as

$$w_P(x) := \frac{1_I(x)}{|I|^{1/2}} w_n(\frac{x}{|I|}).$$

Clearly $\|w_P\|_{L^2} = 1$. It is left as an exercise to check that

$$\langle w_P, w_{P'}\rangle := \int_0^\infty w_P \cdot w_{P'} = 0, \qquad \text{if } P \cap P' = \varnothing.$$

Hence:

Lemma 1.5. *If $\mathbb{P}$ is a disjoint collection of tiles, then $\{w_P : P \in \mathbb{P}\}$ is an orthonormal basis of a subspace of $L^2(0,\infty)$, and*

$$\sum_{P\in\mathbb{P}} \langle f, w_P\rangle w_P$$

is the orthogonal projection of f onto this substpace.

In the wave packet formalism, $w_n 1_{[0,1)} = w_{[0,1)\times[n,n+1)}$. The partial Walsh sums can be written as

$$S_N f(x) = \sum_{n=0}^{N-1} \langle f, w_n\rangle w_n(x) = \sum_{\substack{P \text{ tile} \\ I_P=[0,1) \\ \min \omega_P < N}} \langle f, w_P\rangle w_P(x).$$

For the further analysis, it is useful to rewrite this:

Proposition 1.2 (Key identity).

$$\sum_{\substack{P \text{ tile} \\ I_P=[0,1) \\ \min \omega_P < N}} \langle f, w_P\rangle w_P(x) = \sum_{\substack{P \text{ bitile} \\ I_P\subseteq[0,1) \\ \omega_{P_u} \ni N}} \langle f, w_{P_d}\rangle w_{P_d}(x).$$

Note that, even though the summation variable in the second sum is a *bitile*, inside the sum we have its down-tile P_d, which is a tile, as in the original sum.

To prove the above identity, a key observation is this:

Lemma 1.6. *The collection of tiles P in the first sum, and the collection of down-tiles P_d in the second sum, are both disjoint, and they both cover the same domain $[0,1)\times[0,N)$ of the phase plane.*

You can best convince yourself of this by drawing a picture of the tiles in a special case like $N=7$, and figuring from there what is going on in general. A detailed proof follows further below.

Hence, both sides of the key identity represent orthogonal projections of f to certain subspaces of $L^2(0,\infty)$. It only remains to check that these subspaces are the same, for then so are the projections. This is a special case of the following:

Proposition 1.3. *Let $\mathbb{P}_i$, $i=1,2$, be two finite collections of tiles such that*

$$\bigcup_{P\in\mathbb{P}_1} P \subseteq \bigcup_{P\in\mathbb{P}_2} P$$

as subsets of $\mathbb{R}^2_+$. Then also

$$\operatorname{span}\{w_P : P\in\mathbb{P}_1\} \subseteq \operatorname{span}\{w_P : P\in\mathbb{P}_2\}$$

as subspaces of $L^2(\mathbb{R}_+)$. In particular, if there is equality on the first line, then also on the second.

It is convenient to handle two special cases first:

Lemma 1.7.

$$w_{I\times\omega} \in \operatorname{span}\{w_{I_i\times\omega^{(1)}} : i=0,1\},$$

where I_i, $i=0,1$, are the two children of I, and $\omega^{(1)}$ is the parent of ω.

Proof. Let $\omega=|I|^{-1}[n,n+1)$ and $\omega^{(1)}=2|I|^{-1}[m,m+1)$. Then $\omega\subset\omega^{(1)}$ implies that

$$2m\leq n, \qquad n+1\leq 2(m+1) \qquad \Rightarrow \qquad n=2m+r, \qquad r\in\{0,1\}.$$

Thus

$$w_{I\times\omega} = |I|^{-1/2}1_I w_n(\frac{\cdot}{|I|}) = |I|^{-1/2}1_I w_{2m+r}(\frac{\cdot}{|I|}) = |I|^{-1/2}1_I w_m(2\frac{\cdot}{|I|}) w_r(\frac{\cdot}{|I|}).$$

Here $w_r(\cdot/|I|)\equiv(-1)^{ir}$ on I_i, $i=0,1$, and hence

$$|I|^{-1/2}1_I w_m(2\frac{\cdot}{|I|}) w_r(\frac{\cdot}{|I|}) = \sum_{i=0}^{1}\frac{2^{-1/2}}{|I_i|^{1/2}}1_{I_i} w_m(\frac{\cdot}{|I_i|})(-1)^{ir} = \sum_{i=0}^{1}\frac{(-1)^{ir}}{\sqrt{2}} w_{I_i\times\omega^{(1)}}. \qquad \square$$

Lemma 1.8.

$$w_{I\times\omega} \in \operatorname{span}\{w_{I^{(1)}\times\omega_i} : i=0,1\},$$

where $I^{(1)}$ is the parent of I, and ω_i, $i=0,1$, are the two children of ω.

Proof. Left as an exercise. $\square$

Proof of Proposition 1.3. It suffices to show that every w_{P_1}, $P_1\in\mathbb{P}_1$, belongs to $\operatorname{span}\{w_P : P\in\mathbb{P}_2\}$. Here $\mathbb{P}_2$ is a collection of tiles which covers P_1. It suffices to show that $w_{P_1}\in\operatorname{span}\{w_P : P\in\mathbb{P}\}$, whenever the tiles of $\mathbb{P}$ cover P_1, and in fact (by removing some tiles), we may assume that $\mathbb{P}$ is minimal, in the sense that if we remove any tile from $\mathbb{P}$, then the new collection no longer covers P_1. We will make an induction on the number of elements in $\mathbb{P}$. If $\#\mathbb{P}=1$, then $\mathbb{P}=\{P_1\}$, and the claim is trivial. Otherwise, $P_1\notin\mathbb{P}$. Then every tile P that intersects P_1 either has $I_P\supsetneq I_{P_1}$, or $\omega_P\supsetneq\omega_{P_1}$. We call the first kind of tiles horizontal, and the second kind of tiles vertical.

We claim that $\mathbb{P}$ contains either horizontal tiles only, or vertical tiles only. Suppose first that the vertical tiles in $\mathbb{P}$ cover P_1. Then the horizontal tiles do not occur in $\mathbb{P}$ by minimality. Suppose then that the vertical tiles do not cover P_1, and let $(x, y) \in P_1$ be a point not covered by them. But any vertical tile covers a part of P_1 of the form $I \times \omega_{P_1}$, so if (x, y) is not covered by vertical tiles, then no point of the segment $\{x\} \times \omega_{P_1}$ is covered by vertical tiles. Hence all of this segment must be covered by horizontal tiles. But any horizontal tile covers a part of P_1 of the form $I_{P_1} \times \omega$, so if the horizontal tiles cover all of $\{x\} \times \omega_{P_1}$, they must cover all of $I_{P_1} \times \omega_{P_1} = P_1$. And then, as in the first part, no vertical tiles are needed to cover P_1, and hence they do not occur in $\mathbb{P}$.

Suppose for example that P_1 is covered by horizontal tiles, where no tile can be removed from the cover. Let $P = I \times \omega$ be the longest tile in the cover (any of them, if there are many). Consider its 'sibling' $P' = I \times \omega'$, where $\omega^{(1)} = \omega \cup \omega'$. We claim that also $P' \in \mathbb{P}$. In fact, some tiles in $\mathbb{P}$ must cover the set $P' \cap P_1 = I_{P_1} \times \omega'$. Since P is the longest tile, these must be shorter than or equal to P. But if any such tile is strictly shorter than P, then its frequency interval is strictly longer than ω_P. By the properties of dyadic intervals, this frequency interval would then cover ω_P, and the hypothetical tile would cover $P \cap P_1$. Then P would be redundant, which is a contradiction. Hence the only possibility is that a tile in $\mathbb{P}$ which covers $P_1 \cap P'$ has equal length with P, and then in fact it must be P'. We are done with the claim that also $P' \in \mathbb{P}$.

Next, consider the tile $\hat{P} := \hat{I} \times \omega^{(1)}$, where $\hat{I} \supseteq I_{P_1}$ is one of the children of I. Then $\hat{P}$ covers the same part of P_1 as P and P' together. Thus $\hat{\mathbb{P}} := (\mathbb{P} \setminus \{P, P'\}) \cup \{\hat{P}\}$ is also a cover of P_1. It has one element less than $\mathbb{P}$, so by induction assumption, we have that

$$w_{P_1} \in \operatorname{span}\{w_{P''} : P'' \in (\mathbb{P} \setminus \{P, P'\}) \cup \{\hat{P}\}\}$$

On the other hand, by Lemma 1.8, we have $w_{\hat{P}} \in \operatorname{span}\{w_P, w_{P'}\}$, and hence finally

$$w_{P_1} \in \operatorname{span}\{w_{P''} : P'' \in \mathbb{P}\},$$

completing this branch of the induction.

If, instead, P_1 is covered by vertical tiles, then the induction step is exactly analogous, only using Lemma 1.7 instead of Lemma 1.8. □

To complete the proof of the key identity of Proposition 1.2, we give:

Proof of Lemma 1.6. On the left side, the tiles are of the form $P = [0, 1) \times [n, n + 1)$, where $0 \le n < N$, so clearly they cover disjointly the domain $[0, 1) \times [0, N)$.

We then analyse the right side. Since $I_P \subseteq [0, 1)$ has length at most 1, the frequency intervals ω_{P_d} and ω_{P_u} have length $2^k \ge 1$. Let P be one of the bitiles appearing in the sum with $P_u = I_P \times |I_P|^{-1}[n, n + 1)$, where $|I_P| = 2^{-k}$. The fact that $|I_P|^{-1}[n, n + 1) = 2^k[n, n + 1)$ is the upper half of a dyadic interval means that n is odd. For the parameter N of S_N, let us make the binary expansion

$$N = \sum_{i=0}^{\infty} N_i 2^i.$$

Now the condition that $N \in 2^k[n, n + 1)$ says that

$$n \le 2^{-k}N = \sum_{i=k}^{\infty} N_i 2^{i-k} + \sum_{i=0}^{k-1} N_i 2^{i-k} < n + 1,$$

where $\sum_{i=k}^{\infty} N_i 2^{i-k}$ is an integer, and $\sum_{i=0}^{k-1} N_i 2^{i-k} \in [0, 1 - 2^{-k}]$, where the lower bound corresponds to $N_i \equiv 0$, the upper bound to $N_i \equiv 1$. Hence

$$\sum_{i=k}^{\infty} N_i 2^{i-k} - 1 \le \sum_{i=k}^{\infty} N_i 2^{i-k} + \sum_{i=0}^{k-1} N_i 2^{i-k} - 1 < n \le \sum_{i=k}^{\infty} N_i 2^{i-k} + \sum_{i=0}^{k-1} N_i 2^{i-k} < \sum_{i=k}^{\infty} N_i 2^{i-k} + 1,$$

which shows that (given $k \in \mathbb{N}$) there is a unique admissible value of n, namely, $n = \sum_{i=k}^{\infty} N_i 2^{i-k}$. This is odd exactly when $N_k = 1$. If $\omega_{P_u} = 2^k[n, n + 1)$, then ω_{P_d} is the "previous" interval of the same length, namely, $\omega_{P_d} = 2^k[n-1,n)$, and we have

$$n-1 = \sum_{i=k}^{\infty} N_i 2^{i-k} - 1 = \sum_{i=k+1}^{\infty} N_i 2^{i-k}, \qquad N_k = 1.$$

We conclude that the frequency intervals ω_{P_d} appearing on the right side of the key identity are exactly those of the form

$$2^k[n-1,n) = \Big[\sum_{i=k+1}^{\infty} N_i 2^i, \sum_{i=k}^{\infty} N_i 2^i\Big), \qquad k \in \mathbb{N}, \quad N_k = 1.$$

Notice that if $N_k = 0$, then the lower and upper limits of the above interval coincide, so the interval would be just $\varnothing$.

For the time intervals on the right of the key identity, the only condition is that $I_P \subseteq [0,1)$. Thus for a fixed k, we get all the intervals of the form $2^{-k}[j,j+1)$, $0 \le j \le 2^k - 1$. So altogether, the down-tiles P_d appearing on the right of the key identity are those of the form

$$P_d = 2^{-k}[j,j+1) \times \Big[\sum_{i=k+1}^{\infty} N_i 2^i, \sum_{i=k}^{\infty} N_i 2^i\Big), \qquad k \in \mathbb{N}, \quad N_k = 1, \quad j = 0, \dots, 2^k - 1.$$

For different k, the frequency intervals are disjoint, and for equal k but different j, the time intervals are disjoint. Hence all these tiles are pairwise disjoint.

For a fixed k but variable j, the time intervals cover exactly all of $[0,1)$. And for variable k, it is easy to check that the time intervals cover exactly all of

$$\Big[\lim_{k\to\infty} \sum_{i=k}^{\infty} N_i 2^i, \sum_{i=0}^{\infty} N_i 2^i\Big) = [0,N).$$

This completes the proof. □

2. Approach to Billard's theorem

We have already seen that to prove the pointwise convergence of the Walsh series, $S_N f(x) \to f(x)$, it suffices to estimate the maximal operator $S^* f(x) = \sup_{N\ge 0} |S_N f(x)|$ in the $L^{2,\infty}$ norm. By monotone convergence, it suffices to estimate $S^*_{N_0} f(x) : - = \max_{0\le N\le N_0} |S_N f(x)|$, as long as the bound is independent of the finite but arbitrarily large number N_0. This approximating operator has the advantage that we know *a priori* that it is finite. It is still a nonlinear operator, but we can linearize it as follows:

Lemma 2.1. *Given $f \in L^2(0,1)$, we can choose a measurable function $x \in [0,1) \mapsto N(x) \in \{0,1,\dots,N_0\}$ such that*

$$S^*_{N_0} f(x) = |S_{N(x)} f(x)| \qquad \forall x \in [0,1).$$

Proof. Consider the measurable sets

$$E_n := \{x \in [0,1) : |S_n f(x)| = S^*_{N_0} f(x)\}, \qquad F_n := E_n \setminus \bigcup_{k<n} E_k.$$

The sets F_n form a disjoint cover of $[0,1)$. Then

$$S^*_{N_0} f(x) = \Big(\sum_{n=0}^{N_0} 1_{F_n}(x)\Big) S^*_{N_0} f(x) = \sum_{n=0}^{N_0} 1_{F_n}(x) |S_n f(x)| = \Big|\sum_{n=0}^{N_0} 1_{F_n}(x) S_n f(x)\Big|.$$

If we know define uniquely $N(x) := n$ if $x \in F_n$, then we see that the right side is equal to $|S_{N(x)} f(x)|$. □

So it suffices to prove that

$$\|Cf\|_{L^{2,\infty}} := \|S_{N(x)} f\|_{L^{2,\infty}} \lesssim \|f\|_{L^2}$$

for all $f \in L^2$, where we abbreviate $C := S_{N(x)}$ — C for Carleson. For the estimation of the $L^{2,\infty}$ norm, the following lemma is useful:

Lemma 2.2. *For $p \in (1,\infty)$ and $g \in L^1_{\mathrm{loc}}$, we have*

$$\|g\|_{L^{p,\infty}} := \sup_{\lambda>0} \lambda |\{|g|>\lambda\}|^{1/p} \lesssim A \tag{2.1}$$

if and only if

$$\Big|\int_E g\Big| \lesssim A|E|^{1/p'}, \qquad \frac{1}{p}+\frac{1}{p'}=1, \tag{2.2}$$

for all bounded measurable sets E.

Proof. By considering the positive and negative parts of the real and imaginary parts of g separately, it is easy to see that it suffices to prove this for $g \geq 0$. We only prove that (2.2) implies (2.1) (which is the direction relevant for our approach to Billard's theorem), and leave the other direction as an exercise.

So let $g \geq 0$, assume (2.2), fix $\lambda > 0$, and let $E \subseteq \{g > \lambda\}$ be a bounded subset. Then

$$|E| = \int_E 1 \leq \int_E \frac{g}{\lambda} = \frac{1}{\lambda}\int_E g \leq \frac{1}{\lambda} A|E|^{1/p'},$$

thus $\lambda|E|^{1/p} = \lambda|E|^{1-1/p} \leq A$. Letting $E \subseteq \{g > \lambda\}$ approach $\{g > \lambda\}$, we deduce that $\lambda|\{g > \lambda\}|^{1/p} \leq A$, which was to be proven. □

Hence, we are reduced to proving that

$$\Big|\int_E Cf\Big| = |\langle Cf, 1_E\rangle| \lesssim \|f\|_{L^2}|E|^{1/2}.$$

Using the key identity for $Cf(x) = S_{N(x)}f(x)$, we find that

$$\begin{aligned}
\int_E Cf(x)\,\mathrm{d}x &= \int_E \sum_{\substack{P \text{ bitile} \\ I_P\subseteq[0,1) \\ \omega_{P_u}\ni N(x)}} \langle f, w_{P_d}\rangle w_{P_d}(x)\,\mathrm{d}x \\
&= \int_E \sum_{\substack{P \text{ bitile} \\ I_P\subseteq[0,1)}} \langle f, w_{P_d}\rangle w_{P_d}(x) 1_{\omega_{P_u}}(N(x))\,\mathrm{d}x \\
&= \sum_{\substack{P \text{ bitile} \\ I_P\subseteq[0,1)}} \langle f, w_{P_d}\rangle \langle w_{P_d}, 1_{E_{P_u}}\rangle, \qquad E_{P_u} := E \cap \{x : N(x) \in \omega_{P_u}\}.
\end{aligned}$$

As the final reduction, rather than the quantity as above, we simply choose to estimate

$$\sum_{P\in\mathbb{P}} \langle f, w_{P_d}\rangle \langle w_{P_d}, 1_{E_{P_u}}\rangle, \qquad E_{P_u} := E \cap \{x : N(x) \in \omega_{P_u}\},$$

where $\mathbb{P}$ is an arbitrary finite collection of bitiles. The actual estimate of interest is then obtained by specializing to the particular collection appearing earlier.

The estimation of this object will depend on partitioning the arbitrary collection $\mathbb{P}$ into smaller collections with more structure. For this purpose we introduce:

2.1. **Order for the tiles and bitiles.** A partial order "$\leq$" between either two tiles or two bitiles is defined as follows:

$$P \leq P' \qquad \text{if and only if} \qquad I_P \subseteq I_{P'} \quad \text{and} \quad \omega_P \supseteq \omega_{P'}.$$

One can check (exercise) that

$$P \leq P' \qquad \text{if and only if} \qquad P_d \leq P'_d \quad \text{or} \quad P_u \leq P'_u,$$

as well as (exercise): any two tiles or bitiles satisfy one of the following:

$$P \leq P' \quad \text{or} \quad P' \leq P \quad \text{or} \quad P \cap P' = \varnothing.$$

Definition 2.1. A collection $\mathbb{T}$ of bitiles is called a *tree* if there exists a *top* bitile T (possibly, but not necessarily an element of $\mathbb{T}$) such that $P \leq T$ for all $P \in \mathbb{T}$. The tree is called an up-tree, if $P_u \leq T_u$ for all $P \in \mathbb{T}$, and a down-tile id $P_d \leq T_d$ for all $P \in \mathbb{T}$.

The top is not unique but (exercise) there always exists a minimal top.

The following result says something useful (as seen later) about the structure of the wavepackets associated to a tree:

Lemma 2.3. *Let $\mathbb{T}$ be an up-tree with top T. Then for all $P \in \mathbb{T}$, we have*

$$w_{P_d}(x) = \epsilon_{PT} \cdot w^{\infty}_{T_u}(x) \cdot h_{I_P}(x),$$

where

- $\epsilon_{PT} \in \{-1, +1\}$ *are constants depending on P and T,*
- $w^{\infty}_{T_u}(x) := |I_{T_u}|^{1/2} w_{T_u}(x)$ *is the Walsh wave packet with L^∞ normalization (so that its absolute value is one on its support), and*
- h_{I_P} *are so-called* Haar functions*, defined by*

$$h_I(x) := \frac{1_I(x)}{|I|^{1/2}} r_0(\frac{x}{|I|}).$$

Proof. We have $T_u = I_T \times |I_T|^{-1}[n_T, n_T + 1)$, with odd n_T. Consider an element $P \in \mathbb{T}$ with $P_u = I_P \times |I_P|^{-1}[n_P, n_P + 1)$, again with odd n_P, and let $2^{-k} := |I_P|/|I_T|$. Then $P_u \leq T_u$ says that

$$I_P \subseteq I_T \qquad \text{and} \qquad 2^{-k}(n_T + 1) - 1 \leq n_P \leq 2^{-k} n_T.$$

If $n_T = \sum_{i=0}^{\infty} 2^i n_i$, then the unique integer value of n_P in the given range is

$$n_P = \sum_{i=k}^{\infty} 2^{i-k} n_i,$$

which is odd if and only if $n_k = 1$. For those values of k, we have $P_d = I_P \times |I_P|^{-1}[n_P - 1, n_P)$, where

$$n_P - 1 = \sum_{i=k+1}^{\infty} 2^{i-k} n_i.$$

Hence

$$\begin{aligned}
w_{P_d}(x) &= \frac{1_{I_P}(x)}{|I_P|^{1/2}} w_{n_P - 1}\Big(\frac{x}{|I_P|}\Big) = \frac{1_{I_P}(x)}{|I_P|^{1/2}} w_{2^k(n_P-1)}\Big(\frac{x}{|I_T|}\Big) \\
&= \frac{1_{I_P}(x)}{|I_P|^{1/2}} \prod_{i=k+1}^{\infty} r_i\Big(\frac{x}{|I_T|}\Big)^{n_i} \\
&\overset{(*)}{=} \frac{1_{I_P}}{|I_P|^{1/2}} \prod_{i=0}^{\infty} r_i\Big(\frac{x}{|I_T|}\Big)^{n_i} \times r_k\Big(\frac{x}{|I_T|}\Big) \times \prod_{i=0}^{k-1} r_i\Big(\frac{x}{|I_T|}\Big)^{n_i} \\
&= 1_{I_T}(x) w_{n_T}\Big(\frac{x}{|I_T|}\Big) \times \frac{1_{I_P}}{|I_P|^{1/2}} r_0\Big(\frac{x}{|I_P|}\Big) \times \prod_{i=0}^{k-1} r_i\Big(\frac{x}{2^k |I_P|}\Big)^{n_i} \\
&= w^{\infty}_{T_u}(x) \times h_{I_P}(x) \times \prod_{i=0}^{k-1} r_i\Big(\frac{x}{2^k |I_P|}\Big)^{n_i}.
\end{aligned}$$

Note that $n_k = 1$ was used in $(*)$, together with $r_i^2 \equiv 1$. Notice that the last product takes a constant value on I_P, as r_i is constant over dyadic intervals of length 2^{-i-1}; this is our ϵ_{PT}. □

3. Tree analysis

Recall that our task is to estimate

$$\sum_{P \in \mathbb{P}} \langle f, w_{P_d} \rangle \langle w_{P_d}, 1_{E_{P_u}} \rangle, \qquad E_{P_u} := E \cap \{x : N(x) \in \omega_{P_u}\},$$

where $\mathbb{P}$ is an arbitrary finite collection of bitiles. The plan is to divide $\mathbb{P}$ into more structured subcollections (trees), and to estimate the sum over each of them separately. The function $f \in L^2$ and the measurable set E are considered to be fixed throughout the argument, so we do not always indicate the dependence of some quantities on these objects.

For every collection $\mathbb{P}$ of bitiles, we introduce two quantities to measure its "size":

$$\operatorname{density}(\mathbb{P}) := \sup_{P\in\mathbb{P}}\sup_{P'\geq P}\frac{|I_{P'}\cap E_{P'}|}{|I_{P'}|}, \qquad E_{P'} := E\cap\{x\ :\ N(x)\in\omega_{P'}\},$$

$$\operatorname{energy}(\mathbb{P}) := \sup_{\mathbb{T}\subseteq\mathbb{P}\text{ up-tree}}\Big(\frac{1}{|I_{\mathbb{T}}|}\sum_{P\in\mathbb{T}}|\langle f,w_{P_d}\rangle|^2\Big)^{1/2},$$

where $I_{\mathbb{T}}$ is the minimal time interval of a top of $\mathbb{T}$. Note that a more precise notation would be

$$\operatorname{density}_E(\mathbb{P}), \qquad \operatorname{energy}_f(\mathbb{P}),$$

as these quantities depend on E and f. The "density" has also been called "mass", but "density" is more descriptive, since involves the ratio of the measures of $I_{P'}\cap E_{P'}$ and $I_{P'}$; another name found in the literature is "support-size". The "energy" is also called "coefficient-size" or just "size". The name "energy" is motivated by the fact that such square-sums often describe energy in physical applications. However, this is only a philosophical motivation, and there isn't any deep physical meaning behind our chosen terminology.

3.1. **The tree lemma.**

Proposition 3.1. *For each tree $\mathbb{T}$, we have*

$$\Big|\sum_{P\in\mathbb{T}}\langle f,w_{P_d}\rangle\langle w_{P_d},g1_{E_{P_u}}\rangle\Big|\leq 10\cdot\operatorname{energy}(\mathbb{T})\operatorname{density}(\mathbb{T})|I_{\mathbb{T}}|,$$

where

$$E_{P_u} := E\cap\{x\ :\ N(x)\in\omega_{P_u}\}.$$

We start the proof by choosing $\mathscr{J}$ as the collection of maximal dyadic intervals $J\subseteq I_{\mathbb{T}}$ which do not contain any I_P, $P\in\mathbb{T}$. These intervals cover the set $I_{\mathbb{T}}$. Hence, observing that all w_{P_d} are supported on $I_{\mathbb{T}}$,

$$\begin{aligned}\Big|\sum_{P\in\mathbb{T}}\langle f,w_{P_d}\rangle\langle w_{P_d},g1_{E_{P_u}}\rangle\Big| &\leq \Big\|\sum_{P\in\mathbb{T}}\epsilon_P\langle f,w_{P_d}\rangle w_{P_d}1_{E_{P_u}}\Big\|_{L^1(\mathbb{R}_+)}\\ &= \sum_{J\in\mathscr{J}}\Big\|\sum_{P\in\mathbb{T}}\epsilon_P\langle f,w_{P_d}\rangle w_{P_d}1_{E_{P_u}}\Big\|_{L^1(J)}\\ &= \sum_{J\in\mathscr{J}}\Big\|\sum_{\substack{P\in\mathbb{T}\\ I_P\supsetneq J}}\epsilon_P\langle f,w_{P_d}\rangle w_{P_d}1_{E_{P_u}}\Big\|_{L^1(J)},\end{aligned}\tag{3.1}$$

where the last step is based on the following observations: First, only those w_{P_d}, whose support $\operatorname{supp} w_{P_d}=I_P$ intersects J contribute to the $L^1(J)$ norm. Second, if the dyadic intervals I_P and J intersect, then one is contained in the other. But J cannot contain any I_P by the definition of $\mathscr{J}$, so the only possibility is that $I_P\supsetneq J$. And this is exactly what we have written in the sum above.

Note that the function inside the $L^1(J)$ norm in (3.1) is supported on the set G_J defined below:

Lemma 3.1. *For a fixed $J\in\mathscr{J}$, the subset*

$$G_J := J\cap\bigcup_{\substack{P\in\mathbb{T}\\ I_P\supsetneq J}}E_{P_u}$$

satisfies $|G_J|\leq 2\operatorname{density}(\mathbb{T})|J|$.

Proof. Consider the dyadic parent $\hat{J}$ of J. By maximality of J, we have $\hat{J}\supseteq I_{\tilde{P}}$ for some $\tilde{P}\in\mathbb{T}$. Let $\hat{\omega}$ be the dyadic interval of size $2/|\hat{J}|$ such that $\omega_{\tilde{P}}\supseteq\hat{\omega}\supseteq\omega_T$, where T is the top of $\mathbb{T}$, so that the bitile $\hat{P}:=\hat{J}\times\hat{\omega}$ satisfies $\tilde{P}\leq\hat{P}\leq T$. Now we claim that

$$G_J\subseteq J\cap E_{\hat{P}}.\tag{3.2}$$

Indeed, consider one of the P appearing in G_J. Then $P \in \mathbb{T}$, thus $I_P \subseteq I_T$ and $\omega_P \supseteq \omega_T$, and also $I_P \supsetneq J$, thus $I_P \supseteq \hat{J}$. We also have $|\omega_P| = 2/|I_P| \leq 2/|\hat{J}| = |\hat{\omega}|$, and $\omega_P \cap \hat{\omega} \supseteq \omega_T \neq \varnothing$, hence $\omega_P \subseteq \hat{\omega}$. But this means that

$$E_{P_u} = E \cap \{N \in \omega_{P_u}\} \subseteq E \cap \{N \in \omega_P\} \subseteq E \cap \{N \in \hat{\omega}\} = E_{\hat{P}},$$

which proves the claim (3.2).

The proof is completed as follows, recalling that $\hat{P} \geq \tilde{P} \in \mathbb{T}$:

$$\begin{aligned} |G_J| \leq |J \cap E_{\hat{P}}| \leq |\hat{J}| \frac{|\hat{J} \cap E_{\hat{P}}|}{|\hat{J}|} &= 2|J| \frac{|I_{\hat{P}} \cap E_{\hat{P}}|}{|I_{\hat{P}}|} \\ &\leq 2|J| \sup_{P' \geq \tilde{P}} \frac{|I_{P'} \cap E_{P'}|}{|I_{P'}|} \leq 2|J| \operatorname{density}(\mathbb{T}). \end{aligned}$$ □

Next, divide $\mathbb{T}$ into the down- and up-trees

$$\mathbb{T}_d := \{P \in \mathbb{T} : P \leq_d T\}, \qquad \mathbb{T}_u := \mathbb{T} \setminus \mathbb{T}_d,$$

and write

$$F_{jJ} := \sum_{\substack{P \in \mathbb{T}_j \\ I_P \supsetneq J}} \epsilon_J \langle f, w_{P_d} \rangle w_{P_d} 1_{E_{P_u}}, \qquad j \in \{d, u\}.$$

To continue the estimate (3.1), we want to control $\|F_{dJ} + F_{uJ}\|_{L^1(J)}$. We will use the simple estimate $\|F\|_{L^1(J)} \leq |\operatorname{supp}(1_J F)| \|F\|_{L^\infty(J)}$, where we already controlled the size of the support in the previous lemma. The L^∞ norms of F_{dJ} and F_{uJ} will be controlled because of different reasons.

Lemma 3.2.

$$\|F_{dJ}\|_{L^\infty(J)} \leq \operatorname{energy}(\mathbb{T}).$$

Proof. Here the key observation is:

> The ω_{P_u} appearing in F_{dJ} are pairwise disjoint. Hence, so are the sets $E_{P_u} = E \cap \{N \in P_u\}$.

To prove this, suppose that $P, P' \in \mathbb{T}_d$ appear in the same sum F_{dJ}. Then $\omega_{P_d}, \omega_{P'_d} \supseteq \omega_{T_d}$. If ω_{P_d} is the larger of the two, then $\omega_{P_d} \supsetneq \omega_{P'_d}$ and hence $\omega_{P_d} \supseteq \omega_{P'}$. Thus ω_{P_u} is disjoint from $\omega_{P'}$ and in particular from $\omega_{P'_u}$.

Thanks to this disjointness of the supports, the L^∞ norm of the sum is the maximum of the individual terms:

$$\|F_{dJ}\|_\infty = \max_{\substack{P \in \mathbb{T}_d \\ I_P \supsetneq J}} \|\langle f, w_{P_d} \rangle w_{P_d} 1_{E_{P_u}}\|_\infty \leq \max_{\substack{P \in \mathbb{T}_d \\ I_P \supsetneq J}} \frac{|\langle f, w_{P_d} \rangle|}{|I_P|^{1/2}} \leq \operatorname{energy}(\mathbb{T}),$$

where the last step used the observation that every collection $\{P\}$ consisting of a single bitile is always an up-tree! □

Lemma 3.3.

$$\|F_{uJ}\|_{L^\infty(J)} \leq 2 \inf_{z \in J} M\tilde{f}(z),$$

where

$$\tilde{f} := \sum_{P \in \mathbb{T}_u} \langle f, w_{P_d} \rangle w_{P_d},$$

and M is the dyadic maximal operator, defined by

$$Mg(x) := \sup_{\substack{I \in \mathscr{D} \\ I \ni x}} \frac{1}{|I|} \int_I |f(y)| \, dy.$$

Proof. Here, instead of disjointness, we have nestedness. To be more precise, consider a fixed $x \in J$ with $F_{uJ}(x) \neq 0$; hence $1_{E_{P_u}}(x) = 1_E(x) 1_{\omega_{P_u}}(N(x)) \neq 0$ for some $P \in \mathbb{T}_u$ with $I_P \supsetneq J$. Since $\mathbb{T}_u$ is an up-tree, all ω_{P_u} contain ω_{T_u} for the top T, and hence they form a nested sequence of dyadic intervals. If $\omega_{P'_u}$ is the smallest interval among these, which contains the point $N(x)$, then all the bigger intervals ω_{P_u} also contain $N(x)$. Thus $1_{E_{P_u}}(x) \neq 0$ holds for if and only if $|\omega_{P_u}| \geq |\omega_{P'_u}|$, or equivalently, if and only if $|I_P| \leq |I_{P'}|$. Since F_{uJ} involves the summation condition $I_P \supsetneq J$, all the appearing time intervals also form a nested sequence, and the last condition can be written as $I_P \subset I_{P'}$. Thus

$$\begin{aligned} F_{uJ}(x) &= \sum_{\substack{P \in \mathbb{T}_u \\ J \subsetneq I_P \\ 1_{E_{P_u}}(x) \neq 0}} \langle f, w_{P_d} \rangle w_{P_d}(x) \\ &= \sum_{\substack{P \in \mathbb{T}_u \\ J \subsetneq I_P \subseteq I_{P'}}} \langle f, w_{P_d} \rangle w_{P_d}(x) \\ &= w_{T_u}^{\infty}(x) \sum_{\substack{P \in \mathbb{T}_u \\ J \subsetneq I_P \subseteq I_{P'}}} \epsilon_{PT} \langle f, w_{P_d} \rangle h_{I_P}(x), \end{aligned}$$

where we used Lemma 2.3 in the last step.

To proceed, we make an observation about the dyadic averages

$$\mathsf{E}_K g := \frac{1_K}{|K|} \int_K g, \qquad K \in \mathscr{D},$$

of the Haar functions:

$$\mathsf{E}_K h_I = \begin{cases} 1_K h_I & K \subsetneq I \\ 0 & \text{else.} \end{cases} \tag{3.3}$$

Indeed, h_I is constant on the dyadic intervals strictly contained in K, so for $K \subsetneq I$, the average of h_I on K is just the value of h_I on any point of K. The remaining cases are that either $K \cap I = \varnothing$, or $K \supseteq I$. In the first case, since h_I is supported on I, its average outside is zero. And in the last case, we notice that also the average of h_I on its support is zero, since it takes the same positive and negative value on sets of equal size.

From (3.3) it follows that for all $x \in J \subseteq I_{P'}$,

$$\begin{aligned} \sum_{\substack{P \in \mathbb{T}_u \\ J \subsetneq I_P \subseteq I_{P'}}} \epsilon_{PT} \langle f, w_{P_d} \rangle h_{I_P}(x) &= \Big(\sum_{\substack{P \in \mathbb{T}_u \\ J \subsetneq I_P}} - \sum_{\substack{P \in \mathbb{T}_u \\ I_{P'} \subsetneq I_P}} \Big) \epsilon_{PT} \langle f, w_{P_d} \rangle h_{I_P}(x) \\ &= (\mathsf{E}_J - \mathsf{E}_{I_{P'}}) \sum_{P \in \mathbb{T}_u} \epsilon_{PT} \langle f, w_{P_d} \rangle h_{I_P}(x) = (\mathsf{E}_J - \mathsf{E}_{I_{P'}}) w_{T_u}^{\infty} \sum_{P \in \mathbb{T}_u} \langle f, w_{P_d} \rangle w_{P_d}(x), \end{aligned}$$

where we used again Lemma 2.3 in the last step, together with $\epsilon_{PT}^2 = 1$.

Let $z \in J$. Since both J and $I_{P'}$ contain any z, both $\mathsf{E}_J g(x)$ and $\mathsf{E}_{I_{P'}} g(x)$ are among the averages that occur in the definition of $Mg(z)$. Thus

$$|(\mathsf{E}_J - \mathsf{E}_{I_{P'}}) g(x)| \leq 2Mg(z).$$

We apply this to the results of the above computations, recalling that $\|w_{T_u}^{\infty}\|_\infty = 1$:

$$\begin{aligned} |F_{uJ}(x)| &\leq \Big| (\mathsf{E}_J - \mathsf{E}_{I_{P'}}) w_{T_u}^{\infty} \sum_{P \in \mathbb{T}_u} \langle f, w_{P_d} \rangle w_{P_d}(x) \Big| \\ &\leq 2M \Big(\sum_{P \in \mathbb{T}_u} \langle f, w_{P_d} \rangle w_{P_d} \Big)(z). \end{aligned}$$

Taking the supremum over $x \in J$ and the infimum over $z \in J$ proves the claim. $\square$

Proof of the Tree Lemma (Proposition 3.1). We substitute the estimates from Lemmas 3.1, 3.2 and 3.3 to (3.1):

$$\begin{aligned}\Big|\sum_{P\in\mathbb{T}}\langle f,w_{P_d}\rangle\langle w_{P_d},g1_{E_{P_u}}\rangle\Big| &\leq \sum_{J\in\mathscr{J}}\|F_{dJ}+F_{uJ}\|_{L^1(J)}\\ &\leq \sum_{J\in\mathscr{J}}|G_J|\big(\|F_{dJ}\|_\infty+\|F_{uJ}\|_\infty\big)\\ &\leq \sum_{J\in\mathscr{J}}2\,\mathrm{density}(\mathbb{T})|J|\Big(\mathrm{energy}(\mathbb{T})+2\inf_J M\tilde f\Big)\\ &\leq 2\,\mathrm{density}(\mathbb{T})\,\mathrm{energy}(\mathbb{T})|I_T|+4\,\mathrm{density}(\mathbb{T})\int_{I_T}M\tilde f(x)\,\mathrm{d}x,\end{aligned}$$

where we observed that

$$|J|\inf_J M\tilde f=\int_J\inf_J M\tilde f\,\mathrm{d}x\leq\int_J M\tilde f(x)\,\mathrm{d}x$$

and used the disjointness of the intervals $J\in\mathscr{J}$, which are all contained in I_T, to conclude that

$$\sum_{J\in\mathscr{J}}|J|\leq|I_T|,\qquad\sum_{J\in\mathscr{J}}\int_J M\tilde f(x)\,\mathrm{d}x\leq\int_{I_T}M\tilde f(x)\,\mathrm{d}x.$$

To conclude, we use Cauchy–Schwarz and the L^2-boundedness of the dyadic maximal operator (with norm 2; see Proposition 3.2 below):

$$\int_{I_T}M\tilde f(x)\,\mathrm{d}x\leq|I_T|^{1/2}\Big(\int M\tilde f(x)^2\,\mathrm{d}x\Big)^{1/2}\leq 2|I_T|^{1/2}\Big(\int|\tilde f(x)|^2\,\mathrm{d}x\Big)^{1/2}.$$

Recall that $\tilde f=\sum_{P\in\mathbb{T}_u}\langle f,w_{P_d}\rangle w_{P_d}$, and use the orthonormality of the functions w_{P_d} for $P\in\mathbb{T}_u$ to see that

$$\Big(\int|\tilde f(x)|^2\,\mathrm{d}x\Big)^{1/2}=\Big(\sum_{P\in\mathbb{T}_u}|\langle f,w_{P_d}\rangle|^2\Big)^{1/2}\leq|I_T|^{1/2}\,\mathrm{energy}(\mathbb{T}).$$

Thus

$$4\,\mathrm{density}(\mathbb{T})\int_{I_T}M\tilde f(x)\,\mathrm{d}x\leq 4\,\mathrm{density}(\mathbb{T})2|I_T|\,\mathrm{energy}(\mathbb{T}),$$

and altogether

$$\Big|\sum_{P\in\mathbb{T}}\langle f,w_{P_d}\rangle\langle w_{P_d},g1_{E_{P_u}}\rangle\Big|\leq(2+4\cdot 2)\,\mathrm{density}(\mathbb{T})\,\mathrm{energy}(\mathbb{T})|I_T|,$$

where $2+4\cdot 2=10$, as claimed. □

In the proof we used the following result:

Proposition 3.2. *The dyadic maximal operator*

$$Mf(x):=\sup_{I\in\mathscr{D}}\frac{1_I(x)}{|I|}\int_I|f(y)|\,\mathrm{d}y$$

satisfies $\|Mf\|_{L^2}\leq 2\|f\|_{L^2}$.

A proof is indicated in the exercises.

3.2. **The density lemma.** Recall that our final goal is to estimate

$$\sum_{P\in\mathbb{P}}\langle f,w_{P_d}\rangle\langle w_{P_d},1_{E_{P_u}}\rangle,$$

where $\mathbb{P}$ is any finite collection of bitiles. From the Tree Lemma, we know how to make such an estimate if $\mathbb{P}$ is a tree. This is not the case in general, so our task is to efficiently decompose an arbitrary collection into trees. This is accomplished by means of two decomposition results, where in both cases we extract some trees, and make sure that either the density or the energy of the remaining collection decreases.

We first address the somewhat easier case of density:

Proposition 3.3. *Every finite set $\mathbb{P}$ of bitiles has a disjoint decomposition*

$$\mathbb{P} = \mathbb{P}_{\text{sparse}} \cup \bigcup_j \mathbb{T}_j,$$

where each $\mathbb{T}_j$ is a tree, and

$$\operatorname{density}(\mathbb{P}_{\text{sparse}}) \leq \frac{1}{4}\operatorname{density}(\mathbb{P}), \qquad \sum_j |I_{\mathbb{T}_j}| \leq 4\operatorname{density}(\mathbb{P})^{-1}|E|.$$

Proof. We make $\mathbb{P}_{\text{sparse}}$ as big as possible by setting

$$\mathbb{P}_{\text{sparse}} := \Big\{P \in \mathbb{P} : \sup_{P' \geq P} \frac{|I_{P'} \cap E_{P'}|}{|I_{P'}|} \leq \frac{1}{4}\operatorname{density}(\mathbb{P})\Big\}.$$

Thus $\mathbb{P} \setminus \mathbb{P}_{\text{sparse}}$ is as small as possible, and we should now decompose it into tiles.

For every $P \in \mathbb{P} \setminus \mathbb{P}_{\text{sparse}}$, we pick some bitile P' such that

$$\frac{|I_{P'} \cap E_{P'}|}{|I_{P'}|} > \frac{1}{4}\operatorname{density}(\mathbb{P}).$$

Let T_j be the maximal bitiles (with respect to their partial order $\leq$) among these chosen P', and let

$$\mathbb{T}_j := \{P \in \mathbb{P} : P \leq T_j\}$$

be the tree in $\mathbb{P}$ with top T_j. Then

$$\mathbb{P} \setminus \mathbb{P}_{\text{sparse}} = \bigcup_j \mathbb{T}_j.$$

Observe that the sets $I_{T_j} \cap E_{T_j} = I_{T_j} \cap E \cap \{N \in \omega_{T_j}\}$, which are all contained in E, are pairwise disjoint. Indeed, if

$$[I_{T_j} \cap E \cap \{N \in \omega_{T_j}\}] \cap [I_{T_k} \cap E \cap \{N \in \omega_{T_k}\}] \neq \varnothing,$$

then necessarily

$$I_{T_j} \cap I_{T_k} \neq \varnothing, \qquad \text{and} \qquad \omega_{T_j} \cap \omega_{T_k} \neq \varnothing,$$

thus

$$T_j \cap T_k = [I_{T_j} \cap I_{T_k}] \times [\omega_{T_j} \cap \omega_{T_k}] \neq \varnothing,$$

and then one of T_j and T_k could not be maximal. Thus we have

$$\sum_j |I_{T_j}| \leq 4\operatorname{density}(\mathbb{P})^{-1} \sum_j |I_{T_j} \cap E_{T_j}| \leq 4\operatorname{density}(\mathbb{P})^{-1}|E|,$$

where the last estimate used the fact that all $I_{T_j} \cap E_{T_j}$ are disjoint and contained in E. □

3.3. **The energy lemma.** The analogous statement for the energy is as follows:

Proposition 3.4. *Every finite set $\mathbb{P}$ of bitiles has a disjoint decomposition*

$$\mathbb{P} = \mathbb{P}_{\text{low}} \cup \bigcup_j \mathbb{T}_j,$$

where each $\mathbb{T}_j$ is a tree, and

$$\operatorname{energy}(\mathbb{P}_{\text{low}}) \leq \frac{1}{2}\operatorname{energy}(\mathbb{P}), \qquad \sum_j |I_{\mathbb{T}_j}| \leq 4\operatorname{energy}(\mathbb{P})^{-2}\|f\|_{L^2}^2.$$

Proof. For every tree $\mathbb{T}$, let

$$\Delta(\mathbb{T}) := \Big(\frac{1}{|I_T|}\sum_{P\in\mathbb{T}_u} |\langle f, w_{P_d}\rangle|^2\Big)^{1/2},$$

where T is a minimal top of $\mathbb{T}$, and $\mathbb{T}_u := \{P \in \mathbb{T} : P \leq_u T\}$ is the up-tree supported by the same top.

Let $\mathscr{E} :=$ energy$(\mathbb{P})$. We extract the trees $\mathbb{T}_j$ recursively as follows: Consider all maximal trees $\mathbb{T} \subseteq \mathbb{P}$ among the ones with $\Delta(\mathbb{T}) > \frac{1}{2}\mathscr{E}$. Among them, let $\mathbb{T}_1$ be one whose top frequency interval $\omega_{\mathbb{T}}$ has the minimal center $c(\omega_{\mathbb{T}})$. Replace $\mathbb{P}$ by $\mathbb{P} \setminus \mathbb{T}_j$, and iterate. When no trees can be chosen anymore, the remaining collection $\mathbb{P}_{\text{low}}$ satisfies energy$(\mathbb{P}_{\text{low}}) \leq \frac{1}{2}\mathscr{E}$ by definition.

The sum over the top intervals is immediately estimated by

$$\sum_j |I_{T_j}| \leq \Big(\frac{2}{\mathscr{E}}\Big)^2 \sum_j \sum_{P\in\mathbb{T}_{j,u}} |\langle f, w_{P_d}\rangle|^2. \tag{3.4}$$

We would like to identify the double summation with the L^2 norm of the orthogonal projection of f onto the span of the appearing functions w_{P_d}. For this it is necessary to check that any two w_{P_d} above are orthogonal to each other, equivalently, that any two P_d above are disjoint. This is automatically the case if both bitiles belong to the same up-tree $\mathbb{T}_{j,u}$, so it remains to consider the case where $P_j \in \mathbb{T}_{j,u}$, $P_i \in \mathbb{T}_{i,u}$, and $i \neq j$. We argue by contradiction and assume that two such bitiles intersect. And then it follows that, for example, $P_{j,d} \leq P_{i,d}$, and hence $\omega_{P_{i,d}} \subseteq \omega_{P_{j,d}}$. Since $P_i \neq P_j$ are different bitiles, we must in fact have $\omega_{P_{i,d}} \subsetneq \omega_{P_{j,d}}$ and hence $\omega_{P_i} \subseteq \omega_{P_{j,d}}$. Thus, we have

$$\omega_{T_i} \subseteq \omega_{P_i} \subseteq \omega_{P_{j,d}}, \qquad \omega_{T_{j,u}} \subseteq \omega_{P_{j,u}},$$

and therefore

$$c(\omega_{T_j}) = \inf \omega_{T_{j,u}} \geq \inf \omega_{P_{j,u}} = \sup \omega_{P_{j,d}} > c(\omega_{T_i}).$$

This means that the tree $\mathbb{T}_i$ was chosen first, thus $i < j$. But $P_{j,d} \leq P_{i,d}$ implies $P_j \leq P_i \leq T_i$, so that P_j should have been taken to $\mathbb{T}_i$ by maximality. This gives a contradiction, proving the falsity of our counterassumption.

Hence any two w_{P_d} appearing in (3.4) are disjoint, and thus we continue with

$$\sum_j |I_{T_j}| \leq \Big(\frac{2}{\mathscr{E}}\Big)^2 \sum_j \sum_{P\in\mathbb{T}_{j,u}} |\langle f, w_{P_d}\rangle|^2 \leq \frac{4}{\mathscr{E}^2}\|f\|_{L^2}^2. \qquad \square$$

3.4. **Iterating the decomposition lemmas.** By using the density and energy lemmas consecutively, we obtain the following result, whose verification is left as an exercise:

Lemma 3.4. *Suppose that*

$$\text{density}(\mathbb{P}_n) \leq 4^n|E|, \qquad \text{energy}(\mathbb{P}_n) \leq 2^n\|f\|_{L^2}.$$

Then

$$\mathbb{P}_n = \mathbb{P}_{n-1} \cup \bigcup_j \mathbb{T}_{n,j}, \qquad \sum_j |I_{\mathbb{T}_{j,n}}| \leq C4^{-n},$$

where $\mathbb{P}_{n-1}$ satisfies estimates similar to $\mathbb{P}_n$ with $n-1$ in place of n.

The following result goes one step further:

Proposition 3.5. *For any finite collection of bitiles $\mathbb{P}$, there is a decomposition*

$$\mathbb{P} = \bigcup_{n\in\mathbb{Z}} \bigcup_j \mathbb{T}_{n,j} \cup \mathbb{P}_{-\infty},$$

where each $\mathbb{T}_{n,j}$ is a tree, and we have the estimates

$$\text{density}(\mathbb{T}_{n,j}) \leq 4^n|E|, \qquad \text{energy}(\mathbb{T}_{n,j}) \leq 2^n\|f\|_{L^2}, \qquad \sum_j |I_{\mathbb{T}_{n,j}}| \leq C4^{-n}$$

and density$(\mathbb{P}_{-\infty}) =$ energy$(\mathbb{P}_{-\infty}) = 0$.

Proof. Since the density and energy of every finite collection are some finite numbers, $\mathbb{P}$ satisfies the conditions of $\mathbb{P}_n$ in Lemma 3.4 for some possibly large $n_0 \in \mathbb{Z}$. By iterating that Lemma, we obtain

$$\begin{aligned}\mathbb{P} = \mathbb{P}_{n_0} &= \bigcup_j \mathbb{T}_{n_0,j} \cup \mathbb{P}_{n_0-1} = \bigcup_j \mathbb{T}_{n_0,j} \cup \bigcup_j \mathbb{T}_{n_0-1,j} \cup \mathbb{P}_{n_0-2} \\ &= \sum_{n=n_1}^{n_0} \bigcup_j \mathbb{T}_{n,j} \cup \mathbb{P}_{n_1-1}, \qquad n_1 < n_0,\end{aligned}$$

where $\mathbb{P}_{n_0} \supseteq \mathbb{P}_{n_0-1} \supseteq \cdots \supseteq \mathbb{P}_{n_1} \supseteq \mathbb{P}_{n_1-1}$. Iterating indefinitely, we observe that every bitile in $\mathbb{P}$ either gets chosen to some $\mathbb{T}_{n,j}$, or else it belongs to every residual colllection $\mathbb{P}_n$, hence also to their intersection $\mathbb{P}_{-\infty} := \bigcap_{n \leq n_0} \mathbb{P}_n$. But then

$$\text{density}(\mathbb{P}_{-\infty}) \leq \text{density}(\mathbb{P}_n) \leq 4^n |E|$$

for every $n \leq n_0$, and letting $n \to -\infty$ we see that $\text{density}(\mathbb{P}_{-\infty}) = 0$. The argument for $\text{energy}(\mathbb{P}_{-\infty}) = 0$ is similar.

Finally, since $\mathbb{T}_{n,j} \subseteq \mathbb{P}_n$, it is clear that it satisfies the same density and energy bounds, and the bound for $\sum_j |I_{\mathbb{T}_{n,j}}|$ is part of Lemma 3.4. □

3.5. **Completion of the proof of Theorem 1.4.** Recall that the proof has been reduced to controlling

$$\sum_{P \in \mathbb{P}} \langle f, w_{P_d} \rangle \langle w_{P_d}, 1_{E_{P_u}} \rangle,$$

where $\mathbb{P}$ is an arbitrary finite collection of bitiles. By Proposition 3.5, we have

$$\sum_{P \in \mathbb{P}} \langle f, w_{P_d} \rangle \langle w_{P_d}, 1_{E_{P_u}} \rangle = \sum_{n \in \mathbb{Z}} \sum_j \sum_{P \in \mathbb{T}_{n,j}} \langle f, w_{P_d} \rangle \langle w_{P_d}, 1_{E_{P_u}} \rangle + \sum_{P \in \mathbb{P}_{-\infty}} \langle f, w_{P_d} \rangle \langle w_{P_d}, 1_{E_{P_u}} \rangle.$$

From the fact that $\text{energy}(\mathbb{P}_{-\infty}) = 0$, it follows that $\langle f, w_{P_d} \rangle = 0$ for all $P \in \mathbb{P}_{-\infty}$, so the last sum may be ignored. And then we simply estimate with the Tree Lemma, observing that every collection has the trivial density bound $\text{density}(\mathbb{P}) \leq 1$:

$$\begin{aligned}&\sum_{n \in \mathbb{Z}} \sum_j \Big| \sum_{P \in \mathbb{T}_{n,j}} \langle f, w_{P_d} \rangle \langle w_{P_d}, 1_{E_{P_u}} \rangle \Big| \\ &\lesssim \sum_{n \in \mathbb{Z}} \sum_j \text{density}(\mathbb{T}_{n,j})\, \text{energy}(\mathbb{T}_{n,j}) |I_{T_{n,j}}| \\ &\lesssim \sum_{n \in \mathbb{Z}} \min\{1, 4^n |E|\} \times 2^n \|f\|_{L^2} \times \sum_j |I_{T_{n,j}}| \\ &\lesssim \sum_{n \in \mathbb{Z}} \min\{1, 4^n |E|\} \times 2^n \|f\|_{L^2} \times 4^{-n} \\ &\lesssim \sum_{n : 2^n \leq |E|^{-1/2}} 4^n |E| \times 2^n \|f\|_{L^2} \times 4^{-n} + \sum_{n : 2^n > |E|^{-1/2}} 1 \times 2^n \|f\|_{L^2} \times 4^{-n} \\ &\lesssim \sum_{n : 2^n \leq |E|^{-1/2}} |E| \|f\|_{L^2} 2^n + \sum_{n : 2^n > |E|^{-1/2}} \|f\|_{L^2} \times 2^{-n}.\end{aligned}$$

The two convergent geometric series give us

$$|E| \|f\|_{L^2} |E|^{-1/2} + \|f\|_{L^2} |E|^{1/2} \lesssim \|f\|_{L^2} |E|^{1/2},$$

which was the required estimate to prove that

$$|\langle S_{N(\cdot)} f, 1_E \rangle| \lesssim \|f\|_{L^2} |E|^{1/2}$$

and therefore

$$\|S_{N(\cdot)} f\|_{L^{2,\infty}} \lesssim \|f\|_{L^2}.$$

This completes the proof of the pointwise convergence of Walsh series.

4. The Fourier transform

We now leave the Walsh model and return to classical Fourier analysis. However, instead of the Fourier series from the motivating discussion in the beginning, we now investigate the *Fourier transform* defined for $f \in L^1(\mathbb{R})$ and $\xi \in \mathbb{R}$ as

$$\mathscr{F} f(\xi) \equiv \hat{f}(\xi) := \int_{\mathbb{R}} f(x) e^{-i2\pi x\cdot\xi} \, \mathrm{d}x.$$

By taking absolute values inside the norm, it is clear that $|\hat{f}(\xi)| \leq \|f\|_1$ for all ξ. If $\xi \to \xi_0$, the continuity of the exponential function implies that $e^{-i2\pi x\cdot\xi} \to e^{-i2\pi x\cdot\xi_0}$ for every x, and it follows from dominated convergence that $\hat{f}(\xi) \to \hat{f}(\xi_0)$. So the function $\hat{f}$ is continuous, and in particular measurable. If $f, g \in L^1(\mathbb{R})$, then it is easy to check that their *convolution*

$$f * g(x) := \int_{\mathbb{R}} f(x-y) g(y) \, \mathrm{d}y = \int_{\mathbb{R}} f(y) g(x-y) \, \mathrm{d}y$$

satisfies $\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$.

Important properties of the Fourier transform follow just from the knowledge of the transform of one particular function:

Lemma 4.1. *The function $\varphi(x) := e^{-\pi x^2}$ satisfies $\hat{\varphi} = \varphi$.*

Proof. Note that $\hat{\varphi}(0)$ is given by the familiar Gaussian integral,

$$\hat{\varphi}(0)^2 = \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-\pi(x^2+y^2)} \, \mathrm{d}x \, \mathrm{d}y = \int_0^\infty \int_0^{2\pi} e^{-\pi r^2} \, \mathrm{d}\theta \, r \, \mathrm{d}r = \int_0^\infty 2\pi r e^{-\pi r^2} \, \mathrm{d}r = \Big|_0^\infty - e^{-\pi r^2} = 1,$$

and hence $\hat{\varphi}(0) = 1$, since it is clearly positive. Now there are (at least) two ways to finish the proof.

(*By Cauchy's theorem for complex path integrals.*) Completing the square

$$\hat{f}(\xi) = \int_{-\infty}^\infty e^{-\pi(x^2+i2x\cdot\xi+(i\xi)^2-(i\xi)^2)} \, \mathrm{d}x = \int_{-\infty}^\infty e^{-\pi(x+i\xi)^2} \, \mathrm{d}x \cdot e^{-\pi\xi^2} = \int_{-\infty+i\xi}^{\infty+i\xi} e^{-\pi z^2} \, \mathrm{d}z \cdot e^{-\pi\xi^2}.$$

By Cauchy's theorem, it is easy to check that one can shift the integration path back to the real axis, and this was just computed above.

(*By the uniqueness theory of ordinary differential equations.*) Notice that $\varphi'(x) = -2\pi x\varphi(x)$. Taking the Fourier transform of both sides and integrating by parts, it follows that $i2\pi\xi\hat{\varphi}(\xi) = -i\hat{\varphi}'(\xi)$. Hence both φ and $\hat{\varphi}$ are solutions of the differential equation

$$u'(x) = -2\pi x u(x), \qquad u(0) = 1,$$

and therefore must be equal. □

By Lemma 4.1 (interchanging the roles of x and ξ)

$$\varphi(x) = \hat{\varphi}(x) = \int_{\mathbb{R}} \varphi(\xi) e^{-i2\pi x\cdot\xi} \, \mathrm{d}\xi.$$

Since φ is real-valued, taking complex conjugates of both sides one can replace $-i$ by $+i$ in the exponent. Substituting x/ε in place of x and changing integration variables, one further obtains

$$\varphi_\varepsilon(x) := \frac{1}{\varepsilon}\varphi\Big(\frac{x}{\varepsilon}\Big) = \int_{\mathbb{R}} \varphi(\varepsilon\xi) e^{i2\pi x\cdot\xi} \, \mathrm{d}\xi.$$

Lemma 4.2. *For $f \in L^p(\mathbb{R})$, $p \in [1,\infty)$, we have $\varphi_\varepsilon * f(x) \to f(x)$ as $\varepsilon \to 0$ for almost every x.*

Proof. This is a toy variant of the pointwise convergence questions that we study in this course. First, we check that the convergence holds for a dense class of functions, say $g \in C_c(\mathbb{R})$ (continuous with compact support). We have

$$\varphi_\varepsilon * g(x) = \int_{\mathbb{R}} \frac{1}{\varepsilon}\varphi\Big(\frac{y}{\varepsilon}\Big) g(x-y) \, \mathrm{d}y = \int_{\mathbb{R}} \varphi(y) g(x-\varepsilon y) \, \mathrm{d}y.$$

As $\varepsilon \to 0$, we have $g(x - \varepsilon y) \to g(x)$ for every y, and also that $|\varphi(y)g(x - \varepsilon y)| \leq |\varphi(y)|\|g\|_\infty$, which is integrable with respect to y. Hence $\varphi_\varepsilon * g(x) \to \int \varphi(y)\,\mathrm{d}y \cdot g(x) = g(x)$ by dominated convergence.

Next, we need to control the maximal operator $\sup_{\varepsilon>0} |\varphi_\varepsilon * f(x)|$. But it is easy to check (exercise) that this is dominated by the Hardy–Littlewood maximal function $Mf(x)$, which satisfies $\|Mf\|_{L^{p,\infty}} \leq C\|f\|_{L^p}$. Hence the pointwise convergence follows for every $f \in L^p(\mathbb{R})$ by the standard procedure. □

Theorem 4.1 (Fourier inversion)**.** *Suppose that both $f \in L^1(\mathbb{R})$ and $\hat{f} \in L^1(\mathbb{R})$. Then, a.e.,*

$$f(x) = \int_{\mathbb{R}} \hat{f}(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi.$$

Accordingly, we call

$$\mathscr{F}^{-1} g(x) := \check{g}(x) = \int_{\mathbb{R}} g(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi$$

the *inverse Fourier transform* of $g \in L^1(\mathbb{R})$.

Proof. By Lemma 4.2, $\varphi_\varepsilon * f(x) \to f(x)$ for a.e. $x \in \mathbb{R}$, and then

$$\begin{aligned} f(x) &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}} \varphi_\varepsilon(y) f(x-y)\,\mathrm{d}y \\ &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(\varepsilon\xi) e^{i2\pi y\cdot\xi}\,\mathrm{d}\xi f(x-y)\,\mathrm{d}y \\ &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}} \varphi(\varepsilon\xi) \int_{\mathbb{R}} e^{-i2\pi(x-y)\cdot\xi} f(x-y)\,\mathrm{d}y\, e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi \\ &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}} \varphi(\varepsilon\xi)\hat{f}(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi = \int_{\mathbb{R}} \hat{f}(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi, \end{aligned}$$

where the second to last equality was the definition of $\hat{f}(\xi)$, and the last one was dominated convergence based on the fact that $\varphi(\varepsilon\xi) \to \varphi(0) = 1$ at every ξ. □

Corollary 4.1 (L^2 isometry)**.** *If $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then $\hat{f} \in L^2(\mathbb{R})$ and $\|\hat{f}\|_2 = \|f\|_2$.*

Proof. Let first $\hat{f} \in L^1(\mathbb{R})$. Then

$$\begin{aligned} \|\hat{f}\|_2^2 = \int_{\mathbb{R}} \hat{f}(\xi)\overline{\hat{f}(\xi)}\,\mathrm{d}\xi &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x) e^{-i2\pi x\cdot\xi}\,\mathrm{d}x\,\overline{\hat{f}(\xi)}\,\mathrm{d}\xi \\ &= \int_{\mathbb{R}} f(x)\overline{\int_{\mathbb{R}} \hat{f}(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi}\,\mathrm{d}x = \int_{\mathbb{R}} f(x)\overline{f(x)}\,\mathrm{d}x = \|f\|_2^2. \end{aligned}$$

In general, let again $\varphi(x) = e^{-\pi x^2}$. Then $\varphi_\varepsilon * f \in L^1 \cap L^2$, and its Fourier transform is $\varphi(\varepsilon\xi)\hat{f}(\xi) \in L^1(\mathbb{R})$. Thus $\|\varphi(\varepsilon\xi)\hat{f}(\xi)\|_2 = \|\varphi_\varepsilon * f\|_2$. As $\varepsilon \searrow 0$, we have $\varphi_\varepsilon * f \to f$ in the L^2 norm. (We have this convergence pointwise by Lemma 4.2, and the L^2 convergence follows easily by dominated convergence, since $|\varphi_\varepsilon * f| \lesssim Mf \in L^2$). Moreover, $|\varphi(\varepsilon\xi)\hat{f}(\xi)| \nearrow |\hat{f}(\xi)|$, so

$$\|\hat{f}\|_2 = \lim_{\varepsilon\searrow 0} \|\varphi(\varepsilon\xi)\hat{f}(\xi)\|_2 = \lim_{\varepsilon\searrow 0} \|\varphi_\varepsilon * f\|_2 = \|f\|_2$$

by monotone convergence. □

4.1. **The Fourier transform on $L^2(\mathbb{R})$.** The previous result allows to extend the Fourier transform for all $f \in L^2(\mathbb{R})$: Note that $L^1 \cap L^2$ is dense in L^2, since it contains for instance all continuous compactly supported functions. If $f \in L^2(\mathbb{R})$ is arbitrary, let $f_n \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ be a sequence with $f_n \to f$ in the L^2 norm. Then

$$\|\hat{f}_n - \hat{f}_m\|_2 = \|\mathscr{F}(f_n - f_m)\|_2 = \|f_n - f_m\|_2 \to 0,$$

and hence $\hat{f}_n$ form a Cauchy sequence in $L^2(\mathbb{R})$. We denote its limit by $\hat{f} := \lim_{n\to\infty} \hat{f}_n$. It is routine to check that this definition of $\hat{f}$ is independent of the chosen approximating sequence f_n, and that we have $\|\hat{f}\|_2 = \|f\|_2$.

Since $\mathscr{F}$ and $\mathscr{F}^{-1}$ only differ by the sign of i in the exponent, it is clear that the same is true for the inverse transform $\mathscr{F}^{-1}$: it also extends to an isometry $\mathscr{F}^{-1} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$. If $g \in L^1(\mathbb{R})$ and $\hat{g} \in L^1(\mathbb{R})$, then $\mathscr{F}^{-1}\mathscr{F}g = \mathscr{F}\mathscr{F}^{-1}g = g$. Since such functions are dense in $L^2(\mathbb{R})$, and both $\mathscr{F}$ and $\mathscr{F}^{-1}$ are continuous on $L^2(\mathbb{R})$, these identities remain valid for all $g \in L^2(\mathbb{R})$. So $\mathscr{F}$ and $\mathscr{F}^{-1}$ are both bijections, and inverses to each other, as operators on $L^2(\mathbb{R})$.

4.2. **Pointwise inversion for $f \in L^2(\mathbb{R})$.** For a general $f \in L^2(\mathbb{R})$, we can write the identities

$$f = \mathscr{F}^{-1}\hat{f} = \lim_{n\to\infty} \mathscr{F}^{-1}(1_{[-n,n]}\hat{f}),$$

where the limit is in the sense of L^2 norm convergence. Since $\hat{f} \in L^2 \subset L^1_{\mathrm{loc}}$, we have $1_{[-n,n]}\hat{f} \in L^1(\mathbb{R})$, and we can write the inverse Fourier transform explicitly as

$$\mathscr{F}^{-1}(1_{[-n,n]}\hat{f})(x) = \int_{\mathbb{R}} 1_{[-n,n]}(\xi)\hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi = \int_{-n}^{n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi,$$

and these converge to $f(x)$ in the sense of the L^2 norm. The question we would like to address is whether we can also have the pointwise convergence. This is the content of Carleson's theorem for the Fourier transform:

Theorem 4.2 (Carleson 1966)**.** *For all $f \in L^2(\mathbb{R})$, the following convergence takes place at almost every $x \in \mathbb{R}$:*

$$f(x) = \lim_{n\to\infty} \int_{-n}^{n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi.$$

To prove it, we use the usual procedure. First check the convergence for a dense class of functions. A suitable dense class is given by the function $f \in L^1(\mathbb{R})$ with also $\hat{f} \in L^1(\mathbb{R})$, since for them we already proved that

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi = \lim_{n\to\infty} \int_{-n}^{n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi$$

by Theorem 4.1 and dominated convergence. After this observation, the proof of Carleson's theorem depends on the control of Carleson's maximal operator

$$\begin{aligned}\sup_{n>0}\Big|\int_{-n}^{n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi\Big| &= \sup_{n>0}\Big|\int_{-\infty}^{n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi - \int_{-\infty}^{-n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi\Big| \\ &\leq 2\sup_{n\in\mathbb{R}}\Big|\int_{-\infty}^{n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi\Big| =: 2Cf(x).\end{aligned}$$

Theorem 4.3 (Carleson)**.** *For all $f \in L^2(\mathbb{R})$, we have $\|Cf\|_{L^{2,\infty}(\mathbb{R})} \lesssim \|f\|_{L^2(\mathbb{R})}$.*

The proof of this will depend on setting up an appropriate time-frequency analysis, after the model from the Walsh case. However, let us first discuss some intrinsic difficulties in the construction of wave packets for the Fourier transform.

4.3. **The uncertainty principle.** A key difficulty is that both the function and its Fourier transform cannot be arbitrarily well localized at the same time. Different versions of this phenomenon are known as *uncertainty principles.*

Proposition 4.1 (A toy uncertainty principle)**.** *If $\operatorname{supp}\hat{f}$ is compact for some $f \in L^2(\mathbb{R})$, then $\operatorname{supp} f = \mathbb{R}$ (unless f is identically zero).*

By symmetry, we also have a similar conclusion with the roles of f and $\hat{f}$ reversed.

Proof. If $\operatorname{supp}\hat{f} \subseteq [-n,n]$, then

$$f(x) = \int_{-n}^{n} \hat{f}(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi$$

extends to a holomorphic function

$$f(z) = \int_{-n}^{n} \hat{f}(\xi)e^{i2\pi z\xi}\,\mathrm{d}\xi.$$

Indeed, each $z \in \mathbb{C} \mapsto e^{i2\pi z\xi}$ is a holomorphic function, and complex differentiation under the integral is easily justified by dominated convergence. But it is well-known from complex analysis that the zeros of a holomorphic function are isolated points, so in particular f cannot vanish identically on any interval, so its support must contain all of $\mathbb{R}$. $\square$

Proposition 4.2 (Heisenberg's uncertainty principle)**.** *For all $f \in L^2(\mathbb{R})$ and all $x_0, \xi_0 \in \mathbb{R}$, we have*

$$\Big(\int_{\mathbb{R}} (x-x_0)^2 |f(x)|^2 \,\mathrm{d}x\Big)^{1/2} \Big(\int_{\mathbb{R}} (\xi-\xi_0)^2 |\hat{f}(\xi)|^2 \,\mathrm{d}\xi\Big)^{1/2} \geq \frac{1}{4\pi}\|f\|_2^2.$$

In quantum mechanics, the state of a particle is represented by a *wave function* ψ. Then $|\psi(x)|^2$ is interpreted as the probability density of finding the particle in the position x. Similarly $|\hat{\psi}(\xi)|^2$ is the probability density of momentum of the particle. The uncertainty principle gives a lower bound on how well both the position and the momentum can be localized simultaneously. Note that we are using dimensionless units in the above mathematical formulation; with physical units, the Planck constant would appear in this formula.

Proof. We consider the case of f sufficiently nice, so that all the manipulations below are justified. The general case follows easily by density considerations. Observe that

$$\begin{aligned}\xi\hat{f}(\xi) = \xi\int_{\mathbb{R}} f(x)e^{-i2\pi x\xi}\,\mathrm{d}x &= \frac{1}{-i2\pi}\int_{\mathbb{R}} f(x)\partial_x e^{-i2\pi x\xi}\,\mathrm{d}x \\ &= \frac{1}{i2\pi}\int_{\mathbb{R}} \partial_x f(x)e^{-i2\pi x\xi}\,\mathrm{d}x = \mathscr{F}(\frac{\partial_x}{i2\pi}f)(\xi)\end{aligned}$$

using integration by parts, assuming that f decays so that no boundary terms appear at $\pm\infty$. Hence the left side of the claim, with $x_0 = \xi_0 = 0$, is equal to

$$\|xf(x)\|_2\|\xi\hat{f}(\xi)\|_2 = \|xf(x)\|_2\|\frac{\partial_x}{i2\pi}f(x)\|_2.$$

The operators $P = x$ (understood as pointwise multiplication by x) and $Q = \partial_x/i2\pi$ are known as the position and momentum operators in quantum mechanics. Like all operators in quantum mechanics, they are both self-adjoint ($\langle Af, g\rangle = \langle f, Ag\rangle$ for both $A \in \{P, Q\}$ — for Q this amounts to integration by parts, where the boundary terms disappear; note also that we are using the sesqui-linear inner product, and i changes sign under the complex conjugation). The uncertainty principle results from the fact that the two operators do not commute: $[P,Q] := PQ - QP \neq 0$. In fact,

$$[P,Q]f := \frac{1}{i2\pi}(x\partial_x f - \partial_x(xf)) = \frac{1}{i2\pi}(x\partial_x f - f - x\partial_x f) = \frac{i}{2\pi}f.$$

Hence

$$\begin{aligned}\frac{1}{2\pi}\|f\|_2^2 &= |\langle f, [P,Q]f\rangle| = |\langle f, PQf\rangle - \langle f, QPf\rangle| = |\langle Pf, Qf\rangle - \langle Qf, Pf\rangle| \\ &= |\langle Pf, Qf\rangle - \overline{\langle Pf, Qf\rangle}| = |2i\operatorname{Im}\langle Pf, Qf\rangle| \leq 2\|Pf\|_2\|Qf\|_2.\end{aligned}$$

Unravelling the definitions, this is the claim for $x_0 = \xi_0 = 0$. The proof for general x_0 and ξ_0 is almost the same, observing that $\|(\xi-\xi_0)\hat{f}(\xi)\|_2 = \|(\partial_x/i2\pi - \xi_0)f(x)\|_2$, and that the shifted operators $P' = P - x_0$, $Q' = Q - \xi_0$ satisfy exactly the same commutation relation $[P', Q'] = i/2\pi$. $\square$

4.4. **The Schwartz test functions.** For the analysis of the Fourier transform, the following test function class is often convenient:

$$\mathscr{S}(\mathbb{R}) := \{f \in C^\infty(\mathbb{R}) : \forall \alpha, \beta \in \mathbb{N}, \sup_{x\in\mathbb{R}} |x^\alpha \partial_x^\beta f(x)| < \infty\}.$$

Notice that $C_c^\infty(\mathbb{R}) \subset \mathscr{S}(\mathbb{R})$, where $C_c^\infty(\mathbb{R})$ is the space of smooth, compactly supported functions. Since $C_c^\infty(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ ($p \in [1,\infty)$), so is the bigger space $\mathscr{S}(\mathbb{R})$. A useful feature of $\mathscr{S}(\mathbb{R})$ compared to $C_c^\infty(\mathbb{R})$ is that the Fourier transform (and the inverse Fourier transform) maps $\mathscr{S}(\mathbb{R})$ into itself. This is left as an exercise. We will often define a function $f \in \mathscr{S}(\mathbb{R})$ by specifying its Fourier transform $\hat{f}$. If $\hat{f} \in \mathscr{S}(\mathbb{R})$, then so is f.

5. Time-frequency analysis for the Fourier transform

5.1. **Wave packets.** Given the restrictions on the possible wave packets dictated by the uncertainty principle, we choose to work with functions that have perfect localization in frequency, and infinite support (but with rapidly decaying tail) in the time domain. Our *basic wave packet* will be a function $\phi \in \mathscr{S}(\mathbb{R})$ defined by the properties of its Fourier transform $\hat{\phi}$ as follows:

$$\operatorname{supp}\hat{\phi} \subseteq [-\tfrac{1}{20}, \tfrac{1}{20}], \qquad \sum_{k\in\mathbb{Z}} |\hat{\phi}(\xi + \tfrac{1}{20}k)|^2 \equiv 1 \quad \forall \xi \in \mathbb{R}. \tag{5.1}$$

A possible construction of such a function is indicated in the exercises.

Lemma 5.1. *The basic wave packet satisfies*

$$\langle T_{20n}\phi, \phi\rangle = \begin{cases} 1/20 & \text{if } n = 0, \\ 0 & \forall n \in \mathbb{Z}\setminus\{0\}, \end{cases}$$

where $T_y f(x) := f(x-y)$.

Proof. Since the Fourier transform is an L^2-isometry, we have

$$\langle T_{20n}\phi, \phi\rangle = \langle \widehat{T_{20n}\phi}, \hat{\phi}\rangle = \langle M_{-20n}\hat{\phi}, \hat{\phi}\rangle = \int_{\mathbb{R}} e^{-i2\pi\cdot 20n\xi}\hat{\phi}(\xi)\overline{\hat{\phi}(\xi)}\,\mathrm{d}\xi, \tag{5.2}$$

where $M_y f(x) := e^{i2\pi yx} f(x)$, and the proof that $\widehat{T_y f} = M_{-y}\hat{f}$ is left as an exercise. We split the domain of integration on the right of (5.2) as follows:

$$\begin{aligned}
\langle T_{20n}\phi, \phi\rangle &= \int_{\mathbb{R}} e^{-i2\pi\cdot 20n\xi}|\hat{\phi}(\xi)|^2\,\mathrm{d}\xi = \sum_{k\in\mathbb{Z}}\int_{k/20}^{(k+1)/20} e^{-i2\pi\cdot 20n\xi}|\hat{\phi}(\xi)|^2\,\mathrm{d}\xi \\
&= \sum_{k\in\mathbb{Z}}\int_0^{1/20} e^{-i2\pi\cdot 20n(\xi+k/20)}|\hat{\phi}(\xi+k/20)|^2\,\mathrm{d}\xi \\
&= \sum_{k\in\mathbb{Z}}\int_0^{1/20} e^{-i2\pi\cdot 20n\xi}|\hat{\phi}(\xi+k/20)|^2\,\mathrm{d}\xi \qquad \text{since } e^{-i2\pi\cdot 20n\cdot k/20} = 1 \\
&= \int_0^{1/20} e^{-i2\pi\cdot 20n\xi}\sum_{k\in\mathbb{Z}}|\hat{\phi}(\xi+k/20)|^2\,\mathrm{d}\xi \\
&= \int_0^{1/20} e^{-i2\pi\cdot 20n\xi}\cdot 1\,\mathrm{d}\xi \qquad \text{by assumption (5.1)},
\end{aligned}$$

and the last integral is easily seen to give the claimed result for all $n \in \mathbb{Z}$. □

Our tiles will be the same geometric objects as in the Walsh model, the only difference is that both time and frequency axes are now equal to $\mathbb{R}$, instead of $\mathbb{R}_+$. Following the literature, we make a slight (inessential) deviation from the Walsh model: instead of tiles and bitiles (dyadic rectangles of area 2), we now work with tiles and *semitiles* (dyadic rectangles of area $\frac{1}{2}$). So for every tile $P = I\times\omega$, we denote by $P_u = I\times\omega_u$ and $P_d = I\times\omega_d$ its up-semitile and down-semitile, respectively. We will also use the upmost quarter-tile $P_{uu} := I\times\omega_{uu}$, where ω_{uu} is the upmost quarter of ω, hence the upper half of ω_u.

The general wave packet on a tile P is built from the basic wave packet by three basic operations. Above, we already encountered the *translations* and *modulations*

$$T_y f(x) := f(x-y), \qquad M_y f(x) := e^{i2\pi yx}f(x), \qquad y\in\mathbb{R}.$$

The third kind of operation is the *dilation*

$$D^p_\lambda f(x) := \lambda^{-1/p} f(x/\lambda), \qquad \lambda > 0.$$

Here $p \in [1,\infty]$ (we interpret $\lambda^{-1/\infty} = \lambda^0 = 1$) is a parameter, which indicated that D^p_λ preserves the L^p norm. We now define:

$$\phi_P := M_{c(\omega_d)}T_{c(I)}D^2_{|I|}\phi, \qquad P = I\times\omega,$$

where $c(\omega_d)$ is the centre of the down-half ω_d of the frequency interval of ω, and $c(I)$ is the centre of the time interval I.

It is left as an exercise to check the following relations for the Fourier transforms of the basic operations:

$$\widehat{T_y f} = M_{-y}\hat{f}, \qquad \widehat{M_y f} = T_y \hat{f}, \qquad \widehat{D^p_\lambda f} = D^{p'}_\lambda \hat{f},$$

where p' is the dual exponent $(1/p + 1/p' = 1)$. Hence

$$\hat{\phi}_P = T_{c(\omega_d)} M_{-c(I)} D^2_{|\omega|}\hat{\phi},$$

where we also used the fact that $|I| \cdot |\omega| = 1$ when $P = I \times \omega$ is a tile.

Note that

$$\operatorname{supp} T_y f = y + \operatorname{supp} f, \quad \operatorname{supp} M_y f = \operatorname{supp} f, \quad \operatorname{supp} D^p_\lambda f = \lambda(\operatorname{supp} f) = \{\lambda x : x \in \operatorname{supp} f\}.$$

Hence

$$\operatorname{supp}\hat{\phi}_P = c(\omega_d) + |\omega| \operatorname{supp}\phi \subseteq c(\omega_d) + |\omega|[-\tfrac{1}{20}, \tfrac{1}{20}].$$

5.2. **The model operator A_ξ.** Recall that the relevant operator for the proof of Carleson's theorem is

$$Cf(x) = \sup_{\xi\in\mathbb{R}} |S_\xi f(x)|, \qquad S_\xi f(x) = \int_{-\infty}^{\xi} \hat{f}(\eta) e^{i2\pi x\eta}\,\mathrm{d}\eta.$$

In the Walsh model, we expressed the analogue of S_ξ, namely as

$$S_n f(x) = \sum_{k=0}^{n-1} \langle f, w_k\rangle w_k(x) = \sum_{\substack{P \text{ bitile} \\ I_P \subseteq [0,1)}} \langle f, w_{P_d}\rangle w_{P_d}(x) 1_{\omega_{P_u}}(n).$$

We would like to have a similar identity for S_ξ, and with this in mind, we define

$$A_\xi f(x) := \sum_{P \text{ tile}} \langle f, \phi_P\rangle \phi_P(x) 1_{\omega_{P_{uu}}}(\xi).$$

(As said, we are now using tiles in place of bitiles, and semitiles instead of tiles. Note that, by definition, ϕ_P is actually in some sense associated with P_d rather than P, and could also be called ϕ_{P_d}, but we prefer the shorter notation. Having $\omega_{P_{uu}}$ instead of ω_{P_u} is an inessential technical convenience later on.)

Now, it is *not true that* $S_\xi = A_\xi$. However, we will later see that there is nevertheless a useful way of expressing S_ξ *in terms of* A_ξ. (This will involve averaging over different choices of the system of dyadic cubes, which are implicitly behind the definition of tiles, and hence of A_ξ.) Before turning to this, let us analyse the operator A_ξ a bit on its own right.

Lemma 5.2. *If a tile P appears in A_ξ, then*

$$\operatorname{supp}\hat{\phi}_P \subseteq [\xi - 0.8|\omega_P|, \xi - 0.45|\omega_P|].$$

Proof. We now from above that, on the one hand,

$$\operatorname{supp}\hat{\phi}_P \subseteq [c(\omega_{P_d}) - 0.05|\omega_P|, c(\omega_{P_d}) + 0.05|\omega_P|],$$

and, on the other hand,

$$\xi \in \omega_{P_{uu}} = [c(\omega_{P_u}), c(\omega_{P_u}) + 0.25|\omega_P|].$$

Also, it is clear that $c(\omega_{P_u}) = c(\omega_{P_d}) + 0.5|\omega_P|$. Thus

$$c(\omega_{P_d}) \in [\xi - 0.75|\omega_P|, \xi - 0.5|\omega_P|],$$

and hence

$$\operatorname{supp}\hat{\phi}_P \subseteq [\xi - 0.75|\omega_P| - 0.05|\omega_P|, \xi - 0.5|\omega_P| + 0.05|\omega_P|],$$

which simplifies to the claim. □

Proposition 5.1. *For any tiles P, P' appearing in A_ξ, with $|\omega_P| \neq |\omega_{P'}|$, we have $\langle \phi_P, \phi_{P'}\rangle = 0$.*

Proof. Suppose for instance that $|\omega_{P'}| < |\omega_P|$. Since these are dyadic intervals, it means that $|\omega_{P'}| \le \frac{1}{2}|\omega_P|$, and then $0.8|\omega_{P'}| \le 0.4|\omega_P| < 0.45|\omega_P|$. Thus the whole interval $\xi - 0.45|\omega_P|$, which is the highest point of the support of $\hat{\phi}_P$, is strictly smaller that $\xi - 0.8|\omega_{P'}|$, which is the lowest point of the support of $\hat{\phi}_{P'}$. □

Proposition 5.2. *All tiles P appearing in A_ξ with given frequency size $|\omega_P| = |\omega|$ can be divided into 20 subcollections, so that for any tiles P, P' in the same subcollection, we have $\langle \phi_P, \phi_{P'}\rangle = 0$.*

$\langle \phi_P, \phi_{P'}\rangle = 0$.

Proof. When the length of the frequency interval $|\omega|$ is fixed, the full interval is uniquely determined by the condition that $\xi \in \omega_{uu}$. Since we are dealing with tiles (area 1), also the length of the time interval is uniquely determined as $|I| = 1/|\omega|$. Since the time intervals are dyadic, their centres satisfy $c(I_P) - c(I_{P'}) = n|I|$, $n \in \mathbb{Z}$. By picking every 20th such interval into a subcollection, we can form 20 subcollections with $c(I_P) - c(I_{P'}) = 20n|I|$ for P, P' in the same subcollection. Then, for such P, P', we have

$$\begin{aligned}
\langle \omega_P, \omega_{P'}\rangle &= \langle M_{c(\omega_{P_d})} T_{c(I_P)} D^2_{|I_P|}\phi, M_{c(\omega_{P'_d})} T_{c(I_{P'})} D^2_{|I_{P'}|}\phi\rangle \\
&= \langle M_{c(\omega_d)} T_{c(I_P)} D^2_{|I|}\phi, M_{c(\omega_d)} T_{c(I_{P'})} D^2_{|I|}\phi\rangle \\
&= \langle M_{c(\omega_d)} D^2_{|I|} T_{c(I_P)/|I|}\phi, M_{c(\omega_d)} D^2_{|I|} T_{c(I_{P'})/|I|}\phi\rangle \\
&= \langle T_{c(I_P)/|I|}\phi, T_{c(I_{P'})/|I|}\phi\rangle = \langle T_{c(I_P)/|I| - c(I_{P'})/|I|}\phi, \phi\rangle = \langle T_{20n}\phi, \phi\rangle = 0,
\end{aligned}$$

where we used the simple-to-check identities $\langle M_y f, M_y g\rangle = \langle f, g\rangle = \langle D^2_\lambda f, D^2_\lambda g\rangle$ and $\langle T_y f, T_z g\rangle = \langle T_{y-z} f, g\rangle$, and finally Lemma 5.1 in the last step. □

Now we easily see:

Theorem 5.1. *For every fixed $\xi \in \mathbb{R}$, A_ξ is a bounded operator on $L^2(\mathbb{R})$, and in fact*

$$\|A_\xi f\|_{L^2} \le \|f\|_{L^2}.$$

Proof. For each fixed side-length $|\omega|$ of the frequency interval, we divide the tiles appearing in A_ξ into 20 subcollections $\mathbb{P}_k(|\omega|)$, $k = 0, 1, \ldots, 19$, as in Proposition 5.2. Then we form $\mathbb{P}_k$ as the union of $\mathbb{P}_k(|\omega|)$ over all different side-lengths $|\omega|$. By Propositions 5.1 and 5.2 together, if both P, P' belong to the same $\mathbb{P}_k$, then $\langle \phi_P, \phi_{P'}\rangle = 0$. Hence each

$$\sum_{P \in \mathbb{P}_k} \langle f, \phi_P\rangle \phi_P = \|\phi\|^2_{L^2} \sum_{P \in \mathbb{P}_k} \langle f, \frac{\phi_P}{\|\phi_P\|_{L^2}}\rangle \frac{\phi_P}{\|\phi_P\|_{L^2}}$$

(we used $\|\phi_P\|_{L^2} = \|\phi\|_{L^2}$) is $\|\phi\|^2_{L^2} = 1/20$ (by Lemma 5.1) times the orthogonal projection of f into a certain subspace. In particular, the L^2 norm of such a sum is bounded by $\frac{1}{20}\|f\|_{L^2}$.

Since

$$A_\xi f = \sum_{k=0}^{19} \sum_{P \in \mathbb{P}_k} \langle f, \phi_P\rangle \phi_P,$$

we find that $\|A_\xi f\|_{L^2} \le 20 \cdot \frac{1}{20}\|f\|_{L^2} = \|f\|_{L^2}$. □

The intermediate results used in the proof of Theorem 5.1 yield us further information about the operator A_ξ. Suppose that we wanted to consider a truncated sum

$$A_\xi^{\le\Omega} f = \sum_{P : |\omega_P| \le \Omega} \langle f, \phi_P\rangle \phi_P \cdot 1_{P_{uu}}(\xi),$$

where Ω is the maximal dyadic length that we allow for our frequency intervals. From Lemma 5.2 we see that

$$\operatorname{supp} \widehat{A_\xi^{\le\Omega} f} \subseteq [\xi - 0.8\Omega, 0] \subseteq [\xi - 0.8\Omega, \xi + 0.8\Omega].$$

The complementary truncated sum

$$A_\xi^{>\Omega} f = \sum_{P : |\omega_P| > \Omega} \langle f, \phi_P\rangle \phi_P \cdot 1_{P_{uu}}(\xi)$$

in fact only contains tiles with $|\omega_P| \geq 2\Omega$ (since the dyadic lengths increase by multiples of 2), and hence, again by Lemma 5.2,

$$\operatorname{supp} \widehat{A_\xi^{\leq\Omega} f} \subseteq (-\infty, \xi - 0.45 \cdot 2\Omega] = (-\infty, \xi - 0.9\Omega].$$

We choose another auxiliary function $\chi \in \mathscr{S}(\mathbb{R})$ as follows: Let $\hat{\chi}$ satisfy

$$\hat{\chi}(\eta) = \begin{cases} 1 & \text{if } |\eta| \leq 0.8, \\ 0 & \text{if } |\eta| \geq 0.85. \end{cases}$$

We check immediately that

$$\hat{\chi}\Big(\frac{\eta - \xi}{\Omega}\Big)\widehat{A_\xi^{\leq\Omega} f}(\eta) = \widehat{A_\xi^{\leq\Omega} f}(\eta), \qquad \hat{\chi}\Big(\frac{\eta - \xi}{\Omega}\Big)\widehat{A_\xi^{>\Omega} f}(\eta) = 0,$$

and hence

$$\hat{\chi}\Big(\frac{\eta - \xi}{\Omega}\Big)\widehat{A_\xi f}(\eta) = \widehat{A_\xi^{\leq\Omega} f}(\eta).$$

Observe that

$$\hat{\chi}\Big(\frac{\eta - \xi}{\Omega}\Big) = T_\xi D_\Omega^\infty \hat{\chi}(\eta) = \mathscr{F}(M_\xi D^1_{1/\Omega}\chi)(\eta),$$

and hence

$$\mathscr{F}(A_\xi^{\leq\Omega} f)(\eta) = \mathscr{F}(M_\xi D^1_{1/\Omega}\chi)(\eta)\mathscr{F}(A_\xi f)(\eta) = \mathscr{F}(M_\xi D^1_{1/\Omega}\chi * A_\xi f)(\eta),$$

where $*$ means convolution — recall that $\widehat{g * f} = \hat{g}\hat{f}$. Thus

$$\begin{aligned}(A_\xi^{\leq\Omega} f)(x) = (M_\xi D^1_{1/\Omega}\chi * A_\xi f)(x) &= \int_{\mathbb{R}} M_\xi D^1_{1/\Omega}\chi(x-y)(A_\xi f)(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}} e^{i2\pi\xi(x-y)} D^1_{1/\Omega}\chi(x-y)(A_\xi f)(y)\,\mathrm{d}y \\ &= e^{i2\pi\xi x}\int_{\mathbb{R}} D^1_{1/\Omega}\chi(x-y)e^{-i2\pi\xi y}(A_\xi f)(y)\,\mathrm{d}y \\ &= M_\xi(D^1_{1/\Omega}\chi * M_{-\xi}A_\xi f)(x).\end{aligned}$$

Since χ is a Schwartz function, we have $|D^1_\lambda \chi * g(x)| \lesssim Mg(x)$, where

$$Mg(x) := \sup_I 1_I(x)\frac{1}{|I|}\int_I |f(y)|\,\mathrm{d}y$$

is the Hardy–Littlewood maximal function. Thus

$$\begin{aligned}|(A_\xi^{\leq\Omega} f)(x)| &= |M_\xi(D^1_{1/\Omega}\chi * M_{-\xi}A_\xi f)(x)| \\ &= |(D^1_{1/\Omega}\chi * M_{-\xi}A_\xi f)(x)| \\ &\lesssim M(M_{-\xi}A_\xi f)(x) = M(A_\xi f)(x).\end{aligned}$$

(Don't confuse the modulation — M_ξ with a subscript — with the maximal operator — M without any subscript!)

Recall that

$$A_\xi^\Omega f(x) = \sum_{P:|\omega_P|\leq\Omega} \langle f, \phi_P\rangle \phi_P(x) \cdot 1_{\omega_{P_{uu}}}(\xi)$$

was the one-sided truncated sum of $A_\xi f(x)$. A two-sided truncation

$$\sum_{P:\Omega_1<|\omega_P|\leq\Omega_2} \langle f, \phi_P\rangle \phi_P \cdot 1_{\omega_{P_{uu}}}(\xi) = \Big(\sum_{P:|\omega_P|\leq\Omega_2} - \sum_{P:|\omega_P|\leq\Omega_1}\Big)\langle f, \phi_P\rangle \phi_P \cdot 1_{\omega_{P_{uu}}}(\xi)$$

is a difference of two one-sided truncation, and hence satisfies a similar bound (just multiplied by 2).

Combining these considerations, we have shown:

Proposition 5.3. *The truncated sums*

$$A_\xi^{\Omega_1,\Omega_2} f(x) := \sum_{P:\Omega_1<|\omega_P|\leq\Omega_2} \langle f,\phi_P\rangle\phi_P\cdot 1_{\omega_{P_{uu}}}(\xi)$$

satisfy the uniform pointwise bound

$$|A_\xi^{\Omega_1,\Omega_2} f(x)| \lesssim M(A_\xi f)(x).$$

This will be useful in combination with the dominated convergence theorem, observing that $M(A_\xi f)\in L^2$ for $f\in L^2$ by Theorem 5.1 and the maximal inequality.

6. Recovering Carleson's operator from the model operators

6.1. **Motivation; translated and dilated dyadic intervals.** In analogy to the Walsh formula

$$S_n f(x) = \sum_{k=0}^{n-1}\langle f,w_k\rangle w_k(x) = \sum_{\substack{P \text{ bitile} \\ I_P\subseteq[0,1)}} \langle f, w_{P_d}\rangle w_{P_d}(x) 1_{\omega_{P_u}}(n),$$

we would like to recover the Carleson operator S_ξ from the model operators A_ξ. There are simple reasons that $S_\xi f\neq A_\xi f$. Note that $\operatorname{supp}\widehat{S_\xi f}\subseteq(-\infty,\xi]$ (and there can be equality for suitable f), whereas

$$\operatorname{supp}\widehat{A_\xi f}\subseteq \bigcup_{k\in\mathbb{Z}}\big[\xi-0.8\cdot 2^k|\omega|,\xi-0.45\cdot 2^k|\omega|\big]\subsetneq(-\infty,\xi],$$

where $|\omega|$ is a length of one chosen frequency interval, so that all other frequency intervals have length $2^k|\omega|$ for some $k\in\mathbb{Z}$. The reason for the strict containment $\subsetneq$ is that the union of the intervals above contains certain gaps, for instance, $(\xi-0.9|\omega|,\xi-0.8|\omega|)$.

However, if we vary the value of $|\omega|$ continuously over $(0,\infty)$, then the gaps change position, and every point of $(-\infty,\xi)$ will be covered by a suitable choice of $|\omega|$. With standard dyadic intervals, we always have $|\omega|=2^j$, so we cannot vary it continuously, but the above remark motivates the consideration of *dilated* dyadic intervals

$$r\mathscr{D} := \{rI : I\in\mathscr{D}\},\qquad rI := \{rx : x\in I\}.$$

It suffices to consider $r\in[1,2)$, since $r=2^j$ would just map $\mathscr{D}$ into itself.

Another reason for $S_\xi f\neq A_\xi f$ is that the model operator does not possess the correct invariance with respect to translations. Namely, it is easy to check that $T_y S_\xi = S_\xi T_y$ for all $y\in\mathbb{R}$, but this is not true with A_ξ in place of S_ξ. To overcome this obstacle, we also want to consider different *translations* of $\mathscr{D}$. For definiteness, let us write

$$\mathscr{D} := \{2^{-k}[m,m+1) : k,m\in\mathbb{Z}\}$$

for the standard dyadic system, and then

$$I\dot{+}\beta := I + \sum_{j:2^{-j}<|I|} 2^{-j}\beta_j,\qquad \mathscr{D}\dot{+}\beta := \{I\dot{+}\beta : I\in\mathscr{D}\},$$

where

$$\beta = (\beta_j)_{j\in\mathbb{Z}}\in\{0,1\}^{\mathbb{Z}}$$

is a binary sequence. The parameter space $\{0,1\}^{\mathbb{Z}}$ is equipped with a natural probability measure P so that all coordinates β_j are independent and satisfy $\mathsf{P}(\beta_j=0)=\mathsf{P}(\beta_j=1)=\frac12$. We use the probabilistic notation

$$\mathsf{E}\Phi := \int_{\{0,1\}^{\mathbb{Z}}}\Phi(\beta)\,\mathrm{d}\mathsf{P}(\beta)$$

for the *expectation* of a *random variable* Φ—a function on the probability space $\{0,1\}^{\mathbb{Z}}$. We will use two independent copies of the same probability space $\{0,1\}^{\mathbb{Z}}$ to parameterize the translation of the time intervals and the frequency intervals. We denote the variables by β,β', and the expectations by E,E'.

Combining both translations and dilations, we define general family of tiles as

$$\mathbb{P}_\upsilon := \{P \in \frac{1}{r}(\mathscr{D}\dot{+}\beta) \times r(\mathscr{D}\dot{+}\beta') : |P| = 1\}, \qquad \upsilon := (\beta, \beta', r) \in \{0,1\}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}} \times [1,2)$$

Note that the translations are independent, but we use reciprocal dilations in time and frequency to ensure the existence of rectangles with area 1. For the interval $[1,2)$, we use the measure $\mathrm{d}r/r$.

6.2. **General model operators and averaging.** We define the model operator A_ξ^υ for every $\upsilon = (\beta, \beta', r)$ as

$$A_\xi^\upsilon f := \sum_{P\in\mathbb{P}_\upsilon} 1_{\omega_{P_{uu}}}(\xi)\langle f, \phi_P\rangle\phi_P = \sum_{k\in\mathbb{Z}} \sum_{\substack{P:|\omega_P|=r2^k \\ |I_P|=1/(r2^k)}} 1_{\omega_{P_{uu}}}(\xi)\langle f, \phi_P\rangle\phi_P.$$

It is easy to check that everything we proved about $A_\xi = A_\xi^{(0,0,1)}$ remains true for every A_ξ^υ, with uniform bounds. The recovery of S_ξ from the model operators is expressed by the following:

Theorem 6.1. *For $f \in \mathscr{S}(\mathbb{R})$ and almost every $x \in \mathbb{R}$, we have*

$$\int_1^2 \mathsf{E}\mathsf{E}' A_\xi^\upsilon f(x) \frac{\mathrm{d}r}{r} = c\int_{-\infty}^{\xi} \hat{f}(\eta) e^{i2\pi x\eta}\,\mathrm{d}\eta,$$

where c is a positive constant.

Lemma 6.1. *For almost every $x \in \mathbb{R}$,*

$$\int_1^2 \mathsf{E}\mathsf{E}' A_\xi^\upsilon f(x)\frac{\mathrm{d}r}{r} = \lim_{\substack{m\to-\infty \\ n\to+\infty}} \sum_{k=m}^{n} \int_1^2 \mathsf{E}\mathsf{E}' \sum_{\substack{P:|\omega_P|=r2^k \\ |I_P|=1/(r2^k)}} 1_{\omega_{P_{uu}}}(\xi)\langle f, \phi_P\rangle\phi_P(x)\frac{\mathrm{d}r}{r}.$$

Proof. Clearly we have

$$A_\xi^\upsilon f(x) = \lim_{\substack{m\to-\infty \\ n\to+\infty}} \sum_{k=m}^{n} \sum_{\substack{P:|\omega_P|=r2^k \\ |I_P|=1/(r2^k)}} 1_{\omega_{P_{uu}}}(\xi)\langle f, \phi_P\rangle\phi_P(x),$$

so the claim amounts to exchanging the limit and the integrals $\int_1^2 \mathsf{E}\mathsf{E}'$. Since each truncated sum above is dominated by $M(A_\xi^\upsilon f)(x)$, it suffices by dominated convergence to check that $\int_1^2 \mathsf{E}\mathsf{E}' M(A_\xi^\upsilon f)(x)\,\mathrm{d}r/r < \infty$. We show that this is true for almost every x by showing that in fact this quantity is in $L^2(\mathrm{d}x)$. Indeed,

$$\Big\| \int_1^2 \mathsf{E}\mathsf{E}' M(A_\xi^\upsilon f)(x)\frac{\mathrm{d}r}{r}\Big\|_{L^2(\mathrm{d}x)} \leq \int_1^2 \mathsf{E}\mathsf{E}' \|M(A_\xi^\upsilon f)\|_{L^2}\frac{\mathrm{d}r}{r} \lesssim \int_1^2 \mathsf{E}\mathsf{E}'\|f\|_{L^2}\frac{\mathrm{d}r}{r} \leq \log 2\cdot\|f\|_{L^2},$$

since $\|M(A_\xi^\upsilon f)\|_{L^2} \leq C\|A_\xi f\|_{L^2} \leq C\|f\|_{L^2}$. □

Proof of Theorem 6.1. Thanks to the Lemma, we can concentrate on

$$\begin{aligned}
&\int_1^2 \mathsf{E}\mathsf{E}' \sum_{\substack{P\in\mathbb{P}_\upsilon:|\omega_P|=r2^k \\ |I_P|=1/(r2^k)}} 1_{\omega_{P_{uu}}}(\xi)\langle f, \phi_P\rangle\phi_P(x)\frac{\mathrm{d}r}{r} \\
&\qquad = \int_1^2 \mathsf{E}' \sum_{\substack{\omega\in r(\mathscr{D}\dot{+}\beta') \\ |\omega|=r2^k}} 1_{\omega_{uu}}(\xi)\; \mathsf{E} \sum_{\substack{I\in\frac{1}{r}(\mathscr{D}\dot{+}\beta) \\ |I|=1/(r2^k)}} \langle f, \phi_{I\times\omega}\rangle\phi_{I\times\omega}(x)\frac{\mathrm{d}r}{r}
\end{aligned}$$

for each fixed $k \in \mathbb{Z}$. Let us denote $t := r2^k$.

The time interval sums $\mathsf{E}\sum_I$. Write

$$\langle f,\phi_{I\times\omega}\rangle\phi_{I\times\omega}(x)=\int_{\mathbb{R}} f(y)e^{i2\pi c(\omega_d)(x-y)}\frac{1}{|I|}\bar{\phi}\Big(\frac{y-c(I)}{|I|}\Big)\phi\Big(\frac{x-c(I)}{|I|}\Big)\,\mathrm{d}y.$$

With $|I|=1/t$ fixed, let u/t, $u\in[0,1)$, denote the first end-point of an interval immediately to the right from the origin. Then all the centre points have the form $(u+1/2+k)/t$, $k\in\mathbb{Z}$. With this parameterization, we have

$$\sum_{\substack{I\in(\mathscr{D}\dot{+}\beta)/r\\|I|=1/t}}=\sum_{k\in\mathbb{Z}}.$$

On the other hand, the expectation E also takes a simpler form. Let us consider the particular interval $\frac{1}{r}([0,2^{-k})\dot{+}\beta)$ with left end-point immediately right from the origin. By definition,

$$\frac{1}{r}([0,2^{-k})\dot{+}\beta)=\frac{1}{r}\Big([0,2^{-k})+\sum_{j:2^{-j}<2^{-k}}\beta_j2^{-j}\Big),$$

so the left end-point, which we have also denoted by u/t, takes the form

$$\frac{1}{r}\sum_{j:2^{-j}<2^{-k}}\beta_j2^{-j}=\frac{1}{t}\sum_{j=k+1}^{\infty}\beta_j2^{k-j}.$$

When the binary digits β_j are chosen randomly from $\{0,1\}$, the value of the binary series above is uniformly distributed on the interval $[0,1]$. Hence E can be replaced by $\int_0^1\,\mathrm{d}u$.

Substituting the above observations, we get

$$\begin{aligned}
&\mathsf{E}\sum_{\substack{I\in\frac{1}{r}(\mathscr{D}\dot{+}\beta)\\|I|=1/(r2^k)}}\frac{1}{|I|}\bar{\phi}\Big(\frac{y-c(I)}{|I|}\Big)\phi\Big(\frac{x-c(I)}{|I|}\Big)\\
&=\int_0^1\sum_{k\in\mathbb{Z}}t\bar{\phi}(ty-k-u-\tfrac12)\phi(tx-k-u-\tfrac12)\,\mathrm{d}u=\int_{-\infty}^{\infty}t\bar{\phi}(ty-v)\phi(tx-v)\,\mathrm{d}v\\
&=\int_{-\infty}^{\infty}t\tilde{\bar{\phi}}(v)\phi(t(x-y)-v)\,\mathrm{d}v,\qquad\tilde{\phi}(v):=\phi(-v),\\
&=D^1_{1/t}(\phi*\tilde{\bar{\phi}})(x-y).
\end{aligned}$$

The frequency interval sum $\mathsf{E}'\sum_\omega$. Arguing in a similar way, denoting by ut the first end-point of the intervals $\omega\in r(\mathscr{D}\dot{+}\beta')$, $|\omega|=t$, to the right of the origin, we find that E' amounts to $\int_0^1\,\mathrm{d}u$. Then all points $c(\omega_d)$ have the form $t(u+1/4+k)$, $k\in\mathbb{Z}$, the intervals ω_{uu} are given by $c(\omega_d)+[1/2,3/4]|\omega|=[t(u+3/4+k),t(u+1+k)]$, and

$$\sum_{\substack{\omega\in r(\mathscr{D}\dot{+}\beta')\\|\omega|=t}}=\sum_{k\in\mathbb{Z}}.$$

Thus we can compute

$$\begin{aligned}
&\mathsf{E}'\sum_{\substack{\omega\in r(\mathscr{D}\dot{+}\beta')\\|\omega|=t}}1_{\omega_{uu}}(\xi)e^{i2\pi c(\omega_d)(x-y)}\\
&=\int_0^1\sum_{k\in\mathbb{Z}}1_{[t(u+3/4+k),t(u+1+k)]}(\xi)e^{i2\pi t(u+1/4+k)(x-y)}\,\mathrm{d}u\\
&=\int_{-\infty}^{\infty}1_{[t(v+1/2),t(v+3/4)]}(\xi)e^{i2\pi tv(x-y)}\,\mathrm{d}v\\
&=\int_{\xi/t-3/4}^{\xi/t-1/2}e^{i2\pi v(x-y)}\,\mathrm{d}v=e^{i2\pi\xi(x-y)}\int_{1/2}^{3/4}e^{-i2\pi vt(x-y)}\,\mathrm{d}v.
\end{aligned}$$

A combination of the time and frequency computions shows that

$$\begin{aligned}
&\mathsf{E}' \sum_{\substack{\omega\in r(\mathscr{D}\dot{+}\beta') \\ |\omega|=r2^k}} 1_{\omega_{uu}}(\xi)\ \mathsf{E} \sum_{\substack{I\in\frac{1}{r}(\mathscr{D}\dot{+}\beta) \\ |I|=1/(r2^k)}} \int_{\mathbb{R}} f(y)\bar{\phi}_{I\times\omega}(y)\,\mathrm{d}y\phi_{I\times\omega}(x) \\
&= \int_{1/2}^{3/4}\int_{\mathbb{R}} f(y)D^1_{1/t}(\phi*\tilde{\bar{\phi}})(x-y)e^{i2\pi\xi(x-y)}e^{-i2\pi vt(x-y)}\,\mathrm{d}y\,\mathrm{d}v, \qquad t=r2^k, \\
&= e^{i2\pi\xi x}\int_{1/2}^{3/4}\int_{\mathbb{R}}(M_{-\xi}f)(y)D^1_{1/t}M_{-v}(\phi*\tilde{\bar{\phi}})(x-y)\,\mathrm{d}y\,\mathrm{d}v,
\end{aligned}$$

where the last step just amounted to writing the complex exponentials in terms of the modulation operators.

Fourier transform. Next, we want to have the Fourier transform $\hat{f}$ appear. To this end, we apply the formula $\langle g,h\rangle = \langle\hat{g},\hat{h}\rangle$ to the y-integral above. Recall that $\langle g,h\rangle = \int_{\mathbb{R}} g(x)\overline{h(x)}\,\mathrm{d}x$ is the sesqui-linear inner product, so we need to rewrite $D^1_{1/t}M_{-v}(\phi*\tilde{\bar{\phi}})(x-y)$ as a complex conjugate. Using simple identities for the reflection $\tilde{g}(x)=g(-x)$ and the pointwise complex conjugation $\bar{g}(x)=\overline{g(x)}$, we have

$$\begin{aligned}
D^1_{1/t}M_{-v}(\phi*\tilde{\bar{\phi}})(x-y) &= [D^1_{1/t}M_{-v}(\phi*\tilde{\bar{\phi}})]\tilde{\ }(y-x) \\
&= D^1_{1/t}M_v(\tilde{\phi}*\bar{\phi})(y-x) = T_xD^1_{1/t}M_v(\tilde{\phi}*\bar{\phi})(y) \\
&= \overline{T_xD^1_{1/t}M_{-v}(\tilde{\bar{\phi}}*\phi)(y)}.
\end{aligned}$$

Thus

$$\begin{aligned}
&\int_{\mathbb{R}}(M_{-\xi}f)(y)D^1_{1/t}M_{-v}(\phi*\tilde{\bar{\phi}})(x-y)\,\mathrm{d}y = \int_{\mathbb{R}}(M_{-\xi}f)(y)\overline{T_xD^1_{1/t}M_{-v}(\tilde{\bar{\phi}}*\phi)(y)}\,\mathrm{d}y \\
&= \int_{\mathbb{R}}\widehat{M_{-\xi}f}(\eta)\overline{\mathscr{F}[T_xD^1_{1/t}M_{-v}(\tilde{\bar{\phi}}*\phi)](\eta)}\,\mathrm{d}\eta \\
&= \int_{\mathbb{R}}\hat{f}(\eta+\xi)\overline{[M_{-x}D^\infty_tT_v\mathscr{F}(\tilde{\bar{\phi}})\mathscr{F}(\phi)](\eta)}\,\mathrm{d}\eta \\
&= \int_{\mathbb{R}}\hat{f}(\eta+\xi)M_xD^\infty_tT_v|\hat{\phi}|^2(\eta)\,\mathrm{d}\eta = \int_{\mathbb{R}}\hat{f}(\eta+\xi)e^{i2\pi\eta x}|\hat{\phi}|^2\Big(\frac{\eta}{t}+v\Big)\,\mathrm{d}\eta,
\end{aligned}$$

where we used the identity $\mathscr{F}(\tilde{\bar{\phi}})=\bar{\hat{\phi}}$, and hence $\mathscr{F}(\tilde{\bar{\phi}})\mathscr{F}(\phi)=\bar{\hat{\phi}}\hat{\phi}=|\hat{\phi}|^2$.

Synthesis. Collecting the results of different steps, we have seen that (recall the abbreviation $t=2^kr$)

$$\begin{aligned}
&\int_1^2 \mathsf{E}\mathsf{E}'A^v_\xi f(x)\frac{\mathrm{d}r}{r} \\
&= \lim_{\substack{m\to-\infty \\ n\to+\infty}}\sum_{k=m}^n\int_1^2 e^{i2\pi\xi x}\int_{1/2}^{3/4}\int_{\mathbb{R}}\hat{f}(\eta+\xi)e^{i2\pi\eta x}|\hat{\phi}|^2\Big(\frac{\eta}{2^kr}+v\Big)\,\mathrm{d}\eta\,\mathrm{d}v\frac{\mathrm{d}r}{r}
\end{aligned}$$

Now, in the r-integral, we substitute $t=2^kr$ to continue with

$$\begin{aligned}
&= \lim_{\substack{m\to-\infty \\ n\to+\infty}}\sum_{k=m}^n\int_{2^k}^{2^{k+1}} e^{i2\pi\xi x}\int_{1/2}^{3/4}\int_{\mathbb{R}}\hat{f}(\eta+\xi)e^{i2\pi\eta x}|\hat{\phi}|^2\Big(\frac{\eta}{t}+v\Big)\,\mathrm{d}\eta\,\mathrm{d}v\frac{\mathrm{d}t}{t} \\
&= \lim_{\substack{m\to-\infty \\ n\to+\infty}}\int_{2^m}^{2^{n+1}} e^{i2\pi\xi x}\int_{1/2}^{3/4}\int_{\mathbb{R}}\hat{f}(\eta+\xi)e^{i2\pi\eta x}|\hat{\phi}|^2\Big(\frac{\eta}{t}+v\Big)\,\mathrm{d}\eta\,\mathrm{d}v\frac{\mathrm{d}t}{t}.
\end{aligned}$$

We claim that the t-integral actually exists as a proper Lebesgue integral over $(0,\infty)$, and that we can use Fubini's theorem to change the order of integration. This is more easily justified a little later, after proceeding formally first:

$$\begin{aligned}&= \int_0^\infty e^{i2\pi\xi x} \int_{1/2}^{3/4} \int_{\mathbb{R}} \hat{f}(\eta+\xi) e^{i2\pi\eta x} |\hat{\phi}|^2\Big(\frac{\eta}{t}+v\Big)\,\mathrm{d}\eta\,\mathrm{d}v \frac{\mathrm{d}t}{t} \\ &= e^{i2\pi\xi x} \int_{1/2}^{3/4} \int_{\mathbb{R}} \hat{f}(\eta+\xi) e^{i2\pi\eta x} \int_0^\infty |\hat{\phi}|^2\Big(\frac{\eta}{t}+v\Big) \frac{\mathrm{d}t}{t}\,\mathrm{d}\eta\,\mathrm{d}v.\end{aligned} \tag{6.1}$$

The integral over $\mathrm{d}t/t$. This looks potentially dangerous due to the singularity as $t \to 0$. However, if $\eta \geq 0$, then $\eta/t + v \geq v \geq 1/2$, whereas $\operatorname{supp}\hat{\phi} \subseteq [-1/20, 1/20]$, so

$$\int_0^\infty |\hat{\phi}|^2\Big(\frac{\eta}{t}+v\Big)\frac{\mathrm{d}t}{t} = 0 \qquad \text{if } \eta \geq 0.$$

Let then $\eta < 0$. We make the change of variable $s = \eta/t + v$. So that $s \to \operatorname{sgn}(\eta)\infty = -\infty$ as $t \to 0$ and $s \to v$ as $t \to \infty$. Moreover, $t = \eta/(s-v)$ and $\mathrm{d}t = -\eta\,\mathrm{d}s/(s-v)^2$. Hence

$$\int_0^\infty |\hat{\phi}|^2\Big(\frac{\eta}{t}+v\Big)\frac{\mathrm{d}t}{t} = \int_{-\infty}^v |\hat{\phi}|^2(s)\frac{\mathrm{d}s}{v-s} = \int_{-1/20}^{1/20} |\hat{\phi}|^2(s)\frac{\mathrm{d}s}{v-s} \qquad \text{if } \eta < 0.$$

The last identity follows again from $\operatorname{supp}\hat{\phi} \subseteq [-1/20, 1/20] \subseteq (-\infty, v]$ for any $v \geq 1/2$.

Conclusion. These computations show that the t-integral in (6.1) exists with a uniform upper bound independent of η and v. For $\hat{f} \in \mathscr{S}(\mathbb{R})$, the η-integral also converges absolutely, and so does the v-integral over the finite interval $[1/2, 3/4]$. In retrospect, this legitimates the application of Fubini's theorem in (6.1), and we find that

$$\begin{aligned}(6.1) &= e^{i2\pi\xi x} \int_{1/2}^{3/4} \int_{\mathbb{R}} \hat{f}(\eta+\xi) e^{i2\pi\eta x} 1_{(-\infty,0)}(\eta) \Big(\int_{-1/20}^{1/20} |\hat{\phi}|^2(s)\frac{\mathrm{d}s}{v-s}\Big)\,\mathrm{d}\eta\,\mathrm{d}v \\ &= c\int_{-\infty}^0 \hat{f}(\eta+\xi) e^{i2\pi(\eta+\xi)x}\,\mathrm{d}\eta, \qquad c := \int_{1/2}^{3/4} \Big(\int_{-1/20}^{1/20} |\hat{\phi}|^2(s)\frac{\mathrm{d}s}{v-s}\Big)\,\mathrm{d}v, \\ &= c\int_{-\infty}^{\xi} \hat{f}(\eta) e^{i2\pi(\eta)x}\,\mathrm{d}\eta = cS_\xi f(x).\end{aligned}$$

This completes the proof of Theorem 6.1, and even provides an explicit expression for the constant $c > 0$. □

6.3. **The road to Carleson's theorem.** We want to show that $(S_\xi - S_{-\xi})f(x) \to f(x)$, which by the familiar procedure is reduced to proving that

$$\|Cf\|_{L^{2,\infty}} \lesssim \|f\|_{L^2}, \qquad Cf(x) := \sup_{\xi\in\mathbb{R}} |S_\xi f(x)|.$$

By linearizing the Carleson maximal operator C (replace ξ by an arbitrary measurable function $N : \mathbb{R} \to \mathbb{R}$), and using the characterization of the $L^{2,\infty}$ as in the Walsh case, we have the further reductions to showing that

$$|\langle S_{N(\cdot)} f(\cdot), 1_E\rangle| \lesssim \|f\|_{L^2} |E|^{1/2}.$$

By substituting $\xi = N(x)$ in the statement of Theorem 6.1, we have

$$S_{N(x)} f(x) = c\int_1^2 \mathsf{E}\mathsf{E}' A^v_{N(x)} f(x) \frac{\mathrm{d}r}{r},$$

and hence

$$\begin{aligned}\langle S_{N(\cdot)} f(\cdot), 1_E\rangle = \int_E S_{N(x)} f(x)\,\mathrm{d}x &= c\int_1^2 \mathsf{E}\mathsf{E}' \int_E A^v_{N(x)} f(x)\,\mathrm{d}x \frac{\mathrm{d}r}{r} \\ &= c\int_1^2 \mathsf{E}\mathsf{E}' \langle A^v_{N(\cdot)} f(\cdot), 1_E\rangle \frac{\mathrm{d}r}{r}.\end{aligned} \tag{6.2}$$

Suppose we can prove that

$$|\langle A^v_{N(\cdot)} f(\cdot), 1_E\rangle| \lesssim \|f\|_{L^2} |E|^{1/2} \tag{6.3}$$

uniformly in the parameter υ which controls the choice of the dyadic systems. By (6.2), this would immediately give the corresponding bound for $S_{N(\cdot)}$ in place of $A^{\upsilon}_{N(\cdot)}$.

So our task is to prove (6.3). In doing so, we will drop the reference to the parameter υ, since the particular dyadic system that is used is unimportant for the argument, and it will be obvious that all estimates are independent of this choice. Written out in full, we hence want to estimate

$$\langle A_{N(\cdot)}f(\cdot), f\rangle = \sum_{P\in\mathbb{P}}\langle f,\phi_P\rangle\langle\phi_P 1_{\omega_{P_{uu}}}(N(\cdot)), 1_E\rangle = \sum_{P\in\mathbb{P}}\langle f,\phi_P\rangle\langle\phi_P, 1_{E_{P_{uu}}}\rangle$$

where

$$E_{P_{uu}} := E\cap\{x : N(x)\in\omega_{P_{uu}}\}.$$

So the goal is now formally very similar to the Walsh case, the only difference being the replacement of the Walsh wave packets w_P by the Fourier wave packets ϕ_P.

7. Density, energy and trees for the Fourier model

Our proof of Carleson's theorem for the Fourier model will be based on the same key concepts as in the Walsh case: density, energy and trees.

7.1. **Density.**

Recall that the Walsh-density of a collection $\mathbb{P}$ was defined by

$$\text{Walsh-density}(\mathbb{P}) := \sup_{P\in\mathbb{P}}\sup_{P'\geq P}\frac{|E_{P'}\cap I_{P'}|}{|I_{P'}|}.$$

Here the ratio of the measures can also be written as

$$\frac{|E_{P'}\cap I_{P'}|}{|I_{P'}|} = \int_{E_{P'}}\frac{1_{I_{P'}}}{|I_{P'}|} = \int_{E_{P'}} T_{c(I_{P'})}D^1_{|I_{P'}|}1_{[-1/2,1/2)},$$

where the integrand $1_{I_{P'}}/|I_{P'}|$ is the L^1-normalized version of $1_{I_{P'}}/|I_{P'}|^{1/2} = |w_{P'}|$ for the Walsh wave packet $w_{P'}$.

The present definition of density will be accordingly adapted to the Fourier wave packets ϕ_P. Since $\phi\in\mathscr{S}(\mathbb{R})$, we have in particular that

$$|\phi(x)| \lesssim (1+|x|)^{-10} =: v(x),$$

and then

$$|\phi_P| = |M_{c(\omega_{P_d})}T_{c(I_P)}D^2_{|I_P|}\phi| = T_{c(I_P)}D^2_{|I_P|}|\phi| \lesssim T_{c(I_P)}D^2_{|I_P|}v.$$

Changing to L^1-normalization, we define

$$v_I(x) := T_{c(I)}D^1_{|I|}v(x) = \frac{1}{|I|}\Big(1+\frac{|x-c(I)|}{|I|}\Big)^{-10},$$

so that $|\phi_P| \lesssim |I_P|^{1/2}v_{I_P}$. Then we define:

$$\text{density}(\mathbb{P}) := \sup_{P\in\mathbb{P}}\sup_{P'\geq P}\int_{E_{P'}} v_{I_{P'}}.$$

Thanks to the next lemma, an immediate reformulation is

$$\text{density}(\mathbb{P}) \eqsim \sup_{P\in\mathbb{P}}\sup_{P'\geq P}\sum_{k=0}^{\infty}2^{-10k}\frac{|E_{P'}\cap 2^k I_{P'}|}{|I_{P'}|},$$

which clearly illustrates the difference compared to the Walsh case: The Walsh-density involves just the zeroth term of the sum, whereas we now need to also handle the rapidly decaying tail.

Lemma 7.1.

$$v_I(x) \eqsim \frac{1}{|I|}\sum_{k=0}^{\infty}2^{-10k}1_{2^k I}(x).$$

Proof. Clearly

$$v_I(x) = 1_I(x)v_I(x) + \sum_{j=1}^{\infty} 1_{2^jI\setminus 2^{j-1}I} v_I(x) \approx \frac{1}{|I|}1_I(x) + \sum_{j=1}^{\infty} \frac{2^{-10j}}{|I|} 1_{2^jI\setminus 2^{j-1}I},$$

since $v_I(x) \approx |I|^{-1}2^{-10j}$ for $x \in 2^jI \setminus 2^{j-1}I$. On the other hand,

$$\begin{aligned}\sum_{k=0}^{\infty} 2^{-10k} 1_{2^kI}(x) &= \sum_{k=0}^{\infty} 2^{-10k}\Big(1_I(x) + \sum_{j=1}^{k} 1_{2^jI\setminus 2^{j-1}I}(x)\Big) \\ &= 1_I(x)\sum_{k=0}^{\infty} 2^{-10k} + \sum_{j=1}^{\infty} 1_{2^jI\setminus 2^{j-1}I}(x) \sum_{k=j}^{\infty} 2^{-10k} \approx 1_I(x) + \sum_{j=1}^{\infty} 1_{2^jI\setminus 2^{j-1}I}(x) 2^{-10j},\end{aligned}$$

since $\sum_{k=j}^{\infty} 2^{-10k} \approx 2^{-10j}$. □

The statement of the density lemma is exactly as in the Walsh case, but the proof requires elaboration due to the mentioned tail.

Proposition 7.1 (Density lemma). *Let $\mathbb{P}$ be a finite collection of tiles. Then there is a disjoint decomposition*

$$\mathbb{P} = \mathbb{P}_{\text{sparse}} \cup \bigcup_j \mathbb{T}_j,$$

where

$$\text{density}(\mathbb{P}_{\text{sparse}}) \le \tfrac{1}{4}\,\text{density}(\mathbb{P}),$$

and each $\mathbb{T}_j$ is a tree with top T_j, such that

$$\sum_j |I_{T_j}| \lesssim \text{density}(\mathbb{P})^{-1}|E|.$$

Proof. Denote $\delta := \text{density}(\mathbb{P})$. As in the Walsh case, the collection

$$\mathbb{P}_{\text{sparse}} := \Big\{P \in \mathbb{P} : \sup_{P'\ge P} \int_{E_{P'}} v_{I_{P'}} \le \frac{\delta}{4}\Big\}$$

satisfies the required bound. For each of the remaining tiles $P \in \mathbb{P}_1 := \mathbb{P} \setminus \mathbb{P}_{\text{sparse}}$, there exists by definition some $P' \ge P$ such that

$$\frac{\delta}{4} < \int_{E_{P'}} v_{I_{P'}} \le C\sum_{k=0}^{\infty} 2^{-10k}\frac{|E_{P'} \cap 2^kI_{P'}|}{|I_{P'}|} \le C\Big(\sup_{k\ge 0} 2^{-9k}\frac{|E_{P'}\cap 2^kI_{P'}|}{|I_{P'}|}\Big)\sum_{k=0}^{\infty} 2^{-k}. \tag{7.1}$$

Among all these P', let T_j be the maximal ones with respect to $\le$. Then every $P \in \mathbb{P}_1$ satisfies $P \le P' \le T_j$ for some P' and T_j, and hence all $P \in \mathbb{P}_1$ can be arranged into trees $\mathbb{T}_j$ with tops T_j. It remains to estimate $\sum |I_{T_j}|$. To this end, we need to define additional decompositions.

Using (7.1) with $P' = T_j$, we find that for each T_j, there is *some* $k \in \mathbb{N}$, such that

$$\frac{\delta}{4} < 2C \cdot 2^{-9k}\frac{|E_{T_j} \cap 2^kI_{T_j}|}{|I_{T_j}|}. \tag{7.2}$$

We define

$$J_k := \{j : (7.2) \text{ holds}\},$$

and conclude that every j belongs to at least one J_k. Hence

$$\sum_j |I_{T_j}| \le \sum_{k=0}^{\infty}\sum_{j\in J_k} |I_{T_j}|. \tag{7.3}$$

In the Walsh case, for $k = 0$, we simply estimated

$$|I_{T_j}| \lesssim \frac{1}{\delta}|E_{T_j} \cap I_{T_j}| = \frac{1}{\delta}|E \cap \{x : N(x) \in \omega_{T_j}\} \cap I_{T_j}| = \frac{1}{\delta}|E \cap \{x : (x, N(x)) \in I_{T_j} \times \omega_{T_j} = T_j\}|,$$

and observed that these sets, by the maximality of the T_j, are pairwise disjoint and contained in E. Now we need additional considerations to ensure such disjointness. Let us write $I_j := I_{T_j}$, $\omega_j := \omega_{T_j}$ for short.

A covering argument. Consider a fix $k \in \mathbb{N}$. We choose a subcollection $\tilde{J}_k \subseteq J_k$ such that all $2^k I_\ell \times \omega_\ell$ for $\ell \in \tilde{J}_k$ are pairwise disjoint rectangles in the phase plance. We use the following recursion:

Choose some $\ell \in J_k$ for which $|I_\ell|$ is maximal, and remove all $j \in J_k$ for which $2^k I_j \times \omega_j$ intersects $2^k I_\ell \times \omega_\ell$. Repeat until all indices have been either chosen or removed.

By construction, for every $j \in J_k$, there exists an $\ell \in \tilde{J}_k$ such that

$$|I_\ell| \geq |I_j| \text{ and } (2^k I_j \times \omega_j) \cap (2^k I_\ell \times \omega_\ell) \neq \varnothing. \tag{7.4}$$

Indeed, if $j \in \tilde{J}_k$, then we can take $\ell = j$. If $j \notin \tilde{J}_k$ is not among the chosen indices, then it must have been removed exactly because $2^k I_j \times \omega_j$ intersected some already chosen $2^k I_\ell \times \omega_\ell$. Moreover, since ℓ was chosen instead of j, it must be that $|I_\ell| \geq |I_j|$.

For every $\ell \in \tilde{J}_k$, we define

$$J_{k\ell} := \{j \in J_k : (7.4) \text{ holds}\},$$

so that

$$J_k = \bigcup_{\ell \in \tilde{J}_k} J_{k\ell}.$$

Consequences of the covering. Let $k \in \mathbb{N}$ and $\ell \in \tilde{J}_k$ be fixed. We claim that all I_j with $j \in J_{k\ell}$ are pairwise disjoint, and contained in $3 \cdot 2^k I_\ell$.

The containment is easy: we have $I_j \subseteq 2^k I_j$, and this intersects with $2^k I_\ell$ by (7.4). Also, $2^k I_j$ is a shorter interval than $2^k I_\ell$, which implies that $2^k I_j \subseteq 3 \cdot 2^k I_\ell$.

To prove the disjointness, let both $j, j' \in J_{k\ell}$. Thus (7.4) holds for both j and j' in place of j. In particular, looking at the frequency intervals, we have

$$\omega_j \cap \omega_\ell \neq \varnothing \neq \omega_{j'} \cap \omega_\ell.$$

Since $|I_j| \leq |I_\ell|$, we have $|\omega_j| \geq |\omega_\ell|$, and similarly for j'. As the intervals are dyadic, the previous equation line takes the more precise form

$$\omega_j \supseteq \omega_\ell, \qquad \omega_{j'} \supseteq \omega_\ell.$$

But then $\omega_j \cap \omega_{j'} \supseteq \omega_\ell \neq \varnothing$. On the other hand, since the tiles T_j and $T_{j'}$ are maximal and hence disjoint, we have

$$\varnothing = T_j \cap T_{j'} = (I_j \cap I_{j'}) \times (\omega_j \cap \omega_{j'}).$$

The second factor on the right is non-empty, so the first factor must be empty, which is the claimed disjointness.

Synthesis. Now we have performed a sufficient analysis, and it is time to collect the pieces together. Continuing from (7.3), we complete the proof of the Density Lemma as follows:

$$\begin{aligned}
\sum_j |I_{T_j}| &\leq \sum_{k=0}^{\infty} \sum_{j \in J_k} |I_j| \leq \sum_{k=0}^{\infty} \sum_{\ell \in \tilde{J}_k} \sum_{j \in J_{k\ell}} |I_j| \\
&\leq \sum_{k=0}^{\infty} \sum_{\ell \in \tilde{J}_k} |3 \cdot 2^k I_\ell| \qquad \text{(since the } I_j \subseteq 3 \cdot 2^k I_\ell \text{ are pairwise disjoint)} \\
&\lesssim \sum_{k=0}^{\infty} 2^k \sum_{\ell \in \tilde{J}_k} |I_\ell| \lesssim \sum_{k=0}^{\infty} 2^k \sum_{\ell \in \tilde{J}_k} \frac{1}{\delta} 2^{-9k} |E_{T_\ell} \cap 2^k I_{T_\ell}| \qquad \text{(definition of } J_k \supseteq \tilde{J}_k) \\
&\lesssim \frac{1}{\delta} \sum_{k=0}^{\infty} 2^{-8k} \sum_{\ell \in \tilde{J}_k} |E \cap \{x : (x, N(x)) \in 2^k I_{T_\ell} \times \omega_{T_\ell}\}| \qquad \text{(definition of } E_{T_\ell}) \\
&\lesssim \frac{1}{\delta} \sum_{k=0}^{\infty} 2^{-8k} |E| \qquad \text{(since the } 2^k I_{T_\ell} \times \omega_{T_\ell} \text{, for } \ell \in \tilde{J}_k \text{, are pairwise disjoint)} \\
&\lesssim \frac{1}{\delta} |E|.
\end{aligned}$$

□

7.2. **Energy.** The definition of energy looks exactly the same as in the Walsh case, just replacing the Walsh wave packets w_P by the Fourier wave packets ϕ_P:

$$\text{energy}(\mathbb{P}) := \sup_{\substack{\mathbb{T}\subseteq\mathbb{P}\\ \text{up-tree}}} \Big(\frac{1}{|I_T|}\sum_{P\in\mathbb{T}} |\langle f,\phi_P\rangle|^2\Big)^{1/2}$$

The difficulties arise from the fact that the orthogonality and localization properties of the ϕ_P's are not as good as those of the w_P's, and we again have some tails to estimate.

Proposition 7.2 (Energy lemma). *Let $\mathbb{P}$ be a finite collection of tiles. Then there is a disjoint decomposition*

$$\mathbb{P} = \mathbb{P}_{\text{low}} \cup \bigcup_j \mathbb{T}_j,$$

where

$$\text{energy}(\mathbb{P}_{\text{low}}) \le \tfrac{1}{2}\,\text{energy}(\mathbb{P}),$$

and each $\mathbb{T}_j$ is a tree with top T_j, such that

$$\sum_j |I_{T_j}| \lesssim \text{energy}(\mathbb{P})^{-2}\|f\|_{L^2}^2. \tag{7.5}$$

Proof assuming an intermediate estimate. The general strategy is the same as in the Walsh case. For a tree $\mathbb{T}$, let

$$\Delta(\mathbb{T}) := \Big(\frac{1}{|I_T|}\int_{P\in\mathbb{T}_u} |\langle f,\phi_P\rangle|^2\Big)^{1/2},$$

where $\mathbb{T}_u := \{P\in\mathbb{T} : P_u \le T_u\}$ is the maximal up-tree supported by the same top T as $\mathbb{T}$.

The trees $\mathbb{T}_j$ are chosen recursively as follows: Let $\mathscr{E} := \text{energy}(\mathbb{P})$. Consider all maximal trees $\mathbb{T}\subseteq\mathbb{P}$ among those with the property that $\Delta(T) > \mathscr{E}/2$. Among them, let $\mathbb{T}_1$ be one with the minimal $c(\omega_T)$. Remove $\mathbb{T}$ from $\mathbb{P}$ and iterate. When no trees $\mathbb{T}_j$ can be extracted anymore, the remaining collection qualifies for $\mathbb{P}_{\text{low}}$, and it remains to estimate $\sum_j |I_{T_j}|$.

As in the Walsh case, we get

$$\sum_j |I_{T_j}| \le \frac{4}{\mathscr{E}^2}\sum_j\sum_{P\in\mathbb{T}_u} |\langle f,\phi_P\rangle|^2.$$

However, unlike there, we cannot readily identify the double sum as the norm of an orthogonal projection of f, and actually such an estimate is not in general valid. Instead, there is the following weaker substitute, were the sum on the left, the one to be estimated, reappears:

$$\sum_j\sum_{P\in\mathbb{T}_u} |\langle f,\phi_P\rangle|^2 \lesssim \|f\|_{L^2}^2 + \Big(\mathscr{E}^2\sum_j |I_{T_j}|\Big)^{1/3}\|f\|_{L^2}^{4/3}. \tag{7.6}$$

Assuming this (nontrivial) estimate for the moment, we can conclude the proof: First, we get

$$\sum_j |I_{T_j}| \lesssim \mathscr{E}^{-2}\|f\|_{L^2}^2 + \mathscr{E}^{-4/3}\Big(\sum_j |I_{T_j}|\Big)^{1/3}\|f\|_{L^2}^{4/3} =: A + B.$$

If $A \ge B$, then $A + B \le 2A$, which gives exactly the bound (7.5) that we wanted. Else, if $A \le B$, then $A + B \le 2B$, and we get

$$\sum_j |I_{T_j}| \lesssim \mathscr{E}^{-4/3}\Big(\sum_j |I_{T_j}|\Big)^{1/3}\|f\|_{L^2}^{4/3}.$$

However, dividing both sides by $\big(\sum_j |I_{T_j}|\big)^{1/3}$ and then raising to power $3/2$ gives (7.5) again. This completes the proof of the Energy Lemma, aside from the key estimate (7.6). □

The rest of this subsection will be concerned with the proof of (7.6). We first collect some geometric lemmas about the structure of the relevant collection of three appearing in this estimate.

Lemma 7.2. *The up-trees $\mathbb{T}_{j,u}$ constructed in the energy lemma satisfy the following disjointness property:*

$$\text{If } P \in \mathbb{T}_{j,u} \text{ and } P' \in \mathbb{T}_{j',u} \text{ satisfy } \omega_P \subseteq \omega_{P'_d}, \text{ then } I_{P'} \cap I_{\mathbb{T}_j} = \varnothing. \tag{7.7}$$

Proof. Let P, P' be two such tiles. Then $c(\omega_{\mathbb{T}_j}) \in \omega_{\mathbb{T}_j} \subseteq \omega_P \subseteq \omega_{P'_d}$, and hence $c(\omega_{\mathbb{T}_j}) < \sup \omega_{P'_d} = c(\omega_{P'})$. On the other hand, since $P' \in \mathbb{T}_{j',u}$, we have $\omega_{\mathbb{T}_{j',u}} \subseteq \omega_{P'_u}$, and thus $c(\omega_{\mathbb{T}_{j'}}) = \inf \omega_{\mathbb{T}_{j',u}} \geq \inf \omega_{P'_u} = c(\omega_{P'})$. Altogether, $c(\omega_{\mathbb{T}_j}) < c(\omega_{P'}) \leq c(\omega_{\mathbb{T}_{j'}})$, and hence $\mathbb{T}_j$ was chosen before $\mathbb{T}_{j'}$ in the construction.

Suppose contrary to the claim that $I_{P'} \cap I_{\mathbb{T}_j} \neq \varnothing$. But we also have $\omega_{P'} \supseteq \omega_P \subseteq \omega_{\mathbb{T}_j}$, hence $|\omega_{P'}| \geq |\omega_{\mathbb{T}_j}|$, thus $|I_{P'}| \leq |I_{\mathbb{T}_j}|$, and therefore $I_{P'} \subseteq I_{\mathbb{T}_j}$. But this means that $P' \leq I_{\mathbb{T}_j} \times \omega_{\mathbb{T}_j}$, and hence P' would have been included in the maximal tree $\mathbb{T}_j$, and removed before the construction of $\mathbb{T}_{j'}$. This contradiction proves the claim that $I_{P'} \cap I_{\mathbb{T}_j} \neq \varnothing$. □

Lemma 7.3. *Property* (7.7) *implies that all down-halves P_d of all*

$$P \in \bigcup_{\mathbb{T} \in \mathscr{T}} \mathbb{T} =: \mathbb{P}$$

are pairwise disjoint.

Proof. Let $P \in \mathbb{T}$, $P' \in \mathbb{T}'$ be two different tiles in $\mathbb{P}$. If $\omega_{P_d} \cap \omega_{P'_d} = \varnothing$, then $P_d \cap P'_d = \varnothing$. So assume that $\omega_{P_d} \cap \omega_{P'_d} \neq \varnothing$, and then for example $\omega_{P_d} \subseteq \omega_{P'_d}$. If $\omega_{P_d} = \omega_{P'_d}$ two tiles, then I_P and $I_{P'}$ are either equal (which cannot be in our case, since this would imply that $P = P'$) or disjoint (in which case $P \cap P' = \varnothing$). We are left with the case that $\omega_{P_d} \subsetneq \omega_{P'_d}$, and then $\omega_P \subsetneq \omega_{P'_d}$. But this implies that $I_{P'} \cap I_{\mathbb{T}} = \varnothing$ by (7.7), hence $I_{P'} \cap I_P = \varnothing$ and thus $P \cap P' = \varnothing$ also in this final case. □

Lemma 7.4. *Suppose that $\mathscr{T}$ satisfies* (7.7)*, and fix a $P \in \mathbb{T} \in \mathscr{T}$. Then among the tiles $P' \in \mathbb{P}$ with $\omega_{P'_d} \supseteq \omega_P$, the time intervals $I_{P'}$ are pairwise disjoint and contained in $I_{\mathbb{T}}^c$.*

Proof. Let P', P'' be two such tiles. So in particular $\omega_{P'_d} \cap \omega_{P''_d} \supseteq \omega_P \neq \varnothing$. Since $P'_d \cap P''_d = \varnothing$, it must be that $I_{P'} \cap I_{P''} = \varnothing$. The fact that $I_{P'} \subseteq I_{\mathbb{T}}^c$ is immediate from (7.7) □

Proposition 7.3. *Let $\mathscr{T}$ be a disjoint collection of trees with the property* (7.7)*. Then*

$$\Big(\sum_{P \in \mathbb{P}} |\langle f, \phi_P \rangle|^2\Big)^{1/2} \lesssim \|f\|_{L^2} + \Big(\sup_{P \in \mathbb{P}} \frac{|\langle f, \phi_P \rangle|}{|I_P|^{1/2}} \Big[\sum_{\mathbb{T} \in \mathscr{T}} |I_{\mathbb{T}}|\Big]^{1/2}\Big)^{1/3} \|f\|_{L^2}^{2/3},$$

where

$$\mathbb{P} := \bigcup_{\mathbb{T} \in \mathscr{T}} \mathbb{T}.$$

Note that

$$\sup_{P \in \mathbb{P}} \frac{|\langle f, \phi_P \rangle|}{|I_P|^{1/2}} \leq \text{energy}(\mathbb{P}),$$

so that Proposition 7.3 proves the estimate (7.6) required to complete the proof of the Energy lemma.

Proof. We start with the estimate

$$\begin{aligned} S^2 := \sum_{P \in \mathbb{P}} |\langle f, \phi_P \rangle|^2 &= \Big\langle \sum_{P \in \mathbb{P}} \langle f, \phi_P \rangle \phi_P, f \Big\rangle \\ &\leq \Big\| \sum_{P \in \mathbb{P}} \langle f, \phi_P \rangle \phi_P \Big\|_{L^2(\mathbb{R};H)} \|f\|_{L^2(\mathbb{R};H)}. \end{aligned}$$

Here

$$\begin{aligned}\Big\| \sum_{P\in\mathbb{P}} \langle f,\phi_P\rangle\phi_P\Big\|^2_{L^2(\mathbb{R};H)} &= \sum_{P,P'\in\mathbb{P}} \langle f,\phi_P\rangle\langle\phi_P,\phi_{P'}\rangle\langle\phi_{P'},f\rangle \\ &= \Big(\sum_{\substack{P,P'\in\mathbb{P}\\ \omega_P=\omega_{P'}}} + 2\sum_{\substack{P,P'\in\mathbb{P}\\ \omega_P\subsetneq\omega_{P'_d}}}\Big)\langle f,\phi_P\rangle\langle\phi_P,\phi_{P'}\rangle\langle\phi_{P'},f\rangle \\ &=: S_1+2S_2,\end{aligned}$$

where the middle line follows from the fact that $\operatorname{supp}\hat{\phi}_P \subseteq \omega_{P_d}$, so that $\langle\phi_P,\phi_{P'}\rangle \neq 0$ only if $\omega_{P_d}\cap\omega_{P_d} \neq \varnothing$, which means that these intervals either coincide, or one is strictly contained in the other.

To proceed further, we need the following estimate, which is left as an exercise:

$$|\langle\phi_P,\phi_{P'}\rangle| \lesssim \Big(\frac{|I_P|}{|I_{P'}|}\Big)^{1/2}\|v_{I_P}1_{I_{P'}}\|_1, \qquad |I_{P'}|\leq|I_P|, \tag{7.8}$$

where $v_I(x) := \dfrac{1}{|I|}\Big(1+\dfrac{|x-c(I)|}{|I|}\Big)^{-10}$.

In S_1, we can use (7.8) with either order of I_P and $I_{P'}$, to get

$$\begin{aligned}S_1 &\leq \sum_{\substack{P,P'\in\mathbb{P}\\ \omega_P=\omega_{P'}}} \frac{1}{2}(|\langle f,\phi_P\rangle|^2+|\langle f,\phi_{P'}\rangle|^2)\min\{\|v_{I_P}1_{I_{P'}}\|_1,\|v_{I_{P'}}1_{I_P}\|_1\} \\ &\leq \sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle|^2 \sum_{\substack{P'=I'\times\omega'\in\mathbb{P}\\ \omega'=\omega_P}} \|v_{I_P}1_{I'}\|_1 \leq \sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle|^2 \sum_{\substack{I'\in\mathscr{D}\\ |I'|=|I_P|}}\|v_{I_P}1_{I'}\|_1 \\ &= \sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle|^2\|v_{I_P}\|_1 \lesssim \sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle|^2 = S^2.\end{aligned}$$

We used the fact that the intervals $I'\in\mathscr{D}$ with $|I'|=|I_P|$ form a partition of $\mathbb{R}$.

We turn our attention to S_2:

$$\begin{aligned}S_2 &\lesssim \sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle| \sum_{\substack{P'\in\mathbb{P}\\ \omega_{P'_d}\supsetneq\omega_P}} \Big(\frac{|I_P|}{|I_{P'}|}\Big)^{1/2}\|v_{I_P}1_{I_{P'}}\|_1|\langle\phi_{P'},f\rangle| \\ &\leq \Big(\sup_{P'\in\mathbb{P}}\frac{|\langle\phi_{P'},f\rangle|}{|I_{P'}|^{1/2}}\Big)\sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle||I_P|^{1/2}\sum_{\substack{P'\in\mathbb{P}\\ \omega_{P'_d}\supsetneq\omega_P}}\|v_{I_P}1_{I_{P'}}\|_1 \\ &\leq \Big(\sup_{P'\in\mathbb{P}}\frac{|\langle\phi_{P'},f\rangle|}{|I_{P'}|^{1/2}}\Big)\sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle||I_P|^{1/2}\|v_{I_P}1_{I^c_{\mathbb{T}(P)}}\|_1.\end{aligned}$$

In the last line, we denoted by $\mathbb{T}(P)$ the the unique tree with $P\in\mathbb{T}(P)\in\mathscr{T}$, and used Lemma 7.4 which guarantees that the intervals $I_{P'}$ appearing in the inner sum on the penultimate line are pairwise disjoint and contained in $I^c_{\mathbb{T}(P)}$.

We use Cauchy–Schwarz to get

$$\begin{aligned}&\sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle||I_P|^{1/2}\|v_{I_P}1_{I^c_{\mathbb{T}(P)}}\|_1 \\ &\qquad\leq \Big(\sum_{P\in\mathbb{P}}|\langle f,\phi_P\rangle|^2\Big)^{1/2}\Big(\sum_{P\in\mathbb{P}}|I_P|\|v_{I_P}1_{I^c_{\mathbb{T}(P)}}\|_1^2\Big)^{1/2} \\ &\qquad\lesssim S\Big(\sum_{P\in\mathbb{P}}|I_P|\|v_{I_P}1_{I^c_{\mathbb{T}(P)}}\|_1\Big)^{1/2},\end{aligned}$$

where we estimated $\|v_{I_P}1_{I^c_{\mathbb{T}(P)}}\|_1 \leq \|v_{I_P}\|_1 \lesssim 1$.

Finally, we write

$$\sum_{P\in\mathbb{P}} |I_P| \|v_{I_P} 1_{I^c_{\mathbb{T}(P)}}\|_1 = \sum_{\mathbb{T}\in\mathscr{T}} \sum_{P\in\mathbb{T}} |I_P| \|v_{I_P} 1_{I^c_{\mathbb{T}}}\|_1$$
$$\leq \sum_{\mathbb{T}\in\mathscr{T}} \sum_{P=I\times\omega\leq I_{\mathbb{T}}\times\omega_{\mathbb{T}}} |I| \|v_I 1_{I^c_{\mathbb{T}}}\|_1 \leq \sum_{\mathbb{T}\in\mathscr{T}} \sum_{\substack{I\in\mathscr{D}\\ I\subseteq I_{\mathbb{T}}}} |I| \|v_I 1_{I^c_{\mathbb{T}}}\|_1,$$

where the last step follows from the fact that the summand depends only on the time interval I, and that the frequency interval ω of $P = I\times\omega$ is uniquely determined by I, since $|\omega| = 1/|I|$ and $\omega \supseteq \omega_{\mathbb{T}}$.

We still need to estimate

$$\sum_{\substack{I\in\mathscr{D}\\ I\subseteq I_{\mathbb{T}}}} |I| \|v_I 1_{I^c_{\mathbb{T}}}\|_1 = \sum_{k=0}^{\infty} \sum_{\substack{I\subseteq I_{\mathbb{T}}\\ |I|=2^{-k}|I_{\mathbb{T}}|}} |I| \int_{I^c_{\mathbb{T}}} \frac{1}{|I|}\Big(1+\frac{|x-c(I)|}{|I|}\Big)^{-10} \mathrm{d}x$$
$$= \sum_{k=0}^{\infty} \int_{I^c_{\mathbb{T}}} \sum_{\substack{I\subseteq I_{\mathbb{T}}\\ |I|=2^{-k}|I_{\mathbb{T}}|}} \Big(1+\frac{|x-c(I)|}{|I|}\Big)^{-10} \mathrm{d}x.$$

For a fixed $x\in I^c_{\mathbb{T}}$, the smallest possible value of $|x-c(I)|$ is at least $\operatorname{dist}(x, I_{\mathbb{T}})$ (since $c(I)\in I_{\mathbb{T}}$), and other possible values increase in integer multiples of $|I| = 2^{-k}|I_{\mathbb{T}}|$. Thus

$$\sum_{\substack{I\subseteq I_{\mathbb{T}}\\ |I|=2^{-k}|I_{\mathbb{T}}|}} \Big(1+\frac{|x-c(I)|}{|I|}\Big)^{-10} \lesssim \sum_{n=0}^{\infty} \Big(1+\frac{|\operatorname{dist}(x,I_{\mathbb{T}})|}{2^{-k}|I_{\mathbb{T}}|}+n\Big)^{-10} \lesssim \Big(1+\frac{|\operatorname{dist}(x,I_{\mathbb{T}})|}{2^{-k}|I_{\mathbb{T}}|}\Big)^{-9},$$

for example estimating the sum by an integral in the last step. Hence we have proved the first step below, and the second is left as an exercise:

$$\sum_{\substack{I\in\mathscr{D}\\ I\subseteq I_{\mathbb{T}}}} |I| \|v_I 1_{I^c_{\mathbb{T}}}\|_1 \lesssim \sum_{k=0}^{\infty} \int_{I^c_{\mathbb{T}}} \Big(1+2^k\frac{|\operatorname{dist}(x,I_{\mathbb{T}})|}{|I_{\mathbb{T}}|}\Big)^{-9} \mathrm{d}x. \lesssim |I_{\mathbb{T}}|,$$

Then, combining all the estimates, we have shown that

$$S^2 \leq \sqrt{S_1+2S_2}\|f\|_2 \lesssim \sqrt{S^2+AS}\|f\|_2,$$

where

$$A := \Big(\sup_{P'\in\mathbb{P}} \frac{|\langle\phi_{P'}, f\rangle|}{|I_{P'}|^{1/2}}\Big)\Big(\sum_{\mathbb{T}\in\mathscr{T}} |I_{\mathbb{T}}|\Big)^{1/2} \|f\|_2.$$

If $S^2 \geq AS$, we get $S^2 \lesssim S\|f\|_2$, and hence $S\lesssim \|f\|_2$. If $S^2 < AS$, then $S^2 \lesssim A^{1/2}S^{1/2}\|f\|_2$, thus $S^{3/2} \lesssim A^{1/2}\|f\|_2$, and hence $S \lesssim A^{1/3}\|f\|_2^{2/3}$. So in any case we deduce that

$$S \lesssim \|f\|_2 + A^{1/3}\|f\|_2^{2/3},$$

which is the asserted bound. □

7.3. **Trees.**

Proposition 7.4 (Tree lemma)**.** *Let $\mathbb{T}$ be a finite tree of tiles. Then*

$$\Big|\sum_{P\in\mathbb{T}} \langle f, \phi_P\rangle \langle \phi_P, 1_{E_{P_{uu}}}\rangle\Big| \lesssim \operatorname{density}(\mathbb{T}) \operatorname{energy}(\mathbb{T}) |I_{\mathbb{T}}|.$$

Let T denote a top of $\mathbb{T}$. Let $\mathscr{J}$ be the collection of maximal dyadic intervals J with the property that $2J \not\supseteq I_P$ for any $P\in\mathbb{T}$. We use this collection to reorganize the integrals $\langle\phi_P, 1_{E_{P_{uu}}}\rangle = \int \phi_P 1_{E_{P_{uu}}}$. Note that the support of the Fourier wave-packets ϕ_P is all $\mathbb{R}$ (in contrast to the Walsh wave-packets w_P with $\operatorname{supp} w_P = I_P \subseteq I_T$, and therefore we need to choose $\mathscr{J}$ slightly differently now.

Lemma 7.5. *$\mathscr{J}$ is a partition of $\mathbb{R}$.*

Proof. Exercise. □

Lemma 7.6. *Every $J \in \mathscr{J}$ satisfies either $J \subseteq 5I_T$, or $|J| \geq 2|I_T|$ and $\frac{1}{2}|J| \leq \operatorname{dist}(J, I_T) \leq 2|J|$.*

Proof. Suppose first that $J \subseteq 5I_T$ and $|J| > |I_T|$. Note that there are exactly three dyadic intervals J with this property: $I_T^{(1)}$, $I_T^{(2)}$, and $I_T^{(2)} \setminus I_T^{(1)}$. But in each case it is easy to check that $2J \supseteq I_T \supseteq I_P$ for every $P \in \mathbb{T}$, and therefore such J cannot be in $\mathscr{J}$. Thus all $J \in \mathscr{J}$ with $J \subseteq 5I_T$ satisfy $|J| \leq |I_T|$.

Suppose then that $J \not\subseteq 5I_T$ and $|J| \leq |I_T|$. Thus J is disjoint from $5I_T$, so that $\operatorname{dist}(J, I_T) \geq 2|I_T| \geq 2|J|$. But then $2J^{(1)} \subset 5J$ is still disjoint from I_T, so cannot contain any I_P. Thus J is not a maximal interval with this property, and hence $J \notin \mathscr{J}$. Thus all $J \in \mathscr{J}$ with $J \not\subseteq 5I_T$ satisfy $|J| \geq 2|I_T|$.

Let then $J \in \mathscr{J}$ satisfy $|J| \geq 2|I_T|$. Let $J' \subset 2J$ be a dyadic interval with $|J'| = \frac{1}{2}|J| \geq |I_T|$. If $J' \cap I_T \neq \varnothing$, then $J' \supseteq I_T$, and therefore $2J \supset I_P$ for all P, which is a contradiction. Thus every such J' is disjoint from I_T, and hence $2J \cap I_T = \varnothing$. Hence $\operatorname{dist}(J, I_T) \geq \frac{1}{2}|J|$. On the other hand, by maximality, we have that $2J^{(1)} \supseteq I_P$ for some I_P. Since the biggest distance from J to a point in $2J^{(1)} \setminus J$ is $2|J|$, we find that $\operatorname{dist}(J, I_T) \leq 2|J|$. □

By Lemma 7.5, we analyse the quantity in the Tree lemma as follows

$$\begin{aligned}
&\Big|\sum_{P\in\mathbb{T}}\langle f,\phi_P\rangle\langle\phi_P,1_{E_{P_{uu}}}\rangle\Big| = \Big|\int_{\mathbb{R}}\sum_{P\in\mathbb{T}}\langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\,\mathrm{d}x\Big| \\
&\leq \Big\|\sum_{P\in\mathbb{T}}\langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_{L^1(\mathbb{R})} = \sum_{J\in\mathscr{J}}\Big\|\sum_{P\in\mathbb{T}}\langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_{L^1(J)} \\
&\leq \sum_{J\in\mathscr{J}}\Big(\sum_{\substack{P\in\mathbb{T}\\|I_P|\leq|J|}}|\langle f,\phi_P\rangle|\|\phi_P 1_{E_{P_{uu}}}\|_{L^1(J)} + \Big\|\sum_{\substack{P\in\mathbb{T}\\|I_P|>|J|}}\langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_{L^1(J)}\Big).
\end{aligned}$$

Part $|I_P| \leq |J|$. A typical coefficient here can be bounded as follows:

$$\begin{aligned}
|\langle f,\phi_P\rangle|\|\phi_P 1_{E_{P_{uu}}}\|_{L^1(J)} &= \frac{|\langle f,\phi_P\rangle|}{|I_P|^{1/2}}\Big\|\frac{\phi_P}{|I_P|^{1/2}}1_{E_{P_{uu}}}\Big\|_{L^1(J)}|I_P| \\
&\lesssim \operatorname{energy}(\mathbb{T})\int_{J\cap E_{P_{uu}}}\frac{1}{|I_P|}\Big(1+\frac{|x-c(I_P)|}{|I_P|}\Big)^{-20}\mathrm{d}x|I_P| \\
&\lesssim \operatorname{energy}(\mathbb{T})\Big(1+\frac{\operatorname{dist}(J,I_P)}{|I_P|}\Big)^{-10}\int_{E_{P_{uu}}}\frac{1}{|I_P|}\Big(1+\frac{|x-c(I_P)|}{|I_P|}\Big)^{-10}\mathrm{d}x|I_P| \\
&\lesssim \operatorname{energy}(\mathbb{T})\Big(1+\frac{\operatorname{dist}(J,I_P)}{|I_P|}\Big)^{-10}\operatorname{density}(\mathbb{T})|I_P|.
\end{aligned}$$

We then turn to the sum

$$\sum_{\substack{P\in\mathbb{T}\\|I_P|\leq|J|}} = \sum_{k:2^k\leq|J|}\sum_{\substack{P\in\mathbb{T}\\|I_P|=2^k}}$$

of these coefficients. Note first that all I_P in the sum must be disjoint from J: if $I_P \cap J \neq \varnothing$ and $|I_P| \leq |J|$, then $I_P \subseteq J \subset 2J$, a contradiction. So all relevant I_P are either on the left or right of J. The distance $\operatorname{dist}(J, I_P)$ attains some minimal value (possible for two different intervals I_P on opposite sides of J), and all other values (again, each one attained for at most two different I_P) are equal to this smallest value plus $n|I_P|$ for some $n \in \mathbb{N}$. Moreover, the smallest value can be estimated from below by $\operatorname{dist}(J, I_P)/|I_P| \geq \operatorname{dist}(J, I_T)/|I_T|$, and hence

$$\sum_{\substack{P\in\mathbb{T}\\|I_P|=2^k}}\Big(1+\frac{\operatorname{dist}(J,I_P)}{|I_P|}\Big)^{-10} \leq 2\sum_{n=0}^{\infty}\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}+n\Big)^{-10} \lesssim \Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-9}.$$

Since $\sum_{k:2^k\leq|J|}2^k=2\cdot|J|$, we have

$$\sum_{\substack{P\in\mathbb{T}\\|I_P|\leq|J|}}|\langle f,\phi_P\rangle|\|\phi_P 1_{E_{P_{uu}}}\|_{L^1(J)}\lesssim \operatorname{energy}(\mathbb{T})\operatorname{density}(\mathbb{T})|J|\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-9},$$

and it remains to sum this over $J\in\mathscr{J}$. We have

$$\sum_{J\in\mathscr{J}}|J|\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-9}\leq\sum_{\substack{J\in\mathscr{J}\\J\subseteq 5I_T}}|J|+\sum_{\substack{J\in\mathscr{J}\\J\not\subseteq 5I_T}}|J|\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-9},$$

where the first sem is immediately dominated by $5|I_T|$, since the intervals J are pairwise disjoint.

For the second sum, we use Lemma 7.6. If $z\in J$ is arbitrary, we have $\operatorname{dist}(J,I_T)\leq\operatorname{dist}(z,I_T)\leq\operatorname{dist}(J,I_T)+|J|$. By Lemma 7.6, $\operatorname{dist}(I_T,J)\eqsim|J|$, so in fact $\operatorname{dist}(z,I_T)\eqsim|J|\eqsim\operatorname{dist}(J,I_T)$ for all $z\in J$. Thus

$$|J|\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-9}\lesssim\int_J\mathrm{d}x\cdot\inf_{z\in J}\Big(1+\frac{\operatorname{dist}(z,I_T)}{|I_T|}\Big)^{-9}\leq\int_J\Big(1+\frac{\operatorname{dist}(x,I_T)}{|I_T|}\Big)^{-9}\mathrm{d}x.$$

Finally, if $x\in J\in\mathscr{J}$ and $J\not\subseteq 5I_T$, Lemma 7.6 further implies that $\operatorname{dist}(x,I_T)\geq\operatorname{dist}(J,I_T)\geq\frac{1}{2}|J|\geq|I_T|$. Thus

$$\begin{aligned}\sum_{\substack{J\in\mathscr{J}\\J\not\subseteq 5I_T}}|J|\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-9}&\lesssim\int_{(3I_T)^c}\Big(1+\frac{\operatorname{dist}(x,I_T)}{|I_T|}\Big)^{-9}\mathrm{d}x\\&\leq 2\int_{|I_T|}^{\infty}\Big(1+\frac{y}{|I_T|}\Big)^{-9}\mathrm{d}y=2|I_T|\int_1^\infty(1+u)^{-9}\,\mathrm{d}u\lesssim|I_T|.\end{aligned}$$

This completes the estimate

$$\sum_{J\in\mathscr{J}}\sum_{\substack{P\in\mathbb{T}\\|I_P|\leq|J|}}|\langle f,\phi_P\rangle|\|\phi_P 1_{E_{P_{uu}}}\|_{L^1(J)}\lesssim\operatorname{density}(\mathbb{T})\operatorname{energy}(\mathbb{T})|I_T|.$$

Part $|I_P|>|J|$. Here we should investigate the functions

$$F_J:=1_J\sum_{\substack{P\in\mathbb{T}\\|I_P|>|J|}}\langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}.$$

Observe first that

$$\operatorname{supp}F_J\subseteq J\cap\bigcup_{\substack{P\in\mathbb{T}\\|I_P|>|J|}}E_{P_{uu}}\subseteq J\cap\bigcup_{\substack{P\in\mathbb{T}\\|I_P|>|J|}}E_{P_u}=:G_J.$$

Lemma 7.7.

$$|G_J|\lesssim\operatorname{density}(\mathbb{T})|J|.$$

Proof. Assume that the union in the definition of G_J is nonempty, for otherwise the claim is trivial.

The idea is to construct a particular tile $\hat{P}$ so that $\hat{P}\geq\tilde{P}$ for some $\tilde{P}\in\mathbb{T}$ and $E_{P_u}\subseteq E_{\hat{P}}$ for all P in the union defining G_J. Then we would have $G_J\subseteq J\cap E_{\hat{P}}$.

For details, let $\hat{J}:=J^{(1)}$ be the dyadic parent of J. Thus $|\hat{J}|\leq|I_P|$ for all P in the union, and hence $|\hat{J}|\leq|I_T|$. By maximality of J, we know that $2\hat{J}$ contains some $I_{\tilde{P}}$. Notice that the biggest *dyadic* interval contained in $2\hat{J}$ is $\hat{J}$ itself, so that $|I_{\tilde{P}}|\leq|\hat{J}|$.

We define a tile $\hat{P}=I_{\hat{P}}\times\omega_{\hat{P}}$ as follows: $I_{\hat{P}}$ is the unique dyadic interval with $|I_{\hat{P}}|=|\hat{J}|$ so that $I_{\hat{P}}\supseteq I_{\tilde{P}}$, and $\omega_{\hat{P}}$ is the unique dyadic interval of length $|\omega_{\hat{P}}|=1/|\hat{J}|$ so that $\omega_{\hat{P}}\supseteq\omega_T$. Since we have $I_{\tilde{P}}\subseteq I_T$ and $\omega_{\tilde{P}}\supseteq\omega_T$, it follows by comparing the sizes that $I_{\tilde{P}}\subseteq I_{\hat{P}}\subseteq I_T$ and $\omega_{\tilde{P}}\supseteq\omega_{\hat{P}}\supseteq\omega_T$. Thus $\tilde{P}\leq\hat{P}\leq T$.

We make an observation about the location of the interval $I_{\hat{P}}$. Recall that $I_{\hat{P}} \cap 2\hat{J} \supseteq I_{\tilde{P}} \neq \varnothing$, and $|I_{\hat{P}}| = |\hat{J}|$. So $I_{\hat{P}}$ is either $\hat{J}$ or one of its dyadic neighbours of the same length. Since J is a dyadic child of $\hat{J}$, this in turn implies that

$$v_{I_{\hat{P}}}(x) = \frac{1}{|I_{\hat{P}}|}\Big(1 + \frac{|x - c(I_{\hat{P}})|}{|I_{\hat{P}}|}\Big)^{-10} \gtrsim \frac{1_J(x)}{|J|}.$$

We can complete the argument: Let P be one of the tiles with $P \in \mathbb{T}$ and $|I_P| > |J|$, thus $|I_P| \geq |\hat{J}|$. Then $\omega_P \supseteq \omega_T$ and $|\omega_P| \leq 1/|\hat{J}|$. But also $\omega_{\hat{P}} \supseteq \omega_T$, and $|\omega_{\hat{P}}| = 1/|\hat{J}|$. This implies that $\omega_P \subseteq \omega_{\hat{P}}$, and hence $E_P = E \cap \{x : N(x) \in \omega_P\} \subseteq E \cap \{x : N(x) \in \omega_{\hat{P}}\} = E_{\hat{P}}$. Therefore $G_J \subseteq J \cap E_{\hat{P}}$, and then

$$|G_J| \leq |J \cap E_{\hat{P}}| = |J| \int_{E_{\hat{P}}} \frac{1_J}{|J|} \lesssim |J| \int_{E_{\hat{P}}} v_{I_{\hat{P}}} \leq |J| \,\mathrm{density}(\mathbb{T}),$$

since $\hat{P} \geq \tilde{P} \in \mathbb{T}$ is one of the tiles appearing in the supremum in the definition of density. □

After this, the strategy will be to estimate

$$\|F_J\|_{L^1(J)} \leq |G_J| \cdot \|F_J\|_\infty \lesssim \mathrm{density}(\mathbb{T})|J| \cdot \|F_J\|_\infty,$$

and we should find for F_J a bound that involves the energy of the tree $\mathbb{T}$.

Splitting the tree $\mathbb{T}$. Recall that

$$P \leq T \quad \Leftrightarrow \quad \big(P_u \leq T_u \quad \text{or} \quad P_d \leq T_d\big)$$

and applying this to the up-parts P_u in place of P,

$$P_u \leq T_u \quad \Leftrightarrow \quad \big(P_{uu} \leq T_{uu} \quad \text{or} \quad P_{ud} \leq T_{ud}\big).$$

(Actually, in the Walsh context, we used these results with tiles and bitiles, but they obviously extend to semi-tiles and quarter-tiles.) Accordingly, we can split

$$\mathbb{T} = \mathbb{T}_{uu} \cup \mathbb{T}_{ud} \cup \mathbb{T}_d,$$

where

$$\begin{aligned}
\mathbb{T}_{uu} &:= \{P \in \mathbb{T} : \omega_{P_{uu}} \supseteq \omega_{T_{uu}}\},\\
\mathbb{T}_{ud} &:= \{P \in \mathbb{T} \setminus \mathbb{T}_{uu} : \omega_{P_{ud}} \supseteq \omega_{T_{ud}}\},\\
\mathbb{T}_d &:= \{P \in \mathbb{T} \setminus (\mathbb{T}_{uu} \cup \mathbb{T}_{ud}) : \omega_{P_d} \supseteq \omega_{T_d}\}.
\end{aligned}$$

The disjoint parts $\mathbb{T}_d$ *and* $\mathbb{T}_{ud}$.

Lemma 7.8. *For tiles* $P, P' \in \mathbb{T}_d$, *if* $\omega_P \neq \omega_{P'}$, *then* $\omega_{P_u} \cap \omega_{P'_u} = \varnothing$. *Similarly, for* $P, P' \in \mathbb{T}_{ud}$, *if* $\omega_P \neq \omega_{P'}$, *then* $\omega_{P_{uu}} \cap \omega_{P'_{uu}} = \varnothing$

Proof. Consider the first case. We have $\omega_{P_d} \cap \omega_{P'_d} \supseteq \omega_{T_d} \neq \varnothing$; therefore e.g. $\omega_{P'_d} \subseteq \omega_{P_d}$, and since the tiles are unequal, in fact $\omega_{P'} \subseteq \omega_{P_d}$. But then $\omega_{P'_u} \cap \omega_{\omega_{P_u}} \subseteq \omega_{P'} \cap \omega_{P_u} = \varnothing$.

The second case may be proven similarly, or as a an application of the first case to the semi-tiles $P_u, P'_u \in (\mathbb{T}_u)_d$ with $\omega_{P_u} \neq \omega_{P'_u}$, hence $\omega_{(P_u)_u} \cap \omega_{(P'_u)_u} = \varnothing$. □

Lemma 7.9. *We have*

$$\Big\| \sum_{\substack{P \in \mathbb{T}_d \\ |I_P| > |J|}} \langle f, \phi_P \rangle \phi_P 1_{E_{P_{uu}}} \Big\|_\infty \lesssim \mathrm{energy}(\mathbb{T}),$$

as well as the similar estimate with $\mathbb{T}_{ud}$ *in place of* $\mathbb{T}_d$.

Proof. We write

$$\sum_{\substack{P \in \mathbb{T}_d \\ |I_P| > |J|}} \langle f, \phi_P \rangle \phi_P 1_{E_{P_{uu}}} = \sum_{k : 2^k > |J|} \sum_{\substack{P \in \mathbb{T}_d \\ |I_P| = 2^k}} \langle f, \phi_P \rangle \phi_P 1_{E_{P_{uu}}}. \tag{7.9}$$

Note that if $|I_P| \neq |I_{P'}|$, then $\omega_P \neq \omega_{P'}$, and hence by Lemma 7.8, $\omega_{P_{uu}} \cap \omega_{P'_{uu}} = \varnothing$. Thus $E_{P_{uu}} \cap E_{P'_{uu}} = E \cap \{x : N(x) \in \omega_{P_{uu}} \cap \omega_{P'_{uu}}\} = \varnothing$. We conclude that the k-summands on the right of (7.9) are disjointly supported, and hence

$$\Big\| \sum_{\substack{P\in\mathbb{T}_d\\ |I_P|>|J|}} \langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_\infty = \sum_{k:2^k>|J|} \Big\| \sum_{\substack{P\in\mathbb{T}_d\\ |I_P|=2^k}} \langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_\infty .$$

To estimate the L^∞-norm for a fixed k, we have

$$\begin{aligned}\sum_{\substack{P\in\mathbb{T}_d\\ |I_P|=2^k}} |\langle f,\phi_P\rangle\phi_P(x) 1_{E_{P_{uu}}}(x)| &\leq \sum_{\substack{P=\omega_P\times I_P\in\mathbb{T}_d\\ |I_P|=2^k}} \frac{|\langle f,\phi_P\rangle|}{|I_P|^{1/2}} \times |I_P|^{1/2}|\phi_P(x)| \\ &\lesssim \operatorname{energy}(\mathbb{T}) \sum_{\substack{I\in\mathscr{D}\\ |I|=2^k}} \Big(1+\frac{|x-c(I)|}{|I|}\Big)^{-10} \\ &\lesssim \operatorname{energy}(\mathbb{T}) \sum_{n=0}^\infty (1+n)^{-10} \lesssim \operatorname{energy}(\mathbb{T}).\end{aligned}$$

We used (as often before) that $c(I)$ assumes values which are increase and decrease in integer multiples of $|I|$.

The proof for $\mathbb{T}_{ud}$ is almost the same, just using the other part of Lemma 7.8. □

Now we can already complete the estimate for

$$\sum_{J\in\mathscr{J}} \Big\| \sum_{\substack{P\in\mathbb{T}_d\cup\mathbb{T}_{ud}\\ |I_P|>|J|}} \langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_{L^1(J)} .$$

Observe that if the inner sum is non-empty, then we have $|J| < |I_P|$ for some $P \in \mathbb{T}$, and hence $|J| < |I_P| \leq |I_T|$. By Lemma 7.6, this implies that $J \subseteq 5I_T$.

$$\begin{aligned}&\sum_{J\in\mathscr{J}} \Big\| 1_J \sum_{\substack{P\in\mathbb{T}_d\cup\mathbb{T}_{ud}\\ |I_P|>|J|}} \langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_1 \\ &\quad\leq \sum_{\substack{J\in\mathscr{J}\\ J\subseteq 5I_T}} |G_J| \times \Big\| 1_J \sum_{\substack{P\in\mathbb{T}_d\cup\mathbb{T}_{ud}\\ |I_P|>|J|}} \langle f,\phi_P\rangle\phi_P 1_{E_{P_{uu}}}\Big\|_\infty \\ &\quad\lesssim \sum_{\substack{J\in\mathscr{J}\\ J\subseteq 5I_T}} \operatorname{density}(\mathbb{T})|J| \times \operatorname{energy}(\mathbb{T}) \\ &\quad\lesssim \operatorname{density}(\mathbb{T})\operatorname{energy}(\mathbb{T})|I_T|,\end{aligned}$$

since the intervals J in the sum are pairwise disjoint and contained in $5I_T$.

The nested part $\mathbb{T}_{uu}$. It only remains to consider

$$\begin{aligned}\sum_{\substack{P\in\mathbb{T}_{uu}\\ |I_P|>|J|}} \langle f,\phi_P\rangle\phi_P(x) 1_{E_{P_{uu}}}(x) &= 1_E(x) \sum_{\substack{P\in\mathbb{T}_{uu}\\ |I_P|>|J|}} \langle f,\phi_P\rangle\phi_P(x) 1_{\omega_{P_{uu}}}(N(x)) \\ &= 1_E(x) \sum_{\substack{P\in\mathbb{T}_{uu}\\ |I_P|>|J|\\ \omega_{P_{uu}}\ni N(x)}} \langle f,\phi_P\rangle\phi_P(x).\end{aligned}$$

The condition that $P \in \mathbb{T}_{uu}$ means that $\omega_{P_{uu}} \supseteq \omega_{T_{uu}}$. Hence any two $\omega_{P_{uu}}$ appearing in the sum intersect, and hence the larger interval contains the smaller. In particular, if $\omega_{P_{uu}} \ni N(x)$ for some P in this sum, then this containment also holds for all bigger $\omega_{P_{uu}}$. So the condition $\omega_{P_{uu}} \ni N(x)$ can be equivalently written as $|\omega_P| > |\omega_x|$, where ω_x is frequency interval depending on x. On the other hand, the condition that $|I_P| > |J|$ is equivalent to $|I_P| \geq 2|J|$, and then further to $|\omega_P| \leq (2|J|)^{-1} =: |\omega_J|$. Thus, we have

$$\sum_{\substack{P\in\mathbb{T}_{uu}\\|I_P|>|J|\\\omega_{P_{uu}}\ni N(x)}} \langle f,\phi_P\rangle\phi_P(x) = \sum_{\substack{P\in\mathbb{T}_{uu}\\|\omega_x|<|\omega_P|\leq|\omega_J|}} \langle f,\phi_P\rangle\phi_P(y)\Big|_{y=x}.$$

Lemma 7.10. *For a fixed x we have the identity*

$$\sum_{\substack{P\in\mathbb{T}_{uu}\\|\omega_x|<|\omega_P|\leq|\omega_J|}} \langle f,\phi_P\rangle\phi_P(y)$$
$$= M_{c(\omega_{T_{uu}})}\Big((D^1_{2|J|}\chi - D^1_{1/|\omega_x|}\chi) * M_{-c(\omega_{T_{uu}})}\sum_{P\in\mathbb{T}_{uu}}\langle f,\phi_P\rangle\phi_P\Big)(y),$$

where $\chi \in \mathscr{S}(\mathbb{R})$ is an auxiliary function whose Fourier transform satisfies

$$1_{[-0.8,0.8]} \leq \hat{\chi} \leq 1_{[-0.85,0.85]}.$$

Proof. We investigate the Fourier transforms. Recall that

$$\phi_P = M_{c(\omega_{P_d})}T_{c(I_P)}D^2_{|I_P|}\phi,$$

where $\operatorname{supp}\hat{\phi} \subseteq [-\frac{1}{20},\frac{1}{20}]$, has

$$\operatorname{supp}\hat{\phi}_P \subseteq c(\omega_{P_d}) + [-0.05, 0.05]|\omega_P|.$$

On the other hand, we have

$$c(\omega_{T_{uu}}) \in \omega_{T_{uu}} \subseteq \omega_{P_{uu}} = c(\omega_{P_d}) + [0.5, 0.75]|\omega_P|.$$

Expressing the support of $\operatorname{supp}\hat{\phi}_P$ relative to $c(\omega_{T_{uu}})$, we thus have

$$\operatorname{supp}\hat{\phi}_P \subseteq c(\omega_{T_{uu}}) + [-0.8, -0.45]|\omega_P|.$$

Then it follows, if χ is as defined, and Ω is a dyadic side-length, that

$$\hat{\chi}\Big(\frac{\eta - c(\omega_{T_{uu}})}{\Omega}\Big)\hat{\phi}_P(\eta) = \begin{cases}\hat{\phi}_P(\eta), & |\omega_P| \leq \Omega,\\ 0 & |\omega_P| > \Omega.\end{cases}$$

Observe that $|\omega_P| > \Omega$ implies that $|\omega_P| \geq 2\Omega$, and hence

$$\operatorname{supp}\hat{\phi}_P \subseteq c(\omega_{T_{uu}}) + (-\infty, 0.45]|\omega_P| \subseteq c(\omega_{T_{uu}}) + (-\infty, 0.9]\Omega.$$

Hence

$$\sum_{\substack{P\in\mathbb{T}_{uu}\\|\omega_x|<|\omega_P|\leq|\omega_J|}} \langle f,\phi_P\rangle\hat{\phi}_P(\eta) = \Big[\hat{\chi}\Big(\frac{\eta - c(\omega_{T_{uu}})}{\omega_J}\Big) - \hat{\chi}\Big(\frac{\eta - c(\omega_{T_{uu}})}{\omega_x}\Big)\Big]\sum_{P\in\mathbb{T}_{uu}}\langle f,\phi_P\rangle\hat{\phi}_P(\eta).$$

Moreover,

$$\hat{\chi}\Big(\frac{\eta - c}{\Omega}\Big)\hat{\phi}_P = T_cD^\infty_\Omega\hat{\chi}(\eta)\hat{\phi}_P = \mathscr{F}(M_cD^1_{1/\Omega}\chi)(\eta)\mathscr{F}\phi_P(\eta) = \mathscr{F}(M_cD^1_{1/\Omega}\chi * \phi_P)(\eta)$$

and thus, identifying the functions under the equal Fourier transforms, we have

$$\sum_{\substack{P\in\mathbb{T}_{uu}\\|\omega_x|<|\omega_P|\leq|\omega_J|}} \langle f,\phi_P\rangle\phi_P(y) = \Big(M_{c(\omega_{T_{uu}})}(D^1_{2|J|}\chi - D^1_{1/|\omega_x|}\chi) * \sum_{P\in\mathbb{T}_{uu}}\langle f,\phi_P\rangle\phi_P\Big)(y).$$

The right side may be expressed as in the statement of the lemma by the simple identity

$$M_cg * h(x) = \int_{\mathbb{R}} M_cg(x-y)h(y)\,\mathrm{d}y = \int_{\mathbb{R}} e^{i2\pi c(x-y)}g(x-y)h(y)\,\mathrm{d}y$$
$$= e^{i2\pi cx}\int_{\mathbb{R}} g(x-y)e^{-i2\pi cy}h(y)\,\mathrm{d}y = M_c(g * M_{-c}h)(x). \qquad \square$$

We combine the identity of Lemma 7.10 with the following estimate:

Lemma 7.11. *For $x \in I$ (an interval), we have*

$$|D^1_{|I|}\chi * g(x)| \lesssim \inf_I Mg.$$

Proof. We write out

$$D^1_{|I|}\chi * g(x) = \int_{\mathbb{R}} \frac{1}{|I|}\chi\Big(\frac{x-y}{|I|}\Big)g(y)\,\mathrm{d}y = \Big(\int_I + \sum_{k=0}^{\infty}\int_{2^{k+1}I\setminus 2^k I}\Big)\frac{1}{|I|}\chi\Big(\frac{x-y}{|I|}\Big)g(y)\,\mathrm{d}y.$$

For $x, y \in 2I$, we simply estimate $|\chi((x-y)/|I|)| \lesssim 1$. For $x \in I$ and $y \in 2^{k+1}I \setminus 2^k I$ and $k \geq 1$, we have $|x-y|/|I| \gtrsim 2^k$, and hence $|\chi((x-y)/|I|)| \lesssim 2^{-10k}$. So altogether

$$\begin{aligned}
|D^1_{|I|}\chi * g(x)| &\lesssim \frac{1}{|I|}\int_I |g(y)|\,\mathrm{d}y + \sum_{k=0}^{\infty} 2^{-10k}\frac{1}{|I|}\int_{2^{k+1}I\setminus 2^k I}|g(y)|\,\mathrm{d}y \\
&\lesssim \frac{1}{|I|}\int_I |g(y)|\,\mathrm{d}y + \sum_{k=0}^{\infty} 2^{-9k}\frac{1}{|2^{k+1}I|}\int_{2^{k+1}I}|g(y)|\,\mathrm{d}y,
\end{aligned}$$

and each of the averages here appears when evaluating the maximal function $Mg(z)$ at any $z \in I$; hence

$$|D^1_{|I|}\chi * g(x)| \leq Mg(z) + \sum_{k=0}^{\infty} 2^{-9k} Mg(z) \lesssim Mg(z),$$

and it remains to take infimum over $z \in I$. □

We apply Lemma 7.11 to the right side of the identity of Lemma 7.10 with $I = 2J$ and $I = J_x \supseteq 2J$, where $|J_x| = 1/|\omega_x|$. Note that we may assume without loss of generality that $|J_x| > 2|J|$, since otherwise the summation of the left of Lemma 7.10 is empty and the result is zero.

Thus, for $y = x \in J \subset 2J \subset J_x$, we have

$$\begin{aligned}
&\Big|\sum_{\substack{P\in\mathbb{T}_{uu}\\ |I_P|>|J|}}\langle f,\phi_P\rangle\phi_P(x)1_{E_{P_{uu}}}(x)\Big| \leq \Big|\sum_{\substack{P\in\mathbb{T}_{uu}\\ |\omega_x|<|\omega_P|\leq|\omega_J|}}\langle f,\phi_P\rangle\phi_P(x)\Big| \\
&= \Big|\Big((D^1_{2|J|}\chi - D^1_{1/|\omega_x|}\chi) * M_{-c(\omega_{T_{uu}})}\sum_{P\in\mathbb{T}_{uu}}\langle f,\phi_P\rangle\phi_P\Big)(x)\Big| \\
&\lesssim \inf_{2J} M\Big(\sum_{P\in\mathbb{T}_{uu}}\langle f,\phi_P\rangle\phi_P\Big) + \inf_{J_x} M\Big(\sum_{P\in\mathbb{T}_{uu}}\langle f,\phi_P\rangle\phi_P\Big) \\
&\lesssim \inf_{J} M\Big(\sum_{P\in\mathbb{T}_{uu}}\langle f,\phi_P\rangle\phi_P\Big),
\end{aligned}$$

since the infimum over a bigger set is smaller.

We can complete the proof of the remaining part of the Tree Lemma:

$$\begin{aligned}
&\sum_{\substack{J\in\mathscr{J}\\ J\subseteq 5I_T}} \Big\| 1_J \sum_{\substack{P\in\mathbb{T}_{uu}\\ |I_P|>|J|}} \langle f,\phi_P\rangle \phi_P 1_{E_{P_{uu}}} \Big\|_1 \\
&\qquad \le \sum_{\substack{J\in\mathscr{J}\\ J\subseteq 5I_T}} |G_J| \times \Big\| 1_J \sum_{\substack{P\in\mathbb{T}_{uu}\\ |I_P|>|J|}} \langle f,\phi_P\rangle \phi_P 1_{E_{P_{uu}}} \Big\|_\infty \\
&\qquad \lesssim \sum_{\substack{J\in\mathscr{J}\\ J\subseteq 5I_T}} \operatorname{density}(\mathbb{T})|J| \times \inf_J M\Big(\sum_{P\in\mathbb{T}_{uu}} \langle f,\phi_P\rangle\phi_P \Big) \\
&\qquad \le \operatorname{density}(\mathbb{T}) \int_{5I_T} M\Big(\sum_{P\in\mathbb{T}_{uu}} \langle f,\phi_P\rangle\phi_P \Big)(x)\,\mathrm{d}x \\
&\qquad \le \operatorname{density}(\mathbb{T})|5I_T|^{1/2} \Big[\int_{\mathbb{R}} M\Big(\sum_{P\in\mathbb{T}_{uu}} \langle f,\phi_P\rangle\phi_P \Big)(x)^2\,\mathrm{d}x \Big]^{1/2} \\
&\qquad \lesssim \operatorname{density}(\mathbb{T})|I_T|^{1/2} \Big[\int_{\mathbb{R}} \Big(\sum_{P\in\mathbb{T}_{uu}} \langle f,\phi_P\rangle\phi_P \Big)(x)^2\,\mathrm{d}x \Big]^{1/2}.
\end{aligned}$$

Recall from an exercise that the functions ϕ_P, where $P\in\mathbb{T}_{uu}$, split into 20 pairwise orthogonal subcollections. This allows to continue the previous estimate with

$$\begin{aligned}
&\lesssim \operatorname{density}(\mathbb{T})|I_T|^{1/2} \Big[\sum_{P\in\mathbb{T}_{uu}} |\langle f,\phi_P\rangle|^2 \Big]^{1/2} \\
&= \operatorname{density}(\mathbb{T}) \Big[\frac{1}{|I_T|} \sum_{P\in\mathbb{T}_{uu}} |\langle f,\phi_P\rangle|^2 \Big]^{1/2} |I_T| \\
&\le \operatorname{density}(\mathbb{T})\operatorname{energy}(\mathbb{T})|I_T|.
\end{aligned}$$

This is the desired bound, and completes the proof of the Tree Lemma in the Fourier model.

7.4. **Carleson's theorem: summary.** With the three basic lemmas at hand, the completion of the proof of Carleson's theorem proceeds exactly as in the Walsh case. Let us briefly recall the main lines. First, the original convergence question, $S_\xi f(x)\to f(x)$, was reduced to the study of the corresponding maximal operator $\sup_{\xi\in\mathbb{R}}|S_\xi f(x)|$, and then via linearization, the operator $S_{N(x)}f(x)$. We used an averaging trick over translated and dilated dyadic systems to replace the partial Fourier integrals $S_\xi f(x)=\int_{-\infty}^{\xi}\hat f(\eta)e^{i2\pi\eta x}\,\mathrm{d}\eta$ by the model operators $A_\xi f(x)=\sum_{P\in\mathbb{P}}\langle f,\phi_P\rangle\phi_P(x)1_{\omega_{P_{uu}}}(\xi)$. After the linearization and a property of the weak L^p norms, the main inequality $\|S_{N(\cdot)}f\|_{L^{2,\infty}}\lesssim\|f\|_{L^2}$ was seen to follows from $|\sum_{P\in\mathbb{P}}\langle f,\phi_P\rangle\langle\phi_P,1_{E_{P_{uu}}}\rangle|\lesssim\|f\|_2|E|^{1/2}$, where $\mathbb{P}$ is any finite collection of tiles. The Tree Lemma allows to control the left side by $\operatorname{density}(\mathbb{T})\operatorname{energy}(\mathbb{T})|I_T|$ if $\mathbb{P}=\mathbb{T}$ is a tree with top T. On the other hand, the Density and Energy Lemmas allow to extract trees out of any finite collection $\mathbb{P}$, in such a way that the energy and the density of the remaining collection are strictly smaller than the original. As in the Walsh case, a recursive use of these lemmas gives the decomposition

$$\mathbb{P}=\bigcup_{n\in\mathbb{Z}}\bigcup_j \mathbb{T}_{n,j}\cup\mathbb{P}_{\text{residual}},$$

where

$$\operatorname{density}(\mathbb{T}_{n,j})\le 2^{2n}|E|,\qquad \operatorname{energy}(\mathbb{T}_{n,j})\le 2^n\|f\|_2,\qquad \sum_j|T_{n,j}|\lesssim 2^{-2n},$$

and $\operatorname{density}(\mathbb{P}_{\text{residual}})=\operatorname{energy}(\mathbb{P}_{\text{residual}})=0$. There is also the universal bound $\operatorname{density}(\mathbb{T}_{n,j})\lesssim 1$. With these estimates and the Tree Lemma, it follows that

$$\Big|\sum_{P\in\mathbb{P}}\langle f,\phi_P\rangle\langle\phi_P,1_{E_{P_{uu}}}\rangle\Big|\lesssim\sum_{n\in\mathbb{Z}}\min\{1,2^{2n}|E|\}\times 2^n\|f\|_2\times 2^{-2n}.$$

This series decays geometrically for both $n \to \infty$ and $n \to -\infty$, hence converges, and a more careful computation reveals the correct bound $\|f\|_2 |E|^{1/2}$. Thus the proof of Carleson's theorem is complete.

7.5. **Bibliography.** Carleson's original paper, where he proved his celebrated theorem on Fourier series, is [2]. Billard's result on Walsh series appeared in the same year [1]. Another classical proof of Carleson's theorem is due to Fefferman [3].

However, rather than following these classical sources, the present lecture notes are essentially an expanded version of the paper of Lacey and Thiele [4], where they give a third proof for Carleson's theorem. The lecturer's ongoing collaboration with Michael Lacey on related topics has also influenced the presentation.

References

[1] P. Billard. Sur la convergence presque partout des séries de Fourier-Walsh des fonctions de l'espace $L^2(0, 1)$. *Studia Math.*, 28:363–388, 1966/1967.

[2] Lennart Carleson. On convergence and growth of partial sums of Fourier series. *Acta Math.*, 116:135–157, 1966.

[3] Charles Fefferman. Pointwise convergence of Fourier series. *Ann. of Math. (2)*, 98:551–571, 1973.

[4] Michael Lacey and Christoph Thiele. A proof of boundedness of the Carleson operator. *Math. Res. Lett.*, 7(4):361–370, 2000.