# Wavelet Packets and Walsh Functions

Chris B Dock

**Abstract**

It was discovered by Mallat and Meyer that the celebrated Haar and Shannon orthonormal wavelet bases for $L^2(\mathbb{R})$ could be produced via a formal process now known as the multi resolution analysis (MRA). Moreover, given an MRA, a wavelet basis can be explicitly constructed. A natural generalization of the MRA yields, for a given wavelet, a family of orthonormal bases for $L^2(\mathbb{R})$ known as wavelet packets. The first $N$ components of a given function in a wavelet basis give rise to more than $2^N$ wavelet packet components, using which a "best basis" can often be selected.[2] The MRA would thus appear to constitute a natural beginning for any historical work on the development of wavelet packets, except for the fact that as sometimes happens in mathematics a specific example of a wavelet packet manifested itself long before the more general theory springing from the MRA had been developed. Specifically, the wavelet packets one would obtain via "filling out the MRA tree" for the Haar wavelet were first set out in 1923 by Walsh (not long in fact after Haar put forth the first wavelet decomposition in his 1909 thesis). Walsh functions were well known prior to the development of the MRA because of their applicability to coding theory, radar systems, image transmission, and speech recognition.[1] It is with this in mind that I will aim in this exposition to follow the pathway from the Haar function, to the MRA, to the wavelet packet formulation of Walsh functions. I conclude, following [7], with a Carleson type theorem for Walsh functions.

## I. The Haar Wavelet

**Definition 1.** ***The Haar Function.*** *Let $\psi : \mathbb{R} \to \mathbb{R}$ be given by*

$$\psi(t) := \begin{cases} -1 & 0 \le t < \frac{1}{2} \\ 1 & \frac{1}{2} \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Then we have that, following [5]

**Theorem 1.** ***(Haar 1909)*** *$\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$ where*

$$\psi_{m,n}(t) := 2^{m/2}\psi(2^m t - n) \tag{2}$$

*is the family of dyadic dilations and translates of $\psi$.*

*Proof:* First note that $\psi_{m,n}(t)$ takes value $-2^{m/2}$ for $t \in [\frac{n}{2^m}, \frac{n}{2^m} + \frac{1}{2^{m+1}}) = [\frac{n}{2^m}, \frac{n+\frac{1}{2}}{2^m})$ and value $2^{m/2}$ on $[\frac{n+\frac{1}{2}}{2^m}, \frac{n+1}{2^m})$ and has support $[\frac{n}{2^m}, \frac{n+1}{2^m}]$. Denote therefore the half open dyadic interval $[\frac{n}{2^m}, \frac{n+1}{2^m})$ by $I_{m,n}$, then by definition we have that

$$\psi_{m,n} = 2^{m/2}(\mathbb{1}_{Im+1,2n+1} - \mathbb{1}_{Im+1,2n}) \tag{3}$$

We will therefore employ the following Lemma to prove that $\psi_{m,n}$ is an orthonormal set:

**Lemma 1.** *Given $I_{m,n}$ and $I_{p,q}$. If $I_{m,n} \cap I_{p,q} \neq \emptyset$ then either $I_{m,n} \subseteq I_{p,q}$ or $I_{p,q} \subseteq I_{m,n}$.*

The Lemma is fairly self-evident geometrically, nevertheless we proceed. Assume $[\frac{n}{2^m}, \frac{n+1}{2^m}) \cap [\frac{q}{2^p}, \frac{q+1}{2^p}) \neq \emptyset$. We may assume without loss of generality that

$$\frac{n}{2^m} \le \frac{q}{2^p} < \frac{n+1}{2^m} \tag{4}$$

since if the intersection is non-empty then the starting point of one of the two intervals must lie inside the other. There are several cases. First, if $m = p$, then

$$\begin{aligned} &[\frac{n}{2^m}, \frac{n+1}{2^m}) \cap [\frac{q}{2^m}, \frac{q+1}{2^m}) \neq \emptyset \\ &\Leftrightarrow [n, n+1) \cap [q, q+1) \neq \emptyset \\ &\Leftrightarrow n = q \\ &\Leftrightarrow I_{m,n} = I_{p,q} \end{aligned} \tag{5}$$

Which obviously satisfies the conclusion of the Lemma. Assume instead that $m < p$ or equivalently $1 \le p-m$. Then $2 \le 2^{p-m}$, which can be used in combination with (4) since

$$\frac{n2^{p-m}}{2^p} = \frac{n}{2^m} \le \frac{q}{2^p} < \frac{n+1}{2^m} = \frac{(n+1)2^{p-m}}{2^p} \tag{6}$$

From which it follows that

$$\begin{aligned} &n2^{p-m} \le q < (n+1)2^{p-m} \\ &\Rightarrow q+1 \le (n+1)2^{p-m} \\ &\Rightarrow \frac{q+1}{2^p} \le \frac{n+1}{2^m} \end{aligned} \tag{7}$$

In concert with the fact that $\frac{n}{2^m} \le \frac{q}{2^p}$ this tells us that $I_{p,q} \subseteq I_{m,n}$. The final case is $p < m$ (note that we cannot without loss of generality assume $p \ge m$ because in combination with (4) this would in fact lose generality). The $p < m$ case unfortunately requires us to consider two sub-cases separately: $\frac{n}{2^m} = \frac{q}{2^p}$ and $\frac{n}{2^m} < \frac{q}{2^p}$ ($\frac{n}{2^m} > \frac{q}{2^p}$ violates (4)). If $\frac{n}{2^m} = \frac{q}{2^p}$ then $1 \le m-p \Rightarrow 2 \le 2^{m-p}$ hence

$$\begin{aligned} &\frac{1}{2^m} \le \frac{2^{m-p}}{2^{m+1}} = \frac{1}{2^{p+1}} < \frac{1}{2^p} \\ &\Rightarrow \frac{n+1}{2^m} = \frac{n}{2^m} + \frac{1}{2^m} < \frac{n}{2^m} + \frac{1}{2^p} = \frac{q}{2^p} + \frac{1}{2^p} = \frac{q+1}{2^p} \end{aligned} \tag{8}$$

In combination with $\frac{n}{2^m} = \frac{q}{2^p}$ we conclude $I_{m,n} \subseteq I_{p,q}$. Finally assume $p < m$ and $\frac{n}{2^m} < \frac{q}{2^p}$. Then

$$0 < \frac{q}{2^p} - \frac{n}{2^m} = \frac{q2^{m-p} - n}{2^m} \tag{9}$$

Noting that $q2^{m-p} - n$ is an integer since $p < m$ this implies that $q2^{m-p} - n \ge 1$ hence $\frac{q}{2^p} \ge \frac{n+1}{2^m}$ which precisely contradicts our hypothesis that $I_{m,n} \cap I_{p,q} \ne \emptyset$, hence in fact this second subcase cannot occur. This concludes the proof of Lemma 1. We will use it to show that $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$ is an orthonormal set. First note that

$$||\psi_{m,n}||_{L^2(\mathbb{R})} = 2^m \int_{\frac{n}{2^m}}^{\frac{n+1}{2^m}} 1dt = 1 \tag{10}$$

If $m = p$ and $n \ne q$ then in fact $I_{m,n} \cap I_{p,q} = \phi$, hence $\langle \psi_{m,n}, \psi_{p,q} \rangle_{L^2(\mathbb{R})} = 0$. The interesting case is when $m \ne p$. In this case, we will still have $I_{m,n} \cap I_{p,q} = \phi$ when $q$ is sufficiently large or sufficiently negative, making the corresponding inner products automatically zero, but in between there will be cases where the intersection is non-empty. By Lemma 1, however, we have that either $I_{m,n} \subseteq I_{p,q}$ or $I_{p,q} \subseteq I_{m,n}$. Suppose without loss of generality that $I_{p,q} \subseteq I_{m,n}$ (if not use symmetry of the inner product to interchange the two). In this case $\mathbb{1}_{I_{p,q}} \cdot \mathbb{1}_{I_{m,n}} = \mathbb{1}_{I_{m,n}}$. Since $p < m$ is strict it follows that $\psi_{m,n}$ is constant over the support of $\psi_{p,q}$. Denote this constant by $C_{p,q,m,n}$ (it obviously depends on $m$ and $n$ but it's sign will also depend on $p$ and $q$). In this case

$$\langle \psi_{m,n}, \psi_{p,q} \rangle_{L^2(\mathbb{R})} = C_{p,q,m,n} \int_{\text{supp}(\psi_{p,q})} \psi_{p,q} dt = 0 \tag{11}$$

Where the last inequality follows since each $\psi_{m,n}$ has integral over any set containing its support. Thus we conclude that $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$ is an orthonormal set in $L^2(\mathbb{R})$. We employ a standard result from functional analysis without proof, namely:

**Theorem 2.** ***(Goldberg-Gottberg)*** *Let $H$ be a separable Hilbert space and $\{\theta_n\}_{n\in Z} \subseteq H$. Then the following are equivalent:*
*(a) $\{\theta_n\}_{n\in Z}$ is an orthonormal Riesz basis for $H$.*
*(b) If $x \in H$ and $\forall n \langle x, \theta_n \rangle_H = 0$ then $x = 0$.*
*(c) $\overline{span}\{\theta_n\} = H$.*
*(d) $\forall x \in H, ||x||_H^2 = \sum_{n\in\mathbb{Z}} |\langle x, \theta_n \rangle_H|$*

We will therefore show $\overline{\text{span}}\{\psi_{m,n}\} = L^2(\mathbb{R})$, which will conclude the proof of Theorem 1 (later, Theorem 1 will follow from the MRA construction of $\psi$). To this end fix $\epsilon > 0$ and recall that if $g \in L^2(\mathbb{R})$ then $\exists g_N$ such that $\text{supp}(g_N) \subseteq [-2^N, 2^N]$ and $||g - g_N||_{L^2(\mathbb{R})} < \frac{\epsilon}{3}$. Namely, take $g_N = g \cdot \mathbb{1}_{[-2^N, 2^N]}$ then

$$\begin{aligned} &g = g\mathbb{1}_{[-2^N,2^N]} + g\mathbb{1}_{[-2^N,2^N]^C} \\ &\Rightarrow ||g||_{L^2(\mathbb{R})} = ||g\mathbb{1}_{[-2^N,2^N]}||_{L^2(\mathbb{R})} + ||g\mathbb{1}_{[-2^N,2^N]^C}||_{L^2(\mathbb{R})} \\ &\Rightarrow \lim_{N\to\infty} ||g\mathbb{1}_{[-2^N,2^N]^C}||_{L^2(\mathbb{R})} = 0 \end{aligned} \tag{12}$$

Where the last implication follows from the fact that $||g||_{L^2(\mathbb{R})} := \lim_{N\to\infty} ||g\mathbb{1}_{[-2^N,2^N]}||_{L^2(\mathbb{R})}$. But $g\mathbb{1}_{[-2^N,2^N]^C} = g - g_N$ hence $\lim_{N\to\infty} ||g - g_N||_{L^2(\mathbb{R})} = 0$. Obviously this proves the existence of such a $g_N$. Furthermore, we claim that $\exists M > 0$ and $f \in L^2(\mathbb{R})$ such that $\text{supp}(f) \subseteq [-2^N, 2^N]$ and that for each $n \in \{-2^{M+N}, \dots, 2^{M+N} - 1\}$ $f$ is constant on the dyadic interval $I_{M+N,n}$ and $||f - g_N||_{L^2(\mathbb{R})} < \frac{\epsilon}{3}$. Specifically take $f_{M,n} = \frac{1}{|I_{M,n}|}\int_{I_{M,n}} g_N(s)ds$ and let $f = \sum_{n=-2^{M+N}}^{2^{M+N}-1} f_{M,n}\mathbb{1}_{I_{M,n}}$. Then

$$\begin{aligned} g_N(t) - f(t) &= \sum_{n=-2^{M+N}}^{2^{M+N}-1} \mathbb{1}_{I_{M,n}} \frac{1}{|I_{M,n}|}\int_{I_{M,n}} (g_N(s) - g_N(t))ds \\ \Rightarrow ||g_N - f||_{L^2(\mathbb{R})} &= \Big( \sum_{n=-2^{M+N}}^{2^{M+N}-1} \frac{1}{|I_{M,n}|}\int_{I_{M,n}}\int_{I_{M,n}} |g_N(s) - g_N(t)|^2 dtds \Big)^{\frac{1}{2}} \end{aligned} \tag{13}$$

We now use the fact that that the set of smooth functions is dense in $L^2(\mathbb{R})$, hence there exists such a function satisfying $||g_N - h_N||_{L^2(\mathbb{R})} < \frac{\epsilon}{6\sqrt{6}}$ from which it follows that

$$\begin{aligned} |g_N(s) - g_N(t)|^2 &= |g_N(s) - h_N(s) + h_N(s) - h_N(t) + h_N(t) - g_N(t)|^2 \\ &\leq 3(|g_N(s) - h_N(s)|^2 + |h_N(s) - h_N(t)|^2 + |h_N(t) - g_N(t)|^2) \end{aligned} \tag{14}$$

From this and (13) it follows that

$$||g_N - f||_{L^2(\mathbb{R})} \leq \sqrt{6}||g_N - h_N||_{L^2(\mathbb{R})} + \Big( \sum_{n=-2^{M+N}}^{2^{M+N}-1} \frac{1}{|I_{M,n}|}\int_{I_{M,n}}\int_{I_{M,n}} |h_N(s) - h_N(t)|^2 dtds \Big)^{\frac{1}{2}} \tag{15}$$

The point of this exercise was obviously to utilize the continuous differentiability of $h_N$, which we now do. $|h_N(s) - h_N(t)|^2 \leq ||h'_N||_\infty^2 (t-s)^2$ from which it follows that

$$\begin{aligned} ||g_N - f||_{L^2(\mathbb{R})} &\leq \sqrt{6}||g_N - h_N||_{L^2(\mathbb{R})} + \Big( \sum_{n=-2^{M+N}}^{2^{M+N}-1} \frac{1}{12|I_{M,n}|} \cdot |I_{M,n}|^4 \cdot ||h'_N||_\infty^2 \Big)^{1/2} \\ &= \sqrt{6}||g_N - h_N||_{L^2(\mathbb{R})} + (2^{N+1}/12)^{1/2}||h'_N||_\infty |I_{M,n}| \end{aligned} \tag{16}$$

Where we used the fact that the number of terms in the sum is $\frac{2^{N+1}}{|I_{M,n}|}$.The first term is less than $\frac{\epsilon}{6}$ and the second can be made less than $\frac{\epsilon}{6}$ by correct choice of $M$ since $|I_{M,n}| = 2^{-M}$, from which we conclude that $||g_N - f||_{L^2(\mathbb{R})} \leq \frac{\epsilon}{3}$. Long winded detour aside, the proof of the theorem amounts to proving that a function that is constant on each $I_{M,n}$ and supported in $[-2^N, 2^N]$ can be approximated arbitrarily well in $L^2(\mathbb{R})$ by linear combinations of the $\{\psi_{m,n}\}$. In fact, given $\epsilon > 0$ and a function $f$ that is supported on $[-2^N, 2^N]$ and constant on $I_{m,n}$ for $n \in \{2^{-N+m}, \dots, 2^{N+m}-1\}$ we shall prove the existence of $Q < \infty$ and $\{c_{m,n}\}$ such that

$$||f - \sum_{m=0}^{Q} \sum_{n=-2^{N+m}}^{2^{N+m}-1} c_{m,n}\psi_{m,n}||_{L^2(\mathbb{R})} < \frac{\epsilon}{3} \tag{17}$$

Note first that we can split a given interval $I_{m-1,n}$ in half as $I_{m-1,n} = I_{m,2n} \sqcup I_{m,2n+1}$. The given $f$ is of course constant on each $I_{m,n}$, so we would like to design a function which is constant on each $I_{m-1,n}$ for $n \in \{-2^{m+N-1}, \dots, 2^{m+N-1}-1\}$ and that approximates $f$ as well as possible, and then proceed recursively. We will denote this new function as $f_{-1}$ and $f$ as $f_0$ and define $f_{-1}$ as the average of the values of $f$ on the two half intervals:

$$\begin{aligned} f_{-1}(t) &:= \frac{1}{2}(f|_{I_{m,2n}} + f|_{I_{m,2n+1}}) \text{ on } I_{m-1,n} \\ &= \frac{1}{2}(f_0(\frac{2n}{2^m}) + f_0(\frac{2n+1}{2^m})) \text{ on } I_{m-1,n} \end{aligned} \tag{18}$$

The situation is essentially:

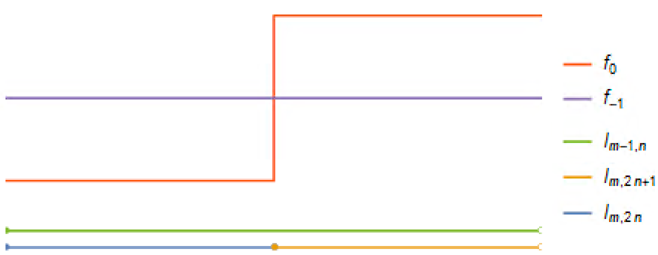



Fig. 1. Setup for our recursion

In this case it isn't too difficult to see that $f_0 = f_{-1} + e_{-1}$ where the error term $e_{-1}$ is given by

$$e_{-1} = \frac{1}{2}(f_0(\frac{2n+1}{2^m}) - f_0(\frac{2n}{2^m}))(\mathbb{1}_{I_{m,2n+1}} - \mathbb{1}_{I_{m,2n}}) \text{ on } I_{m-1,n} \tag{19}$$

Hence $e_{-1}(\frac{2n}{2^m}) = -e_{-1}(\frac{2n+1}{2^m})$. The key observation at this stage is that $\psi_{m-1,n} = -2^{(m-1)/2}$ on $I_{m,2n}$ and $\psi_{m-1,n} = 2^{(m-1)/2}$ on $I_{m,2n+1}$ (it behaves just like $e_{-1}$ up to a possible scaling). Hence on $I_{m-1,n}$ we have

$$\begin{aligned} e_{-1}(t)\psi_{m-1,n}(t) &= \begin{cases} -2^{(m-1)/2} e_{-1}(\frac{2n}{2^m}) & t \in I_{m,2n} \\ 2^{(m-1)/2} e_{-1}(\frac{2n+1}{2^m}) & t \in I_{m,2n+1} \end{cases} \\ &= \begin{cases} 2^{(m-1)/2} e_{-1}(\frac{2n+1}{2^m}) & t \in I_{m,2n} \\ 2^{(m-1)/2} e_{-1}(\frac{2n+1}{2^m}) & t \in I_{m,2n+1} \end{cases} = 2^{(m-1)/2} e_{-1}(\frac{2n+1}{2^m}) \end{aligned} \tag{20}$$

Multiplying both sides of the above by $\psi_{m,n}$ and noting that $\psi_{m,n}^2 = \mathbb{1}_{I_{m,n}}$ for any $n$ and $m$, we obtain that $e_{-1}(t) = 2^{(m-1)/2} e_{-1}(\frac{2n+1}{2^m})\psi_{m-1,n}(t)$ on $I_{m,n}$ hence

$$\begin{aligned} e_{-1}(t) &= \sum_{n=-2^{m+N-1}}^{2^{m+N-1}-1} 2^{(m-1)/2} e_{-1}(\frac{2n+1}{2^m})\psi_{m-1,n}(t) \\ \Rightarrow f = f_0 &= f_{-1} + \sum_{n=-2^{m+N-1}}^{2^{m+N-1}-1} 2^{(m-1)/2} e_{-1}(\frac{2n+1}{2^m})\psi_{m-1,n}(t) \end{aligned} \tag{21}$$

Now for the recursive step: $f_{-1}$ is just like $f_0 = f$ except that it is constant on the larger intervals $I_{m-1,n}$, and so we can write

$$f_{-1} = f_{-2} + e_{-2} = f_{-2} + \sum_{n=-2^{m+N-2}}^{2^{m+N-2}-1} 2^{(m-2)/2} e_{-2}(\frac{2n+1}{2^{m-1}})\psi_{m-2,n}(t) \tag{22}$$

Continuing this way up to $f_{-(N+m)}$ we obtain that

$$f = f_{-(m+N)} + \sum_{l=1}^{m+N} \sum_{n=-2^{m+N-l}}^{2^{m+N-l}-1} c_{m-l,n}\psi_{m-l,n} \tag{23}$$

Where $c_{m-l,n} = 2^{(m-l)/2} e_{-l}(\frac{2n+1}{2^{m+1-l}})$. The double sum term is in the form we want, so our remainder is $f_{-m-N}$. It behooves us therefore to figure out what $f_{-m-N}$ is. At each iteration, $f_{-p}$ is the average of the previous averages, hence we collect $p$ powers of $1/2$ and obtain that for $p <= N + m$

$$f_{-p} = 2^{-p} \sum_{k=0}^{2^p-1} f_0(\frac{2^p n + k}{2^m}) \text{ on } I_{m-p,n} \tag{24}$$

If $p = m + N$ we obtain that

$$f_{-m-N} = \begin{cases} 2^{-(m+N)} \sum_{k=0}^{2^{m+N}-1} f_0(\frac{-2^{m+N}+k}{2^m}) & \text{on } [-2^N, 0) \text{ (this is } n = -1) \\ 2^{-(m+N)} \sum_{k=0}^{2^{m+N}-1} f_0(\frac{k}{2^m}) & \text{on } [0, 2^N) \text{ (this is } n = 0) \end{cases} \tag{25}$$

Which tells us precisely that $f_{-m-N} = f_+\mathbb{1}_{[0,2^N)} + f_-\mathbb{1}_{[-2^N,0)}$ where $f_+$ and $f_-$ are the average values of $f$ on $[0, 2^N)$ and $[-2^N, 0)$ respectively. At this point, clearly, the norm of the remainder $||f_{-m-N}||_{L^2(\mathbb{R})}$ is not particularly small. The trick, however, is that there is no need to stop at $p = m + N$. For $p = m + N + P$, the averages will be taken over $[0, 2^{N+P}$ and $[-2^{N+P}, 0)$ respectively. In each case, however, the function is nonzero only on a fraction $2^{-p}$ of the interval where it takes the value $f_+$ or $f_-$ respectively. Hence,

$$\begin{aligned} &f_{-(m+N+P)} = 2^{-p}(f_+\mathbb{1}_{[0,2^{N+P})} + f_-\mathbb{1}_{[-2^{N+P},0)}) \\ &\Rightarrow ||f - \sum_{l=1}^{m+N+P} \sum_{n=-2^{m+N+P-l}}^{2^{m+N+P-l}-1} c_{m-l,n}\psi_{m-l,n}||^2_{L^2(\mathbb{R})} = ||f_{-(m+N+P)}||^2_{L^2(\mathbb{R})} \leq 2^{-2p} \int_{-2^{N+P}}^{2^{N+P}} |f_+|^2 + |f_-|^2 dt \\ &<= 2^{N+2-P} \max(|f_+|^2, |f_-|^2) \end{aligned} \tag{26}$$

Thus the error can be made arbitrarily small in $L^2$ norm by choosing $P$ large enough, and we are done. □

The other somewhat staggering theorem Haar proved in his thesis is as follows[5]

**Theorem 3.** ***(Haar 1909)*** *Let $f \in L^1(\mathbb{T})$ and let $H(f) = \sum_{m=0}^{\infty} \sum_{n=0}^{2^m-1} \langle f, \psi_{m,n} \rangle \psi_{m,n} + \langle f, \mathbb{1}_{\mathbb{T}} \rangle \mathbb{1}_{\mathbb{T}}$. Then*
*(a) $H(f) = f$ almost everywhere.*
*(b) If $f$ is continuous at $\gamma \in \mathbb{T}$ then $H(f)(\gamma) = f(\gamma)$.*
*(c) If $f$ is uniformly continuous on $I \subset \mathbb{T}$ then $H(f) = f$ uniformly on I (i.e. the convergence of the partial sums is uniform).*

Part $(a)$ of the above theorem is false for Fourier series $S(f)$ of functions in $L^1(\mathbb{T})$ (there exist counter examples). A corollary of part (a) is that if $f$ is continuous $H(f) = f$ a.e. For Fourier series, the analogous result is Carleson's Theorem, the celebrated positive answer to Lusin's conjecture. The proof of Carleson's Theorem is quite arduous and as such it is remarkable that an apparently stronger version of it was proved much earlier for the Haar wavelet basis. As we shall see, the Haar wavelet orthonormal basis is one example of a Walsh basis, all of which have the property that if $f$ is continuous $W_n(f) = f$ almost everywhere (Where $W_f$ is the nth Walsh series for $f$).

## II. The MRA

It is of course possible to obtain the Walsh functions without passing through the MRA (as Walsh did in 1923). One way is to use Hadamard matrices (see the Section IV), but it is convenient to use the MRA because doing so automatically provides a proof that each collection of Walsh functions is an orthonormal basis for $L^2(\mathbb{R})$. In everything that follows we rely heavily on the following definitons[6]:

**Definition 2.** ***Multi Resolution Analysis.*** *Let $\{V_j\}_{j\in\mathbb{Z}}$ be a collection of closed linear sub-spaces of $L^2(\mathbb{R})$, and $\phi \in V_0$. Definite the translation operator $\tau$ as $\tau_k\phi(t) = \phi(t-k)$. Then the pair $(\{V_j\}_{j\in\mathbb{Z}}, \phi)$ is called a multi resolution analysis (MRA) if*
*(a) $\forall j$ we have $V_j \subseteq V_{j+1}$*
*(b) $\overline{\cup V_j} = L^2(\mathbb{R})$*
*(c) $f(\cdot) \in V_j \Leftrightarrow f(2\cdot) \in V_{j+1}$*
*(d) $f \in V_0 \implies \forall k \in \mathbb{Z} \tau_k f \in V_0$*
*(e) $\{\tau_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $V_0$*

**Definition 3.** ***Conjugate Mirror Filter.*** *A function $H \in L^1(\mathbb{T})$ is called a conjugate mirror filter (CMF) if $|H(\gamma)|^2 + |H(\gamma + \frac{1}{2})|^2 = 2$ almost everywhere.*

Then we have the following

**Theorem 4.** ***(Mallat, Meyer)*** *Given an MRA, there exists a function $\psi$ (the wavelet) constructed from the MRA such that $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$. Specifically, define the family of sub-spaces $\{W_j\}_{j\in\mathbb{Z}}$ via $V_j \oplus W_j = V_{j+1}$ (meaning that $W_j = V_j^{\perp} \cap V_{j+1}$). Then the following hold*
*(a) $\exists\psi \in W_0$, constructed from $\phi$, such that $\{\tau_k\phi\}_{k\in Z}$ is an orthonormal basis for $W_0$.*
*(b) $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$.*
*(c) $\exists h_0 : \mathbb{Z} \to \mathbb{C}$ such that $\phi(t) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_0[n]\phi(2t-n) \Leftrightarrow \sqrt{2}\hat{\phi}(2\gamma) = H_0(\gamma)\hat{\phi}(\gamma)$ where $H_0(\gamma) = \sum_{n\in\mathbb{Z}} h_0[n]e^{-2\pi in\gamma}$ and $H_0$ is a CMF. Moreover, $\psi$ is defined in terms of $\phi$ via $\psi(t) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_1[n]\phi(2t-n) \Leftrightarrow \sqrt{2}\hat{\psi}(2\gamma) = H_1(\gamma)\hat{\phi}(\gamma)$ where $H_1(\gamma) = \sum_{n\in\mathbb{Z}} h_1[n]e^{-2\pi in\gamma} = e^{-2\pi i\gamma}\overline{H_0(\gamma + \frac{1}{2})}$ and $h_1[n] = (-1)^n\overline{h_0[-n+1]}$ and $H_1$ is also a CMF.*

*Proof:* Suppose $(\{V_j\}_{j\in\mathbb{Z}}, \phi)$ is an MRA. Then $\overline{\text{span}}\{\tau_k\phi\}_{k\in\mathbb{Z}} = V_0$ and $V_0 \subseteq V_1$ hence $\phi \in V_1 \Rightarrow \phi(\frac{\cdot}{2}) \in V_0$, implying the existence of coefficients $\{h_0[n]\}_{n\in\mathbb{Z}}$ such that

$$\phi(\frac{t}{2}) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_0[n]\phi(t-n) \Leftrightarrow \phi(t) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_0[n]\phi(2t-n) \tag{27}$$

In that case define $H_0(\gamma) := \sum_{n\in\mathbb{Z}} h_0[n]e^{-2\pi in\gamma}$. Then (27) is a discrete convolution with the Fourier coefficients of $H_0$, from which it follows:

$$\begin{aligned}
\hat{\phi}(\gamma) &= \int e^{-2\pi i\gamma t}\sqrt{2}\sum_{n\in\mathbb{Z}} h_0[n]\phi(2t-n)dt \\
&= \frac{1}{\sqrt{2}}\int \sum_{n\in\mathbb{Z}} e^{-2\pi i\frac{\gamma}{2}(s+n)}h_0[n]\phi(s)ds \\
&= \frac{1}{\sqrt{2}}H_0(\frac{\gamma}{2})\int e^{-2\pi i\frac{\gamma}{2}s}\phi(s)ds = \frac{1}{\sqrt{2}}H_0(\frac{\gamma}{2})\hat{\phi}(\frac{\gamma}{2})
\end{aligned} \tag{28}$$

The next claim is that $H_0$ as defined above is a CMF. We will invoke the following Lemma:

**Lemma 2.** *Let $\phi \in L^2(\mathbb{R})$ and $V = \overline{span}\{\tau_k\phi\}_{k\in\mathbb{Z}}$. Denote $\sum_{k\in\mathbb{Z}} |\hat{\phi}(\gamma+k)|^2$ as $\Phi(\gamma)$. Then $\{\tau_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $V$ if and only if $\Phi = 1$ almost everywhere.*

Note that $\Phi \in L^1(\mathbb{T})$ since $\int \Phi(\gamma)d\gamma = ||\hat{\phi}||^2_{L^2(\hat{R})} = ||\phi||_{L^2(\mathbb{R})} < \infty$. The proof of the Lemma then begins by invoking Parseval:

$$\delta_{k,0} = \int \phi(t)\overline{\phi(t-k)} = \int |\hat{\phi}(\gamma)|^2 e^{2\pi ik\gamma}d\gamma = \sum_{n\in\mathbb{Z}} \int_{\mathbb{T}} |\hat{\phi}(\gamma+n)|^2 e^{2\pi ik\gamma}d\gamma \tag{29}$$

We then interchange the sum and the integral via the Lebesgue dominated convergence theorem ($\Phi \in L^1(\mathbb{R})$) and obtain

$$\delta_{k,0} = \int_{\mathbb{T}} \Phi(\gamma)e^{2\pi ik\gamma}d\gamma \tag{30}$$

Obviously this relation holds if $\Phi = 1$ almost everywhere, but moreover if this relation holds then every fourier coefficient of $\Phi$ is zero except for the constant term, which can only be one, hence $\Phi = 1$ almost everywhere. This concludes the proof of the Lemma.

Since $\{\tau_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $V_0$ we conclude that $\sum_{k\in\mathbb{Z}} |\hat{\phi}(\gamma+k)|^2 = 1$ almost everywhere. We also know that $\hat{\phi}(\gamma+n) = \frac{1}{\sqrt{2}}H_0(\frac{\gamma}{2}+\frac{n}{2})\hat{\phi}(\frac{\gamma}{2}+\frac{n}{2})$. Hence,

$$2 = \sum_{n\in\mathbb{Z}} |H_0(\frac{\gamma}{2}+\frac{n}{2})|^2|\hat{\phi}(\frac{\gamma}{2}+\frac{n}{2})|^2 \tag{31}$$

Decompose the sum into a sum over the odd integers and another over the even integers, and let $\lambda = \frac{\gamma}{2}$ to obtain

$$2 = |H_0(\lambda)|^2\Phi(\lambda) + |H_0(\lambda+\frac{1}{2})|^2\Phi(\lambda) = |H_0(\lambda)|^2 + |H_0(\lambda+\frac{1}{2})|^2 \tag{32}$$

Thus $H_0$ is a CMF. With this in mind, we will invoke the following Lemma:

**Lemma 3.** ***CMF Lemma*** *Suppose $H$ is a CMF and $F \in L^1(\mathbb{T})$, then $F(\gamma)\overline{H(\gamma)} = -F(\gamma+\frac{1}{2})\overline{H(\gamma+\frac{1}{2})}$ if and only if $\exists K_F \in L^1(\mathbb{T})$ such that $F(\gamma) = K_F(\gamma)\overline{H(\gamma+\frac{1}{2})}$ and $K_F(\gamma+\frac{1}{2}) = -K_F(\gamma)$*

Let's prove the if direction first. Assume that there exists such a $K_F$. Then

$$\begin{aligned} F(\gamma)\overline{H(\gamma)} &= K_F(\gamma)\overline{H(\gamma+\frac{1}{2})}\,\overline{H(\gamma)} \\ &= -K_F(\gamma+\frac{1}{2})\overline{H(\gamma+\frac{1}{2}+\frac{1}{2})H(\gamma+\frac{1}{2})} \\ &= -F(\gamma+\frac{1}{2})\overline{H(\gamma+\frac{1}{2})} \end{aligned} \tag{33}$$

Note that we didn't use the fact that $H$ is a CMF in this direction. Now for the only if direction. Assume $F(\gamma)\overline{H(\gamma)} = -F(\gamma+\frac{1}{2})\overline{H(\gamma+\frac{1}{2})}$. Then by the fact that $H$ is a CMF we obtain

$$\begin{aligned} 2F(\gamma) &= F(\gamma)\Big(\overline{H(\gamma)}H(\gamma) + \overline{H(\gamma+\frac{1}{2})}H(\gamma+\frac{1}{2})\Big) \\ &= -F(\gamma+\frac{1}{2})\overline{H(\gamma+\frac{1}{2})}H(\gamma) + F(\gamma)\overline{H(\gamma+\frac{1}{2})}H(\gamma+\frac{1}{2}) \\ &= (-F(\gamma+\frac{1}{2})H(\gamma) + F(\gamma)H(\gamma+\frac{1}{2}))\overline{H(\gamma+\frac{1}{2})} \end{aligned} \tag{34}$$

Therefore define $K_F(\gamma) = \frac{1}{2}(F(\gamma)H(\gamma+\frac{1}{2}) - F(\gamma+\frac{1}{2})H(\gamma))$ such that $F(\gamma) = K_F(\gamma)\overline{H(\gamma+\frac{1}{2})}$. Then $K_F \in L^1(\mathbb{T})$ since $H \in L^\infty(\mathbb{T})$ and moreover $K_F(\gamma+\frac{1}{2}) = \frac{1}{2}(F(\gamma+\frac{1}{2}H(\gamma) - F(\gamma)H(\gamma+\frac{1}{2})) = -K_F(\gamma)$. This concludes the proof of the CMF Lemma. Using these preliminaries we would now like to construct a basis for $W_0$. We begin by characterizing an arbitrary function $f \in V_1$. $f \in V_1 \Rightarrow f(\frac{\cdot}{2}) \in V_0 \Rightarrow f(\frac{t}{2}) = \sum_{n\in\mathbb{Z}}\langle f(\frac{\cdot}{2}), \phi(\cdot - n)\rangle\phi(t-n)$ since $\{\tau_k\phi\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $V_0$. Changing variable in both the inner product integration and the variable t,

$$\begin{aligned} f(t) &= 2\sum_{n\in\mathbb{Z}}\langle f, \phi(2\cdot -n)\rangle\phi(2t-n) \\ \Leftrightarrow \hat{f}(\gamma) &= \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})\sum_{n\in\mathbb{Z}}\langle f, \sqrt{2}\phi(2\cdot -n)\rangle e^{-2\pi in\frac{\gamma}{2}} \end{aligned} \tag{35}$$

Thus $f \in V_1 \Leftrightarrow \hat{f}(\gamma) = \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})F_f(\frac{\gamma}{2})$ where

$$F_f(\gamma) = \sum_{n\in\mathbb{Z}} \langle f, \sqrt{2}\phi(2\cdot -n)\rangle e^{-2\pi i n\gamma} \tag{36}$$

Note that the fact that $\{\phi(2\cdot -n)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for $V_1$ implies that $\{\langle f, \sqrt{2}\phi(2\cdot -n)\rangle\}_{n\in\mathbb{Z}} \in l^2(\mathbb{Z})$ so that in fact $F_f \in L^2(\mathbb{T})$. Now that we have a Fourier characterization of functions in $V_1$ we can use the fact that $f \in W_0 \Leftrightarrow f \in V_1$ and $f \perp V_0$. We use Parseval to obtain a criterion for $f \perp V_0$ as follows:

$$\begin{aligned}
&\int f(t)\overline{\phi(t-n)}dt = 0 \qquad \forall n \\
\Leftrightarrow &\int \hat{f}(\gamma)\overline{\hat{\phi}(\gamma)}e^{2\pi i n\gamma}d\gamma = 0 \qquad \forall n \\
\Leftrightarrow &\sum_{k\in\mathbb{Z}}\int_0^1 \hat{f}(\gamma+k)\overline{\hat{\phi}(\gamma+k)}e^{w\pi i n\gamma}d\gamma = 0 \qquad \forall n \\
\Leftrightarrow &\int_0^1 G_f(\gamma)e^{2\pi i n\gamma}d\gamma = 0 \qquad \forall n
\end{aligned} \tag{37}$$

Where $G_f(\gamma) = \sum_{k\in\mathbb{Z}} \hat{f}(\gamma+k)\overline{\hat{\phi}(\gamma+k)}$ (Lebesgue dominated convergence justifies the interchange of sum and integral here). Then $G_f \in L^1(\mathbb{T})$ since $||G_f||_{L^1(\mathbb{T})} \leq ||f||_{L^2(\mathbb{R})}||\phi||_{L^2(\mathbb{R})}$. Therefore $G_f$ is an $L^1(\mathbb{T})$ function with identically zero Fourier coefficients, implying that $G_f = 0$ almost everywhere. Thus our criterion for $f \in W_0$ is $\hat{f}(\gamma) = \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})F_f(\frac{\gamma}{2})$ and $G_f(\gamma) = 0$ (where all equalities are in $L^1$ sense). Now we also know that $\hat{\phi}(\gamma) = \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})H_0(\frac{\gamma}{2})$. Plugging that and the $F_f$ criterion into the $G_f = 0$ criterion we obtain

$$\begin{aligned}
&0 = \sum_{k\in\mathbb{Z}} \hat{f}(\gamma+k)\overline{\hat{\phi}(\gamma+k)} \\
\Leftrightarrow\ &0 = \frac{1}{2}\sum_{k\in\mathbb{Z}} F_f(\frac{\gamma+k}{2})|\hat{\phi}(\frac{\gamma+k}{2})|^2\overline{H_0(\frac{\gamma+k}{2})}
\end{aligned} \tag{38}$$

Let $\lambda = \frac{\gamma}{2}$ and split the above sum into a sum of even terms and a sum of odd terms as before, and we obtain

$$0 = F_f(\lambda)\overline{H_0(\lambda)}\sum_{m\in\mathbb{Z}}|\hat{\phi}(\lambda+m)|^2 + F_f(\lambda+\frac{1}{2})\overline{H_0(\lambda+\frac{1}{2})}\sum_{m\in\mathbb{Z}}|\hat{\phi}(\lambda+\frac{1}{2}+m)|^2 \tag{39}$$

By Lemma 2 both sums in the above are 1 almost everywhere. Hence $0 = F_f(\lambda)\overline{H_0(\lambda)} + F_f(\lambda+\frac{1}{2})\overline{H_0(\lambda+\frac{1}{2})}$. By the CMF Lemma therefore $\exists K_{F_f} \in L^2(\mathbb{T})$ such that $F_f(\gamma) = K_{F_f}(\gamma)\overline{H_0(\gamma+\frac{1}{2})}$ and $K_{F_f}(\gamma) = -K_{F_f}(\gamma+\frac{1}{2})$. Finally, the $F_f$ criterion tells us that

$$f \in W_1 \Leftrightarrow \hat{f}(\gamma) = \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})K_{F_f}(\frac{\gamma}{2})\overline{H_0(\frac{\gamma}{2}+\frac{1}{2})} \tag{40}$$

Where $K_{F_f}(\gamma) = -K_{F_f}(\gamma+\frac{1}{2})$. To make this last condition slightly cleaner define $L_f(\gamma) = e^{\pi i\gamma}K_{F_f}(\frac{\gamma}{2})$. Then $K_{F_f}(\gamma) = -K_{F_f}(\gamma+\frac{1}{2})$ for all $\gamma$ if and only if $L_f$ is 1 periodic:

$$L(\gamma+1) = L(\gamma) \qquad \forall\gamma \tag{41}$$

$$\Leftrightarrow -e^{\pi i\gamma}K(\frac{\gamma}{2}+\frac{1}{2}) = e^{\pi i\gamma}K(\frac{\gamma}{2}) \qquad \forall\gamma \tag{42}$$

$$\Leftrightarrow K(\lambda)+K(\lambda+\frac{1}{2}) = 0 \qquad \forall\lambda = \frac{\gamma}{2} \tag{43}$$

Therefore we have

$$f \in W_1 \Leftrightarrow \hat{f}(\gamma) = \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})e^{-\pi i\gamma}L_f(\gamma)\overline{H_0(\frac{\gamma}{2}+\frac{1}{2})} \text{ for some } L_f \in L^2(\mathbb{T}) \tag{44}$$

We define the wavelet $\psi$ on the frequency domain by taking $L_f = 1$ so that

$$\hat{\psi}(\gamma) := \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})e^{-\pi i\gamma}\overline{H_0(\frac{\gamma}{2}+\frac{1}{2})} := \frac{1}{\sqrt{2}}\hat{\phi}(\frac{\gamma}{2})H_1(\frac{\gamma}{2}) \tag{45}$$

This defines $H_1(\gamma) = e^{-2\pi i\gamma}\overline{H_0(\gamma + \frac{1}{2})}$ and therefore $h_1[n] = (-1)^n h_0[-n+1]$. We now have that an arbitrary function $f \in W_0$ satisfies $\hat{f}(\gamma) = L_f(\gamma)\hat{\psi}(\gamma)$ for some $L_f \in L^2(\mathbb{T})$. We can take the Fourier transform of this condition of this condition and obtain that any $f$ in $W_0$ has a decomposition

$$f = \sum_{k\in\mathbb{Z}} \hat{L_f}[k]\psi(t-k) \tag{46}$$

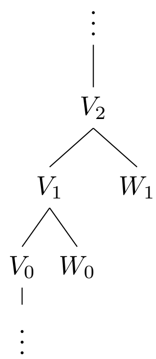
Thus indeed, $W_0 = \overline{\text{span}}\{\tau_k\psi\}_{k\in Z}$. It remains to show that $\langle \tau_k\psi, \tau_j\psi\rangle = \delta_{kj}$, after which we will have proved (by Theorem 2) that $\{\tau_k\psi\}_{k\in Z}$ is an orthonormal basis for $W_0$. By Lemma 2, it suffices to show that $\Psi(\gamma) := \sum_{k\in\mathbb{Z}} |\hat{\psi}(\gamma+k)|^2 = 1$ almost everywhere:

$$\Psi(\gamma) = \frac{1}{2}\sum_{k\in Z} |\hat{\phi}(\frac{\gamma}{2} + \frac{k}{2})H_0(\frac{\gamma+1}{2} + \frac{k}{2})|^2 \tag{47}$$

Split the sum into a sum over the even integers and another over the odd integers as usual and we find that

$$\Psi(\gamma) = \frac{1}{2}(|H_0(\frac{\gamma}{2} + \frac{1}{2})|^2)\Phi(\gamma) + |H_0(\frac{\gamma}{2})|^2\Phi(\gamma)) = 1 \tag{48}$$

Where in the last equality we used the fact that $\Phi = 1$ almost everywhere and that $H_0$ is a CMF. Thus, we have proved that $\{\tau_k\psi\}_{k\in Z}$ is an orthonormal basis for $W_0$. What we have in fact done is hop over a single leaf from the leftmost branch of the wavelet packet tree, and obtained a basis for the space that we found there ($W_0$). The situation is essentially as follows




Where the siblings form an orthogonal decomposition of the parent. This picture will be crucial in Section III. We will now set out to prove that $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}} = \{\tau_n D_m\psi\}_{m,n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$ where $D_m\psi = 2^{-m/2}\psi(2^m t)$. First note that $\{\psi_{n,m}\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $W_m$ since

$$\begin{aligned} g \in W_m &\Leftrightarrow g \perp f \forall f \in V_m \\ &\Leftrightarrow \langle g, \phi(2^m \cdot - n)\rangle = 0 \qquad \forall n \\ &\Leftrightarrow \langle g(2^m \cdot), \phi(\cdot - n)\rangle = 0 \qquad \forall n \\ &\Leftrightarrow g(2^{-m}\cdot) \in W_0 \Leftrightarrow \exists \{c_n\}_{n\in\mathbb{Z}} \text{ st. } g(2^{-m}t) = \sum_{n\in\mathbb{Z}} c_n\psi(t-n) \\ &\Leftrightarrow g = \sum_{n\in\mathbb{Z}} \tilde{c_n}\psi_{m,n} \end{aligned} \tag{49}$$

Orthogonality follows from the fact that $D_m$ is an isometry. Not further that $W_k \perp W_j$ if $j \neq k$ since either $W_k \subset V_j$ or $W_j \subset V_k$ and $V_j \perp W_j$ and $V_k \perp W_k$. Thus $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$ is an orthonormal subset of $L^2(\mathbb{R})$. By Theorem 2 all that remains in the proof of the MRA Theorem is to show that $\overline{\text{span}}\{\psi_{m,n}\}_{m,n\in\mathbb{Z}} = L^2(\mathbb{R})$, or actually just that $\overline{\bigoplus_{m\in Z} W_m} = L^2(\mathbb{R})$. Note that this direct sum is well defined since if $g_m, f_m \in W_m$ and $\sum_{m\in Z} f_m = \sum_{m\in Z} f_m$ in the sense of $L^2$ then

$$\begin{aligned} &\sum_{j\neq m} g_j - f_j = f_m - g_m \\ &\Rightarrow ||f_m - g_m||^2_{L^2(\mathbb{R})} = \sum_{j\neq m} \langle g_j - f_j, f_m - g_m\rangle = 0 \end{aligned} \tag{50}$$

Hence $g_m = f_m$ for each $m$ and the decomposition is uniquely defined. We shall prove that $\overline{\bigoplus_{m=-\infty}^{j} W_m} = V_{j+1}$, from which the result follows immediately since $V_j \subseteq V_{j+1}$ and $\overline{\bigcup_{j\in Z} V_j} = L^2(\mathbb{R})$ (these conditions come straight from the definition of the MRA). One direction is easy; $\overline{\bigoplus_{m=-\infty}^{j} W_m} \subseteq V_{j+1}$ since each $W_m \subseteq V_{j+1}$ for $m \leq j$ and $V_{j+1}$ is closed. To show the reverse inclusion, consider an arbitrary $f \in V_{j+1}$. Then there exists a unique decomposition $f = f_j + g_j$ where $f_j \in V_j$ and

$g_j \in W_j$. Continuing inductively, we may write $f$ uniquely as $g_j + g_{j-1} + \ldots g_{j-J} + f_{j-J}$ where $g_l \in W_l$ for each $l$ and $f_{j-J} \in V_{j-J} = \bigcap_{l=j-J}^{J} V_l$. Define $g = \sum_{j=-\infty}^{J} g_j$. Then $g \in L^2(\mathbb{R})$ since by Pythagorus since for every $m < J$

$$||f||^2_{L^2(\mathbb{R})} = ||f_m||^2_{L^2(\mathbb{R})} + \sum_{j=m}^{J} ||g_j||^2_{L^2(\mathbb{R})} \geq \sum_{j=m}^{J} ||g_j||^2_{L^2(\mathbb{R})} \tag{51}$$

Moreover $f - g \in V_m$ for every $m \leq J$ implies that $f - g \in \bigcap_{m \leq J} V_m = \{0\} \implies f = g$ in $L^2$ sense, and $g$ is manifestly an element of $\overline{\bigoplus_{m=-\infty}^{j} W_m}$ so we're done. $\square$
Note that if we take $\phi = \mathbb{1}_{[0,1)}$ and define $V_j$ recursively via the definition of the MRA, then $(\{V_j\}_{j\in\mathbb{Z}}, \phi)$ is an MRA. Moreover $\phi(2t) = \mathbb{1}_{[0,\frac{1}{2})}$ and $\phi(2t-1) = \mathbb{1}_{[\frac{1}{2},1)}$ hence $\phi(t) = \sqrt{2}(\frac{1}{\sqrt{2}}\phi(2t) + \frac{1}{\sqrt{2}}\phi(2t-1)$. Thus $h_0[0] = h_0[1] = \frac{1}{\sqrt{2}}$ and all other $h_0[n]$ are zero. Hence $h_1[0] = \frac{1}{\sqrt{2}} = -h_1[1]$, thus $\psi(t) = \sqrt{2}(\frac{1}{\sqrt{2}}\phi(2t) - \frac{1}{\sqrt{2}}\phi(2t-1) = \mathbb{1}_{[0,\frac{1}{2})} - \mathbb{1}_{[\frac{1}{2},1)}$, which is just $-1$ times the Haar function. This provides an alternate proof that the dilations and translations of the Haar function are an orthonormal basis for $L^2(\mathbb{R})$.

## III. Filling out the tree

We saw earlier in the proof of the MRA theorem that the Fourier multipliers $H_0$ and $H_1$ allowed us to decompose a given function in $V_j$ into two functions, lying in $V_{j-1}$ and $W_{j-1}$ respectively. Continuing recursively, starting from $\phi \in V_0$ we could obtain an orthonormal basis for $V_{-p}$ and $W_{-}p$ via

$$\begin{aligned} \hat{\phi}(\gamma) = 2^{-\frac{p}{2}} H_0(\frac{\gamma}{2}) \cdots H_0(\frac{\gamma}{2^p}) \hat{\phi}(\frac{\gamma}{2^p}) &\Leftrightarrow \hat{\phi}(2^p\gamma) = 2^{-\frac{p}{2}} H_0(2^{p-1}\gamma) \cdots H_0(\gamma)\hat{\phi}(\gamma) \\ \hat{\psi}(\gamma) = 2^{-\frac{p}{2}} H_0(\frac{\gamma}{2}) \cdots H_0(\frac{\gamma}{2^{p-1}}) H_1(\frac{\gamma}{2^p}) \hat{\phi}(\frac{\gamma}{2^p}) &\Leftrightarrow \hat{\psi}(2^p\gamma) = 2^{-\frac{p}{2}} H_0(2^{p-1}\gamma) \cdots H_0(\frac{\gamma}{2}) H_1(\gamma)\hat{\phi}(\gamma) \end{aligned} \tag{52}$$

The integer translates of the Fourier transforms of the functions on the right give orthonormal bases for $V_{-p}$ and $W_{-p}$ respectively. Squint slightly and you might spot the binary expansions $0 \cdots 0$ and $0 \cdots 01$ above, and ask the natural question as to what the other $2^{p-1} - 2$ combinations of the Fourier multipliers $H_0$ and $H_1$ produce. This is essentially the motivation for wavelet packets: The MRA corresponds to the left most branch of the tree generated by the dyadically dilated Fourier multipliers $H_0$ and $H_1$ (choosing either $H_0$ or $H_1$ at each level of dilation). Specifically, given $\phi \in L^2(\mathbb{R})$ that generates an MRA, define $\Theta_{\epsilon_1 \cdots \epsilon_p}$ for $\epsilon_i \in \{0, 1\}$ on the frequency domain via

$$\hat{\Theta}_{\epsilon_1 \cdots \epsilon_p}(\gamma) := 2^{-\frac{p}{2}} H_{\epsilon_1}(2^{p-1}\gamma) \cdots H_{\epsilon_p}(\gamma)\hat{\phi}(\gamma) \tag{53}$$

Then define the closed sub spaces of $L^2(\mathbb{R})$ $X_{\epsilon_1 \cdots \epsilon_p} := \overline{\text{span}}\{\tau_k \Theta_{\epsilon_1 \cdots \epsilon_p}\}_{k\in\mathbb{Z}}$. This constitutes an extension of the MRA "tree" in the following sense: If we position each of the spaces $X_{\epsilon_1 \cdots \epsilon_p}$ on a tree such that $\epsilon_i = 0$ means that the space is on the left branch of its parent and $\epsilon_i = 1$ means that the space is on the right branch of its parent (ie the sequence $\epsilon_1, \ldots, \epsilon_p$ tells you what path to take from the root to a given space) then the tree looks like:

$$X = V_0$$

$$X_0 = V_{-1} \qquad X_1 = W_{-1}$$

$$X_{00} = V_{-2} \quad X_{01} = W_{-2} \quad X_{10} \quad X_{11}$$

$$\vdots$$

Moreover, the definitions of $X_{\vec{\epsilon}}$ ensure that the siblings still form an orthogonal decomposition of their parent, meaning that any $W_j$ in the decomposition $L^2(\mathbb{R}) = \overline{\bigoplus_{j\in Z} W_j}$ can be replaced by the direct sum of its child nodes. Moreover, each child node can be replaced in a similar manner. Supposing we were only allowed to make $N$ such moves on a single $W_j$, we would still obtain $2^N$ different orthogonal bases for $L^2(\mathbb{R})$! This is the advantage of wavelet packets; it was much harder to get the *first* wavelet decomposition of $L^2(\mathbb{R})$ than it was to obtain the next exponentially many. I should note that the labelling given to the spaces $X_{\vec{\epsilon}}$ need not be quite so complicated; instead of describing the path from the root to a given space we could describe the level of the space and how far to the right it is in that level. Namely if $X^{p,n}$ is the space at level $p$ in the tree and index $n$ from the left side of the tree, then $X^{p,n} = X_{\vec{\epsilon}}$ where $\epsilon_1 \cdots \epsilon_p$ is precisely the binary expansion of $n$ with $p$ digits: $n = \sum_{j=0}^{p-1} \epsilon_j 2^j$. For reasons that are unclear to me, the convention in the literature appears to be to label the spaces via the bit reversal of $\vec{\epsilon}$ rather than $\vec{\epsilon}$; namely define $X^p_m = X_{\vec{\epsilon}}$ where $m = \sum_{j=0}^{p} \alpha_j 2^j$ and $\vec{\alpha} = \text{rev}_p(\vec{\epsilon})$. Obviously, a given naming convention for the spaces $X$ corresponds to the same naming convention for the functions $\Theta$ so that $X^{p,n} = \overline{\text{span}}\{\tau_k \Theta^{p,n}\}_{k\in\mathbb{Z}}$ and $X^p_m = \overline{\text{span}}\{\tau_k \Theta^p_m\}_{k\in\mathbb{Z}}$.

## IV. The Walsh Functions

The Walsh functions are defined immediately from the Haar MRA ($\phi = \mathbb{1}_{[0,1)}$) as

$$W_n(t) := \Theta^{\lfloor \log_2(n+1) \rfloor}_{n-\lfloor \log_2(n+1) \rfloor}(t) \tag{54}$$

so that $W_0 = \phi, W_1 = -\psi$ and $W_n$ satisfy the pair of recursion relations

$$\begin{aligned} W_{2n}(x) &= W_n(2x) + W_n(2x-1) \\ W_{2n+1}(x) &= W_n(2x) - W_n(2x-1) \end{aligned} \tag{55}$$

For clarity, I've plotted $W_0$ up to $W_{20}$ below.

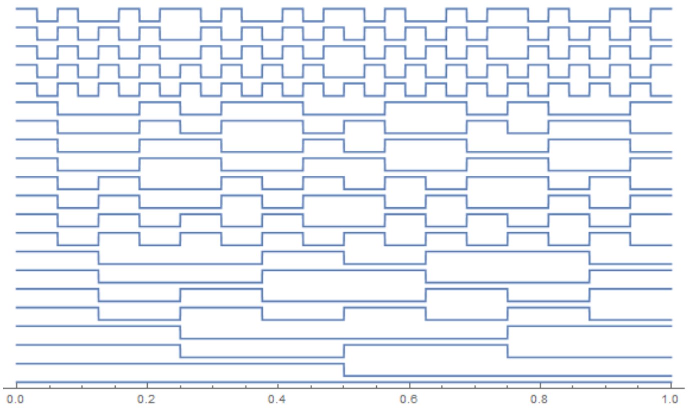



Fig. 2. A plot of $3n + W_n(x)$ from $n = 0$ to $n = 20$.

One interesting fact about Walsh Functions is that $\{W_n\}_{n=0}^{\infty}$ is closed under pointwise multiplication for each $p$ (in fact $\{W_n\}_{n=0}^{2^p-1}$ is closed under pointwise multiplication for each $p$). This is difficult to see from the wavelet packet perspective, but it is clearer in an alternative formulation in terms of representation theory.[3] Namely, define the Cantor Group $G$ as $\prod_{j=1}^{\infty}(\mathbb{Z}/2\mathbb{Z})$ endowed with the product topology (the topology on each copy of $\mathbb{Z}/2\mathbb{Z}$ is discrete). By Tychonoff's theorem the Cantor group is compact. Recall that a representation is a homomorphism from the group in question to the group of matrices $M(n)$, and that the character of a representation is its trace. In this case since $G$ is abelian and each of its elements has order 2 we must have $n = 1$ and characters functions $\chi_\alpha : G \to M(1)$ taking only the values $+1$ and $-1$. Following [3] we note that one obvious set of characters is the set of functions $\chi_n$ that are given by $\chi_n(\{t_1, \ldots\}) = 1$ if $t_n = 1$ and $\chi_n(\{t_1, \ldots\}) = -1$ if $t_n = 0$. The prod $\chi_{n_1} \cdots \chi_{n_j}$ of a finite number of characters is of course also a character of $G$. We claim that in fact any character of $G$ may be written this way. Namely, if $\chi$ is an arbitrary character of $G$ then $\chi(\{t_1, t_2, t_3, \ldots\}) = \chi(\{t_1, 0, \ldots\} + \{0, t_2, 0, \ldots\} + \{0, 0, t_3, 0, \ldots\} + \cdots) = \chi(\delta_1)\chi(\delta_2)\chi(\delta_3)\cdots$ where $\delta_k$ is 0 except at entry $k$ where it is $t_k$. Moreover the sequence $\delta_k$ converges to the identity in the topology on $G$ and $\chi$ is continuous, hence there exists $M$ such that $\chi(\delta_m) = 1 \forall m > M$. Thus $\chi(\{t_1, \ldots\}) = \chi(\delta_1)\cdots\chi(\delta_M) = \chi_1(\{t_1, \ldots\})\cdots\chi_M(\{t_1, \ldots\})$. The last step is to consider, for $x \in [0, 1]$, the binary expansion $x = \sum_{j=0}^{\infty} s_j 2^{-j}$ (we may safely the ignore the ambiguity introduced here since what follows works just as well for finite approximations). Then define $r_n(x) := \chi_n(\{s_1, \ldots\})$. These are the Rademacher functions. Then define the correspondence between the Walsh functions and the full set of characters of $G$ as the closure under pointwise multiplication of the Rademacher functions (it is not difficult to convince oneself that this correspondence holds, albeit the ordering is tricky). At this point it is clear that the Walsh functions are closed under pointwise multiplication since they are exactly the characters of $\prod_{j=1}^{\infty}(\mathbb{Z}/2\mathbb{Z})$.

Yet another way to describe the Walsh functions is in terms of Hadamard matrices:

**Definition 4.** *A matrix $H$ is called Hadamard if it has entries either $+1$ or $-1$ such that the $c_i \cdot c_j = \delta_{ij} \forall i \neq j$ and $r_i \cdot r_j = \delta_{ij} \forall i \neq j$ where $c_i$ is the ith column vector of $H$ and $r_j$ is the jth row vector of $H$.*

For example the matrix

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{56}$$

is Hadamard. Moreover, if we define $H_{2^{n+1}} := H_{2^n} \otimes H_{2^n}$ (where $\otimes$ is the Kronecker product) then each $H_{2^n}$ is a $2^n \times 2^n$ Hadamard matrix. Moreover, it isn't difficult to see that the first $2^p$ Walsh functions are given by the rows of $H_{2^p}$ in bit reversal order (in the sense that the sequence of values in row $\text{rev}_p(r)$ is exactly the sequence of values taken by $W_r$ as one increases from $x = 0$ to $x = 1$ on intervals of length $2^{-p}$).

Finally we have the following notable Carleson type theorem, proved by Gosselin in the extremely general setting of Vilenkin Fourier series: [4][7]

**Theorem 5.** ***(Gosselin)*** *If $f \in C([0,1])$ then the Walsh series $\sum_{n=0}^{\infty} \langle f, W_n \rangle W_n(x)$ converges to $f(x)$ at almost every $x \in [0,1]$.*

## References


[1] Kenneth George Beauchamp and CK Yuen. Walsh functions and their applications. *IEEE Transactions on Systems, Man, and Cybernetics*, 1(11):794–795, 1976.

[2] Ronald R Coifman, Yves Meyer, Steven Quake, and M Victor Wickerhauser. Signal processing and compression with wavelet packets. In *Wavelets and their applications*, pages 363–379. Springer, 1994.

[3] Nathan Jacob Fine. On the walsh functions. *Transactions of the American Mathematical Society*, 65(3):372–414, 1949.

[4] John Gosselin. Almost everywhere convergence of vilenkin-fourier series. *Transactions of the American Mathematical Society*, 185:345–370, 1973.

[5] A Haar and G Zimmermann. On the theory of orthogonal function systems. In C Heil and D Walnut, editors, *Fundamental Papers in Wavelet Theory*, pages 155–188. Princeton University Press, 01 2009.

[6] Stephane G Mallat. Multiresolution approximations and wavelet orthonormal bases of l2). *Transactions of the American mathematical society*, 315(1):69–87, 1989.

[7] Mladen Victor Wickerhauser. Basis and convergence properties of wavelet packets. *Proceedings of the Conference on wavelet analysis and applications*, 2002.