1
Introduction

Prerequisites

See our suggestions in the Preface and in the following tutorial. What if, you say, I don't have the prerequisites for reading the prerequisites? Start the book anyway! You will pick them up along the way.

Overture: Why wavelets?

The first wavelet was discovered by Alfred Haar long ago, but its use was limited since it was based on step-functions, and the step-functions jump from one step to the next. The implementation of Haar's wavelet in the approximation problem for continuous functions was therefore rather bad, and for differentiable functions it is atrocious, and so Haar's method was forgotten for many years. And yet it had in it the one idea which proved so powerful in the recent rebirth (since the 1980's) of wavelet analysis: the idea of a *multiresolution*. You see it in its simplest form by noticing that a box function B of (1.1) may be scaled down by a half such that two copies B' and B'' of the smaller box then fit precisely inside B. See (1.1).

$$(1.1)$$

$$(1.2)$$
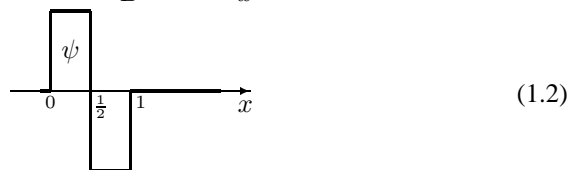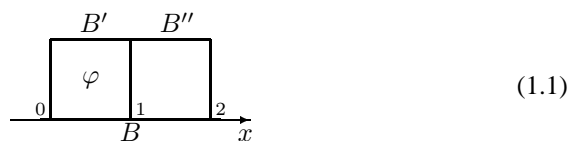
This process may be continued if you scale by powers of 2 in both directions, i.e., by 2^k for integral k, $-\infty < k < \infty$. So for every $k \in \mathbb{Z}$, there is a finer resolution, and if you take an up- and a shifted mirror image down-version of the dyadic scaling as in (1.2), and allow all linear combinations, you will notice that arbitrary functions f on the line $-\infty < x < \infty$, with reasonable integrability properties, admit a representation

$$f\left(x\right) = \sum_{k,n} c_{k,n} \psi\left(2^k x - n\right), \tag{1.3}$$

where the summation is over all pairs of integers $k, n \in \mathbb{Z}$, with k representing scaling and n translation. The very simple idea of turning this construction into a multiresolution ("multi" for the variety of scales in (1.3)) leads not only to an algorithm for the analysis/synthesis problem,

$$f\left(x\right) \longleftrightarrow c_{k,n}, \tag{1.4}$$

in (1.3), but also to a construction of the single functions ψ which solve the problem in (1.3), and which can be chosen differentiable, and yet with support contained in a fixed finite interval. These two features, the algorithm and the finite support (called *compact* support), are crucial for computations: Computers do algorithms, but they do not do infinite intervals well. Computers do summations and algebra well, but they do not do integrals and differential equations, unless the calculus problems are discretized and turned into algorithms.

The prerequisites for our present book include some familiarity with function spaces and with rudimentary ideas from harmonic analysis. And, of course, if you have had experience with Hilbert space, or integration theory, then this is a help; but you will be able to pick up what you need along the way. You will notice that the multiresolution analysis viewpoint is dominant, which increases the role of algorithms; for example, the so-called pyramid algorithm for analyzing signals, or shapes, using wavelets, is an outgrowth of multiresolutions.

One way to quickly brush up on basic ideas that you will meet along the way through the chapters of our book is to consult one or both of the two books [Hub98] and [KaLe95]. The first is a wonderful summary of the fundamental ideas behind the wave of wavelet analysis which began in the mid-1980's, and it is written with a minimal use of mathematical formulas and a maximal use of good writing. You will also learn of the history, and you will meet some of the researchers who gave the early push to the subject. The second book [KaLe95] has somewhat of the same flavor, but it is written for mathematicians. It has formulas, and in addition a lot of excellent writing. Like [Hub98], it stresses the intuitive ideas behind the formulas. Actually, it is two books: the first one (primarily by Kahane) is classical Fourier analysis, and the second one (primarily by P.-G. Lemarié-Rieusset) is the wavelet book. It will help you, among other things, to get a better feel for the French connection, the Belgian connection, and the diverse and early impulses from applications in the subject. Enjoy!

We mention two fast and friendly guides through the basic ideas of harmonic analysis terminating in multiresolution analysis: Chapter 11 in [McWe99], and

Chapter 15 in [DaDo02]. Then there are two popular articles in the *Notices of the American Mathematical Society* [Bri95, Wal97] which serve as entries to the subject. Both are written for students and the general (mathematical) public. C.M. Brislawn explains in [Bri95] how the idea of the two-channel multirate filter bank and the wavelet packets have now turned into a commercial algorithm which is used by the F.B.I. in digitizing, compressing, and storing fingerprints. The second *Notices* article, by J.S. Walker, introduces in [Wal97] several discrete wavelet algorithms to the public, and compares the classical Fourier approach to the analysis/synthesis problem with the one based on wavelets and multirate analysis. Both *Notices* articles include lots of examples and motivation which are ideally appropriate for the viewpoint in our book: the Daubechies method, the vanishing moments, the multiresolutions and their more recent variations and generalizations.

Returning to (1.1) and (1.2) (see also (1.13)), we see that the scaling function φ itself may be expanded in the wavelet basis which is defined from ψ, and we arrive at the infinite series

$$\varphi(x) = \sum_{k=1}^{\infty} 2^{-k} \psi\left(2^{-k} x\right) \qquad (1.5)$$

which is pointwise convergent for $x \in \mathbb{R}$. (It is a special case of the expansion (1.3) when $f = \varphi$.) In view of the picture ($\diamondsuit$) below, (1.5) gives an alternative meaning to the traditional concept of a *telescoping* infinite sum. If, for example, $0 < x < 1$, then the representation (1.5) yields $\varphi(x) = 1 = \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots$, while for $1 < x < 2$, $\varphi(x) = 0 = -\frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \cdots$. More generally, if $n \in \mathbb{N}$, and $2^{n-1} < x < 2^n$, then

$$\varphi(x) = 0 = -\left(\frac{1}{2}\right)^n + \sum_{k>n} \left(\frac{1}{2}\right)^k.$$

So the function φ is itself in the space $\mathcal{V}_0 \subset L^2(\mathbb{R})$, and φ represents the *initial resolution*. The tail terms in (1.5) corresponding to

$$\sum_{k>n} 2^{-k} \psi\left(2^{-k} x\right) = \frac{1}{2^n} \varphi\left(\frac{x}{2^n}\right) \qquad (1.6)$$

represent the *coarser resolution*. The finite sum

$$\sum_{k=1}^{n} 2^{-k} \psi\left(2^{-k} x\right)$$

represents the *missing detail* of φ as a "bump signal". While the sum on the left-hand side in (1.6) is *infinite*, i.e., the summation index k is in the range $n < k < \infty$, the expression $2^{-n} \varphi(2^{-n} x)$ on the right-hand side is merely a coarser scaled version of the original function φ from the subspace $\mathcal{V} \subset L^2(\mathbb{R})$ which specifies the initial resolution. Infinite sums are *analysis problems* while a scale operation is a single simple *algorithmic step*. And so we have encountered a first (easy) instance of the magic of a resolution algorithm; i.e., an instance of a transcendental step (the analysis problem) which is converted into a programmable operation, here the operation of scaling. (Other more powerful uses of the scaling operation may be found in the recent book [Mey98, especially Ch. 5] by Yves Meyer and [HwMa94].)

The sketch below allows you to visualize more clearly this resolution versus detail concept which is so central to the wavelet algorithms, also for general wavelets which otherwise may be computationally more difficult than the Haar wavelet.
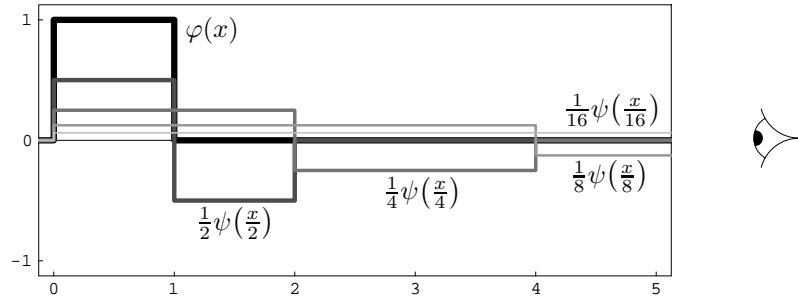
The wavelet decomposition of Haar's bump function φ in (1.1) and (1.5)

Using the sketch we see for example that the simple step function

$$f(x) = a\varphi(x) + b\varphi(x-1) = a\chi_{[0,1)}(x) + b\chi_{[1,2)}(x) \tag{1.7}$$
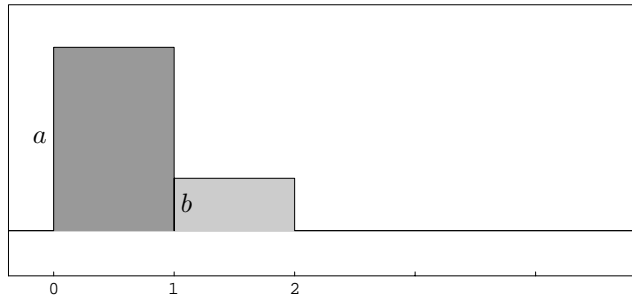
has the wavelet decomposition into a sum of a *coarser resolution* and an *intermediate detail* as follows:

$$f(x) = \underbrace{\frac{a-b}{2}\psi\left(\frac{x}{2}\right)}_{\text{intermediate detail}} + \underbrace{\frac{a+b}{2}\varphi\left(\frac{x}{2}\right)}_{\text{coarser version}}, \qquad x \in \mathbb{R}. \tag{1.8}$$

Thus the details are measured as differences. This is a general feature that is valid for other functions and other wavelet resolutions. See, for instance, Example 2.5.3, Lemma 2.2.2, and Exercise 1–12.

Combining (1.5) and (1.8), we note that the complete orthogonal wavelet expansion for the function f in (1.7) is

$$f(x) = \frac{a-b}{2}\psi\left(\frac{x}{2}\right) + (a+b)\sum_{k=2}^{\infty}\frac{1}{2^k}\psi\left(\frac{x}{2^k}\right). \tag{1.9}$$

While this observation is immediate, it highlights a different feature of the wavelet expansion which makes it far superior to the classical Fourier series: It is strikingly effective in its analysis of point singularities, of which (1.7) is a natural case in point. A glance at (1.9) reveals that for all $n \in \mathbb{Z}_+$, the number of wavelet coefficients exceeding $\frac{1}{n}$ is of the order $\log n$, and this count is thus asymptotically

$\ll n$ for large n. This turns out to be valid generally for scale-based wavelets. The corresponding count of the number of Fourier coefficients is typically at least of the order n.

In summary, point singularities are localized in the world of wavelets, while they are not for the typical classical orthogonal expansions. The effect of the point singularity is encoded in a small number of big wavelet coefficients, and the rest can be ignored. The recognition that wavelets analyze effectively functions which are smooth away from isolated singularities has made them popular in applications such as image coding and data compression.

An added significance of (1.8): It serves to illustrate the use of a quantum computing algorithm for the implementation of multiresolution analysis for Haar's wavelet. In the realm of qubits (denoted $|0\rangle$ and $|1\rangle$), we work with quantum states $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|-\rangle$, where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. So the assignment $\varphi \mapsto |0\rangle$, $\varphi(x-1) \mapsto |1\rangle$, $\frac{1}{\sqrt{2}}\varphi\left(\frac{x}{2}\right) \mapsto |+\rangle$, and $\frac{1}{\sqrt{2}}\psi\left(\frac{x}{2}\right) \mapsto |-\rangle$, turns (1.8) into the beginning of a quantum computing algorithm. Have a look at Exercise 1–19. The two quantum-mechanical states $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ are written in the lingo of Dirac, i.e., as ket vectors, and they are taken up in much more detail in Section 1.3 below. We note here that they are basic to quantum information theory, where they are called *coherent states*. It is a basic fact that coherent states are one of the sources of the speedup of quantum computer programs over classical ones; the other source of speedup is called *quantum entanglement*. Both are discussed in Section 1.3.

With the work of P. Shor and others, it has been demonstrated that exponential speedup of algorithms can be realized in the quantum realm, i.e., when registers of qubits are used in place of the classical bits, and when quantum gates, in the form of unitary matrices acting on tensor slots, take the role of the classical logic gates. Unitary matrix factorizations are then used in the algorithms. If the quantum algorithms are shorter than the best available (analogous) classical ones, there is of course a gain. But the dictates of quantum theory introduce new and serious sources of "error" called decoherence, i.e., when some qubits that are part of the program degenerate and behave classically. Clifford analysis is used in error correction. Wavelet algorithms split functions in a fixed resolution subspace into components, a coarser one, and detail parts. This can be turned into a quantum algorithm, and the factorization problem for the resulting unitary matrix can be implemented effectively, as we show in Section 1.3 and a number of the exercises in this chapter.

Wavelets really have an early start in modern mathematical history, going back almost a hundred years, with a delayed reaction. In fact, the current attention they are receiving in mathematics and its applications dates only from the mid-eighties, where Daubechies's discovery of the wavelets that have both compact support and a degree of smoothness regularity stands out. Others had a hand in it as well; see [Dau92]. One reason for a delay in the discovery of the "good" wavelets, long after the Haar one in [Haa10], is

that the variety from which they are selected is a large one, with most of the territory parameterizing rather ragged $L^2(\mathbb{R})$-wavelets. Here we take the view that the variety is defined from the masking coefficients, or equivalently, the quadrature mirror filters of electrical engineering. These filters in turn have a history of their own going back before the eighties, in signal processing, and independently in operator algebra theory. One aim of this book is to outline the algebraic variety of points which label the wavelets, identify structural features of this variety, such as its connected components, and indicate spectral-theoretic selection criteria which help us to pick out the regular wavelets from the vast landscape of the more ragged or fractal-looking ones.

A second aim is to show that the space of scaling/wavelet functions with a fixed compact support obtained by multiresolution analysis of scale N has a finite number of connected components which can be computed by using the winding number of the determinant of the continuous function from $\mathbb{T}$ into $\mathrm{GL}(N)$ defining the scaling/wavelet functions. Here $\mathbb{T}$ denotes the torus, viewed alternately as $\{z \in \mathbb{C} \mid |z| = 1\}$ or $\mathbb{R}/2\pi\mathbb{Z}$; and $\mathrm{GL}(N)$ is the general linear group. In order to analyze the connected components of wavelets in stronger topologies we have to use spectral properties of the associated transfer operator (see (1.19) below), and we show that operators of this type, as operators on $L^2(\mathbb{T})$, have the open unit disc in their point spectra, although it is well known that the restrictions of these operator to the spaces

$$\mathcal{E}_\alpha = \left\{ \sum_{n \in \mathbb{Z}} x_n e^{int} \ \bigg| \ \sum_n |x_n|^2 e^{2\alpha|n|} < \infty \right\} \tag{1.10}$$

for $\alpha > 0$ are of trace class.

We also study the spectrum of the transfer operators in other function spaces on $\mathbb{T}$, and establish similar properties. Chapters 4 and 5 are devoted to a detailed study of peripheral spectral properties of the transfer operator in nongeneric cases, in the setting of biorthogonal wavelet bases. (Nongeneric means that the eigenvalue 1 occurs with multiplicity greater than 1 for the transfer operator, in the setting of Proposition 4.3.1, or that there are other eigenvalues of absolute value 1.) It is demonstrated that the space of fixed points of the transfer operator R as an operator on $C(\mathbb{T})$ has the structure of a finite-dimensional commutative C^*-algebra in a product different from the usual pointwise product of functions in $C(\mathbb{T})$. Thus the R-invariant positive functionals ω on $C(\mathbb{T})$ with $\omega(1) = 1$ form a finite-dimensional simplex [Alf71], and the extremal points in this simplex are certain measures on cycles on $\mathbb{T}$ under the map $z \rightarrow z^N$ which are characterized. These cycles are important in wavelet theory because of their role in the study of frame properties and of the orthogonality relations which go into the construction of multiresolutions.

Before introducing systematically the concept of *multiresolution analysis* in Section 2.2, we sketch it briefly for the simple example in (1.2), i.e., for the Haar wavelet. Since the Haar wavelet is so simple and transparent, it can well be understood without first in-

troducing explicitly the multiresolution; our viewpoint is instead to use Haar's intuitive and easy construction for explaining the multiresolutions. The point is that the general mechanics of multiresolution analysis may then be used much more widely for the purpose of generating a variety of wavelets with general properties from a preassigned list of specifications. The Haar wavelet itself may in fact not have all the properties that are needed in a particular application.

Consider first the subspace $\mathcal{V}_0$ of $L^2(\mathbb{R})$ which is the closed linear span of the functions $\{\varphi(x-k) \mid k \in \mathbb{Z}\}$ where φ is the unit bump function of (1.1). Then $\mathcal{V}_0$ is the space of all step functions in $L^2(\mathbb{R})$ with jump points, or knots, at the integral points on $\mathbb{R}$. Let $U \colon f(x) \mapsto \frac{1}{\sqrt{2}} f\left(\frac{x}{2}\right)$ be the unitary operator on $L^2(\mathbb{R})$ which scales functions by the factor 2. Then $U(\mathcal{V}_0)$ ($\subset \mathcal{V}_0$) is a more coarse version of the resolution $\mathcal{V}_0$. It consists of the more special step functions in $L^2(\mathbb{R})$ with knots only on $2\mathbb{Z}$, i.e., only on the even integers; and similarly, $U^{-1}(\mathcal{V}_0)$ is the larger resolution subspace consisting of all functions in $L^2(\mathbb{R})$ which are step functions with knots on the finer grid $\frac{1}{2}\mathbb{Z}$. Since the constant (nonzero) functions are not in $L^2(\mathbb{R})$, it is then clear that

$$\bigwedge_{j \in \mathbb{Z}} U^j(\mathcal{V}_0) = \{0\} \tag{1.11}$$

and

$$\bigvee_{j \in \mathbb{Z}} U^j(\mathcal{V}_0) = L^2(\mathbb{R}), \tag{1.12}$$

where the symbol $\bigwedge$ applied to some closed subspaces in $L^2(\mathbb{R})$ refers to intersection, while $\bigvee$ refers to the closed linear span of the spaces, in the present case

$$\overline{\bigcup_{j<0} U^j(\mathcal{V}_0)}$$

with the overbar denoting closure in $L^2(\mathbb{R})$. More generally, we say that a closed subspace $\mathcal{V}_0$ in $L^2(\mathbb{R})$ which satisfies $U(\mathcal{V}_0) \subset \mathcal{V}_0$ and (1.11)–(1.12) is a *resolution*. The vectors in

$$\mathcal{V}_0 \ominus U(\mathcal{V}_0) = \{f \in \mathcal{V}_0 \mid \langle f \mid g \rangle = 0 \text{ for all } g \in U(\mathcal{V}_0)\}$$

then represent the *detail* which must be added to the *coarser resolution* $U(\mathcal{V}_0)$ in order to recover the original one $\mathcal{V}_0$.

The geometry of the situation is again illustrated nicely by Haar's example. To see this, return to the functions φ and ψ of the sketches (1.1)–(1.2). It is immediate that

$$\varphi\left(\frac{x}{2}\right) = \varphi(x) + \varphi(x-1), \qquad \psi\left(\frac{x}{2}\right) = \varphi(x) - \varphi(x-1). \tag{1.13}$$

It follows that the second function ψ is the zig-zag function of (1.2), that ψ is in the *detail space* $\mathcal{D} := U^{-1}(\mathcal{V}_0) \ominus \mathcal{V}_0$, and finally that the translates $\{\psi(x-k)\}_{k \in \mathbb{Z}}$ form

an *orthonormal basis* for $\mathcal{D}$. The last assertion means that the translated functions are mutually orthogonal, that they have $L^2(\mathbb{R})$-norm equal to 1, and that their closed linear span is $\mathcal{D}$. Introducing

$$m_0(t) := m_0\left(e^{-it}\right) = \frac{1}{\sqrt{2}}\left(1 + e^{-it}\right), \tag{1.14}$$

called the low-pass filter, and

$$m_1(t) := m_1\left(e^{-it}\right) = \frac{1}{\sqrt{2}}\left(1 - e^{-it}\right), \tag{1.15}$$

called the high-pass filter, and applying the Fourier transform to (1.13), we get the equivalent relations

$$\sqrt{2}\hat{\varphi}(2t) = m_0(t)\hat{\varphi}(t), \qquad \sqrt{2}\hat{\psi}(2t) = m_1(t)\hat{\varphi}(t). \tag{1.16}$$

But a direct computation also yields

$$\hat{\varphi}(2t) = e^{-it}\frac{\sin t}{t} =: e^{-it}\operatorname{sinc}(t) = 1 - it - \frac{2t^2}{3} + \cdots \tag{1.17}$$

and

$$\hat{\psi}(2t) = ie^{-it}\tan\left(\frac{t}{2}\right)\operatorname{sinc}(t) = ie^{-it}\frac{1 - \cos t}{t} = \frac{it}{2} + \frac{t^2}{2} + \cdots \tag{1.18}$$

(where we have indicated the start of the power series); and we get, from (1.16), as a bonus, an independent verification of the classical infinite product formula

$$\operatorname{sinc}(t) = \prod_{k=1}^{\infty}\cos\left(\frac{t}{2^k}\right).$$

In this book we will also consider scaling by a general $N \in \{2, 3, 4, \dots\}$ in addition to scaling by 2 as above. We will consider a *transfer operator* R constructed from a function $m_0(z)$ which is a polynomial or a Lipschitz function of $z \in \mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$,

$$R_{m_0}(f)(z) = R_{m_0}f(z) = \frac{1}{N}\sum_{\substack{w \in \mathbb{T} \\ w^N = z}}|m_0(w)|^2 f(w). \tag{1.19}$$

It is defined on suitable functions f on $\mathbb{T}$. The spectral theory of R_{m_0} depends on the space of functions f which is chosen (see Sections 3.2 and 3.4). For example, if this space is the set of Fourier polynomials in z (see Definition 3.2.1), R_{m_0} has a finite spectrum, and it keeps the same finite spectrum if it is extended to Hilbert spaces of functions which are analytic in an open annulus containing $\mathbb{T}$; see Proposition 3.5.1 or [Dau95]. However, we will see in Section 3.2 that if R_{m_0} is extended to $L^2(\mathbb{T})$,

the eigenvalue spectrum of R_{m_0} contains an open disc around the origin. We will call the spectrum of the restriction of R_{m_0} to the two former spaces the Perron–Frobenius spectrum. The spectrum of R_{m_0} is of interest in multiresolution wavelet analysis for the following reason: if

$$m_0(z) = \sum_n a_n z^n \qquad (1.20)$$

is the Fourier decomposition of the low-pass filter m_0, then the *scaling function* associated to m_0 is a function or distribution φ on $\mathbb{R}$ satisfying the identity

$$\varphi(x) = \sqrt{N} \sum_{n \in \mathbb{Z}} a_n \varphi(Nx - n), \qquad (1.21)$$

which is called the *scaling identity*. In general, a solution φ of this equation may just be a distribution, and φ may even be a function which is not locally integrable. (We will see an example of this in Exercise 1–7, and, in the context of a slightly different scaling identity, in (3.5.30)–(3.5.32).) Solutions of (1.21) can sometimes be found by iterating the right-hand side in (1.21), starting at

$$\chi(x) = \begin{cases} 1, & 0 \le x < 1, \\ 0, & x \in \mathbb{R} \setminus [0, 1\rangle, \end{cases} \qquad (1.22)$$

by using the *cascade iteration*

$$M_a : \psi \longmapsto \sqrt{N} \sum_n a_n \psi(Nx - n). \qquad (1.23)$$

If m_0 is a polynomial satisfying (1.25) and (1.26) below, then $\|R\| = 1$ on $L^\infty(\mathbb{T})$, and the cascade iteration converges in $L^2(\mathbb{R})$ if and only if the peripheral spectrum ($=$ the intersection of the spectrum with the unit circle $\mathbb{T}$) of R as an operator on $C(\mathbb{T})$ consists of 1 alone, and the eigenvalue 1 is nondegenerate; see Section 2.5 for this and more general results. (It may happen that even though the eigenvalue 1 is nondegenerate for $R|_{C(\mathbb{T})}$, it is degenerate for $R|_{L^\infty(\mathbb{T})}$! See Example 3.5.5.)

A second significance of the eigenspace $\ker(1 - R_{m_0})$ for the transfer operator of (1.19) is that it predicts orthogonality of the $\mathbb{Z}$-translates for a scaling function φ, i.e., a solution φ to (1.21): Suppose $\varphi \in L^2(\mathbb{R})$ satisfies (1.21), and set $c = (c_n)_{n \in \mathbb{Z}}$, where $c_n = \int_{\mathbb{R}} \overline{\varphi(x)} \varphi(x - n) \, dx$ are the correlation coefficients. If R_{m_0} is viewed as an $\infty \times \infty$ matrix relative to the Fourier basis $(z^n)_{n \in \mathbb{Z}}$, then it can be easily checked that $R_{m_0} c = c$. Hence, if c is in a space $\mathcal{C}$ of sequences containing the sequence $(\delta_{0,n})_{n \in \mathbb{Z}}$, and if $\ker(1 - R_{m_0}|_{\mathcal{C}}) = \mathbb{C} (\delta_{0,n})_{n \in \mathbb{Z}}$, then it follows that the set of translates

$$\{\varphi(x - n)\}_{n \in \mathbb{Z}} \qquad (1.24)$$

is an orthogonal family of functions in $L^2(\mathbb{R})$. Even if the $\mathbb{Z}$-translates in (1.24) are not an orthogonal set of functions, an identification of spaces $\mathcal{C}$, and the structure of the corresponding eigenspaces $\ker(1 - R_{m_0}|_{\mathcal{C}})$, predict *a priori* the possible orthogonality relations for the $\mathbb{Z}$-translates of scaling functions φ. Hence these 1-eigenspaces are studied

in detail in Chapters 4 and 5 below. See also the study guide in Section 1.5. We do a similar analysis of the other eigenspaces of R_{m_0} in Chapter 3 below. In Section 3.3 the setting is specialized to a particular two-parameter family of compactly supported wavelets on $\mathbb{R}$, and we indicate, by comparing the spectrum of R_{m_0} with the graph of the corresponding scaling function φ_{m_0}, that the shape of φ_{m_0}, and therefore of the wavelet ψ_{m_0}, depends on the spectrum of R_{m_0} computed on a certain finite-dimensional space. Specifically, if (θ, ρ) are the two parameters, we study the eigenvalues $\lambda_i^{(\theta,\rho)}$ of $R_{m_0}^{(\theta,\rho)}$, and the respective graphs of $x \mapsto \varphi_{m_0}^{(\theta,\rho)}(x)$ on $\mathbb{R}$. The results are summarized in Figure 3.2.

Definitions. A function, or a distribution, φ satisfying (1.21) is said to be *refinable*, the equation (1.21) is called the *refinement equation*, or also, as noted above, the "scaling identity", and φ is called the scaling function. The coefficients a_n of (1.21) are called the *masking coefficients* (see Definition 3.2.1).

We will mainly concentrate on the case when the set $\{a_n\}$ is finite. But in general, a function $\varphi \in L^2(\mathbb{R})$ is said to be refinable with scale number N if $\varphi(x/N)$ is in the L^2-closed linear span of the translates $\{\varphi(x-k)\}_{k\in\mathbb{Z}} \subset L^2(\mathbb{R})$; see, e.g., [HSS96, SSZ99, StZh98, StZh01].

Since there are refinement operations which are more general than scaling (see for example [DLLP01]), there are variations of (1.21) which are correspondingly more general, with regard to both the refinement steps that are used and the dimension of the spaces. The term "scaling identity" is usually, but not always, reserved for (1.21), while more general refinements lead to "refinement equations". However, (1.21) often goes under both names. The vector versions of the identities get the prefix "multi-", for example *multiscaling* and *multiwavelet*; see Section 2.6.

If m_0 satisfies a condition for obtaining orthogonal wavelets,

$$\sum_{w^N=z} |m_0(w)|^2 = N, \tag{1.25}$$

together with the normalization

$$m_0(1) = \sqrt{N}, \tag{1.26}$$

then (1.21) has a solution φ in $L^2(\mathbb{R})$ which can be obtained by taking the inverse Fourier transform of the product expansion

$$\hat{\varphi}(t) = \prod_{k=1}^{\infty} \left(\frac{m_0\left(tN^{-k}\right)}{\sqrt{N}} \right). \tag{1.27}$$

(Here and later we use the convention that if $m(z)$ is a function of $z \in \mathbb{T}$, then $m(t) = m\left(e^{-it}\right)$.) That (1.27) gives a solution φ of (1.21) follows from the relation

$$\hat{\varphi}(t) = \frac{1}{\sqrt{N}} m_0\left(\frac{t}{N}\right) \hat{\varphi}\left(\frac{t}{N}\right), \tag{1.28}$$

which is equivalent to (1.21). Note that in this case the constant function 1 is an eigenvector for R_{m_0} with eigenvalue 1. We mentioned after (1.23) (see also [BrJo99b, Theorem 2.5]) that if all other eigenvalues in the Perron–Frobenius spectrum have absolute value strictly less than 1, then the iteration converges in $L^2(\mathbb{R})$ towards a scaling function φ with $\|\varphi\|_2 = 1$. If $N = 2$, the spectral condition on the transfer operator has been used to show existence of unconditional biorthogonal wavelet bases of compactly supported wavelets under weaker conditions on m_0 than (1.25) and (1.26) above, as was done in [CoRy95, Theorem 4.2]. Instead of (1.25)–(1.26), one then assumes that there exists another trigonometric polynomial $\tilde{m}_0$ such that (when $N = 2$)

$$\overline{m_0(\omega)}\,\tilde{m}_0(\omega) + \overline{m_0(\omega + \pi)}\,\tilde{m}_0(\omega + \pi) = 2 \tag{1.29}$$

and

$$m_0(0) = \tilde{m}_0(0) = \sqrt{2}. \tag{1.30}$$

If then R_{m_0} and $R_{\tilde{m}_0}$ are the associated transfer operators, and m_0 and $\tilde{m}_0$ are Fourier polynomials of order D, then the dual filters m_0 and $\tilde{m}_0$ generate unconditional biorthogonal wavelet bases in $L^2(\mathbb{R})$ if and only if the spectral radii of the restrictions of R_{m_0} and $R_{\tilde{m}_0}$ to the $(4D + 1)$-dimensional Hilbert space

$$\mathcal{F}_D = \left\{ \sum\nolimits_{k=-2D}^{2D} c_k z^k \;\middle|\; \sum\nolimits_{k=-2D}^{2D} c_k = 0 \right\} \tag{1.31}$$

are both strictly less than 1.

We mentioned that there is a direct connection between $m_0 = \sum a_n z^n$ and the scaling function φ on $\mathbb{R}$ given in (1.20)–(1.21) and (1.27). There is a similar correspondence between the high-pass filters m_i and the wavelet generators $\psi_i \in L^2(\mathbb{R})$. In the *biorthogonal* case, there is a second system $\tilde{m}_i \leftrightarrow \tilde{\psi}_i$ and the two systems

$$\left\{ N^{\frac{i}{2}} \psi_i \left(N^j x - k \right) \right\} \quad \text{and} \quad \left\{ N^{\frac{i'}{2}} \tilde{\psi}_{i'} \left(N^{j'} x - k' \right) \right\},$$
$$i, i' \in \{1, 2, \ldots, N-1\}, \; j, j', k, k' \in \mathbb{Z}, \tag{1.32}$$

then form a dual wavelet basis, or dual wavelet frame for $L^2(\mathbb{R})$ in the sense of [Dau92, Chapter 5]. We will consider this biorthogonal case in more detail in Section 6.4.

The idea of constructing maximally smooth wavelets when some side conditions are specified has been central to much of the activity in wavelet analysis and its applications since the mid-1980's. In addition to [Dau92], the survey article [Str93] is enjoyable reading as a backdrop to our book. See also the latter half of the tutorial to Chapter 3. The paper [LaHe96] treats the issue in a more specialized setting and is focussed on the moment method. Some of the early applications to data compression and image coding are done very nicely in [HSS+95], [SHS+99], and [HSW95]. An interesting, related but different, algebraic and geometric approach to the problem is offered in [PeWi99].

The Fourier transform and the Fourier series. Let us record at the outset the following conventions pertaining to Fourier transforms and Fourier series: If f is a function on $\mathbb{R}$, then its *Fourier transform* is

$$\hat{f}(t) = \int_{\mathbb{R}} e^{-itx} f(x) \ dx, \tag{1.33}$$

and the *inverse Fourier transform* is

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{itx} \hat{f}(t) \ dt, \tag{1.34}$$

so *Plancherel's formula* becomes

$$\int_{\mathbb{R}} |f(x)|^2 \ dx = \frac{1}{2\pi} \int_{\mathbb{R}} \left| \hat{f}(t) \right|^2 \ dt. \tag{1.35}$$

If f is a function on the circle $\mathbb{T} = \mathbb{R} / 2\pi\mathbb{Z}$, its *Fourier coefficients* are

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} f(x) \ dx = \int_{\mathbb{T}} z^{-n} f(z) \ d\mu(z), \tag{1.36}$$

where μ is *normalized Haar measure* on the circle, so

$$f(x) = \sum_{k \in \mathbb{Z}} c_k e^{ikx} = \sum_{k \in \mathbb{Z}} c_k z^k, \tag{1.37}$$

and Plancherel's formula becomes

$$\sum_n |c_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \ dx = \int_{\mathbb{T}} |f(z)|^2 \ d\mu(z). \tag{1.38}$$

Illustration. Consider now the families of functions from (1.32) which go into the construction of wavelets through the looking glass of the Fourier transform (1.33). Each of the doubly indexed function systems (1.32) then takes the form

$$\left\{ N^{\frac{j}{2}} e^{iN^j \cdot kt} \hat{\psi} \left(N^j t \right) \right\}_{j,k \in \mathbb{Z}}. \tag{1.39}$$

This makes it clear, for example, that all function systems $\psi_1, \ldots, \psi_{N-1}$ in $L^2(\mathbb{R})$ which generate a wavelet basis, even of the general form (1.32), have the support of the transforms $\hat{\psi}_i$ confined to certain prescribed subsets of $\mathbb{R}$, as noted in Section 1.6 below.

1.1   Subband filters and sieves

AROUND 1986, a radically new method for performing discrete wavelet analysis and synthesis was born, known as *multiresolution analysis*. This method is completely *recursive* and therefore ideal for computations. One begins with a version

$f^0 = \{f^0_n\}_{n \in \mathbb{Z}}$ of the signal *sampled* at regular time intervals $\Delta t = \tau > 0$. f^0 is split into a "blurred" version f^1 at the coarser scale $\Delta t = 2\tau$ and "detail" d^1 at scale $\Delta t = \tau$. This process is repeated, giving a sequence $f^0, f^1, f^2, \ldots$ of more and more blurred versions together with the details $d^1, d^2, d^3, \ldots$ removed at every scale ($\Delta t = 2^m \tau$ in f^m and d^{m-1}). —Gerald Kaiser, *A Friendly Guide to Wavelets* [Kai94], Chapter 7 summary
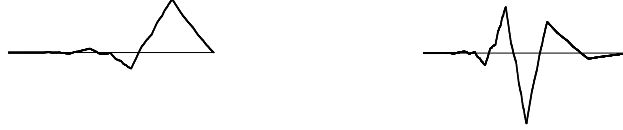
One of Stéphane Mallat's early contributions [Mal89] was to show that *multiresolution analysis* in the context of wavelets could be viewed as (just) another form of what is called the *pyramid algorithm* in computer graphics, thereby making the connection between subband filtering in signal processing, and the theory of subdivision algorithms in computer imaging and data compression (see Exercise 1–12). This was one of the happy marriages of separate fields of science which had not been previously thought to be related, and it was brought about by the focus on wavelet ideas. More such connections and marriages of ideas are outlined in Mallat's lovely book [Mal99], which even contains much more than a review of multiresolution analysis and its applications: it treats, for example, *wavelet packets* and *local cosine bases* which are different but yet share a theoretical foundation with the wavelets. Both are put in a form which brings the geometry of Hilbert space into focus. And then there is the French view of history, starting with the *Fourier Kingdom*, and followed by the *Discrete Revolution* of 1989.

In electrical engineering [Vai93], [VNDS89], [DjVa94], the functions $m_0, \ldots, m_{N-1}$ on $\mathbb{T}$ go under the name of *subband filters*, the subbands are defined relative to the classical Fourier duality of time and frequency, and the *bands* themselves are the subbands of values of the frequency variable. They are used both in classical time series, or time signals, and also in digital signals of optical communications engineering. The estimates of the form (1.5.4) from wavelet theory motivate the term "filter" in the name: if

$$m(\xi) = \sum_{n \in \mathbb{Z}} a_n e^{in\xi}, \tag{1.1.1}$$

then the subband filtering in one of N subbands is an action on time signals $(x_k)_{k \in \mathbb{Z}}$ which has the following form, transforming a signal **x** into a new one **y**:

$$\mathbf{x} \longmapsto \mathbf{y} = F_m \mathbf{x}, \tag{1.1.2}$$

where

$$y_k = \sum_{n \in \mathbb{Z}} a_{Nk-n} x_n. \tag{1.1.3}$$

We then say that m is a *low-pass filter* (the lowest frequency band) if, for some small ε, the values of $|m(\xi)|$ are concentrated on a frequency band of the form $-\varepsilon < \xi < \varepsilon$. This

condition will be refined in the sequel. See for example (1.5.1) where all the weight is placed on the frequency $\xi = 0$, and as a consequence $m = 0$ at the remaining $N-1$ band points. A variant of the filters also arises in number theory, but then under the name of *sieves* [Ten95, p. 62]. For example, we say that m in (1.1.1) is a *large sieve* if, for every $k \in \mathbb{N}$ and every $\delta \in \mathbb{R}_+$, the function

$$m_{p,k}(\xi) = \sum_{p < n \leq p+k} a_n e^{in\xi}, \qquad \xi \in \mathbb{R}, \tag{1.1.4}$$

satisfies the following analogue of the estimate (1.5.4):

$$\sum_j |m_{p,k}(\xi_j)|^2 \leq \left(k + \delta^{-1} - 1\right) \sum_{p < n \leq p+k} |a_n|^2 \tag{1.1.5}$$

for all $p \in \mathbb{Z}$, and all finite sets (ξ_j) of sampled frequencies for which

$$\left|\xi_j - \xi_{j'}\right| \geq \delta \qquad \text{for all } j, j', \ j \neq j'. \tag{1.1.6}$$

The points $\{\xi_j\}$ are called *sample points*. An important special case of (1.1.6) which will be used in the study of wavelets constitutes the special finite arithmetic sets $\{\xi_j\}$ defined as follows: For $N \in \mathbb{N}$ and $\xi \in \mathbb{R}$, the sets $S_{N,\xi}$ have the form $S_{N,\xi} := \left\{\xi + j\frac{2\pi}{N} \mid 0 \leq j < N\right\}$. In this case, $\delta = \frac{2\pi}{N}$, but $\xi \in \mathbb{R}$ is a free variable.

While the engineering applications of sampling methods, scaling identities, and multiresolutions are very impressive and covered widely in the literature, see, e.g., [Vai93], [Wic93], the applications to physics are equally impressive, see, e.g., [vdB99].

When working with functions, or more generally, vectors in some Hilbert space, say $\mathcal{H}$, it will be useful for us, on occasion, to adopt a view from quantum mechanics: for example, a vector $\psi \in \mathcal{H}$ which is normalized by $\|\psi\|^2 = \langle \psi \,|\, \psi \rangle = 1$ represents a quantum-mechanical state. In Dirac's bra-ket terminology, it is denoted $|\psi\rangle$. The *projection E_ψ* onto the one-dimensional subspace $\mathbb{C}_\psi$ is denoted $E_\psi = |\psi\rangle \langle \psi|$. A bonus of Dirac's notation is that many statements made with bras and kets become, so to speak, true by virtue of being grammatically correct (and meaningful). It also emphasizes the probabilistic contents of the Hilbert-space formalism: for instance, if $\{\psi_i\}$ is a fixed orthonormal basis for $\mathcal{H}$, then a given state $|\psi\rangle$ produces a probability distribution by virtue of Bessel's identity

$$1 = \|\psi\|^2 = \sum_i |\langle \psi_i \,|\, \psi \rangle|^2. \tag{1.1.7}$$

In wavelet theory, it turns out that the identity (1.1.7) is also useful for vector systems $\{\psi_i\}$ which are not necessarily orthonormal bases; they are called *tight frames*.

Wavelets in one dimension live in the Hilbert space $L^2(\mathbb{R})$ of all square-integrable functions on the line $-\infty < x < \infty$. The traditional approach is to specify a *scaling number* $N \in \mathbb{Z}_+$, $N \geq 2$, and a lattice of translations $\mathbb{Z}$, called *sample points*, but there

are interesting variations of this setup, including higher dimensions, when the Hilbert space is $L^2\left(\mathbb{R}^d\right)$, $d = 2, 3, \ldots$. Staying with $d = 1$, and N fixed, we will take the viewpoint of what is called *resolutions*, but here understood in a broad sense of closed subspaces: A closed linear subspace $\mathcal{V} \subset L^2\left(\mathbb{R}\right)$ is said to be an N-resolution if it is invariant under the unitary operator

$$U = U_N : f \longmapsto N^{-\frac{1}{2}} f\left(\frac{x}{N}\right), \tag{1.1.8}$$

i.e., if U maps $\mathcal{V}$ into a proper subspace of itself. The subspace $\mathcal{V}$ is said to be *translation invariant* if

$$f \in \mathcal{V} \iff f\left(\cdot - k\right) \in \mathcal{V} \qquad \text{for all } k \in \mathbb{Z}. \tag{1.1.9}$$

If there is a function φ such that $\mathcal{V} = \mathcal{V}_\varphi$ is the closed linear span of

$$\left\{\varphi\left(\cdot - k\right) \mid k \in \mathbb{Z}\right\}, \tag{1.1.10}$$

then clearly $\mathcal{V}$ is translation invariant. The translation-invariant resolution subspaces $\mathcal{V}$ are actively studied and reasonably well understood. If $\mathcal{V}$ is of the form $\mathcal{V}_\varphi$ in (1.1.10), then we say that it is *singly generated*, and that φ is a scaling function of scale N. But the case when $\mathcal{V}$ is not singly generated is also interesting, and these resolution subspaces are frequently called *generalized multiresolution subspaces*. There is much current and very active research on them; see, for example, [BaLa99], [LPT01], [BaMe99], [HLPS99], [HSS01], [SSZ99], and [Jor01a]. The case when $\mathcal{V}$ is not singly generated as a resolution subspace of scale $N > 2$, i.e., when $\mathcal{V}$ is not of the form (1.1.10), occurs in the study of *wavelet sets*. Wavelet sets are measurable subsets $E \subset \mathbb{R}$ of finite measure such that the doubly indexed family $\left\{N^{\frac{j}{2}}\psi_E\left(N^j x - k\right)\right\}_{j,k \in \mathbb{Z}}$ defined from the function

$$\psi_E\left(x\right) = \check{\chi}_E\left(x\right) = \frac{1}{2\pi} \int_E e^{itx} \, dt \tag{1.1.11}$$

forms an orthonormal basis for $L^2\left(\mathbb{R}\right)$. It follows from (1.39) that $E \subset \mathbb{R}$ is a wavelet set for scale N if and only if the four conditions (a)–(d) hold:

(a) $\bigcup_{j \in \mathbb{Z}} N^j E = \mathbb{R}$, except for a set in $\mathbb{R}$ of measure zero, where $N^j E := \left\{N^j x \mid x \in E\right\}$;

(b) $\text{meas}\left(N^j E \cap N^k E\right) = 0$ if $j \neq k$ in $\mathbb{Z}$;

(c) $\bigcup_{k \in \mathbb{Z}}\left(E + 2\pi k\right) = \mathbb{R}$, except for a set in $\mathbb{R}$ of measure zero; and

(d) $\text{meas}\left(\left(E + 2\pi k\right) \cap \left(E + 2\pi n\right)\right) = 0$ if $k \neq n$ in $\mathbb{Z}$.

If, for example, $N = 2$, then the subset of $\mathbb{R}$, $E := [-2\pi, -\pi) \cup [\pi, 2\pi)$, is a wavelet set. (For others, see Exercise 1–46.) In the list (a)–(d), the first two properties state that

E tiles $\mathbb{R}$ under a multiplicative action of $\mathbb{Z}$ on $\mathbb{R}$ (dilations in both small and large scale), while the last two properties refer to a *tiling* in the additive sense. The modification and generalization of the properties to higher dimensions $d > 1$ are clear: the N is an integral $d \times d$ matrix such that its eigenvalues λ satisfy $|\lambda| > 1$. It is known that wavelet sets exist also in higher dimensions; see the references mentioned above.

However, in this book we will primarily be concerned with the resolution subspaces which are singly generated. They are relatively better known, some would say completely understood. Yet we will encounter a number of natural questions which have only incomplete answers as of the present.

We now turn to a group-theoretic formulation of multiresolution analysis, which will be needed in Chapter 2 for charting the connected components of wavelet systems.

1.2 Matrix functions and multiresolutions

The two groups of matrix functions $C(\mathbb{T}, \mathrm{U}_N(\mathbb{C}))$ and $C(\mathbb{T}, \mathrm{GL}_N(\mathbb{C}))$, i.e., the continuous functions from the torus into the respective groups, enter wavelet analysis via the associated wavelet filters $(m_i)_{i=0}^{N-1}$.

In Sections 2.1 and 6.3, we give the details of the multiple correspondence between:

(i) matrix functions, $A \colon \mathbb{T} \to \mathrm{GL}_N(\mathbb{C})$,

(ii) high- and low-pass wavelet filters m_i, $\tilde{m}_{i'}$, $i, i' = 0, 1, \ldots, N-1$, and

(iii) wavelet generators ψ_i, $\tilde{\psi}_{i'}$, $i, i' = 1, \ldots, N-1$, together with scaling functions $\varphi, \tilde{\varphi}$.

In particular,

$$A_{i,j}(z) = \frac{1}{N} \sum_{w^N = z} m_i(w) w^{-j}, \qquad z \in \mathbb{T}, \qquad (1.2.1)$$

$$\left(A^{-1}\right)_{i,j} = \frac{1}{N} \sum_{w^N = z} \overline{\tilde{m}_j(w)} \, w^i, \qquad z \in \mathbb{T}. \qquad (1.2.2)$$

The dependence of the $L^2(\mathbb{R})$-functions in (iii) on the group elements A from (i) gives rise to homotopy properties, and the results in Sections 4.3, 5.4, 6.2, and 6.3 are building up to that, while the final results are stated in Section 2.1. The standard orthogonal wavelets represent the special case when $m_i = \tilde{m}_i$, or equivalently, $A(z) = ((A(z))^*)^{-1}$, $z \in \mathbb{T}$. Hence, the matrix functions are unitary in this case.

The scaling/wavelet functions $\varphi, \psi_1, \ldots, \psi_{N-1}$ with support on a fixed compact interval, say $[0, kN + 1]$, $k = 0, 1, \ldots$, can be parameterized with a finite number of parameters since the unitary-valued function $z \to A(z)$ in (1.2.1) then is a polynomial in z of degree at most $k(N-1)$. It is well-known folklore from computer-generated

pictures that the shape of the scaling/wavelet functions depends continuously on these parameters; see Figures 1.1–1.7 and [Tre01b].

The scaling function $\varphi \in L^2(\mathbb{R})$ of (1.21) is illustrated in Figures 1.1–1.7, in the case $N = 2$, and for orthogonal $\mathbb{Z}$-translates, i.e., the case (1.25). These pictures illustrate the dependence of φ on the masking coefficients (a_n) in the case of [Tre01b]:

$$
\begin{aligned}
a_0 &= (\eta_0 - \eta_1 - \eta_2 + \eta_3 + \eta_4)/4, & a_1 &= (\eta_0 + \eta_1 - \eta_2 + \eta_3 - \eta_4)/4, \\
a_2 &= (\eta_0 - \eta_3 - \eta_4)/2, & a_3 &= (\eta_0 - \eta_3 + \eta_4)/2, \\
a_4 &= (\eta_0 + \eta_1 + \eta_2 + \eta_3 + \eta_4)/4, & a_5 &= (\eta_0 - \eta_1 + \eta_2 + \eta_3 - \eta_4)/4,
\end{aligned}
\tag{1.2.3}
$$

where

$$
\begin{aligned}
\eta_0 &= 1/\sqrt{2}, & \eta_1 &= (\cos 2\theta + \cos 2\rho)/\sqrt{2}, & \eta_2 &= (\sin 2\theta + \sin 2\rho)/\sqrt{2}, \\
& & \eta_3 &= \cos(2\theta - 2\rho)/\sqrt{2}, & \eta_4 &= \sin(2\theta - 2\rho)/\sqrt{2}.
\end{aligned}
\tag{1.2.4}
$$

These formulas arise from an independent pair of rotations by angles θ and ρ of two "spin vectors", i.e., by taking the matrix function A in (1.2.1) unitary, $\mathbb{T} \ni z \to A_{\theta,\rho}(z) \in U_2(\mathbb{C})$, and setting

$$
A(z) = V(Q_\theta^\perp + zQ_\theta)(Q_\rho^\perp + zQ_\rho) = VU_\theta(z)U_\rho(z)
\tag{1.2.5}
$$

with

$$
V = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},
\tag{1.2.6}
$$

$$
Q_\theta = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix} = \frac{1}{2}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix} \right),
\tag{1.2.7}
$$

and the orthogonal complement to the one-dimensional projection Q_θ,

$$
Q_\theta^\perp = Q_{\theta+(\pi/2)}.
\tag{1.2.8}
$$

With the coefficients $a_0, a_1, a_2, a_3, a_4, a_5$ given by (1.2.3), the algorithmic approach to graphing the solution φ to the scaling identity (1.21) is as follows (see [Jor01b], [Tre01b] for details): the relation (1.21) for $N = 2$ is interpreted as giving the values of the left-hand φ by an operation performed on those of the φ on the right, and a binary digit inversion transforms this into the form

$$
\mathbf{f}'_{k+1}\left(x + \frac{1}{2^{k+1}}\right) = \mathbf{A}\mathbf{f}_k(x),
\tag{1.2.9}
$$

where $\mathbf{A}$ is the 2×3 matrix $\mathbf{A}_{i,j} = \sqrt{2}a_{4+i-2j}$ constructed from the coefficients in (1.21), and $\mathbf{f}_j$ and $\mathbf{f}'_j$ are the vector functions

$$
\mathbf{f}_j(x) = \begin{pmatrix} \varphi\left(x - \frac{2}{2^j}\right) \\ \varphi\left(x - \frac{1}{2^j}\right) \\ \varphi(x) \end{pmatrix}, \qquad \mathbf{f}'_j(x) = \begin{pmatrix} \varphi\left(x - \frac{1}{2^j}\right) \\ \varphi(x) \end{pmatrix}.
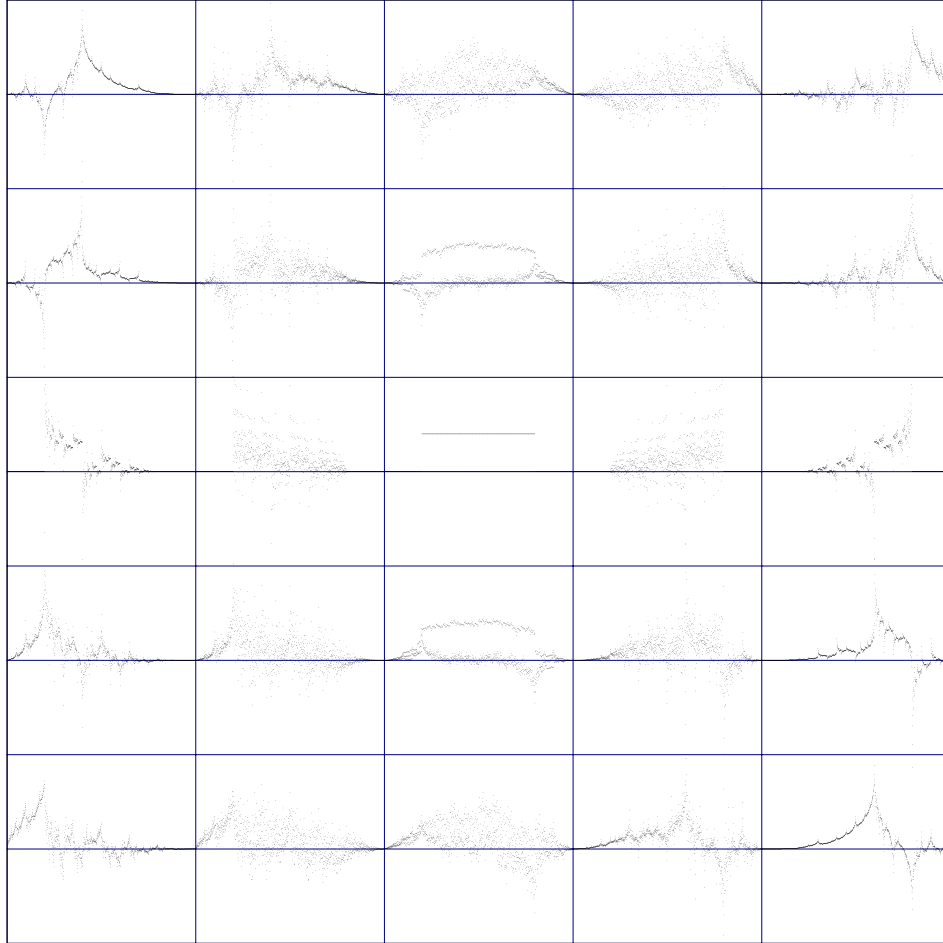\tag{1.2.10}
$$

Figure 1.1. Chaos. Wavelets around $\theta = \rho = 0$. L^2-convergence of cascade approximation fails at the central Haar point, where the spectral condition (2.5.3) fails. At this point the L^2-continuity of the map from the parameter space into the scaling functions fails, although continuity holds in the distribution topology. See also Figure 1.6. The scaling functions in this figure as well as in Figures 1.2–1.7 are actually 8th-order cascade approximants to the real scaling functions (see Section 2.5, in particular Theorem 2.5.1). In particular, the central Haar point here is really the graph of a function which is 1 at $\frac{1}{3}$ of the points in $[1, 4]$ and 0 at all remaining points. See Example 2.5.6 for the coarser approximants. The layered structure of the graphs above and below the Haar point disappears under further iterations of the cascade, and is replaced by a completely chaotic picture; see [BrJo99b, Figures 6–10].

Iterations of this operation give values of an approximation to φ on successively finer dyadic grids in the x variable.
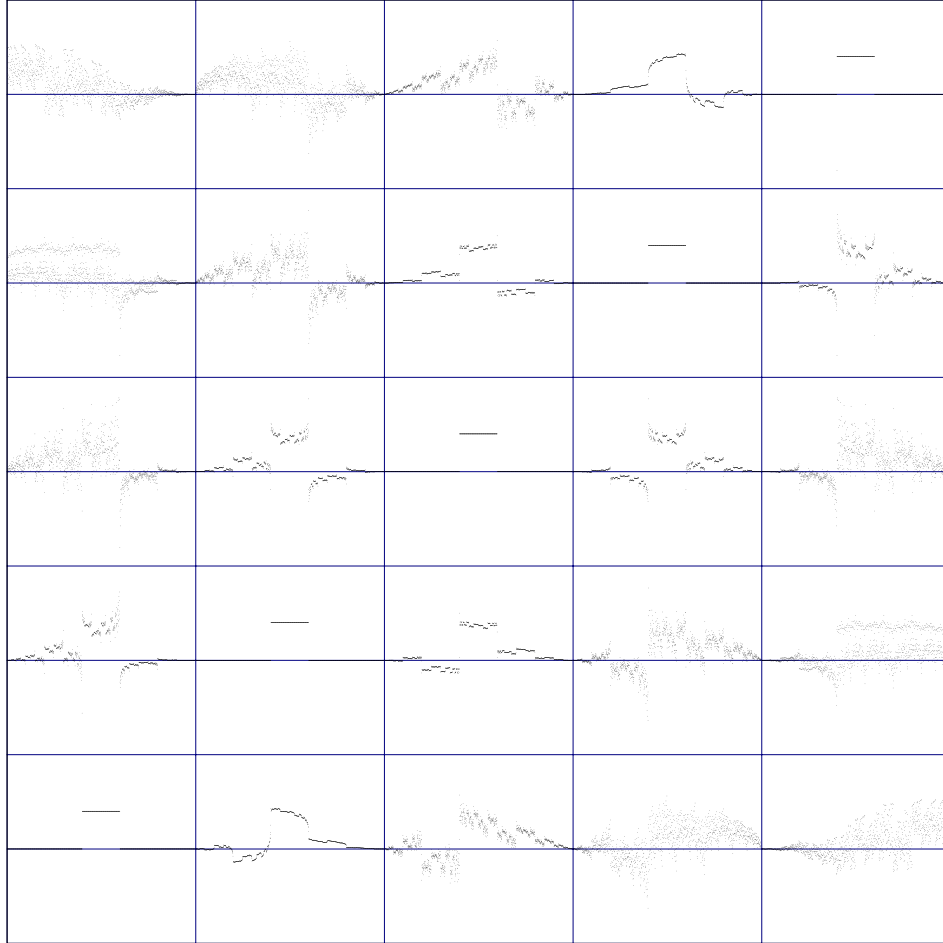
Figure 1.2. Wavelets around $\theta = 0$, $\rho = \pi/2$. One-dimensional subvariety of Haar wavelets; see also Figure 1.5.

Figures 1.2, 1.3, 1.4, and 1.5 include one of the orthogonal Haar scaling functions φ_H, i.e., $\varphi_H = \chi_I$ where I is a unit interval of the form $I = [k, k+1]$, while Figures 1.1 and 1.6 each have one of the nonorthogonal Haar wavelets

$$\varphi_{H'} = \frac{1}{3}\chi_{J'}, \qquad J' = [1, 4] \ \text{(Figure 1.1)} \qquad (1.2.11)$$

or

$$\varphi_{H''} = \frac{1}{5}\chi_{J''}, \qquad J'' = [0, 5] \ \text{(Figure 1.6)}. \qquad (1.2.12)$$

It is in neighborhoods (relative to θ, ρ) of the nonorthogonal scaling functions where the cascade approximation (which is the one used in the sampling of scaling functions shown
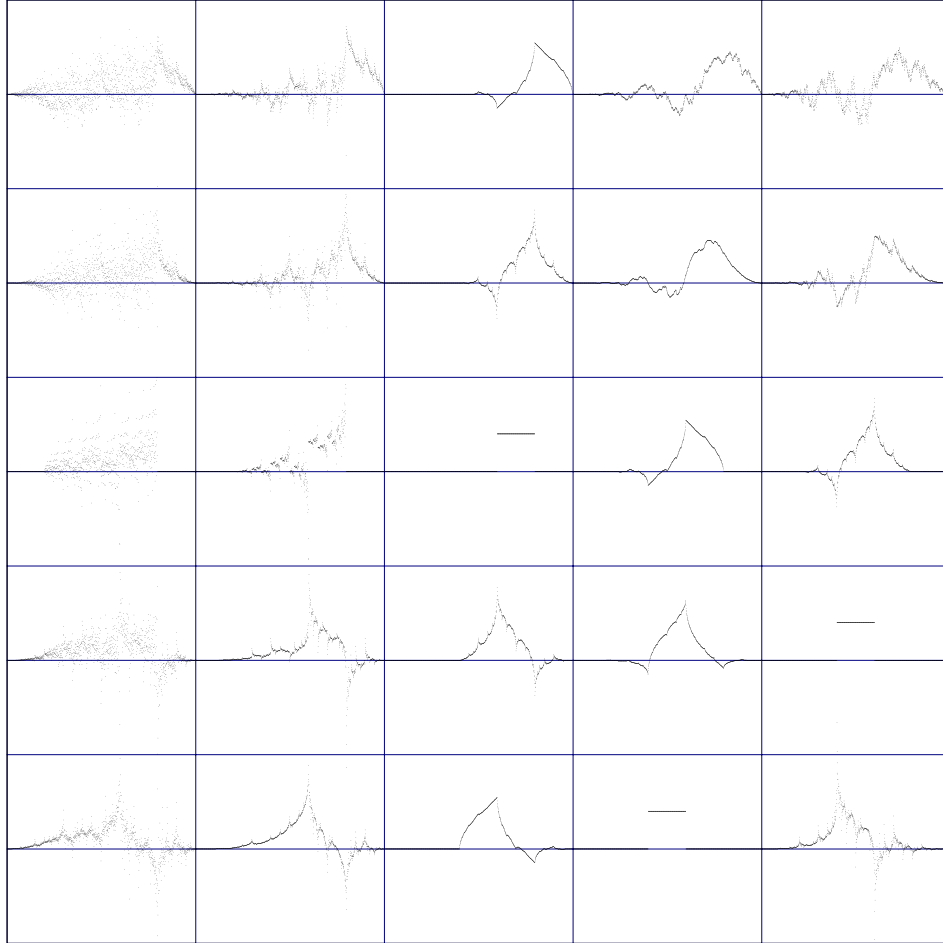
Figure 1.3. Absence of chaos. Wavelets around $\theta = \pi/4$, $\rho = 0$. Both an isolated point and a one-dimensional subvariety of Haar wavelets.

in Figures 1.1–1.7) is especially "fractal-looking". It is also for the nonorthogonal Haar wavelets that an $L^2(\mathbb{R})$-approximation in the form of a Cesaro summation (see Remark 2.5.5) is required. If $N = 2$, and $p \in \mathbb{N}$ is odd, then $m_p(z) = \frac{1}{\sqrt{2}}(1 + z^p)$ satisfies $m_p(z) = m_1(z^p)$, and so the scaling functions φ_1 and φ_p may both be determined from (1.27), and the formula $m_p(e^{it}) = m_1(e^{ipt})$ yields $\hat{\varphi}_p(t) = \hat{\varphi}_1(pt)$ where

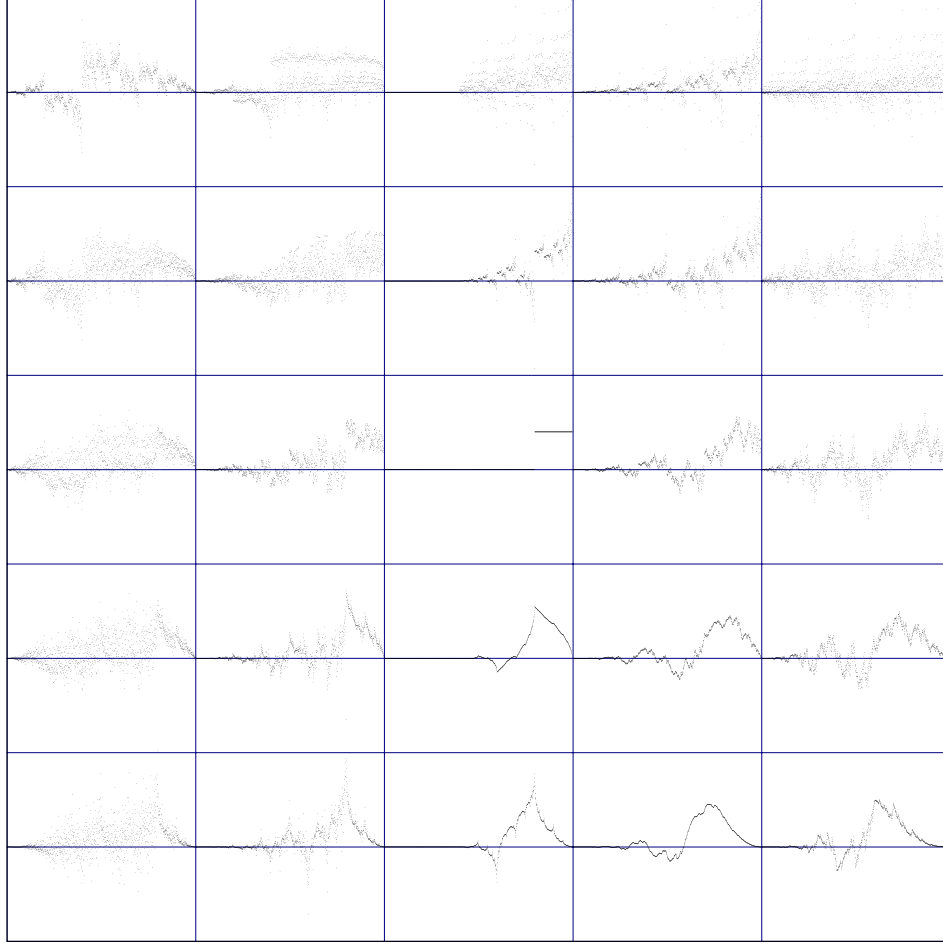$$\hat{\varphi}_1(t) = \frac{e^{-\frac{it}{2}} \sin \frac{t}{2}}{\frac{t}{2}} \tag{1.2.13}$$

Figure 1.4. Wavelets around $\theta = \rho = \pi/4$. An isolated Haar wavelet.

and $\varphi_1(x) = \chi_{[0,1]}(x)$. It follows that

$$\varphi_p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\varphi}_1(pt) e^{itx} dt \tag{1.2.14}$$

$$= \frac{1}{p} \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\varphi}(t) e^{it\frac{x}{p}} dt = \frac{1}{p} \varphi_1\left(\frac{x}{p}\right) = \frac{1}{p} \chi_{[0,p]}(x),$$

and $\|\varphi_p\|_{L^2(\mathbb{R})} = \frac{1}{\sqrt{p}}$. Otherwise we show in Theorem 2.5.1 that the cascades (1.23) converge as sequences. Since this convergence is in $L^2(\mathbb{R})$, there are problems with computer graphics in the case of discontinuous scaling functions $x \mapsto \varphi_{\theta,\rho}(x)$ even if they have compact support; see for example Figures 1.1 and 1.6 for illustrations of
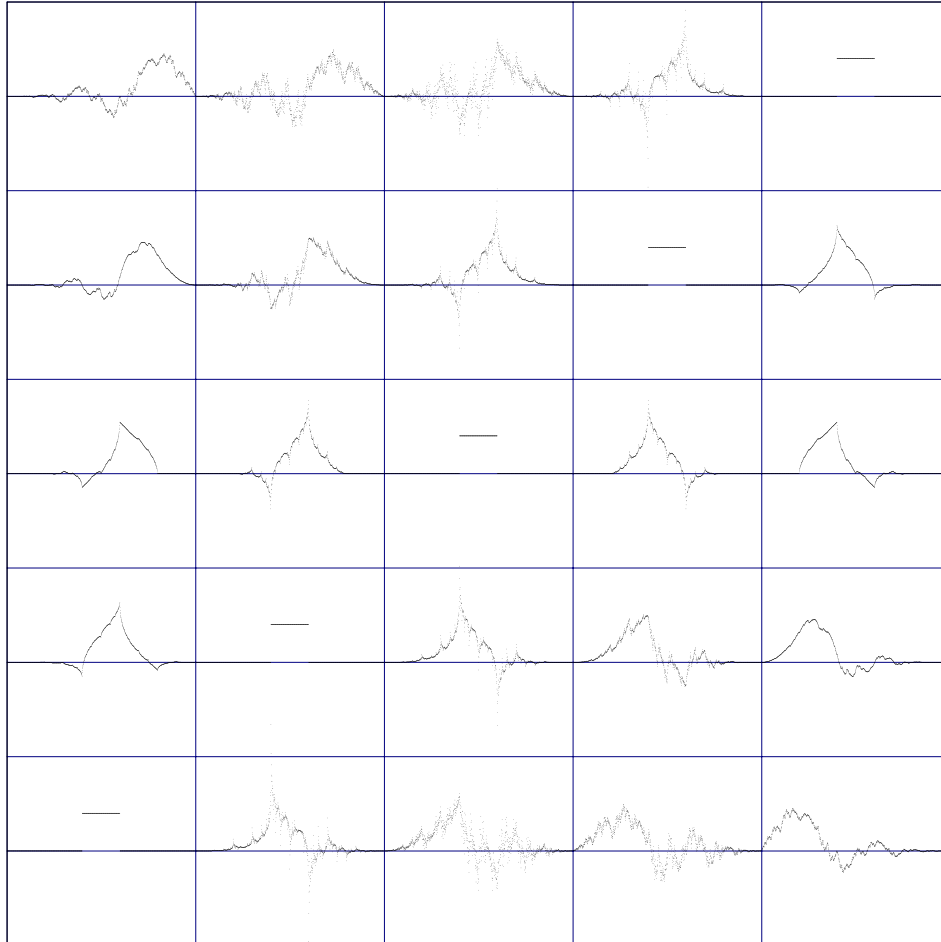
Figure 1.5. Wavelets around $\theta = \pi/2$, $\rho = 0$. One-dimensional subvariety of Haar wavelets.

this point. These pictures are generated by a mathematics program and the use of the subdivision algorithm (1.5.9). For the interpretation and limitations of the subdivision algorithm, the reader is referred to the tutorial in Section 3.3.*

In Section 2.1 we make a more rigorous study of the continuous dependence of the scaling/wavelet functions on the parameters, and the main result in Theorem 2.1.3 gives a parameterization of the connected components in the scaling/wavelet function space. The parameter used is the winding number of the map $z \rightarrow \det(A(z))$, and the re-

---

*An interactive display of the scaling/wavelet functions obtained by putting $\rho = 0$ in (1.2.5) and letting θ vary over the circle can be found in http://cm.bell-labs.com/who/wim/cascade/.
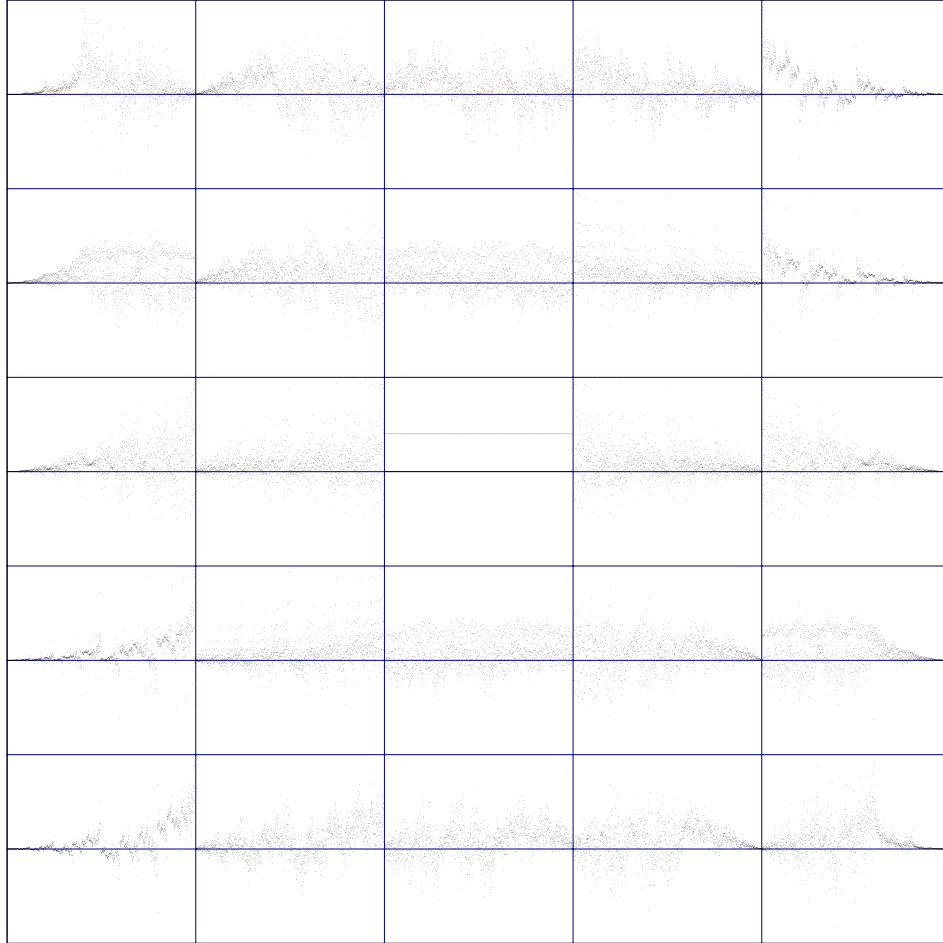
Figure 1.6. Chaos. Wavelets around $\theta = \rho = \pi/2$. L^2-convergence of the cascade approximation and the spectral condition (2.5.3) fails spectacularly at the central Haar function. See also Figure 1.1 for the same phenomenon.

sult says that two scaling/wavelet function N-tuples $\varphi, \psi_1, \ldots, \psi_{N-1}$ with support in $[0, kN + 1]$ can be deformed continuously into each other through scaling/wavelet functions with support in $[0, N^2 k + 1]$ (or any longer interval) if and only if the two winding numbers are equal. This divides the space into exactly $N(N-1)k + 1$ components.

In the examples illustrated in Figures 1.1–1.7 it follows from (1.2.5) that

$$\det(A(z)) = -z^2, \tag{1.2.15}$$

and hence the winding-number invariant in Theorem 2.1.3 is 2. The variety illustrated is a two-dimensional subvariety of $\mathrm{MF}(2) \leftrightarrow \mathrm{WF}(5) \leftrightarrow \mathrm{SF}(5)$; see (2.1.29)–(2.1.35).
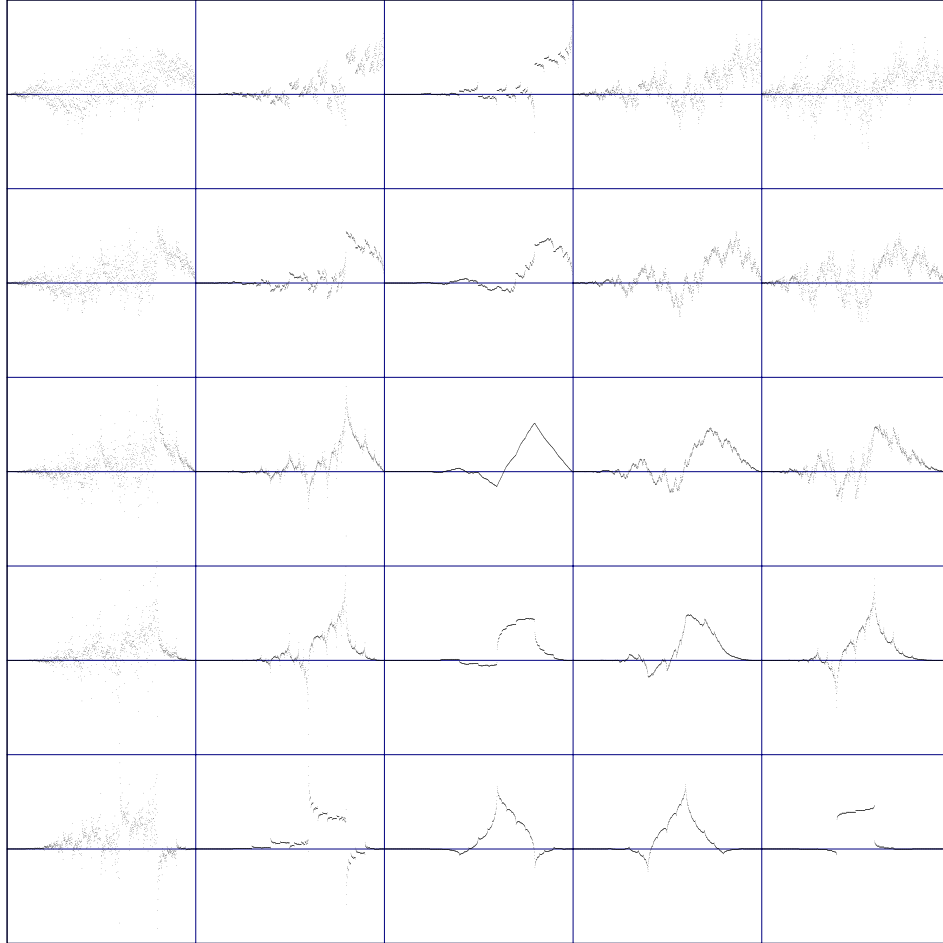
Figure 1.7. Wavelets around the ultrasmooth point $\theta = \cos^{-1} \sqrt[4]{\frac{5}{32}} \approx 0.28\,\pi$, $\rho = \cos^{-1} \sqrt{\frac{5}{4} - \sqrt{\frac{5}{32}}} \approx 0.12\,\pi$. This is the only point, up to symmetry, where the corresponding wavelet function ψ has two vanishing moments. See also [Tre01b, Figure 2], and the latter half of the tutorial to Chapter 3.

This is indeed a subvariety since we restrict to real projections Q_θ in (1.2.7). The space of one-dimensional projections $M_2\,(\mathbb{C})$ is actually homeomorphic to the 2-sphere S^2 [BrRoI, Example 4.2.7], so the variety really has dimension 4.

The continuity of the scaling/wavelet functions as a function of the parameters is in the distribution sense, but generically, outside a subvariety of positive codimension in the parameter space, the continuity is in the $L^2\,(\mathbb{R})$ sense (see Theorem 2.5.8). The latter part

of the monograph, Chapters 4 and 5, is mostly devoted to the study of the singular points in the parameter space where $L^2(\mathbb{R})$-continuity fails. These are the points where the operator R_{m_0} in (1.19) has 1 as a degenerate eigenvalue or R_{m_0} has other eigenvalues of absolute value 1.

In (1.2.1)–(1.2.2) we identified a variety of wavelet filters corresponding to compactly supported wavelets in $L^2(\mathbb{R})$ as an infinite-dimensional semigroup of matrix functions. Much of the theory also applies to the $L^2(\mathbb{R}^d)$ situation. The variety is defined by inequalities, and the orthogonal wavelets correspond to the "boundary" in the sense that the inequalities are identities. The space of biorthogonal wavelets, or equivalently, the space of dual wavelet bases, then is the full variety. The full variety may be realized as matrix functions defined on a torus and taking values in the invertible N-by-N matrices, where N is the scaling of the wavelet at hand. In this representation, the orthogonal wavelets correspond to the matrix functions taking values in $\mathrm{U}_N(\mathbb{C})$.

1.3   Qubits: The oracle of Feynman and the algorithm of Shor

HOW TO GET STARTED: This section, which is somewhat independent from the rest of the book, is meant as an introduction to some issues of quantum computing in connection with wavelet analysis. The reader who gets stuck with the new ideas may wish to review some basic concepts of quantum computing elsewhere. While there is a wealth of more specialized references, H. Pollatsek [Pol01] has just offered a fresh but gentle primer to key mathematical concepts which are used in this section. Our presentation below is also addressed to the novice and it includes much motivation. It is meant as an invitation to an exciting subject whose use in wavelet analysis is so far in its infancy.

While the implementation of the wavelet computations is a vast subject, and we are only able here to touch some high points, it has recently been recognized that computations at the atomic level, i.e., one-atom-per-bit scale, are governed by different kinds of algorithms than classical computations. Granted that physical quantum computers have not yet been built, it is nonetheless understood at a mathematical level that prime factorization can now be done in the quantum realm with polynomial algorithms, while this was known not to be possible on classical computers; see Shor [Sho99]. (Analogously, Grover's theorem [Gro97] states that a quantum search among n objects can be done in a number of steps which goes like a constant times $\sqrt{n}$, a clear gain over the best known classical bound, which is linear in n.) We note here, very sketchily, that the wavelet resolution algorithms mentioned in the tutorial to Chapter 1 share this same efficiency gain.
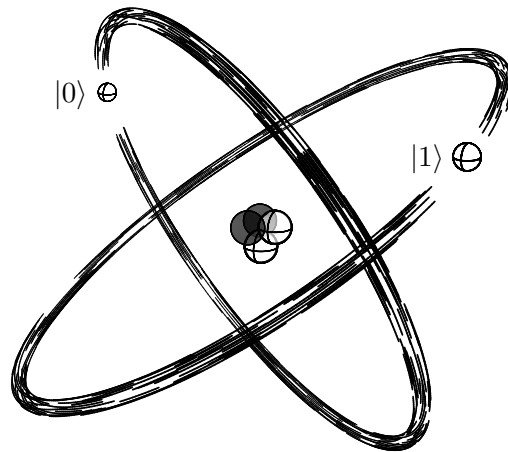
Figure 1.8. Qubits. A two-electron level in an atom (the Rutherford–Bohr model; simplified!) Each electron in the level can exist in either a spin down state $|0\rangle$ or a spin up state $|1\rangle$, thus giving rise to the simplest known toy quantum computer, see [Sho00]. If the atom is placed in a magnetic field pointing down, $|0\rangle$ is the ground state and $|1\rangle$ is the excited state for the internal degrees of freedom of the electron.

The quantum versions of the wavelet resolution algorithms involve the factorization of a unitary matrix built from a given wavelet filter; see (1.3.25) and Exercise 1–22. Readers who wish to follow up on the subject may consult the references we provide below. In addition there is a survey paper from 1998–99 [FiWi99] which is recommended reading. The discussion here is limited to an application of the spin-vector formalism in Section 1.2, which will also be followed up in Chapter 2 below.

Let us first give an overview of the terminology: the dictionary is shown in Table 1. The details will be explained later in the section. The starting point is that wavelet resolution algorithms can be made and implemented both on classical and on quantum computers.

Now, returning to the description of filters in terms of matrix-valued functions from the circle $\mathbb{T}$ into $U_N(\mathbb{C})$ (which we mentioned in Section 1.2, and will treat in more detail in (2.1.11)–(2.1.28)), one may give factorizations of the $U_N(\mathbb{C})$-matrix functions which produce qubit algorithms, with the factors of the $U_N(\mathbb{C})$-functions corresponding to the spin vectors of the quantum algorithm.

The qubit algorithms, i.e., factorization into a string of elementary quantum gates, may be realized as a factorization of a unitary matrix U_A, (1.3.24), of size $2^n \times 2^n$. Varying the spin vectors, one is then able to trace the corresponding wavelets and study their $L^2(\mathbb{R})$-properties. The qubit terminology refers to the bits of quantum computer algorithms, in the sense of R. Feynman [Fey86], [Fey99], [Mil98] and P. Shor [Sho99], [Sho00], [BDMT98]. Its use in wavelet analysis is discussed in [Jor01b], [Tre01b], and in much

Table 1. Wavelet resolution algorithms

	Classical	Quantal
Words of programs	logic gates: AND, OR, XOR, ...	qubit gates: one-qubit gates and CNOT
Program acts on register of	bits	qubits
Input/output:		
encoded	configuration of bits in a register of a Turing machine	configuration of qubits in a quantum channel
decoded	wavelet coefficients $\rightarrow$ signal or image	
Input in wavelet resolution algorithms	A (qu)bit configuration resulting from encoding a given function f chosen from a given resolution subspace.	
Output in wavelet resolution algorithms	A decomposition of the function f in terms of a finite number of intermediate detail components, or equivalently a wavelet decomposition; see the illustration (1.8) in the tutorial of Chapter 1, and also Section 2.2 and Exercise 1–22 for the Haar and the Daubechies wavelets.	
Names of algorithms	pyramid algorithm, Exercise 1–12	U_A, (1.3.24) and p. 36
Factorizations	(1.2.5), (2.1.6)	Exercise 1–22
Error correction	$x \mapsto xxx$ and omit broken strings	protect against decoherence, (1.3.23) and Exercise 1–14

detail in [Kla99]. Quantum computing algorithms involve the factorization of unitary matrices into products of special unitary matrices which are called *quantum gates*. In [Kla99] it is shown how the wavelet qubits yield quantum algorithms and the paper [VoWe00] shows how the two-qubit configuration is special. The gate structure is much easier. See for example (1.3.24) and Exercise 1–22.

The algorithm for reducing wavelet filters to quantum gates is based on two related factorization theorems, one for wavelet filters (see [BrJo01a, Proposition 3.3] or (2.1.6)), and a second one for quantum gates, which we now briefly sketch. In a classical computer, information is represented by binary symbols 0 and 1 (bits), and the bits are manipulated using functions such as AND and NOT. By contrast, a quantum bit, or "qubit", is a microscopic system such as an electron with its spin, or a polarized photon. The Boolean states are then qubit vectors in $\mathbb{C}^2$, $|0\rangle \leftrightarrow$ spin $+\frac{1}{2}$, and $|1\rangle \leftrightarrow$ spin $-\frac{1}{2}$ (see Figure 1.9). For a nice survey for mathematicians, see, e.g., [Man00].

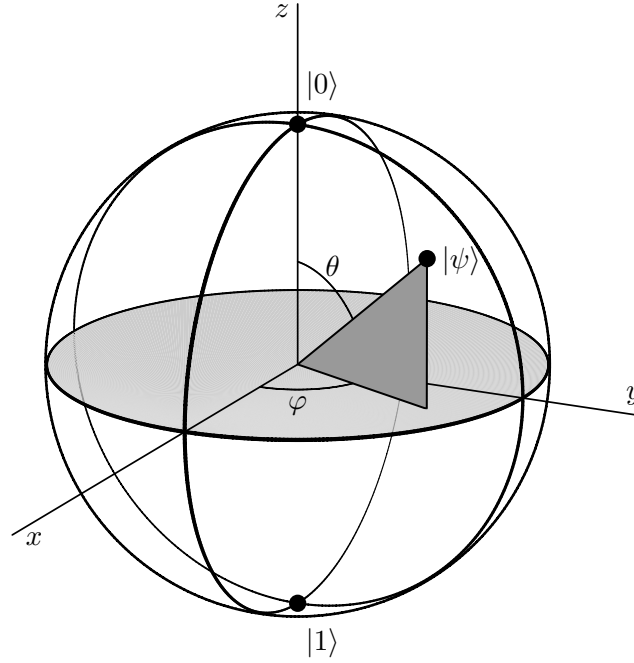Figure 1.9. The Bloch ball of states $\rho(x,y,z) = \frac{1}{2}(\sigma_0 + x\sigma_x + y\sigma_y + z\sigma_z)$ where $\sigma_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $\sigma_x = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\sigma_y = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$, and $\sigma_z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ are the Pauli matrices. The state $\rho(x,y,z)$ is a positive trace-class operator and $\mathrm{Trace}(\rho(x,y,z)) = 1$ if and only if $(x,y,z) \in \mathbb{R}^3$ and $x^2 + y^2 + z^2 \leq 1$, and all such operators have this form (see [BrRoI, Example 4.2.7]). Illustration: $|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\theta}{2}\right)|1\rangle$, where $|0\rangle \sim \frac{1}{2}(\sigma_0 + \sigma_z) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) = |0\rangle\langle 0|$ and $|1\rangle \sim \frac{1}{2}(\sigma_0 - \sigma_z) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) = |1\rangle\langle 1|$. Linear combinations of the vectors $|0\rangle$ and $|1\rangle$ do not correspond to the same linear combinations of the corresponding states. But these linear combinations, called superpositions, represent coherent states in the sense of quantum optics, see [Omn99], and account for a speedup of quantum computing algorithms.
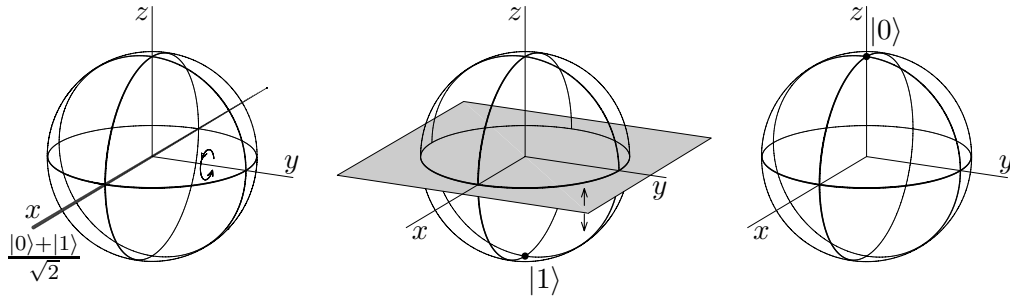

Figure 1.10. The quantum gate $\frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ acting on $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ as a rotation followed by a reflection, i.e., $\frac{|0\rangle+|1\rangle}{\sqrt{2}} \to |1\rangle \to |0\rangle$. A coherent state turning into "classical" ones.

However, according to the laws of quantum mechanics, the states can also exist in a continuum of intermediate or mixed configurations, or superpositions

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \qquad |\alpha|^2 + |\beta|^2 = 1, \tag{1.3.1}$$

which are called *coherent states* if $\alpha\beta \neq 0$; for a concise account of the physics of coherent states, see, e.g., [Omn99, Chapter 17, pp. 196, 227]. Note that the projection $E_\psi = |\psi\rangle \langle\psi|$ has the matrix form

$$\begin{pmatrix} |\alpha|^2 & \alpha\bar{\beta} \\ \bar{\alpha}\beta & |\beta|^2 \end{pmatrix}. \tag{1.3.2}$$

So it follows that the state $|\psi\rangle$ is coherent precisely when the two off-diagonal matrix entries are nonvanishing. (See Exercise 1–31 for the role of the diagonal and the off-diagonal entries of the matrix (1.3.2).) Generally, here *quantum decoherence* means the collapse of some quantum superposition into a single definite state, which presumably is a classical state (see below). For an electron in a magnetic field, or for a polarized photon, it can be the ground state or the excited state. If a state $|\psi\rangle$ is a superposition of, say, $|0\rangle$ and $|1\rangle$, then $|\psi\rangle$ acts as if it can exist simultaneously in these two states.

Quantum decoherence in a system of quantum gates decreases the information, and so must be avoided. The word "coherence" refers to our understanding of quantum states as waves, and hence we may have coherence of phases. This view is possible in any family of nonorthogonal states. In contrast, decoherence results from the selection of a set of special coordinates. States sent through a quantum channel suffer decoherence because of noise in the environment.

A string of n qubits (or an n-qubit memory) may be represented as a unit vector in $\underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n \text{ times}} \simeq \mathbb{C}^{2^n}$, i.e., a unit vector $\|v\| = 1$ in a 2^n-dimensional complex Hilbert space, $v \in \mathbb{C}^{2^n}$. A particular set of basis vectors in $\mathbb{C}^{2^n}$ is the set of tensor monomials of the special form $|i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_n\rangle$ for $i_1, i_2, \ldots, i_n \in \{0, 1\}$, for which the abbreviation $|i_1, i_2, \ldots, i_n\rangle$ is commonplace. These particular tensor monomials and their scalar multiples are said to represent the *classical states*, while all other unit vectors are said to represent *coherent states*.

For $n = 2$, there are some operations in quantum computing algorithms which do not have classical counterparts, such as CONTROL-NOT (CNOT), NOR, NAND; see [BDMT98, p. 48]. The tensors in $\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$ of unit norm are called *registers*; and the unitaries on the individual tensor factors, i.e., $\mathsf{U}_2(\mathbb{C})$, are called the single-qubit (quantum) gates. (See Figures 1.9 and 1.10.) The operations, called gates, are the unitary operators in $\underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n \text{ times}} \simeq \mathbb{C}^{2^n}$.

The important fact about the tensor operation is that it is multilinear, but not linear: hence the *no-cloning theorem* of quantum computation, which will be explained in the

paragraphs around (1.3.19)–(1.3.23) below. The simplest nonclassical gate is a quantum version of what in the classical case is called an XOR gate; see [NiCh00, p. 179]. Labeling the basis vectors in $\mathbb{C}^2$ by the elements in the cyclic group $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z} = \{0,1\}$, and using addition in $\mathbb{Z}_2$, the CNOT gate may be written in the following form:

$$
\begin{aligned}
|0,0\rangle &\longmapsto |0,0\rangle\,, \\
|0,1\rangle &\longmapsto |0,1\rangle\,, \\
|1,0\rangle &\longmapsto |1,1\rangle\,, \\
|1,1\rangle &\longmapsto |1,0\rangle\,,
\end{aligned}
\quad \text{or as a matrix,} \quad
\left(\begin{array}{cc|cc}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{array}\right). \tag{1.3.3}
$$

relative to the basis $|0,0\rangle$, $|0,1\rangle$, $|1,0\rangle$, $|1,1\rangle$. The square of this matrix is the identity matrix, i.e., $(\mathrm{CNOT})^2 = 1_{\mathbb{C}^4}$.
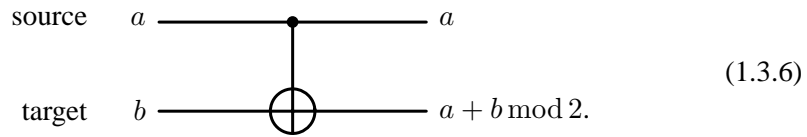
Specifically, on basis vectors, CNOT is

$$
|a\rangle \otimes |b\rangle \longmapsto |a\rangle \otimes |a+b\rangle\,, \tag{1.3.4}
$$

or in abbreviated form,

$$
|a,b\rangle \longmapsto |a, a+b\rangle\,, \tag{1.3.5}
$$

where addition in $\mathbb{Z}_2$ is used in the second tensor slot. Its operational form is

$$(1.3.6)$$

So the CNOT gate is not a single-qubit gate. If the highest significant bit (the control bit) is 1, then the state of the lowest significant bit (the rightmost tensor slot) is flipped; hence the name CNOT for this gate. The "target/source" terminology comes from the notion that a "target" is something that you hit from a "source" of ammunition. The result of hitting b, in this metaphor, is that the bit is flipped.

Compare the CNOT gate in (1.3.6) to the familiar exclusive "or" logic gate, called XOR, and written as a diagram


$$(1.3.7)$$
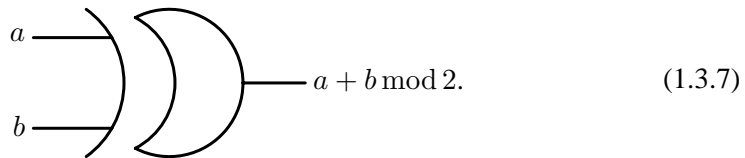
See also Table 2 below. The main difference between the two is that CNOT is invertible, while XOR is not. This means that information (or energy) is not lost in the running of a quantum computer (—it doesn't need to be plugged in!), while the classical computer increases entropy. Information is lost in the logic gate (1.3.7) since the two input bits cannot be recovered from the output $a + b \bmod 2$.

The diagrammatic notation (1.3.6) of the CNOT gate is usually abbreviated to

$$(1.3.8)$$

and the gate acts on $\mathbb{C}^2 \otimes \mathbb{C}^2$ as a permutation of the four base states $|i_1 i_2\rangle$, $i_1, i_2 \in \{0, 1\}$; see (1.3.3). So in physics terms, it is thought of as being realized on a two-electron model, as sketched for example in Figure 1.8. The triple quantum gate diagram

$$(1.3.9)$$
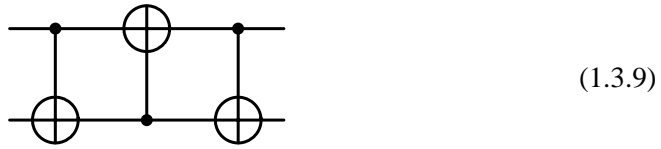
can easily be checked to represent a composition of three gates with the composite diagram resulting in the *qubit swap gate*

$$|a\rangle \otimes |b\rangle \longmapsto |b\rangle \otimes |a\rangle , \qquad (1.3.10)$$

compare (1.3.5), and it is denoted

$$(1.3.11)$$
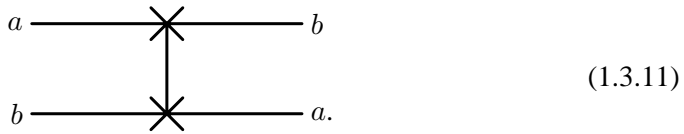
For more details, see Exercises 1–13 and 1–15.

An example of an n-qubit gate which is also a classical n-bit gate is a permutation π in the form

$$|i_1 i_2 \cdots i_n\rangle \mapsto |\pi_1 (i_1) \pi_2 (i_2) \cdots \pi_n (i_n)\rangle , \qquad (1.3.12)$$

or written more fully,

$$|i_1\rangle \otimes \cdots \otimes |i_n\rangle \mapsto |\pi_1 (i_1)\rangle \otimes \cdots \otimes |\pi_n (i_n)\rangle . \qquad (1.3.13)$$

Using then simple geometry of complex unitary matrices, it can be shown that every unitary operator on $\mathbb{C}^{2^n}$, the n-register, can be realized in an explicit form by a composition of single-bit gates and the CNOT gate; i.e., as an operator product involving only these basic operators on $\mathbb{C}^{2^n}$. This is the reason why our formula for the wavelet filters, as polynomial functions $\mathbb{T} \to U_2(\mathbb{C})$, admits efficient quantum computer algorithms; see [Kla99], [FiWi99], [Hol96], and [Jor01b]. Specifically, while $n2^n$ operations would be needed classically in the product representation, only n elementary operations are needed on the quantum computer. The result on the universal quantum gates mentioned

above is proved in detail in [BBC+95], i.e., mathematically, that all unitary operations on n qubits can be expressed in an explicit form as compositions of elementary gates, i.e., as one-qubit gates, and the single two-qubit gate CNOT. (There are other multiple-qubit gates which, when when taken together with certain one-qubit gates, form a complete alphabet for all unitary gates on finite quantum registers. We say that such multiple-qubit gates are *universal*, and the best known examples of universal gates are the CNOT and the Toffoli gates; see Table 2.) It is further proved in [Kla99] that as a consequence the complexity of the corresponding wavelet transform in the qubit algorithm is $\mathcal{O}\big(\log^2 L\big)$ where $L = 2^n$ is the length of the input signal [Kla02]. This in turn is based on the theorem of [BBC+95]. The crucial property of the wavelet filters m_0, m_1, in the form of a QMF system, is that, if $m_0(z) = \sum_{k \in \mathbb{Z}} a_k z^k$, $m_1(z) = \sum_{k \in \mathbb{Z}} b_k z^k$, with masking coefficients (a_k) and detail coefficients (b_k), then, for any orthonormal basis $\{\varepsilon_k\}_{k \in \mathbb{Z}}$, the transformation rules

$$\begin{cases} \xi_{2k} = \displaystyle\sum_{l \in \mathbb{Z}} a_{l-2k} \varepsilon_l, \\ \xi_{2k+1} = \displaystyle\sum_{l \in \mathbb{Z}} b_{l-2k} \varepsilon_l \end{cases} \tag{1.3.14}$$

get us a new orthonormal basis (ONB) for $\ell^2(\mathbb{Z})$, hence a permutation of ONB's, i.e., $U \colon (\varepsilon_k) \mapsto (\xi_k)$. If operators S_0 and S_1 on $L^2(\mathbb{T})$ are defined by

$$(S_i f)(z) = m_i(z) f\big(z^2\big), \qquad f \in L^2(\mathbb{T}), \; z \in \mathbb{T}, \; i = 0, 1, \tag{1.3.15}$$

then the fact that (1.3.14) permutes ONB's in $\ell^2(\mathbb{Z})$ is equivalent to the usual conditions for the QMF's and they take the form of the following familiar operator relations:

$$S_i^* S_j = \delta_{i,j} I, \qquad \sum_i S_i S_i^* = I \tag{1.3.16}$$

(the Cuntz relations [Cun77], see Exercise 1–11). Hence there is a single unitary operator U of $L^2(\mathbb{T})$ onto itself which is induced by the basis change (1.3.14) such that U commutes with $z \mapsto z^2 f(z)$. The Hilbert space $L^2(\mathbb{T})$ is isomorphic to $\ell^2(\mathbb{Z})$ via the Fourier series representation, see (1.36)–(1.38), and the corresponding operator $\hat{U}$ on $\ell^2(\mathbb{Z})$ commutes with the 2-shift $(x_k) \mapsto (x_{k-2})$. The classical code for the example (1.2.5) in the previous section is

$$\tag{1.3.17}$$

and the quantum version is built by turning the unitary matrix factors into quantum gates, as described below, and then using Shor's theorem on the universality of the CNOT gate. Keep in mind that in the single-qubit case, see Figure 1.8, according to

the standard rules of quantum mechanics, a measurement of the system will show that the system is in state $|0\rangle$ or in state $|1\rangle$. For the measurement, we are up against Heisenberg's uncertainty principle. Wait, you say, what then about the state $|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$? When measured, it gives 0 with probability $\cos^2\frac{\theta}{2}$, and 1 with probability $\sin^2\frac{\theta}{2}$, while the phase $e^{i\varphi}$ is immaterial. In information theory, as opposed to quantum mechanics, it is customary to say that the entropy h_θ of the state $|\psi_\theta\rangle$ is $h_\theta = -\left(\cos^2\left(\frac{\theta}{2}\right)\log_2\cos^2\left(\frac{\theta}{2}\right) + \sin^2\left(\frac{\theta}{2}\right)\log_2\sin^2\left(\frac{\theta}{2}\right)\right)$. The entropy looks as follows:
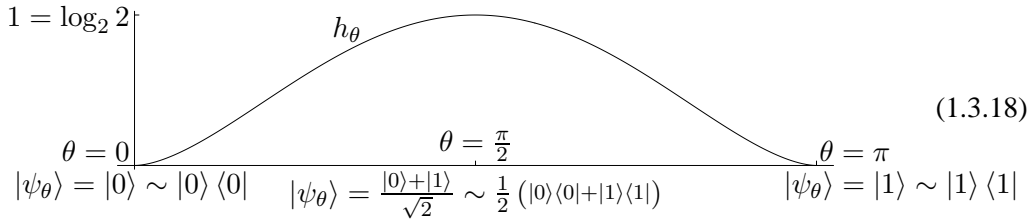
$$(1.3.18)$$

The entropy h_θ refers to the loss of information when the state is subjected to measurement or to noise in the channel. In (1.3.18), $\sim$ refers to a measurement in the sense of von Neumann; see [Omn99, p. 259]. So as the entropy decreases, coherent quantum states turn to classical states. This is "classical" in the sense of "how we are used to bits behaving in a computer, coming out reliably 0 or 1". Usually the term "coherent state" is reserved for the states depicted in (1.3.18) for $0 < \theta < \pi$, and so excluding the endpoints, but of course the two degenerate states $|0\rangle$ and $|1\rangle$ are both classical and quantum states.

In quantum mechanics, as opposed to quantum information theory, the entropy of a vector state is zero. More generally, consider the density matrices $\rho_p = p|0\rangle\langle0| + (1-p)|1\rangle\langle1|$, where $|0\rangle\langle0|$ and $|1\rangle\langle1|$ denote the projections of rank one onto $\mathbb{C}|0\rangle$ and onto $\mathbb{C}|1\rangle$, and $0 \le p \le 1$. The density matrices are interpreted probabilistically, and the quantum-mechanical entropy of the state $\mathrm{Trace}\,(\,\cdot\,\rho)$ is $-\mathrm{Trace}\,(\rho\log\rho)$. The extreme cases $p = 0$ and $p = 1$ are the points of decoherence (See Figure 1.9 and (1.3.18)). (The term "coherence" is from quantum optics.)

Even though we can measure just the values 0 or 1 when subjecting a coherent state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to observation, there is still a continuum of these intermediate (coherent) superpositions. The added computational power in passing from "classical" to "quantum" derives from this continuum of quantum states $|\psi\rangle$, as opposed to just the two classical bits. But the coherent states are highly unstable, and this must be taken very seriously when one tries to build quantum algorithms. The quantum error-correcting codes are there to protect coherence, preventing the coherent states from degenerating into decoherence; see [Omn99, p. 802]. Communication depends on a transmission with coherence, for example interference of the quantum-mechanical phases, which is independent of the distance a wave travels. This interference is recorded in the off-diagonal terms of some density matrix. The decay of these terms is called decoherence; see [SaSa01].

The fact which makes a (hypothetical) quantum computer more efficient than a classical one is that states can be prepared in configurations $a\,|0\rangle + b\,|1\rangle$ for $a, b \in \mathbb{C}$ which are superpositions of the two qubits $|0\rangle$ and $|1\rangle$; the property is called *quantum parallelism*, and of course it does not have a classical analogue. But this is also a fact which makes the quantum computer more sensitive to errors; see [Got96], [MaSl77a], [GrBe00], [CRSS98], [Ger00, p. 225]. And error correction by cloning (which is classical) is not possible in the quantum computer, precisely because of the nonlinearity of the tensor monomials, as opposed to the linearity of superposition.

Classically, the amplifier operator A, or the cloning assignment applied to a bit x, is $Ax = xx$, or $Ax = xxx$, $x \in \{0, 1\}$; see [PeWe72]. So if a test for possible errors yields the outcome 000, 010, and 000, then the register 010 is simply discarded. But imagine the quantum version of the cloning operator A, i.e., A applied to qubits $|\psi\rangle$. It would be $A\,|\psi\rangle = |\psi\rangle \otimes |\psi\rangle$, say, while quantum parallelism and unitarity dictate the linearity (up to phase): $A\left(\alpha\,|\psi_1\rangle + \beta\,|\psi_2\rangle\right) = \alpha A\,|\psi_1\rangle + \beta A\,|\psi_2\rangle$ for $\alpha, \beta \in \mathbb{C}$ and $|\psi_i\rangle \in \mathbb{C}^{2^k}$, $i = 1, 2$. But this is false because of the cross terms of the type

$$\alpha\beta\left(|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle\right) \tag{1.3.19}$$

resulting from the calculation on tensors in $\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathbb{C}^4$. This is the basis for the "no-cloning theorem" for qubits.

The way around the no-cloning theorem which does produce quantum error correction is related again to the S_i-systems of (1.3.15). Corresponding to the Haar (or rather Hadamard) system, Shor suggested the two isometries

$$A_\pm \colon \mathbb{C}^2 \longrightarrow \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \tag{1.3.20}$$

such that

$$A_\pm \colon \alpha\,|0\rangle + \beta\,|1\rangle \longrightarrow \alpha\,|000\rangle \pm \beta\,|111\rangle \tag{1.3.21}$$

for $\alpha, \beta \in \mathbb{C}$. Alternatively, a single linear amplification operator A defined by

$$\begin{aligned}
|0\rangle &\longmapsto \frac{1}{\sqrt{8}}\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right)\left(|000\rangle + |111\rangle\right), \\
|1\rangle &\longmapsto \frac{1}{\sqrt{8}}\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right)\left(|000\rangle - |111\rangle\right),
\end{aligned} \tag{1.3.22}$$

yields a quantum error-correcting operation, in the sense that it preserves the coherent states, and this is Shor's famous nine-qubit correction code. The individual tensor factors, such as $\frac{1}{\sqrt{2}}\left(|000\rangle + |111\rangle\right)$, are called "cat states" after Schrödinger's renowned cat [Omn99, p. 61]. They are some of the most unstable states formed out of several tensor factors of qubits. On the other hand, the argument from (1.8) shows that, unstable or not, they are an essential part of qubit wavelet algorithms.

More generally, let $k \in \mathbb{N}$, $k \geq 3$, and let $\mathcal{C} \subset \underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{k \text{ times}} =: \mathcal{H}_k$ be a subspace, and let P be the projection onto $\mathcal{C}$. Then $\mathcal{C}$ is called a *code* if there is a finite set of

operators $E_i \colon \mathcal{H}_k \to \mathcal{H}_{k \cdot l}$ and a matrix $(a_{i,j})$ such that

$$PE_i^* E_j P = a_{i,j} P. \tag{1.3.23}$$

Then the combined system is called a *quantum error-correcting code*. The idea is that in the quantum realm, we select *subspaces* rather than *bits*. This is how the coherence is protected. Similarly the quantum algorithms must be written as factorization of unitary operators, the factorizations taking the form of a sequence of quantum gates. For quantum error correction, we get subspaces $\{E_i \mathcal{C}\}$ from which we select one, say $E_0 \mathcal{C}$, which does not have errors. While the subspaces $E_i \mathcal{C}$ may not be orthogonal, there are numbers $a_{i,j}$ such that $\langle E_i u \,|\, E_j v \rangle = a_{i,j} \langle u \,|\, v \rangle$ for all $u, v \in \mathcal{C}$; see (1.3.23). The example in (1.3.21) is called the three-bit flip code, but there are others that are nontrivial and effective; see, for example, [CRSS98]. As in the classical case, the classical linear block codes, a quantum error-correcting code encodes the state of each qubit onto a block of, say, k qubits (a register, i.e., a vector in $\underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{k \text{ times}} = \mathbb{C}^{2^k}$). The encoded state is called a *logical* state following von Neumann [vNeu56]; see Exercise 1–14. The code is said to correct d errors if the logical state is recoverable, given that no more than d errors occurred in the k-block. The added difficulty in passing from classical codes to the quantum ones is that quantum *coherence* must be protected, the coherence referring to the continuum $\alpha \,|0\rangle + \beta \,|1\rangle$, $|\alpha|^2 + |\beta|^2 = 1$, as opposed to the classical duality of bits.

The simplest example of a quantum error-correcting code is in the case $k = 3$, where we take the encoded subspace $\mathcal{C} \subset \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$ to be the two-dimensional linear span of the two (logic) states $|000\rangle$ and $|111\rangle$. We then define four error-correcting operators E_i as follows, where $\sigma_x = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$:

$E_0 \colon$ no error $\;\to\;$ no qubit-flip $\;\to\; 1_{\mathbb{C}^2} \otimes 1_{\mathbb{C}^2} \otimes 1_{\mathbb{C}^2}$,

$E_1 \colon$ error on first tensor slot $\;\to\;$ flip first qubit $\;\to\; \sigma_x \otimes 1_{\mathbb{C}^2} \otimes 1_{\mathbb{C}^2}$,

$E_2 \colon$ error on second tensor slot $\;\to\;$ flip second qubit $\;\to\; 1_{\mathbb{C}^2} \otimes \sigma_x \otimes 1_{\mathbb{C}^2}$,

$E_3 \colon$ error on third tensor slot $\;\to\;$ flip third qubit $\;\to\; 1_{\mathbb{C}^2} \otimes 1_{\mathbb{C}^2} \otimes \sigma_x$.

In this case, the four subspaces $E_i \mathcal{C} \subset \mathbb{C}^8$, $i = 0, 1, 2, 3$, are mutually orthogonal, so the operator identity (1.3.23) is then clearly satisfied. The trouble with this error correction is that it is really just "classical", i.e., it only corrects (qu)bit-flips. For nontrivial quantum codes, see for example Exercise 1–14.

Let m_0, m_1 be a dyadic wavelet filter, and let $\mathbb{T} \ni z \mapsto A(z) \in \mathrm{U}_2(\mathbb{C})$ be the corresponding matrix function, $A_{i,j}(z) = \frac{1}{2} \sum_{w^2 = z} w^{-j} m_i(w)$. If the low-pass filter $m_0(z) = a_0 + a_1 z + \cdots + a_{2n+1} z^{2n+1}$, then a choice for $m_1(z) = \sum_{k=0}^{2n+1} b_k z^k$ is $b_k = (-1)^k \bar{a}_{2n+1-k}$. We then have $A(z) = \sum_{k=0}^{n} A_k z^k$ where $A_k = \begin{pmatrix} a_{2k} & a_{2k+1} \\ b_{2k} & b_{2k+1} \end{pmatrix}$,

and the following $2^{n+2} \times 2^{n+2}$ scalar matrix can be checked to be unitary:

$$
\left(
\begin{array}{c|cccccccccc|c}
\begin{matrix}a_1\\b_1\end{matrix} & A_1 & A_2 & \cdots & A_{n-1} & A_n & 0 & & \cdots & & 0 & \begin{matrix}a_0\\b_0\end{matrix} \\
\hline
\begin{matrix}0\\0\end{matrix} & A_0 & A_1 & \cdots & A_{n-2} & A_{n-1} & A_n & 0 & & \cdots & 0 & \begin{matrix}0\\0\end{matrix} \\
\hline
\begin{matrix}0\\0\end{matrix} & 0 & A_0 & \cdots & A_{n-3} & A_{n-2} & A_{n-1} & A_n & 0 & \cdots & 0 & \begin{matrix}0\\0\end{matrix} \\
\hline
\begin{matrix}0\\0\end{matrix} & & & & & & & & & & & \begin{matrix}0\\0\end{matrix} \\
\hline
\vdots & & & \ddots & & & & & \ddots & & & \vdots \\
\hline
\begin{matrix}0\\0\end{matrix} & & & & & & & & & & & \begin{matrix}0\\0\end{matrix} \\
\hline
\begin{matrix}a_{2n+1}\\b_{2n+1}\end{matrix} & 0 & & \cdots & & 0 & A_0 & A_1 & A_2 & \cdots & A_{n-1} & \begin{matrix}a_{2n}\\b_{2n}\end{matrix} \\
\hline
\begin{matrix}a_{2n-1}\\b_{2n-1}\end{matrix} & A_n & 0 & & \cdots & & 0 & A_0 & A_1 & \cdots & A_{n-2} & \begin{matrix}a_{2n-2}\\b_{2n-2}\end{matrix} \\
\hline
\begin{matrix}a_{2n-3}\\b_{2n-3}\end{matrix} & A_{n-1} & A_n & 0 & & \cdots & & 0 & A_0 & \cdots & A_{n-3} & \begin{matrix}a_{2n-4}\\b_{2n-4}\end{matrix} \\
\hline
\vdots & & & \ddots & & & & & \ddots & & & \vdots \\
\hline
\begin{matrix}a_3\\b_3\end{matrix} & A_2 & A_3 & \cdots & A_n & 0 & & \cdots & & 0 & A_0 & \begin{matrix}a_2\\b_2\end{matrix}
\end{array}
\right)
$$

Except for the scalar entries in the two extreme left and right columns, all the other entries of the big combined matrix U_A are taken from the cyclic arrangements of the 2×2 matrices of coefficients $A_0, A_1, \ldots, A_n$ in the expansion of $A(z)$. For the case of $n = 1$ this amounts to the simple 8×8 wavelet matrix

$$
\begin{array}{r}
A_0 \\
\left(
\begin{array}{c|cccc|c}
\begin{matrix}a_1\\b_1\end{matrix} & A_1 & 0 & 0 & \begin{matrix}a_0\\b_0\end{matrix} \\
\hline
\begin{matrix}0\\0\end{matrix} & A_0 & A_1 & 0 & \begin{matrix}0\\0\end{matrix} \\
\hline
\begin{matrix}0\\0\end{matrix} & 0 & A_0 & A_1 & \begin{matrix}0\\0\end{matrix} \\
\hline
\begin{matrix}a_3\\b_3\end{matrix} & 0 & 0 & A_0 & \begin{matrix}a_2\\b_2\end{matrix}
\end{array}
\right) \\
A_1,
\end{array}
\tag{1.3.24}
$$

which is the one that produces the sequence of quantum gates in Exercise 1–22. The quantum algorithm of a wavelet filter is thus represented by a $2^{n+2} \times 2^{n+2}$ unitary

matrix U_A acting on the quantum qubit register $\underbrace{\mathbb{C} \otimes \cdots \otimes \mathbb{C}}_{n+2 \text{ times}} = \mathbb{C}^{2(n+2)}$, i.e., it acts on a configuration of $n + 2$ qubits. The realization of a wavelet algorithm in the quantum realm thus amounts to spelling out the steps in factoring U_A into a product of qubit gates. By Shor's theorem, we know that this can be done, and U_A may be built out of one-qubit gates and CNOT gates following the ideas sketched above. The reader may find more discussion of the matrix U_A in [Fre00, Section 3]

The generalization of classical and quantum wavelet resolution algorithms from $N = 2$ to $N > 2$ is immediate: Then $m_i(z) = \sum_{k \in \mathbb{Z}} a_k^{(i)} z^k$,

$$(S_i f)(z) = m_i(z) f(z^N), \qquad i = 0, \ldots, N - 1, \tag{1.3.25}$$

and the transformation rules

$$\xi_{Nk+i} = \sum_{l \in \mathbb{Z}} a_{l-Nk}^{(i)} \varepsilon_l, \qquad i = 0, 1, \ldots, N - 1, \tag{1.3.26}$$

permute the set of ONB's in $\ell^2(\mathbb{Z})$ and define a unitary commuting with the N-shift. Hence, the standard formulas from [Wic93], [Kla99], and [FiWi99] for the quantum computing algorithm, which are based on (1.3.14), naturally generalize to the case $N > 2$ via (1.3.26). Instead of k-registers $\underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{k \text{ times}} = \mathbb{C}^{2^k}$ over $\mathbb{C}^2$, we will now have to work rather with $\underbrace{\mathbb{C}^N \otimes \cdots \otimes \mathbb{C}^N}_{k \text{ times}} = \mathbb{C}^{N^k}$.

The use of the relations (1.3.16) in engineering and operator algebra theory predates their more recent use in wavelet theory and wavepacket analysis.

Remark 1.3.1. The simplest way of creating entanglement of states, for example two-qubit states of the form $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, are used in electron models (spin $\frac{1}{2}$ particles), see Figure 1.8. (A vector state in $\bigotimes_1^k \mathbb{C}^2$ for some $k > 1$ is said to be *entangled* if it cannot be written as a tensor product of single-qubit states. Entangled states are essential for error correction, and for teleportation, among other things.) Entanglement is generally believed to be one of the essential quantum effects responsible for speedup, i.e., converting slow classical gate models to polynomial quantum gate algorithms. See [Omn99, p. 274]. For universal gates, such as the CNOT gate, we need at least two electrons. But it was recently suggested [MeyD00], [Llo00] that quantum search might be realized without entanglement. To do this [MeyD00] suggests an experiment with single particles having exponentially many states, so not electrons. However, for wavelet models, this is not feasible. The spin is fixed by the scale number N, and so if $N = 2$, we are stuck with spin $\frac{1}{2}$, see Figure 1.8.

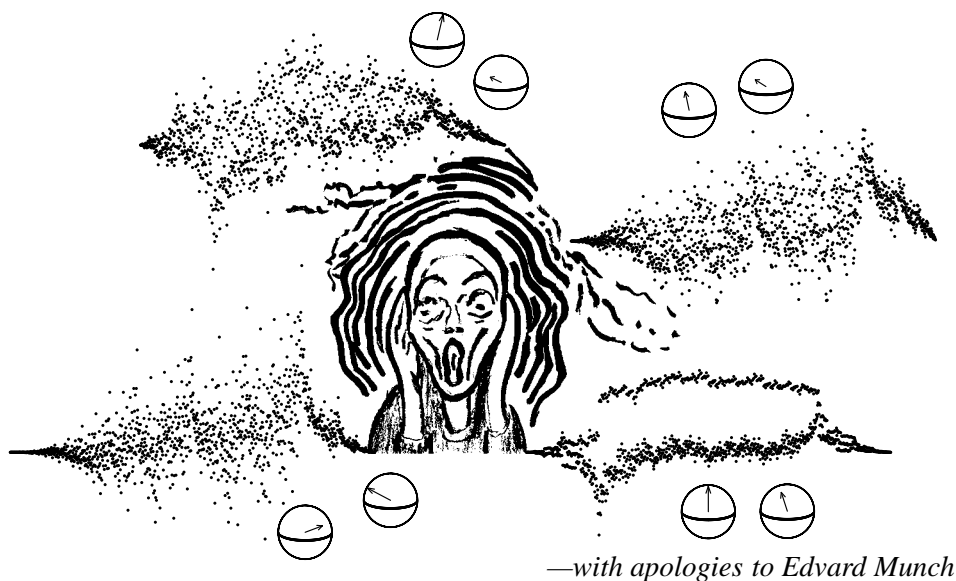*—with apologies to Edvard Munch*

Figure 1.11. Cascading refinements. For the link between spin configurations and scaling functions, see the end of Section 1.4 and Chapter 2.

1.4   Chaos and cascade approximation

The dependence of the $L^2(\mathbb{R})$ theory on the spin vectors exhibits chaos features, in the sense that a small variation of the spin-vector configurations produces "large swings" in the wavelet generators. We will apply a probabilistic approach, in Sections 2.5, 4.4, and 5.1, in trying to take an average of the unpredictable chaos in the cascades of wavelets. The *singular points* on the manifold WF (D) of (2.1.30) of wavelet filters to which we refer are defined relative to the transfer or Ruelle operator $R = R_{m_0}$ defined in (1.19), where m_0 is a point in the variety of filters. The low-pass filter m_0 is said to be a singular point if the multiplicity of the corresponding Perron–Frobenius eigenvalue $\lambda = 1$ is more than one or if R_{m_0} has eigenvalues in $\mathbb{T}$ different from 1. For the examples in Figures 1.1–1.7, this only happens at the points m_0 corresponding to the stretched Haar wavelets, see Figures 1.1 and 1.6, or (1.2.11)–(1.2.12). Let χ be the indicator function (1.22) above. The probabilistic view of the chaotic trajectory $\mathbb{T}^2 \ni (\theta, \rho) \mapsto \varphi^{(\theta,\rho)} \in L^2(\mathbb{R})$ in Figures 1.1 and 1.6 above refers to the approach of determining the cascade approximation $\lim_{n\to\infty} M^n_{A(\theta,\rho)}(\chi) = \varphi^{(\theta,\rho)}$ near one of the chaotic points $(\theta, \rho) = \left(\frac{\pi}{2}, \frac{\pi}{2}\right)$ or $(0,0)$ (or $\left(\frac{\pi}{2} \pm \frac{\pi}{4}, \frac{\pi}{2}\right)$, not shown), not as a direct $L^2(\mathbb{R})$-limit but rather as a suitable average. Note that the $L^2(\mathbb{R})$-limit in fact does not exist at the singular points themselves (Theorem 2.5.1), and that the scaling function $\varphi^{(\theta,\rho)}$, near the singular points, has the least amount of smoothness, as a function, $x \mapsto \varphi^{(\theta,\rho)}(x)$. So the Cesaro averages

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} M^k_{A(\theta,\rho)}(\chi)(x) = \varphi^{(\theta,\rho)}(x) \qquad (1.4.1)$$

may be used instead in a neighborhood in $\mathbb{T}^2$ of the singular points, and it is proved in Corollary 4.5.6, Remark 2.5.5, and Lemma 2.5.7 that (1.4.1) in fact is $L^2(\mathbb{R})$-convergent, also at the singular points.

As explained in (1.2.7), the two unit vectors $\left(\begin{smallmatrix}\cos\theta\\\sin\theta\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}\cos\rho\\\sin\rho\end{smallmatrix}\right)$ correspond to unitary matrix factors in the formula (1.2.5). A matrix function in turn produces masking coefficients by Section 1.2 and therefore a scaling function for a wavelet. This explains how a configuration of two spin points (i.e., a pair of points on S^2) produces a scaling function as one of the four around the head in in Figure 1.11. More generally, unit vectors $\left(\begin{smallmatrix}u\\v\end{smallmatrix}\right) \in \mathbb{C}^2$, $|u|^2 + |v|^2 = 1$, define pure states on the 2×2 complex matrices, and these states in turn are indexed by points on the Bloch sphere S^2 of Figure 1.9. The point on S^2 corresponding to $u\,|0\rangle + v\,|1\rangle$ can be checked to be $(2\,\mathrm{Re}\,(\bar{u}v)\,, 2\,\mathrm{Im}\,(\bar{u}v)\,, |u|^2 - |v|^2)$. So if $u = \cos\theta$, $v = \sin\theta$, then the corresponding point on S^2 is $(\sin 2\theta, 0, \cos 2\theta)$. For more details, see Exercise 1–18.

1.5   Spectral bounds for the transfer and subdivision operators

In Sections 3.2 and 3.4 we study general spectral properties of operators of the form (1.19), or rather more generally (3.2.1), i.e., $|m_0(z)|^2$ is replaced by a general Fourier polynomial $W(z)$. In this case there exists a finite-dimensional space $\mathcal{K}_s$ of Fourier polynomials such that R_{m_0} maps $\mathcal{K}_s$ into $\mathcal{K}_s$; see (3.2.19). For spaces of very regular functions on $\mathbb{T}$, the spectrum of R_{m_0} then coincides with the spectrum of $R_{m_0}|_{\mathcal{K}_s}$; see (3.2.22) and Proposition 3.5.1. For spaces of less regular functions, like $C(\mathbb{T})$ or $L^2(\mathbb{T})$, the spectrum of R_{m_0} in addition contains a disc consisting of eigenvalues; see Theorem 3.2.6, Proposition 3.4.6, and Remark 3.4.7. Nevertheless, if m_0 is a Fourier polynomial, the peripheral point spectrum of R_{m_0} on $C(\mathbb{T})$ is equal to the peripheral spectrum of $R_{m_0}|_{\mathcal{K}_s}$, and the corresponding points in the spectrum have all their eigenvectors in $\mathcal{K}_s$; see Theorem 5.5.4. In the case of $L^2(\mathbb{T})$, none of the peripheral points in the spectrum are eigenvalues; see Corollary 4.2.18. In the case of $L^\infty(\mathbb{T})$, R_{m_0} may have peripheral eigenvectors which are not in $\mathcal{K}_s$, although they are analytic except for one jump discontinuity; see Example 3.5.5.

In Section 4.3 and Chapter 5 we study more closely the eigenspace corresponding to $\lambda = 1$ of the transfer operator. The focus is different from that of Section 3.2, which was L^2-theory. In Sections 4.3 and 5.2, we are concerned with the real-valued continuous functions on $\mathbb{T}$ and their order structure. We show (Proposition 5.3.1) that the transfer operator itself can be obtained from its eigenfunctions in the case that the peripheral spectral properties are nongeneric. Before that, we give in Corollary 4.5.6 a measure of the size of the peripheral spectrum.

Let $R = R_W$ be defined from a polynomial wavelet filter m_0 and $W = |m_0|^2$. If

$$m_0(1) = \sqrt{N} \quad \text{and} \quad \frac{1}{N}\sum_{w^N=z}|m_0(w)|^2 \le 1, \tag{1.5.1}$$

then the eigenspace $\ker\left(1 - R|_{C(\mathbb{T})}\right)$ contains two distinguished functions which are Fourier polynomials. They are studied in Theorems 4.3.1 and 5.1.1. Of course if

$$\dim \ker\left(1 - R|_{C(\mathbb{T})}\right) = 1, \tag{1.5.2}$$

then there is just one function f in $C\left(\mathbb{T}\right)$ satisfying $R\left(f\right) = f$ and $f\left(1\right) = 1$, so the two functions coincide.

The two types of wavelet filters we consider in Section 4.3 are said to be of *orthogonal type* if

$$\frac{1}{N} \sum_{w^N = z} |m_0\left(w\right)|^2 = 1, \tag{1.5.3}$$

or of *biorthogonal type* if

$$\frac{1}{N} \sum_{w^N = z} |m_0\left(w\right)|^2 \le 1. \tag{1.5.4}$$

The two cases (1.5.3) and (1.5.4) are tied directly to the spectral theory of the transfer operator $R = R_W$ through the following observations: Let R_W^* be the formal adjoint,

$$\left(R_W^*\xi\right)\left(z\right) = W\left(z\right)\xi\left(z^N\right), \qquad \xi \in L^1\left(\mathbb{T}\right). \tag{1.5.5}$$

Then

- R_W^* is *isometric* in $L^1\left(\mathbb{T}\right)$ if and only if (1.5.3) holds, (1.5.6)

and

- R_W^* is *contractive* in $L^1\left(\mathbb{T}\right)$ if and only if (1.5.4) holds, where $L^1\left(\mathbb{T}\right)$ (1.5.7)
  is defined from the Haar measure on $\mathbb{T}$.

The operator in (1.5.5) is of independent interest in numerical analysis, where it is called the *subdivision operator* S; see, e.g., [GMW94]. It is considered in a sequence space, such as $\ell^p\left(\mathbb{Z}\right)$, $1 \le p \le \infty$, where the matrix representation is

$$\left(Sx\right)_i = \sum_{j \in \mathbb{Z}} c_{i - Nj} x_j \tag{1.5.8}$$

for $x = \left(x_j\right) \in \ell^p\left(\mathbb{Z}\right)$, $i, j \in \mathbb{Z}$, and $W\left(z\right) = \sum_{k \in \mathbb{Z}} c_k z^k$. Setting $N = 2$, the subdivision scheme (1.5.8) takes the following form of double subband filtering:

$$\left(Sx\right)_{2i} = \sum_{j \in \mathbb{Z}} c_{2j} x_{i - j}, \qquad \left(Sx\right)_{2i+1} = \sum_{j \in \mathbb{Z}} c_{2j+1} x_{i - j}, \tag{1.5.9}$$

which is more familiar in the context of the theory of stationary subdivision, see, e.g., [deR56], [CDM91], i.e., each of the two expressions in (1.5.9) is a convolution, but the first involves only the coefficients of even index, while the second is the analogous

weighted average, but with only the weights of odd index. It is used, for example, in the local cascade algorithm, see Figures 1.1–1.7.

The operator S $(= R_W^*)$ in (1.5.8) is called the subdivision operator, or the *woodcutter operator*, because of its use in computer graphics. Iterations of S will generate a shape which (in the case of one real dimension) takes the form of the graph of a function f on $\mathbb{R}$. If $\xi \in \ell^\infty (\mathbb{Z})$ is given, and if the differences

$$D_n (i) = f \left(\frac{i}{2^n} \right) - (S^n \xi) (i), \qquad i \in \mathbb{Z}, \tag{1.5.10}$$

are small, for example if

$$\lim_{n \to \infty} \sup_{i \in \mathbb{Z}} |D_n (i)| = 0, \tag{1.5.11}$$

then we say that ξ represents *control points*, or a control polygon, and the function f is the limit of the *subdivision scheme*.

It follows that the subdivision operator S on the sequence spaces, especially on $\ell^\infty (\mathbb{Z})$, governs *pointwise approximation* to refinable limit functions. But we will see in Theorem 2.5.1 that the dual version of S, i.e., $R = S^*$ (= the transfer operator) governs the corresponding *mean approximation* problem, i.e., approximation relative to the $L^2 (\mathbb{R})$-norm.

In Scholium 4.1.2, we will consider the eigenvalue problem

$$S\xi = \lambda \xi, \qquad \lambda \in \mathbb{C}, \tag{1.5.12}$$

and $\xi \neq 0$ in some suitably defined space of sequences. The formula (1.5.10) for the limit of a given subdivision scheme S makes it clear that the case (1.5.12) must be excluded. For if (1.5.12) holds, for some $\lambda \in \mathbb{C}$, and some sequence ξ of control points, then there is not a corresponding regular function f on $\mathbb{R}$ with its values given on the finer grids $2^{-n}\mathbb{Z}$, $n = 1, 2, \ldots$, by

$$f_\xi (i2^{-n}) \approx (S^n \xi) (i) = \lambda^n \xi (i), \qquad i \in \mathbb{Z}. \tag{1.5.13}$$

We will show in Example 4.1.3 that there are no such control points ξ in $\ell^2 (\mathbb{Z}) \setminus \{0\}$. Hence the stability of the algorithm!

We will analyze the duality between R_W and R_W^* and their spectra in Sections 4.1 and 4.2.

Setting $W = |m_0|^2$, we note that orthogonal type (referring to orthogonal wavelets) is (1.25) or (3.2.10), while biorthogonal type (referring to the wider class of biorthogonal wavelets) is (1.29) or (3.4.3) in Proposition 3.4.1 or (3.4.25) in Theorem 3.4.4. In the following, we shall refer to these two conditions in connection with a given W which is assumed only to satisfy (3.4.23), i.e., $W \in \mathrm{Lip}_1 (\mathbb{T})$, and (3.4.25), i.e., $W \geq 0$. Other conditions will then be added, such as (3.2.10), or (3.4.25).

While the eigenspace $\{g \in C (\mathbb{T}) \mid R (g) = g\}$ is not an algebra, we show in Sections 5.2, 5.3, and 5.4 that it carries a product in which it induces a finite-dimensional abelian

algebra, and we identify the positive multiplicative functionals on this system as the normalized counting measures on a family of cycles on $\mathbb{T}$. A cycle is a finite orbit on $\mathbb{T}$ under $z \mapsto z^N$. In the orthogonal or biorthogonal cases (1.5.3) or (1.5.4), these cycles are exactly the cycles contained in $\left\{ z \in \mathbb{T} \mid |m_0(z)|^2 = N \right\}$.

1.6   Connections to group theory

In this book, we stress the discrete wavelet transform. But the first line in the two tables of Exercise 1–43 below is the continuous one. It is the only treatment we give to the continuous wavelet transform, and the corresponding *coherent vector decompositions*. But, as is stressed in [Dau92], [Kai94], and [KaLe95], the continuous version came first. A function ψ satisfying the resolution identity is called a *coherent vector* in mathematical physics. The representation theory for the $(ax + b)$-group, i.e., the matrix group $G = \left\{ \left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix} \right) \mid a \in \mathbb{R}_+, \ b \in \mathbb{R} \right\}$, serves as its underpinning. Then the table in Exercise 1–43 illustrates how the $\{\psi_{j,k}\}$ wavelet system arises from a discretization of the following unitary representation of G:

$$\left(U_{\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix} \right)} f \right)(x) = a^{-\frac{1}{2}} f \left(\frac{x - b}{a} \right) \qquad (1.6.1)$$

acting on $L^2(\mathbb{R})$. This unitary representation also explains the discretization step in passing from the first line to the second in the table of Exercise 1–43. The functions $\{\psi_{j,k} \mid j, k \in \mathbb{Z}\}$ which make up a wavelet system result from the choice of a suitable coherent vector $\psi \in L^2(\mathbb{R})$, and then setting

$$\psi_{j,k}(x) = \left(U_{\left(\begin{smallmatrix} 2^{-j} & k \cdot 2^{-j} \\ 0 & 1 \end{smallmatrix} \right)} \psi \right)(x) = 2^{\frac{j}{2}} \psi \left(2^j x - k \right). \qquad (1.6.2)$$

Even though this representation lies at the historical origin of the subject of wavelets (see [DGM86]), the $(ax + b)$-group seems to be now largely forgotten in the next generation of the wavelet community. But [Dau92, Chapters 1–3] still serve as a beautiful presentation of this (now much ignored) side of the subject. It also serves as a link to mathematical physics and to classical analysis.

Since the representation U in (1.6.1) on $L^2(\mathbb{R})$ leaves invariant the Hardy space

$$\mathcal{H}_+ = \left\{ f \in L^2(\mathbb{R}) \mid \mathrm{supp}\,(\hat{f}) \subset [0, \infty) \right\}, \qquad (1.6.3)$$

formula (1.6.2) suggests that it would be simpler to look for wavelets in $\mathcal{H}_+$. After all, it is a smaller space, and it is natural to try to use the causality features of $\mathcal{H}_+$ implied by the support condition in (1.6.3). Moreover, in the world of the Fourier transform, the two operations of the formulas (1.6.1) and (1.6.2) take the simpler forms

$$\hat{f} \longmapsto a^{\frac{1}{2}} e^{-ibt} \hat{f}(at) \quad \text{and} \quad \hat{\psi} \longmapsto 2^{\frac{j}{2}} e^{-i2^j kt} \hat{\psi}\left(2^j t\right). \qquad (1.6.4)$$

So in the early nineties, this was an open problem in the theory, i.e., whether or not there are wavelets in the Hardy space; but it received a beautiful answer in [Aus95]. Auscher showed that there are no wavelet functions ψ in $\mathcal{H}_+$ which satisfy the following mild regularity properties:

(R_0) $\quad \hat{\psi}$ is continuous;

(R_ε) $\quad$ for some $\varepsilon \in \mathbb{R}_+$, $\hat{\psi}(t) = \mathcal{O}\left(|t|^\varepsilon\right)$ and $\hat{\psi}(t) = \mathcal{O}\left((1+|t|)^{-\varepsilon-\frac{1}{2}}\right)$, $t \in \mathbb{R}$.

Comparison of formulas (1.6.1) and (1.6.2) shows that the traditional discrete wavelet transform may be viewed as the restriction to a subgroup H of a classical unitary representation of G. The unitary representations of G are completely understood: the set of irreducible unitary representations consists of two infinite-dimensional inequivalent subrepresentations of the representation (1.6.1) on $L^2(\mathbb{R})$, together with the one-dimensional representations $\left(\begin{smallmatrix} a & b \\ 0 & 1 \end{smallmatrix}\right) \to a^{ik}$ parameterized by $k \in \mathbb{R}$. (The two subrepresentations of (1.6.1) are obtained by restricting to $f \in L^2(\mathbb{R})$ with $\operatorname{supp}\hat{f} \subseteq \langle-\infty, 0]$ and $\operatorname{supp}\hat{f} \subseteq [0, \infty\rangle$, respectively.) (See Exercises 1–41 and 1–42.) However, the subgroup H of G has a rich variety of inequivalent infinite-dimensional representations that do not arise as restrictions of (1.6.1), or of any representation of G. The group H considered in (1.6.2) is a semidirect product (as is G): it is of the form

$$H_N = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \;\middle|\; a = N^j, \; b = \sum_{i \in \mathbb{Z}} n_i N^i, \; j \in \mathbb{Z}, \; n_i \in \mathbb{Z}, \right.$$
$$\left. \text{where the } \sum_i \text{ summation is finite} \right\}. \quad (1.6.5)$$

(In the jargon of pure algebra, the nonabelian group H_N is the semidirect product of the two abelian groups $\mathbb{Z}$ and $\mathbb{Z}\left[\frac{1}{N}\right]$, with a naturally defined action of $\mathbb{Z}$ on $\mathbb{Z}\left[\frac{1}{N}\right]$.)

The papers [DaLa98], [Jor01a], [BaMe99], [HLPS99], [LPT01], and [BreJo91] show that it is possible to use these nonclassical representations of H for the construction of unexpected classes of wavelets, the wavelet sets being the most notable ones. Recall that a subset $E \subset \mathbb{R}$ of finite measure is a *wavelet set* if $\hat{\psi} = \chi_E$ is such that, for some $N \in \mathbb{Z}_+$, $N \geq 2$, the functions $\left\{ N^{\frac{j}{2}}\psi\left(N^j x - k\right) \mid j, k \in \mathbb{Z} \right\}$ form an orthonormal basis for $L^2(\mathbb{R})$. Until the work of Larson and others, see [DaLa98] and [HLPS99], it was not even clear that wavelet sets E could exist in the case $N > 2$. The paper [LPT01] develops and extends the representation theory for the subgroups H_N independently of the ambient group G and shows that each H_N has continuous series of representations which account for the wavelet sets. The role of the representations of the groups H_N and their generalizations for the study of wavelets was first stressed in [BreJo91].

There is a different transform which is analogous to the wavelet transform of (1.6.1)–(1.6.2), but yet different in a number of respects. It is the Gabor transform of Section

1.8, and it has a history of its own; see the details below. Both are special cases of the following construction: Let G be a nonabelian matrix group with center C, and let U be a unitary irreducible representation of G on the Hilbert space $L^2(\mathbb{R})$. When $\psi \in L^2(\mathbb{R})$ is given, we may define a transform

$$(T_\psi f)(\xi) := \langle U(\xi)\psi \,|\, f \rangle, \qquad \text{for } f \in L^2(\mathbb{R}) \text{ and } \xi \in G \diagup C. \qquad (1.6.6)$$

It turns out that there are classes of matrix groups, such as the $ax + b$ group, or the 3-dimensional group of upper triangular matrices, which have transforms T_ψ admitting effective discretizations. This means that it is possible to find a vector $\psi \in L^2(\mathbb{R})$, and a discrete subgroup $\Lambda \subset G \diagup C$, such that the restriction to Λ of the transform T_ψ in (1.6.6) is injective from $L^2(\mathbb{R})$ into functions on Λ.

There are many books on transform theory, and here we are only making the connection to wavelet theory. The book [Per86] contains much more detail on the group-theoretic approach to these continuous and discrete coherent vector transforms.

1.7   Wavelet packets

UNCERTAINTY. The practical gain in passing from wavelets to wavelet packets is an improvement of *localization*, and this localization is useful for a variety of applications. One localization method is the algorithm of Coifman and Wickerhauser for digitizing the fingerprint archive at the F.B.I. As follows from the brief summary below, the wavelet packets will generally not spread through the entire universe; in the case of one dimension, the algorithm allows an adjustment of the position and the size of dyadic intervals which carry the essential portion of the signal, while at the same time controlling the frequency localization. It is known [CoWi93] that the algorithm is also successful in two or three dimensions. Both in modern wavelet packet analysis and in localization problems in quantum mechanics, the obstacle is a variant of the uncertainty relation. In Fourier analysis you have pairs of variables which are in duality, for example position x and momentum p, or energy and time. In Heisenberg's formulation, the uncertainty relation reads $\Delta x \Delta p \geq \frac{\hbar}{2}$ where $\hbar$ is Planck's constant and where Δx and Δp denote the dispersion (uncertainty) in the respective variables. However, as is well known, Fourier analysis allows a scale in one of the two variables of the duality, and so you will frequently see the lower bound in Heisenberg's inequality in a rescaled version.

While Exercises 1–1 and 1–39 below show that Haar's wavelet expansions allow intuitive and directly computable restrictions to finite intervals, this turns out not to be the case for general wavelet expansions. This is related to the known fact ([Dau92, Chapter

8]) that the Haar wavelet is the only wavelet with antisymmetry: the antisymmetry in question is defined for a function f on $\mathbb{R}$ by the condition $f(a - x) = -f(x)$ for some $a \in \mathbb{R}$. Let ψ be a compactly supported orthogonal dyadic wavelet in $L^2(\mathbb{R})$. Then by [Dau92, Theorem 8.1.4], if ψ has an antisymmetry axis, then ψ is the Haar function.

The Haar wavelet is supported in $[0, 1]$, and if $j \in \mathbb{Z}_+$ and $k \in \mathbb{Z}$, then the modified function $x \mapsto \psi(2^j x - k)$ is supported in the smaller interval $\frac{k}{2^j} \leq x \leq \frac{k+1}{2^j}$. When j is fixed, these intervals are contained in $[0, 1]$ for $k \in \{0, 1, \ldots, 2^j - 1\}$. This is not the case for the other wavelet functions. For one thing, the non-Haar wavelets ψ have support intervals of length more than one, and this forces periodicity considerations; see [CDV93]. For this reason, Coifman and Wickerhauser [CoWi93] invented the concept of wavelet packets. They are built from functions with prescribed smoothness, and yet they have localization properties that rival those of the (discontinuous) Haar wavelet.

There are powerful but nontrivial theorems on restriction algorithms for wavelets $\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k)$ from $L^2(\mathbb{R})$ to $L^2(0, 1)$. We refer the reader to [CDV93] and [MiXu94] for the details of this construction. The underlying idea of Exercises 1–1 and 1–39 dates back to Alfred Haar, but it has found a recent renaissance in the work of Wickerhauser [Wic93] on *wavelet packets*. The idea there, which is also motivated by the Walsh function algorithm, is to replace the refinement equation (1.21) by a related recursive system as follows: Let $m_0(z) = \sum_k a_k z^k$, $m_1(z) = \sum_k b_k z^k$, for example $b_k = (-1)^k \bar{a}_{1-k}$, $k \in \mathbb{Z}$, be a given low-pass/high-pass system, $N = 2$. Then consider the following *refinement system* on $\mathbb{R}$:

$$W_{2n}(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} a_k W_n(2x - k), \quad W_{2n+1}(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} b_k W_n(2x - k). \quad (1.7.1)$$

Clearly the function W_0 can be identified with the traditional scaling function φ of (1.21). A theorem of Coifman and Wickerhauser [CoWi93, Theorem 8.1] states that if $\mathcal{P}$ is a partition of $\{0, 1, 2, \ldots\}$ into subsets of the form

$$I_{k,n} = \left\{ 2^k n, 2^k n + 1, \ldots, 2^k(n+1) - 1 \right\},$$

then the function system

$$\left\{ 2^{\frac{k}{2}} W_n\left(2^k x - l\right) \mid I_{k,n} \in \mathcal{P},\ l \in \mathbb{Z} \right\}$$

is an orthonormal basis for $L^2(\mathbb{R})$. Although it is not spelled out in [CoWi93], this construction of bases in $L^2(\mathbb{R})$ divides itself into the two cases, the true orthonormal basis (ONB), and the weaker property of forming a function system which is only a tight frame. As in the wavelet case, to get the $\mathcal{P}$-system to really be an ONB for $L^2(\mathbb{R})$, we must assume the transfer operator $R_{|m_0|^2}$ to have *Perron–Frobenius spectrum* on $C(\mathbb{T})$. This means that the intersection of the point spectrum of $R_{|m_0|^2}$ with $\mathbb{T}$ is the singleton $\lambda = 1$, and that $\dim \ker((1 - R_{|m_0|^2})|_{C(\mathbb{T})}) = 1$; see Table 3 in the next chapter.

More generally, an orthonormal basis for $L^2(\mathbb{R})$ selected from among the functions $2^{\frac{k}{2}} W_n(2^k x - l)$, $n = 0, 1, \ldots, k$, $l \in \mathbb{Z}$, is called a *wavelet packet basis*, and the versatility of these bases derives from the known adaptability of the construction, making both the scale and the "window" choice adjust simultaneously to the synthesis problem at hand, say recovering the fingerprint which corresponds to a digitized wavelet version in the FBI database, or in 3D, reconstructing a sculpture from its wavelet packet coefficients. See [CoWi93] for further details. The construction of the wavelet packets starts with the $L^2(\mathbb{R})$-cascade analysis of the solutions to the refinement system (1.7.1), and it is shown in [Wic93] that the arguments for existence of φ and ψ in the wavelet case (see Sections 1.1–1.2 above) adapt to the existence problem for (1.7.1).

We noted in Exercise 1–43, to be followed up in Lemma 2.2.2 below, that the $L^2(\mathbb{R})$-wavelet expansion may be based on the operator system (1.3.15), i.e., the system of operators S_i acting on the sequence space ℓ^2. These are the Cuntz relations [Cun77], although they are not known under that name in the engineering literature. An examination of the papers [Wic93] and [CoWi93] shows that the wavelet packet algorithm is in fact based on the same relations, i.e., the Cuntz relations. Perhaps surprisingly, this turns out to be also the case for the recent algorithms on ridgelets and curvelets which are used for ridges of singularities in Radon transform/wavelet analysis of problems in $\mathbb{R}^d$, $d \geq 2$, and in neural network problems; see [AyBa01], [CaDo00], and [Can99].

1.8   The Gabor transform

Another useful expansion in signal theory is provided by Gabor analysis. This differs from wavelet analysis in that scaling is replaced by multiplication by $x \to e^{i\omega x}$. Fix a function $g \in L^2(\mathbb{R})$ with $\|g\|_2 = 1$, and, for $\omega, t \in \mathbb{R}$, let $g^{\omega,t}(x) = e^{i\omega x} g(x - t)$. The *windowed Fourier transform*, or *transform*, $T : L^2(\mathbb{R}) \to L^2(\mathbb{R}^2)$ is defined by

$$Tf(\omega, t) = \langle g^{\omega,t} \,|\, f \rangle = \int_{\mathbb{R}} e^{-i\omega x} \overline{g(x - t)} f(x) \, dx \qquad (1.8.1)$$

for $f \in L^2(\mathbb{R})$. Let Λ be a lattice in $\mathbb{R}^2$, i.e., Λ is a discrete subgroup of $\mathbb{R}^2$ of rank 2, so $\Lambda = \{nx + my \mid n, m \in \mathbb{Z}\}$ where $\{x, y\}$ is a basis for $\mathbb{R}^2$. One may ask if every function f is determined by the "discretized" windowed Fourier transform defined by Λ, that is, if f is determined by the restriction of Tf to Λ. Equivalently, the question is whether the linear span of $\{g^{\omega,t} \mid (\omega, t) \in \Lambda\}$ is dense in $L^2(\mathbb{R})$. It was stated without proof by von Neumann that this is the case if $\Lambda = 2\pi\mathbb{Z} \oplus \mathbb{Z}$ and $g(x) = \pi^{-1/4} e^{-x^2/2}$ [vNeu68, p. 217]. It was proved independently by Perelomov [Per71] and Bargmann et al. [BBGK71] that, for any lattice Λ in $\mathbb{R}^2$ and the above Gaussian function g, the set $\{g^{\omega,t} \mid (\omega, t) \in \Lambda\}$ spans $L^2(\mathbb{R})$ if and only if $\mathrm{Vol}(\mathbb{R}^2 / \Lambda) \leq 2\pi$, where $\mathrm{Vol}(\mathbb{R}^2 / \Lambda)$ is the volume of the parallelogram $\{tx + sy \mid t, s \in [0, 1)\}$. Much more generally, the following result is true.

Theorem 1.8.1. *Let Λ be a lattice in $\mathbb{R}^2$. The following two conditions are equivalent:*

There exists a function $g \in L^2(\mathbb{R})$ such that the linear span of $\hspace{2cm}$ (1.8.2)
$\left\{ g^{\omega,t} \mid (\omega,t) \in \Lambda \right\}$ is dense in $L^2(\mathbb{R})$,

and

$$\mathrm{Vol}\left(\mathbb{R}^2 \diagup \Lambda\right) \leq 2\pi. \hspace{2cm} (1.8.3)$$

Moreover, there exists a function $g \in L^2(\mathbb{R})$ such that $\left\{ g^{\omega,t} \mid (\omega,t) \in \Lambda \right\}$ is an orthonormal basis for $L^2(\mathbb{R})$ if and only if $\mathrm{Vol}\left(\mathbb{R}^2 \diagup \Lambda\right) = 2\pi$.

This theorem is an immediate consequence of Theorem 3.2 in [Rie81], and was proved independently in [Bag90]. Both proofs use von Neumann algebra techniques, as does also a recent proof by Bekka [Bek01]. The latter proof is interesting in that it gives an interpretation of the constant $\mathrm{Vol}\left(\mathbb{R}^2 \diagup \Lambda\right) / 2\pi$ as a von Neumann dimension. The proofs are thus outside the scope of this book. For more information on Gabor analysis, see [Dau92, Chapter 2], [FeSt98], [Gro01], [MeyY00], and [Per86].

While there are similarities between wavelet bases and Gabor bases in $L^2(\mathbb{R})$, there are also differences; a notable difference has to do with the respective localization properties, which are typically better for wavelets, as stressed for example in [Dau92, Theorem 4.1.1] or [HeWe96, Theorem 2.1].

As we note in Exercises 1–38 and 1–45, the wavelets and the Gabor bases have in common that they may both be obtained from the discretization of unitary representations of matrix groups G acting on the Hilbert space $L^2(\mathbb{R})$. In the case of the wavelets, the group consists of the 2×2 matrices $\left(\begin{smallmatrix} e^t & b \\ 0 & 1 \end{smallmatrix} \right)$, $t, b \in \mathbb{R}$, while for the Gabor bases, the group is the three-dimensional group G_H of matrices $\left(\begin{smallmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{smallmatrix} \right)$. This latter group is the *Heisenberg group*, and it has a one-dimensional center C. Dividing out with C, the two variables a, b remain, i.e., $G \diagup C \cong \mathbb{R}^2$. From there, the construction of a singly generated basis, in each case, results from a selection of a function ψ in $L^2(\mathbb{R})$, a unitary representation U of one of the groups G such that the respective basis functions have the form $U_\lambda \psi$ with λ varying in a discrete subgroup (as in Section 1.6), or in a lattice $\Lambda \subset \mathbb{R}^2$ as described above for the Gabor case. The reader may check that there is an irreducible representation U of G_H on $L^2(\mathbb{R})$ such that

$$\left(U_{\left(\begin{smallmatrix} 1 & t & 0 \\ 0 & 1 & \omega \\ 0 & 0 & 1 \end{smallmatrix} \right)} g \right)(x) = g^{\omega,t}(x) = e^{i\omega x} g(x - t) \hspace{2cm} (1.8.4)$$

yields the function system in (1.8.1).

Exercises

1–1. (Haar 1909.) Let ψ be the Haar function of (1.2), i.e., $\psi := \chi_{\left[0,\frac{1}{2}\right\rangle} - \chi_{\left[\frac{1}{2},1\right\rangle}$, and set $\psi_{j,k}(x) := 2^{\frac{j}{2}} \psi\left(2^j x - k\right)$ as usual, but instead of having the variables $x \in \mathbb{R}$, $j, k \in \mathbb{Z}$

1–46. For a dyadic wavelet function ψ on $\mathbb{R}$, we define a *dimension function*

$$D_\psi(t) := \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} \left| \hat{\psi}\left(2^j(t + 2\pi k)\right) \right|^2.$$

Let

$$E = [-2\pi, \pi) \cup [\pi, 2\pi),$$

and

$$F = \left[-4\pi - \frac{4\pi}{7}, -4\pi\right) \cup \left[-\pi, -\frac{4\pi}{7}\right) \cup \left[\frac{4\pi}{7}, \pi\right) \cup \left[4\pi, 4\pi + \frac{4\pi}{7}\right).$$

(a) Show that both E and F are wavelet sets, i.e., that both $\psi_E = \check{\chi}_E$ and $\psi_F = \check{\chi}_F$ are dyadic wavelet functions.

(b) Show that
$$D_{\psi_E}(t) \equiv 1, \qquad t \in \mathbb{R}.$$

(c) Show that
$$D_{\psi_F}(t) \equiv 2, \qquad t \in \left[\frac{-2\pi}{7}, \frac{2\pi}{7}\right).$$

(d) Show that the wavelet function ψ_E has a scaling function φ_E and find the refinement equation for φ_E. Find φ_E.

(e) Show that the wavelet function ψ_F does *not* have a scaling function. (See more on this in Section 2.3.)

Terminology

- **mathematics:** "the study of absolutely necessary truths." —David Deutsch, *The Fabric of Reality* [Deu97]

- **multiresolution:** —*real world:* a set of band-pass-filtered component images, assembled into a mosaic of resolution bands, each resolution tied to a finer one and a coarser one.

  —*mathematics:* used in wavelet analysis and fractal analysis, multiresolutions are systems of closed subspaces in a Hilbert space, such as $L^2(\mathbb{R})$, with the subspaces nested, each subspace representing a resolution, and the relative complement subspaces representing the detail which is added in getting to the next finer resolution subspace.

- **matrix function:** a function from the circle, or the one-torus, taking values in a group of N-by-N complex matrices.

- **wavelet:** a function ψ, or a finite system of functions $\{\psi_i\}$, such that for some scale number N and a lattice of translation points on $\mathbb{R}$, say $\mathbb{Z}$, a basis for $L^2(\mathbb{R})$ can be built consisting of the functions $N^{\frac{j}{2}}\psi_i(N^j x - k)$, $j, k \in \mathbb{Z}$.

  > *Then dulcet music swelled*
  > *Concordant with the life-strings of the soul;*
  > *It throbbed in sweet and languid beatings there,*
  > *Catching new life from transitory death;*
  > *Like the vague sighings of a wind at even*
  > *That wakes the wavelets of the slumbering sea. . .*
  > —Shelley, *Queen Mab*

- **subband filter:** —*engineering:* signals are viewed as functions of time and frequency, the frequency function resulting from a transform of the time function; the frequency variable is broken up into bands, and up-sampling and down-sampling are combined with a filtering of the frequencies in making the connection from one band to the next.

  —*wavelets:* scaling is used in passing from one resolution V to the next; if a scale N is used from V to the next finer resolution, then scaling by $\frac{1}{N}$ takes V to a coarser resolution V_1 represented by a subspace of V, but there is a set of functions which serve as multipliers when relating V to V_1, and they are called subband filters.

- **cascades:** —*real world:* a system of successive refinements which pass from a scale to a finer one, and so on; used for example in graphics algorithms: starting with control points, a refinement matrix and masking coefficients are used in a cascade algorithm yielding a cascade of masking points and a cascade approximation to a picture.

  —*wavelets:* in one dimension the scaling is by a number and a fixed simple function, for example of the form ⊓ (from 0 to 1) is chosen as the initial step for the cascades; when the masking coefficients are chosen the cascade approximation leads to a scaling function.

- **scaling function:** a function, or a distribution, φ, defined on the real line $\mathbb{R}$ which has the property that, for some integer $N > 1$, the coarser version $\varphi\left(\frac{x}{N}\right)$ is in the closure (relative to some metric) of the linear span of the set of translated functions $\dots, \varphi(x+1), \varphi(x), \varphi(x-1), \varphi(x-2), \dots$.

- **logic gates:** —*in computation* the classical logic gates are realized as computers, for example as electronic switching circuits with two-level voltages, say high and low. Several gates from Table 2 have two input voltages and one output, each one

allowing switching between high and low: The output of the AND gate is high if and only if both inputs are high. The XOR gate has high output if and only if one of the inputs, but not more than one, is high.

- **qubits:** —*in physics and in computation:* qubits are the quantum analogue of the classical bits 0 and 1 which are the letters of classical computers, the qubits are formed of two-level quantum systems, electrons in a magnetic field or polarized photons, and they are represented in Dirac's formalism $|0\rangle$ and $|1\rangle$; quantum theory allows superpositions, so states $|\psi\rangle = a|0\rangle + b|0\rangle$, $a, b \in \mathbb{C}$, $|a|^2 + |b|^2 = 1$, are also admitted, and computation in the quantum realm allows a continuum of states, as opposed to just the two classical bits.

  —*mathematics:* a chosen and distinguished basis for the two-dimensional Hilbert space $\mathbb{C}^2$ consisting of orthogonal unit vectors, denoted $|0\rangle$, $|1\rangle$.

- **universality:** —*classical computing:* the property of a set of logic gates that they suffice for the implementation of every program; or of a single gate that, taken together with the NOT gate, it suffices for the implementation of every program.

  —*quantum computing:* the property of a set S of basic quantum gates that every (invertible) gate can be written as a sequence of steps using only gates from S. Usually S may be chosen to consist of one-qubit gates and a distinguished tensor gate t. An example of a choice for t is CNOT. An alternative universal one is the Toffoli gate.

  —*mathematics:* the property of a set S of basic unitary matrices that for every n and every $u \in \mathrm{U}_{2^n}(\mathbb{C})$, there is a factorization $u = s_1 s_2 \cdots s_k$, $s_i \in S$, with the understanding that the factors s_i are inserted in a chosen tensor configuration of the quantum register $\underbrace{\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2}_{n \text{ times}}$. Note that the factors s_i, the number k, and the configuration of the s_i's all depend on n and the gate $u \in \mathrm{U}_{2^n}(\mathbb{C})$ to be studied. The quantum wavelet algorithm (1.3.24) is an example of such a matrix u. It may also take the form given in (1E.17) of Exercise 1–22.

- **chaos:** a small variation or disturbance in the initial states or input of some system giving rise to a disproportionate, or exponentially growing, deviation in the resulting output trajectory, or output data. The term is used more generally, denoting rather drastic forms of instability; and it is measured by the use of statistical devices, or averaging methods.

- $\mathrm{GL}_N(\mathbb{C})$: the *general linear group* of all complex $N \times N$ invertible matrices.

- $\mathrm{U}_N(\mathbb{C}) := \{A \in \mathrm{GL}_N(\mathbb{C}) \mid AA^* = 1_{\mathbb{C}^N}\}$ where A^* denotes the adjoint matrix, i.e., $(A^*)_{i,j} = \bar{A}_{j,i}$.

- **transfer operator (transition operator):** —*in probability:* An operator which transforms signals s from input s_{in} to output s_{out}. The signals are represented as functions on some set E. In the simplest case, the operator is linear and given in terms of conditional probabilities $p(x, y)$. The number $p(x, y)$ may represent the probability of a transition from y to x where x and y are points in the set E. Then

$$s_{\text{out}}(x) = \sum_{y \in E} p(x, y)\, s_{\text{in}}(y).$$

—*in computation:* Let X and Y be functions on a set E, both taking values in $\{0, 1\}$. Let Y be the initial state of the bit, and X the final state of the bit. If the process is governed by a probability distribution P, then the transition probabilities $p(x, y) := P(\{X = x \mid Y = y\})$ are conditional probabilities: i.e., $p(x, y)$ is the probability of a final bit value x given an initial value y, and we have

$$P(\{X = x\}) = \sum_{y \in E} p(x, y)\, P(\{Y = y\}).$$

—*in wavelet theory:* Let $N \in \mathbb{Z}_+$, and let W be a positive function on $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$, for example $W = |m_0|^2$ where m_0 is some low-pass wavelet filter with N bands. (Positivity is only in the sense $W \geq 0$, nonnegative, and the function W may vanish on a subset of $\mathbb{T}$.) Then define a function p on $\mathbb{T} \times \mathbb{T}$ as follows:

$$p(z, w) = \begin{cases} \left(\frac{1}{N}\right) W(w) & \text{if } w^N = z, \\ 0 & \text{for all other values of } w. \end{cases}$$

We arrive at the transfer operator R_W which is studied in detail in Chapter 4 below, i.e., the operator transforming functions on $\mathbb{T}$ as follows:

$$s_{\text{out}}(z) = (R_W s_{\text{in}})(z) = \frac{1}{N} \sum_{w^N = z} W(w)\, s_{\text{in}}(w).$$

- **coherence:** —*in mathematics and physics:* The vectors ψ_i of (1.1.7) that make up a tight frame, one which is not an orthonormal basis, are said to be subjected to *coherence*. So coherent vector systems in Hilbert space are viewed as bases which generalize the more standard concept of orthonormal bases from harmonic analysis. A striking feature of the wavelets with compact support, which are based on scaling, is that the varieties of the two kinds of bases can be well understood geometrically. For example, we show in Chapter 2 that the collapse of the wavelet orthogonality relations, degenerating into coherent vectors, happens on a subvariety of a lower dimension.
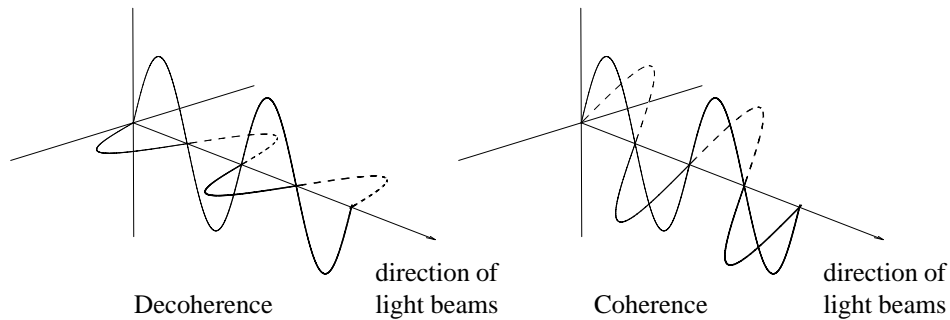
Figure 1.12. Polarized photons visualized as light beams.

More generally, coherent vectors in mathematical physics often arise with a continuous index, as we noted in Exercise 1–43 above, even if the Hilbert space is separable, i.e., has a countable orthonormal basis. This is illustrated by the vector system $\{\psi_{r,s}\}$ in the first line of the tables in Exercise 1–43, which should be thought of as a continuous analogue of (1.1.7), i.e., the version where the sum in (1.1.7) gets replaced with an integral

$$C_\psi^{-1} \iint\limits_{\mathbb{R}^2} \frac{dr\,ds}{r^2}\, |\langle \psi_{r,s}\,|\,f \rangle|^2 = \|f\|^2\,.$$

For more details, see also [Dau92, Section 3.3] and [Kai94, Chapter 3], and Exercise 1–26.
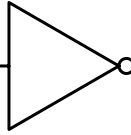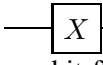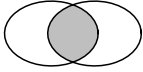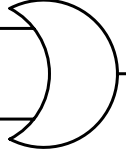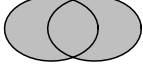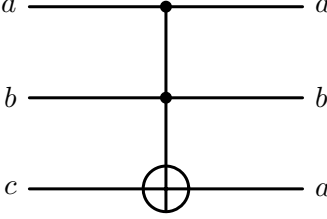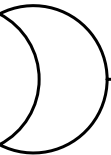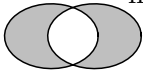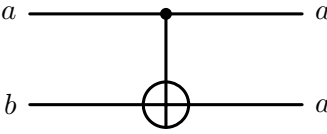
In quantum mechanics, one talks, for example, about coherent states in connection with wavefunctions of the harmonic oscillator. Combinations of stationary wavefunctions from different energy eigenvalues vary periodically in time, and the question is which of the continuously varying wavefunctions one may use to expand an unknown function in without encountering overcompleteness of the basis. The methods of "coherent states" are methods for using these kinds of functions (which fit some problems elegantly) while avoiding the difficulties of overcompleteness. The term "coherent" applies when you succeed in avoiding those difficulties by some means or other. Of course, for students who have just learned about the classic complete orthonormal basis of stationary eigenfunctions, "coherent state" methods at first may seem like a daring relaxation of the rules of orthogonality, so that the term seems to stand for total freedom!

• **decoherence:** *—in quantum computing channels:* the collapse of a pure state $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, $\alpha\beta \neq 0$, i.e., one with quantum parallelism, into one of the two states $|0\rangle$ or $|1\rangle$, where these resulting two states could represent a horizontal or a vertical polarization of a photon; see Figure 1.12.

Table 2. Logic gates

COMPARISONS: The classical NOT is analogous to the quantum X, and the classical XOR is somewhat analogous to the quantum CNOT, while the other pairs of gates are less analogous. The quantum gate Z maps $|0\rangle$ into itself and $|1\rangle$ into $-|1\rangle$, and $-|1\rangle$ defines exactly the same quantum state as $|1\rangle$, so Z does nothing to the classical states, but it affects the coherent states. Thus Z is a quantum gate without a classical counterpart. A gate is *reversible* if its input can be restored from its output, otherwise *irreversible*.

classical gates and Venn diagrams (irreversible except the first one, NOT)	quantum gates (all reversible by virtue of their representation as unitaries)

NOT: (r) $a \longrightarrow$ ▷○ $\longrightarrow \bar{a}$
$:= a + 1 \mod 2$
bit-flip

X: (r) $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ —[X]—
qubit-flip

Z: (r) $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ —[Z]—
phase-reversal

AND: (ir) (u) a —, b — ⊃ $\longrightarrow a \wedge b$
$:= ab \mod 2$

OR: (ir) (u) a —, b — ⊃ $\longrightarrow a \vee b$
$:= a + b + ab \mod 2$

Toffoli: (r) (u) $a \longrightarrow a$
$b \longrightarrow b$
$c \longrightarrow ab + c \mod 2$

XOR: (ir) (nu) a —, b — ⊃) $\longrightarrow a \,\dot\vee\, b$
$:= a + b \mod 2$

CNOT: (r) (u) $a \longrightarrow a$
$b \longrightarrow a + b \mod 2$
see (1.3.3)

$a, b, \cdots \in \mathbb{Z}_2 = \{0, 1\}$; $a + b$ and ab are respectively addition and multiplication $\mod 2$; (r) = reversible; (ir) = irreversible; (u) = universal; (nu) = not universal