Where do wavelets come from? — A personal point of view

Ingrid Daubechies

Department of Mathematics and
Program in Applied and Computational Mathematics
Princeton University
Princeton NJ 08544

The subject area of wavelets, developed mostly over the last fifteen years, is connected to older ideas in many other fields, including pure and applied mathematics, physics, computer science, and engineering. The history of wavelets can therefore be represented as a tree with roots reaching deeply and in many directions. In this picture, the trunk would correspond to the rapid development of "wavelet tools" in the second half of the 1980's, with shared efforts by researchers from many different fields; the crown of the tree, with its many branches, would correspond to different directions and applications in which wavelets are now becoming a standard part of the mathematical tool kit, alongside other more established techniques. The bulk of this issue represents several of these branches; this historical introduction is concerned with the roots and the bottom part of the trunk.

Because of the many roots of the subject, there are many approaches to tackling its history. Lacking the time, space, and expertise (historical research has its tools and techniques too!) to give a truly scholarly history, I have chosen to be resolutely subjective, and to start from the root through which I entered this field. This is not the most ancient or the mathematically deepest root, but it does have one good claim to be at the start of this story: it is the root where wavelets acquired their name. I also believe it is the direction which catalyzed the whole wavelet synthesis. As we proceed up the root system, and simultaneously advance in time, we will come to points where different syntheses were made, with developments from other disciplines; at these points we will peek back for a short description of what happened earlier along those other roots. To emphasize the informality of this story, I decided not to give any precise references.

It is clear that all this results in a highly personal version of the development of wavelets; similar stories told by a harmonic analyst or an approximation theorist or an electrical engineer would all be quite different from mine or from each other. Such other points of view can be found in the introduction to several wavelet books or in book reviews, see e.g., J. Benedetto and M. Frazier's 1994 volume on *Wavelets: Mathematics and Applications*, the book review by Y. Meyer in the Bulletin of the

American Mathematical Society in 1993, or the recent book by M. Vetterli and J. Kovačević.

In the late 1970's, J. Morlet, a geophysical engineer at the French oil company Elf Acquitaine, came up with an alternative for the short time Fourier transform. The standard procedure, which consists of first windowing a signal, and then computing its Fourier coefficients, amounts to taking the inner product of the signal with windowed Fourier functions depending on two parameters — the location of the window and the frequency label of the different coefficients. The signals which Morlet wanted to analyze did consist of different features in time and frequency which he wanted to disentangle, but typically their high frequency components would have a shorter timespan than their low frequency components. In order to gain time resolution for the high frequency transients, he could choose to do a wide-band short time Fourier transform; on the other hand, he also wanted good frequency resolution for the low frequency components and that called for a narrow-band short time Fourier transform. In order to achieve both aims in one single transform, Morlet came up with the idea to generate the transform functions in a different way: he took a windowed cosine wave (using a smooth window — his favorite is a Gaussian window with its width adjusted so that the peaks closest to the central peaks had about half its amplitude), and compressed this in time to obtain a higher frequency function, or spread it out to obtain a lower frequency function. In order to investigate what happened at different times, these functions were shifted in time as well. The transform functions depended thus again on two parameters: their time location and their degree of compression or scale. He then would take the inner product of the signal he wanted to analyze (usually seismic data) with all these transform functions. There was one crucial difference with the transform functions in the standard Fourier transform: in Morlet's approach, the high frequency functions were very narrow, while the low frequency functions were not! Many different choices of reference function existed in other contexts in geophysics, and were called "X's wavelet" (if invented or proposed by X); Morlet chose to call his transform functions "wavelets of constant shape." When, a few years later, the transform of Morlet left the geophysics arena, other researchers soon dropped the qualifier "of constant shape" — the wavelet transform was born! But at the time, Morlet had a hard time convincing his geophysics colleagues that this was a worthwhile mathematical tool — he once paraphrased his audiences' attitude as: "If it were true, then it would be in the math books. Since it isn't in there, it is probably worthless." Undeterred, Morlet looked for help in giving a sounder mathematical footing to his wavelet transform. A friend from his student days referred him to A. Grossmann, a theoretical physicist who had worked extensively in quantum mechanics, where similar problems occur when one tries to pin down local features in a function as well as in its Fourier transform. Grossmann recognized in Morlet's transform something similar to the coherent states formalism — a technique he had used profitably in quantum mechanics (and which he had taught me as well when I worked on my Ph.D. thesis under his direction). He constructed an exact inver-

sion formula for Morlet's integral transform, and they explored several applications together. (Later, it turned out that this particular form of coherent states and the inversion formula had been constructed earlier by E. Aslaksen and J. Klauder in the completely different framework of building a toy model for quantum gravity.) Having worked on different topics as a postdoc, I learned about wavelets on a visit to Marseille in Spring 1985, and started working on wavelet series (as opposed to integrals). Grossmann suggested that we explore the concept of frames (defined in the article by A. Cohen and J. Kovačević in this issue) in this context, which turned out to be a crucial ingredient.

In the meantime, a very important merging of wavelet roots was about to happen. That same Spring (1985), Y. Meyer, a pure mathematician then based at the Ecole Polytechnique near Paris, heard serendipitously (while waiting in line for a photo-copying machine!) about the work of Grossmann and Morlet. When he read their papers, he realized that their analysis and reconstruction formula was a rediscovery of a formula that A. Calderón had introduced in harmonic analysis in the 1960's. "Harmonic analysis" is a discipline in pure mathematics that grew out of Fourier analysis; among the many directions within harmonic analysis, a very important field concerns the study of singularities, integral operators with singular kernels (such as the Hilbert transform), oscillating singular integrals, ... One of the roots of this field is Littlewood-Paley theory, developed in the 1930's, which uses dyadic (i.e., in blocks that scale by factors 2) regroupings of the Fourier transform of a function in order to deal with and characterize more effectively the singularities of this function. This regrouping by scale is a fore-shadowing of the role scaling plays in the wavelet transform. Similarly, Calderón's formula, designed to be a tool in the analysis of certain integral operators with singular integral kernels, used different scales, in a manner similar to the wavelet transform of Grossmann and Morlet. After all, a singularity is really an extremely localized manifestation of very high frequencies, so it is not surprising that a wavelet approach, with its increasingly precise localization as the frequency increases, would be appropriate.

Meyer, a pre-eminent expert in this field, recognized these links with harmonic analysis; aside from the all-too-human reaction of "But we knew all this a long time ago!" (a reaction wavelet researchers would meet again and again as other mergings happened), he was also enthusiastic about this whole new area of applications for harmonic analysis insights, as well as intrigued by the different interpretation that Grossmann and Morlet gave to Calderón's formula. He got in touch with them, and this marked the start of an interaction between pure harmonic analysts and applied researchers that would benefit both communities. Meyer was intrigued especially by the wavelet series; they systematically used redundant families of wavelets because of a subliminal belief that redundancy was unavoidable in order to obtain good time-frequency localization (as is the case for the Gabor transform). (Similar series were in fact developed around that time by harmonic analysts M. Frazier and B. Jawerth as well, independently of the wavelet series development.) After he had identified this subliminal message, Meyer had the typical and healthy mathematician's reflex of

wanting a proof for this belief. A few weeks later, he had not a proof but a beautiful construction of an orthonormal wavelet basis with excellent time-frequency localization properties. With P.G. Lemarié, he soon afterwards generalized this construction to N dimensions. (Later it turned out that J.O. Stromberg, another harmonic analyst, had constructed a different orthonormal wavelet basis a few years earlier, but its importance had not been realized at the time.) Meyer's wavelet basis resulted from a very intriguing and seemingly miraculous cancellation if he set things up just right. A few months later Lemarié and G. Battle came up, independently, and by completely different techniques, with constructions of wavelet bases consisting of spline functions, with better decay (exponential) than Meyer's wavelets, et the price of some loss of regularity (C^k instead of C^∞). (Lemarié, a student of Meyer, is a harmonic analyst; Battle, however, is a mathematical physicist, long interested in quantum field theory. Together with P. Federbush he had built an elaborate machinery, which became technically much simpler after the discovery of smooth and well localized wavelet bases. Battle's constructionwas inspired by renormalization group techniques, a tool used in quantum field theory, that involves studying phenomena at different scales.) These constructions were very ingenious, but they both had an "ad-hoc" feel. A different understanding of these bases was just around the corner. It would lead to another important merging of roots for our tree.

In the summer of 86, S. Mallat, then a graduate student at Penn, had taken a break to go on vacation. There he met an old friend from his undergraduate days, who was now a graduate student of Meyer's, and who mentioned the new wavelet bases to him. (Yes, another serendipitous encounter. This one even took place on a beach.) Mallat was immediately very interested, mainly because he recognized concepts familiar from a very different framework. In Mallat's field of specialization, computer vision and image analysis, it was common wisdom that coarse features in an image are (almost by definition) large-scale objects, whereas fine-scale features should be studied much more locally. (This fails when you talk about textures, where fine scale features can have a very large correlation length.) This principle was at the basis of the scale-space representation of A. Witkin, and it inspired the Laplacian pyramid constructions of P. Burt and E. Adelson. So the "philosophy" of wavelet decompositions, where you use very narrow functions for the fine scale features, and much wider ones for coarse features, fitted beautifully into this view. In vision theory, one way to build multi-scale representations is to strip away the different scales layer by layer. In the Laplacian pyramid scheme, for instance, you compute a blurred version of the picture, and you subtract it from the original; the difference gives you the desired fine scale features which you break up in tiny "elementary" pieces. Successive layers of decreasing spatial resolution are obtained by repeating the procedure on the blurred picture. Thinking about all this after his return to Penn, Mallat conceived of a similar layered structure for wavelet expansions, in which all the terms corresponding to one scale in the wavelet decomposition of a function did indeed give the difference between two successive approximations. Upon hearing that Meyer would be visiting Chicago that Fall, he arranged to meet

him there, and in the next few days, the two of them hammered out all the mathematical details of "multiresolution analysis", a framework that explained all the "miracles" in the wavelet bases constructed up till then, and that made it very easy to construct other orthonormal wavelet bases. (There exist pathological wavelet bases that do not fit into this framework. But in 92, Lemarié and P. Auscher would prove that if the wavelet basis has any decency, i.e. as soon as the wavelets have any reasonable time-frequency localization properties, it necessarily stems from a multiresolution analysis.) More important even, multiresolution analysis led to a simple and recursive filtering algorithm to compute the wavelet decomposition of a function from its finest scale approximation. The filters that corresponded to the bases of Meyer or Battle-Lemarié are infinite and must be truncated for direct implementation. (Although other implementations, via FFT and multiplication in the Fourier domain, also work, without truncation. On the other hand, it was later realized that Stromberg's earlier basis corresponds to IIR filters with rational z-transforms, so that its implementation can be done directly, without truncation.) This begged the question of how to get wavelet bases for which such truncation was not necessary. The answer was to work backwards: instead of deriving the filters from a wavelet basis, construct first a pair of appropriate FIR filter, and then investigate whether they correspond to an orthonormal wavelet basis. This was the point of departure for the construction in early 1987 of orthonormal wavelet bases of compactly supported wavelets. (A complete characterization of filter pairs that give rise to orthonormal wavelet bases would be given by A. Cohen and W. Lawton a few years later.) The use of filters in wavelet decompositions led to the next merging of roots, with subband filtering in electrical engineering.

Electrical engineers, like harmonic analysts, had been long accustomed to the idea of grouping frequencies together in bands with a width proportional to the average frequency in that band. This is constant relative-bandwidth or constant-Q filtering. One way to obtain such a splitting is to work iteratively: first the full range of frequencies (of a bandlimited function) is halved by applying two filters, one high-pass, one low-pass. The lower frequency half can then be halved again, and so on. Since the different components that result from this procedure have different bandwidths, they correspond to different Nyquist sampling rates; an easy way to obtain the "correctly" sampled versions of all the components is to retain only half the output samples at every filtering step. Because the filters used are not perfect, such critical sampling gives rise to aliasing that can lead to unacceptable artefacts at reconstruction if the filters are not designed carefully. In the 70's, A. Croisier, D. Esteban and C. Galland discovered a design procedure that led to exact cancellation of the aliasing: quadrature mirror filters (QMF) were born. Then, about ten years later, in 1983, M. Smith and T. Barnwell, and independently F. Mintzner, discovered QMF-like pairs that gave exact reconstruction. (M. Vetterli made this same discovery independently as well, a little later.) These conjugated quadrature filter (CQF) pairs were exactly the type of filter pairs that researchers in search of orthonormal wavelet bases would rediscover a few years afterwards, from a

completely different angle. (The mathematical properties that were usually asked of the wavelet bases, such as smoothness, also made for different design constraints.) In fact, by the time the wavelet builders became aware of the CQF development, Smith and Barnwell and their students had moved on to more complicated constructions that would design many more filters simultaneously, rather than a structure with two channels at every split, as described above; for many practical EE applications they had in mind, this leads to better performance. (Even within the tree structure, Smith and Barnwell would typically split all the branches, every time, instead of splitting only the lowest frequency band. I have emphasized the latter procedure above to make it as clear as possible how wavelet filtering fitted into established subband filtering practice.)

The nice and easy filter implementation of decompositions into wavelet bases was therefore nothing new to electrical engineers, and the hype that would arise around wavelets caused surprise and some understandable resentment in the subband filtering community. In the end, the mutual awareness of the wavelet and the subband filtering communities nevertheless benefited both. Wavelet reserachers benefited from the work and insights of Vetterli and his students, or of P.P. Vaydyanathan and his collaborators. On the other hand, the different point of view of wavelet theory, which emphasized time or spatial localization as much as frequency localization, on which traditional filtering was more exclusively focused, led to the development of approaches and applications rooted in the mathematical understanding of wavelet bases and their approximation properties, that would not have followed naturally from only the subband filtering algorithms.

The approximation properties of wavelets bring me to yet another merging of roots that took place at about the same time. Approximation theorists became interested in wavelets, which fitted quite naturally into their field after the advent of multiresolution analysis. Many of the insights from harmonic analysis were written by Meyer in the form of inequalities that established correspondences between smoothness and/or decay properies of a function and how well it could be approximated by its wavelet expansion in various norms. This is of course right up the alley of approximation theory. The mathematically important feature of all these inequalities is that they use only the absolute values of the wavelet coefficients, not their signs (for real coefficients) or phases (complex case). In mathematical parlance, wavelets provide "unconditional bases" for large classes of functional spaces; this concept is of interest to both harmonic analysts and approximation theorists. These mathematical properties turn out to be important from the point of view of applications as well, as shown by the work of, e.g., W. Dahmen (numerical analysis), R. DeVore (most recently in nonlinear approximation) and D. Donoho (statistics) and their respective collaborators. For many mathematical applications, it is useful to have wavelets that are spline functions; in that case the corresponding multiresolution analysis of nested spline spaces has of course been known and loved by many generations of approximation theorists.

It would be easy to go on for many more pages about interesting links between

wavelets and numerical analysis, or computer science (especially graphics), or quantum physics, but I don't want to try the reader's patience unduly. (Some of these links will come up in the review articles, or in the summing-up article by W. Sweldens at the end of this issue.)

Moreover, in this tale I have really concentrated on one strand only in the wavelet skein, namely wavelet bases. Rich developments took place also in very redundant wavelet representations, and there too, links with other fields were made. Orthonormal wavelet bases also give rise to wavelet packet bases; it is noteworthy that their construction was inspired by viewing wavelet decompositions as a sequence of filterings. Wavelet bases need not be orthogonal; a straightforward generalization works with two dual wavelet bases ("biorthogonal wavelets"), which are often preferred in applications. Another idea "whose time had come" are the lapped transforms, discovered independently by P. Cassereau, by H. Malvar, and by J. Princen and A. Bradley, and rediscovered a few years later by R. Coifman and Meyer, who found that, like wavelet packet bases, they fit in a scheme where different branches can be regrouped at will, in a search for an ever-better (because more efficient) representation. This leads to the best-basis algorithms of Coifman, Meyer and M. Wickerhauser. All these developments led to more interconnections, more mergings, more rediscoveries with a different twist, but that, as Kipling would say, is another story.

In summary, the development of wavelets is an example where ideas from many different fields combined to merge into a whole that is more than the sum of its parts — many of the applications that will described in the review articles in the remainder of this issue would not have been possible if this merging had not taken place. I have very much enjoyed being a part of this process. I wish there were many more examples of such interconnections — we would all benefit from them.