International Journal of Wireless Information Networks, Vol. 5, No. 2, 1998

Digital Signal Processing in Radio Receivers and Transmitters

fred harris[1]

The interface between analog and digital signal processing paths in radio receivers and transmitters is steadily migrating toward the antenna as engineers learn to combine the unique attributes and capabilities of DSP with those of traditional communication system designs to achieve systems with superior and broadened capabilities while reducing system cost. Digital signal processing (DSP) techniques are rapidly being applied to many signal conditioning and signal processing tasks traditionally performed by analog components and subsystems in RF communication receivers and transmitters [1–4]. The incentive to replace analog implementations of signal processing functions with DSP-based processing includes reduced cost, enhanced performance, improved reliability, ease of manufacturing and maintenance, and operating flexibility and configurability [5]. Technologies that facilitate cost-effective DSP-based implementations include a very large market base supporting high-performance programmable signal processing chips [6], field programmable gate arrays (FPGA), application-specific integrated circuits (ASICs), and high-performance analog-to-digital and digital-to-analog converters (ADC and DAC respectively) [7]. The optimum point for inserting DSP in a signal processing chain is determined by matching the system performance requirements to bandwidth and signal-to-noise ratio (i.e., speed and precision) limitations of the signal processors and the signal converters. In this paper we review how clever algorithmic structures interact with DSP hardware to extend the range and performance of DSP-based processing in RF transmitters and receivers.

1. INTRODUCTION

The cost of implementing traditional signal processing functions by DSP techniques, following Moore's Law, falls an order of magnitude every five years. This is equivalent to a compound interest rate of 58.8% per year. The effect of cost falling at this truly amazing rate cannot be overstated and suggests that only at great peril can one ignore the option of timely DSP insertion. The insertion process is facilitated by a proliferation of technology infrastructure. These include an abundance of design and simulation tools, a rich selection of signal processing conferences, an extensive quantity of texts and publications (including semiconductor supplier's application notes), and a large population of recent graduates with academic exposure to DSP material. The successful application of DSP technology to source coding for the consumer market such as the compact disk (CD), Joint Photographic Expert Group (JPEG) image storage and transportation, Motion Picture Expert Group (MPEG) video, Direct Video Broadcasting (DVB), MUSICAM audio, and VCELP speech has further whetted our appetite for additional high-performance DSP applications.

Applications ripe for DSP insertion are the many communication receivers and transmitters being developed for cable TV and broadband bidirectional digital data links, high-speed modems for twisted pair and coax,

[1] fred harris, Signal Processing Chair, Communications and Signal Processing Center, E&CE Department, San Diego State University,

1068-9605/98/0400-0133$15.00/0

direct broadcast satellite receivers, and dual-use handsets for the wireless and mobile channels. The elements common to all of these receiver applications is an input bandwidth much wider than the particular selected band of interest. Bandwidth translation and bandwidth reduction occurs in the signal processing stream. In most systems these tasks are distributed between the continuous and sampled data domains. Usually, band shifting and band limiting of signals with bandwidths on the order of 100 MHz and above is performed by analog components in the continuous domain while band shifting and band limiting of signals with bandwidths on the order of 10 MHz and below is performed by DSP hardware and software in the sampled data domain. Bandwidths bounded by these extremes can be processed in the analog or digital domain, with a marked tendency toward the DSP solution and with steady pressure to raise the 100 MHz boundary. The interface between the two domains is the ADC at the input port and the DAC at the output port. System performance is dictated by the bandwidth (hence sampling speed) and quantizing noise level (bit precision) required at the interface boundary. The boundary selection process is discussed later in this paper. Part of the selection process is the algorithmic structure of the DSP following or preceding the interface. We will examine a number of useful algorithm structures that minimize the performance requirements on the ADC and DAC as well as the processing burden on the DSP hardware. We will also describe the performance levels required of the ADC and DAC at the interface point.

2. DIGITAL COMMUNICATIONS AND DSP RECEIVERS AND TRANSMITTERS

Let us first resolve possible confusion between the concepts of digital communications and digital signal processing in a communication system. RF communication involves narrowband waveforms propagating through a band-limited noisy channel. Information is transferred by modifying the amplitude, phase, and frequency parameters of the continuous analog waveform. To achieve an efficient balance between bandwidth and signal energy of the waveform we invoke source coding to obtain a discrete description of the signal represented by elements from a short finite alphabet such as binary m-tuples. This finite alphabet is mapped to a binary data stream which, in turn, is mapped to a restricted set of easily distinguishable wave-shape parameters such as the 16 possible combinations of amplitude and phase terms of a 16-QAM signal set. A mapping process, known to the transmitter and receiver, is employed to select, in response to the current (and possibly previous) input, the complex envelope of the waveform to be presented to the channel over the given signaling interval. The mapping process between the digital input and the continuous output waveforms impart upon these continuous waveforms the title "digital waveform" [8]. The waveform is always a continuous signal, and the transmitter and receiver treat it as a digital signal through the input and output mapping procedures.

In a DSP-based transmitter or receiver, the complex envelope of the continuous wave shape is represented by uniformly spaced samples of that envelope at a sample rate sufficiently high to satisfy the Nyquist criteria for its bandwidth. In a transmitter these samples are formed internally and are converted to a continuous waveform by a DAC, a deglitching sample and hold, and an analog smoothing filter. In a receiver, these samples are formed with the aid of an analog anti-aliasing filter, a sample and hold, and an ADC. Processing of the samples substitutes for processing of the continuous waveform associated with the samples. Since the continuous signaling waveforms are essentially band limited, the sampled data representation at the transmitter and receiver contain all the information of the continuous signal. The use of DSP to perform the required signal conditioning and signal processing tasks at the transmitter and receiver is an economic, market-driven choice. It is a choice made in response to the reduced cost of implementing a number of required tasks with easily controlled performance trades such as implementation loss.

To reiterate, the point made in this section is that we should take care to not confuse digital signal processing with digital communications. DSP is a procedure used to implement some of the signal processing tasks of a communication system while digital communication is a philosophy and procedure enabling the delivery of information through noisy band-limited channels subject to satisfying a specified fidelity criterion.

3. RF RECEIVERS

A primary function we require from a digital receiver is the formation of digital samples representing the complex envelope of a carrier located at a specific frequency within a specified band of frequencies. The block diagram and the spectral representation of this task is shown in Fig. 1. At some location in the signal pro-

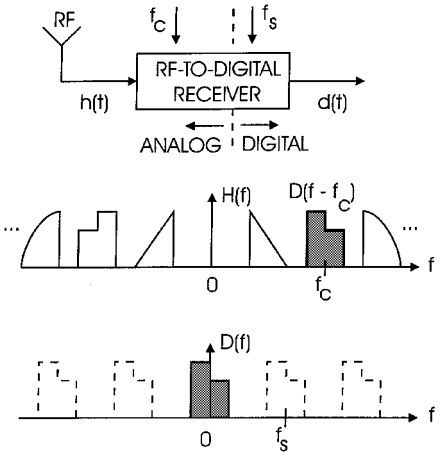

Fig. 1. Block diagram and spectra at input and output of RF-to-digital receiver.

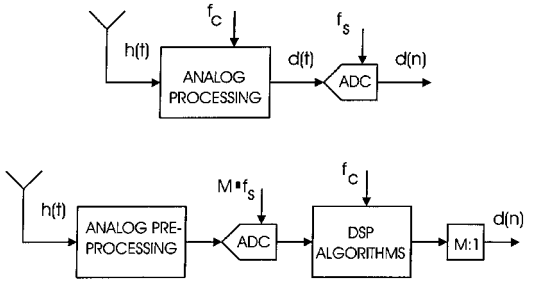

Fig. 2. Digital receiver with ADC at extreme output and input of processing chain.

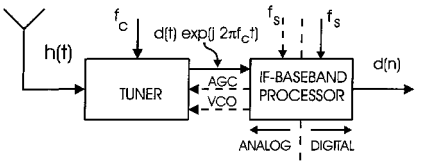

Fig. 3. Partition of digital receiver to tuner and IF-baseband processor.

cessing chain of the RF-to-digital receiver a sample-and-hold and an ADC perform the transformation from the continuous analog representation of the signal to a uniformly sampled and quantized representation of the same signal. Note that if the arrows in the signal flow path of the receiver were reversed, the processing would represent the tasks performed by the transmitter. Thus, any discussion about receiver structures implemented with continuous and discrete sections are equally applicable to transmitter structures.

At one extreme, the ADC can be located at the output port where it operates at the minimum sample rate with the maximum number of bits. At the other extreme, the ADC can be located early in the processing chain, closer to the antenna, where it operates at the maximum sample rate with a reduced number of bits. In the first of these scenarios all the signal processing is performed in the continuous domain using analog subsystems while in the second a significant fraction of the signal processing is performed in the sampled data domain using DSP hardware and software. These very different options for the location of the ADC interface are shown in Fig. 2.

In practical realizations the digital receiver is partitioned into a tuner that performs the RF-to-IF conversion, and an IF and baseband processor. The functions performed by the tuner are image rejection, interference suppression, frequency translation, and power gain while maintaining low levels of added noise and signal distortion. The functions performed by the IF and baseband processor are bandpass filtering, final frequency translation, the ADC process, automatic gain control (AGC), carrier recovery, timing recovery, and demodulation. The IF and baseband processor also performs the channel equalization if required. This is an important partition and the value of this front-end RF processing cannot be overstated. Signal conditioning errors (such as inadequate image suppression and front end overload) in the RF chain cannot be reversed in subsequent processing. This partition is shown in Fig. 3.

4. A DSP RECEIVER SHOULD NOT EMULATE AN ANALOG RECEIVER

The IF and baseband processor can be realized as an all-analog implementation or as a mix of analog and digital hardware. It is tempting to visualize the analog IF and baseband processor as a prototype and model of the signal processing functions to be synthesized in the DSP version. A word of caution: If the separate analog functions are synthesized by DSP hardware and software and then assembled as a replica of the analog process, we obtain a digital emulation of the analog prototype which is not the desired goal. The DSP replica of the prototype may in fact work well, but it is more likely that while eliminating some of the problems of an analog prototype system, it inherits many of its defects and then adds DSP-related defects.

Using DSP to emulate a prototype analog implementation is a common practice among the uninitiated. Of one thing we are sure: An implementation obtained in this manner will be suboptimal in terms of system performance, cost, and power budget, and will require

extravagant capabilities of the DSP engine, of the memory block, and of the ADCs. In addition, the DSP-emulated system does not have access to, or take advantage of, the signal processing options unique to the sampled data world for which there do not exist continuous counterparts. In particular, the process of multirate filtering in which sample rate changes are imbedded in the signal processing algorithms is an essential tool required to obtain cost-effective DSP-based implementations of IF and baseband processors. We address the topic of multirate algorithm structures in the next section, but return here to the unique characteristics of a DSP-based receiver.

Figure 4 presents block diagrams of three different implementations of an IF and baseband processor modeled after an analog prototype. The first implementation is an all-analog system. The next option inserts the ADC after the analog down conversion where it collects complex input samples. The third option inserts the ADC at the output of the final analog IF stage, prior to the down conversion, where it collects real input samples that are digitally downconverted as part of the DSP function. A primary reason for inserting DSP in a receiver system is the performance gains available by implementing the matched filtering and demodulation functions of the output processing block with DSP algorithms. As mentioned in the previous paragraph, the propensity to simply replicate the analog processes in the DSP counterpart

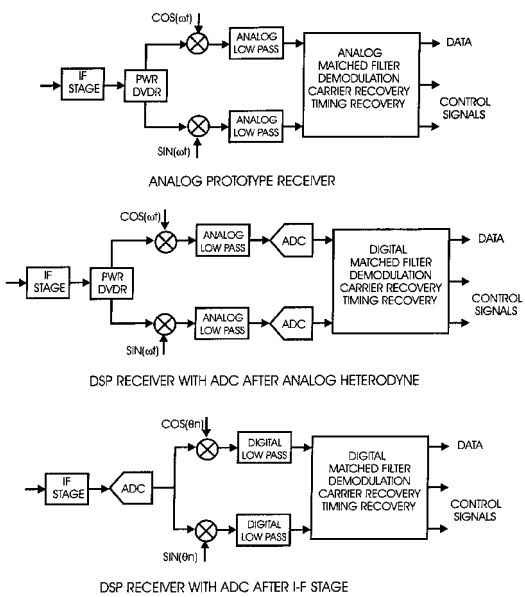

Fig. 4. Locating ADC in DSP-based IF baseband processor.

is not performance- or cost-effective. For instance, the filtering performed by the analog prototype is performed with RLC networks characterized by recursive relationships that almost always exhibit nonuniform group delay. It would be foolish to replace the recursive filters of the analog solution with their recursive counterparts in the digital solution when nonrecursive, linear phase filters are readily available in the discrete world. By the same reasoning, the analog lowpass filters following the analog mixers exhibit a known nonuniform group delay. This group delay can be linearized by imbedding its conjugate phase in a subsequent digital filtering task, such as the matched filter [9, 10]. Linearizing the group delay characteristics of the analog preprocessing filters is made particularly easy if the ADC is followed by an additional digital low-pass filter.

A significant cost-saving option in the transition from an analog to a digital receiver is related to the function and quality of the analog low-pass filters following the final mixer. In the analog system, the analog filters perform the function of eliminating the sum frequencies formed by the mixers as well as limiting the noise bandwidth of the signal. These filters must match in gain and phase over time and temperature to control ghost images due to I-Q mismatch. We require these filters to perform the same tasks in the digital system with the understanding that they are also the system anti-aliasing filters and that their bandwidth and transition bandwidth combine to determine the ADC sample rate as shown in Eq. (1).

$$f_s = 2f_{bandwidth} + f_{transition} \tag{1}$$

5. DSP IMPROVES QUALITY OF DATA COLLECTION PROCESS

In traditional data collection processes, to keep the sample rate close to the Nyquist rate, the transition bandwidth of the anti-alias filter must be a small fraction of the passband bandwidth, and is typically on the order of 10% to 20% of the Nyquist rate. Recall that for the compact disc (CD) and digital audio tape (DAT), the 20 KHz audio bandwidth is sampled at 44.1 KHz and 48 KHz respectively. In a DSP receiver, the sample rate is usually related to the system's symbol rate and is often selected to be 2 samples per symbol to accommodate the excess bandwidth due to spectral shaping and to simplify timing recovery in the receiver. DSP presents the option to replace the high-order, high-cost analog filters characterized by a sharp transition bandwidth with low-

order, lower-cost analog filters that exhibit wide transition bandwidth and operate the ADC at an appropriate, significantly higher sample rate (typically 4 or 8 times the Nyquist rate). The oversampled data are then filtered by finite impulse response filters, with linear or specified compensating phase, to the desired bandwidth by a DSP algorithm and then downsampled to the desired process sample rate (e.g., Nyquist or 2*Nyquist). The combination of two-stage filtering by an initial low-performance analog anti-alias filter, an oversampling process, a digital filter, and a downsampling operation offers a higher-performance data conversion process, usually at a reduced cost. The same reasoning finds application in the inverse task of forming an analog signal from a digital sequence such as in a transmitter modulator or a CD player. By way of example, in order to control the cost of the analog smoothing filter following the DAC in a CD system, the recorded data sequence is digitally upsampled and presented to the DAC at higher sample rates (typically 4 or 8 times Nyquist) and then filtered by a low-order, reduced-cost analog filter with a large transition bandwidth.

In addition to correcting the known imperfect phase and gain characteristics of the analog filters with preset compensating characteristics in subsequent DSP algorithms, we can also correct other analog defects in the receiver described by unknown parameters. We are already familiar with the use of equalizers to correct unknown channel effects and with cancelers to remove structured additive interference. Adaptive filters can also be inserted in the discrete signal flow path to correct defects with unknown parameters introduced in the receiver as well as those introduced by the channel. These include gain and phase imbalance of the analog *I-Q* mixers (which limit achievable ghost rejection) and DC offsets introduced by the ADCs (which bias thresholds in detectors) [11, 12]. The structure of the digital preprocessor performing receiver signal path equalization is shown in Fig. 5. Similar predistortion equalization is performed in a transmitter to remove sin x/x distortion of the DAC and gain and group delay distortion of the smoothing filter following the DAC.

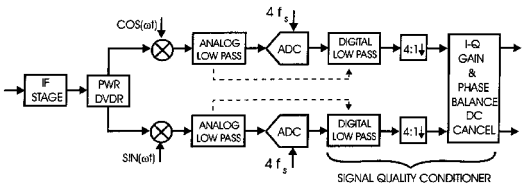

Fig. 5. Improving signal quality with DSP conditioning.

A word about the statistics of conversion errors formed by an ADC. A common representation of an ADC is a memoryless nonlinearity that maps uniformly spaced samples of a continuous analog input $x(n)$ to an approximation $x_q(n)$. This entails mapping the input amplitude to a point on a line that has been partitioned into uniformly spaced intervals over some limited range and then identifying the interval in which the point is located. We visualize this process by a staircase approximation to a linear relationship for which the stair risers delineate the intervals, and the stair treads identify the output corresponding to the intervals. The difference between the input and output of the ADC is the quantization error that we model as additive noise. Under the proper conditions (of sufficient number of output levels and independent dithering) the noise is considered to be zero mean, white, and uncorrelated with the input sequence [13]. The latter properties are easily satisfied so that the white noise model of the quantization error is valid. What is not valid is the zero-mean assumption. Flash and successive approximation converters, when presented with an input, output the highest quantized level not greater than that input level. Quantizers truncate rather than round the measurement to the nearest output level. Thus the quantization errors are all the same polarity, which means that the quantized samples have a DC term in their additive error component. In addition, the analog components of the ADC will exhibit temperature-dependent drift of the zero amplitude level on the order of 2 least significant bits (LSBs).

In addition to the 0.5 bit DC bias due to the quantization process and the slowly varying drift offset of amplitude 2 bits, the amplifiers and sample-and-hold circuit in the signal path exhibit low-frequency $1/f$ noise. These terms experience growth during any accumulation process such as low-pass filtering. This growth occurs independent of the presence of signal, such as the noise-only condition, and results in a bias at the output of the filter. This bias impinges on the dynamic range of data registers and results in degraded detection performance due to signal offset. As an example, if the DSP algorithm is a low-pass filter-reducing bandwidth of the collected data by a factor of 20, the 2.5 bit DC bias term grows by the same factor to amplitude 50, which presets the least significant 6 bits of the output data. Scaling strategies within the algorithm to accommodate numerical growth reduces the observed bias but cannot eliminate it entirely. This DC bias is one of the DSP defects we alluded to earlier. A baseband-centered system with large bandwidth reductions requires a DC estimator subsystem to mea-

sure and cancel this accumulated bias term. This process is performed in the signal conditioning block shown in Fig. 5.

6. INTENTIONAL ALIASING WITH IF SAMPLING

Current design philosophy is that the interface between the analog and digital world should migrate toward the antenna. Following this line of thought, the conversion can take place after the final IF strip with a single ADC as was shown in the third option of Fig. 4. The advantage of this option is that the system no longer requires a pair of matched balanced mixers, matched filters, and a quadrature oscillator. These functions can be provided as DSP operations with significant performance advantages. Note that the DSP equivalent of the mixer is a true multiplier and is not the nonlinear ring modulator or the time-varying Gilbert cell of the analog implementation. The analog implementations have restricted dynamic range related to balancing limitations and to mismatch tolerance between the two paths. The DSP implementation has a dynamic range related only to arithmetic precision where the system performance is purchased at 5 dB per bit. The disadvantage is that the ADC must operate at a higher sample rate. In another variation, called IF subsampling, to be discussed shortly, we will be able to operate the ADC at the reduced rate related to the signal bandwidth rather than its center frequency, providing the sample and hold has the full signal bandwidth and providing the IF filter has sufficient out-of-band attenuation.

Let us consider a simple example for which the signal of interest is at an IF frequency of 450 KHz with a two-sided bandwidth of 25 KHz. Since the signal is real at this point, as indicated in Eq. (2a), an acceptable sample rate is any frequency that exceeds twice the highest frequency by more than 20%. For this case, f_{MAX} is 462.5 KHz, so f_s must be greater than or equal to 1017.5 KHz. After conversion, the center frequency will be translated to baseband by a digital heterodyne. For ease of generating the *I-Q* samples of the digital heterodyne, it is desirable to select the sample rate to be a rational multiple of the center frequency. This restriction, indicated in Eq. (2b), permits the sine–cosine values to be periodic in a short look-up table. Following the digital heterodyne and low-pass digital filtering, the processed data will likely be downsampled to their Nyquist rate. For this condition, indicated in Eq. (2c), it is desirable that the sample rate also be an integer multiple of the output bandwidth or symbol rate.

$$f_S \geq 2.2 \cdot (f_C + f_{BW}) \tag{2a}$$

$$f_S = \frac{P}{Q} f_C \tag{2b}$$

$$f_S = M \cdot (2 \cdot f_{BW}) \tag{2c}$$

Possible sample rate solutions satisfying Eq. (2a) and (2b) are shown in Table I with options satisfying Eq. (2c) (multiples of 25) indicated as a side note. Note that the sine–cosine table has entries of the form $\exp(j2\pi fc/fsn)$ or from (2b), $\exp(j2\pi Q/Pn)$, which for option 1 in Table I is $\exp(j2\pi 9/22n)$, a sequence periodic in 22 samples, hence requiring a look-up table with 22 entries. In a similar fashion, look-up tables with 3 and 4 entries are required for options 2 and 3 respectively. Option 3 is interesting, because the center frequency resides at one-quarter of the sample rate and the cos–sin table only contains ±1 and 0 for which the heterodyne requires no multiplication: This option deserves special attention and we will return to a variation of it shortly.

If we select 1100 KHz (option 1 in Table I) as the sample rate, we can perform the heterodyne and filtering suggested by the digital translation shown as the third architecture in Fig. 4 with the addition of a downsampling operation following the filters. In the next section we couple the downsampling and the filtering, but to finish this section we return to option 3 listed in Table I.

There is a signal processing advantage to placing the center frequency of the narrowband signal at the quarter sample rate of the process. The obvious one is the collapse of the heterodyne from the sequence $\exp[-j\theta n]$

Table I. Some Frequencies Satisfying Equations 2a, 2b (and 2c)

$f_S = \frac{407}{180}\ 450 = 1017.5 \geq 2.2\ (450 + 12.5) = 1017.5$
$f_S = \frac{102}{45}\ 450 = 1020$
$f_S = \frac{24}{10}\ 450 = 1080$
$f_S = \frac{22}{9}\ 450 = 1100$ (option 1)
$f_S = \frac{3}{1}\ 450 = 1350$ (option 2)
$f_S = \frac{4}{1}\ 450 = 1800$ (option 3)

to $\exp[-j(\pi/2)n]$ or $[-j]^n$, or explicitly, the periodic sequence $\{1 + j_0, 0 - j_1, -1 + j_0, \text{ and } 0 + j_1\}$, which requires no multiplication and sets half the data in the subsequent filter to be zero valued. We will take advantage of these zero-valued samples in the next section. We can arrange for the signal to be located at the quarter sample frequency, while maintaining a lower sample rate, by intentionally aliasing the signal as part of the sampling process. The first step in this process is that of limiting the bandwidth of the narrowband signal with the analog IF filter. We now select a sample rate that satisfies Eq. (3). Note we obtain a spectral inversion as well as a spectral translation if the minus sign is used in Eq. (3). The K in Eq. (3) indicates which Nyquist interval folds into the primary interval during the alias process. Table II lists frequencies that satisfy Eq. (3) for different degrees of aliasing.

$$f_C = \left(K \pm \frac{1}{4} \right) f_S \tag{3}$$

If we select a sample rate of 200 KHz to sample the narrowband signal located at a 450 KHz IF frequency, the spectra will experience two spectral folds and will alias to 50 KHz. This aliasing is shown in Fig. 6. Note that the noise spectra for these two intervals also fold into the aliased signal bandwidth and must be sufficiently suppressed by the IF filter. The ADC process proceeds in two distinct steps. The first is the actual sampling by a sample-and-hold network, and the second is the conversion of these samples to the quantized approximation. The sample process occurs at a bandwidth equivalent to the Nyquist rate for the carrier-centered signal and must have a bandwidth exceeding 2.2 f_{MAX} or 1017.5 KHz and must exhibit an aperture uncertainty

Table II. Candidate Sample Rates to Alias 450 KHz to Quarter Sample Rate

	$f_C = \left(K + \frac{1}{4} \right) f_S$	$f_C = \left(K - \frac{1}{4} \right) f_S$
K	f_S	f_S
0	$4f_C = 1800$	$4f_C = 1800$
1	$\frac{4}{5} f_C = 360$	$\frac{4}{3} f_C = 600$
2	$\frac{4}{9} f_C = 200$	$\frac{4}{7} f_C = 257.14\ldots$
3	$\frac{4}{13} f_C = 138.46\ldots$	$\frac{4}{11} f_C = 165.45\ldots$

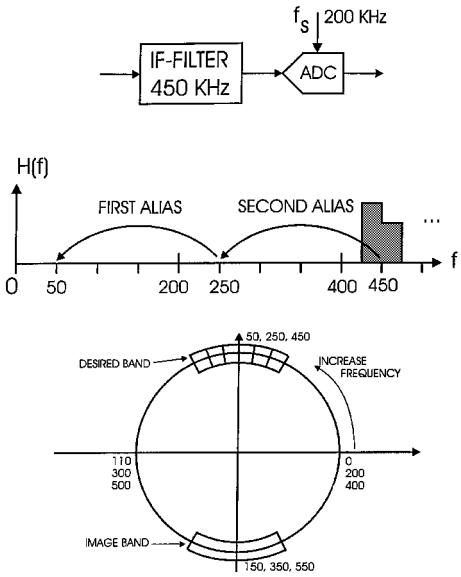

Fig. 6. Spectral translation to quarter sample rate by aliasing IF frequency.

sufficiently small to acquire the data at this full Nyquist rate. In a memoryless quantizer, the aperture uncertainty causes time location errors, which can be viewed as amplitude errors through the slope of the sampled signal. These errors must be less than the LSB of the converter to avoid quantizer degradation. It is the high slope of the full bandwidth carrier–centered signal that drives the aperture constraint as derived in Eq. 4. In Eq. (4a) we define a sampled full-scale sinusoid at maximum input frequency. In Eq. (4b) we offset the nth sample by the time jitter δT to obtain the jittered sample, which we approximate in Eq. (4c). In Eq. (4d) we form the error due to the timing jitter, which we show explicitly in Eq. (4e) and approximate in Eq. (4f). In Eq. (4g) we replace f_{MAX} with the half sample rate and bound the error by $1/2$ bit. Finally in Eq. (4h) we solve the acceptable timing jitter, δT, in terms of the sample rate and number of quantizing bits, b.

$$x(nT) = 2^{b-1} \cos(2\pi f_{MAX} nT) \tag{4a}$$

$$x(nT - \delta T) = 2^{b-1} \cos(2\pi f_{MAX}(nT - \delta T)) \tag{4b}$$

$$\approx 2^{b-1} \cos(2\pi f_{MAX} nT) + 2^{b-1} \sin(2\pi f_{MAX} \delta T) \tag{4c}$$

$$e(nT) = x(nT - \gamma T) - x(nT) \tag{4d}$$

$$= 2^{b-1} \sin(2\pi f_{MAX} \delta T) \quad (4e)$$

$$\approx 2^{b} \pi f_{MAX} \delta T \quad (4f)$$

$$= 2^{b} \pi \frac{1}{2} f_{SMPL} \delta T \leq \frac{1}{2} \quad (4g)$$

$$\delta T \leq \frac{1}{2^{b} \pi f_{SMPL}} \quad (4h)$$

As a reference point, a 16 bit converter sampling a full-amplitude 1 MHz sinusoid requires an aperture uncertainty less than 5×10^{-12} seconds or 5 picosec, the time it takes light to travel 1/16th of an inch. The state of the art in today's technology supports aperture jitter times on the order of 2 to 4 psec and technology shifts of this parameter are glacially slow. It is this parameter that controls the relationship between conversion rate (f_S) and conversion precision (b bits). For every octave increase in sample rate relative to the 16 bit, 2 MHz reference, there must a corresponding decrease of 1 bit. Thus, a 60 MHz converter (increasing f_S by a factor of nearly 2^5) can have at most 11 bits (decreasing b by 5 bits). A 60 MHz converter can have more output pins but cannot have a per sample quantization error below that of an 11 bit converter. Remember, we are now discussing memoryless quantizers with a quantization noise performance of 6 dB per bit on a sample by sample basis. Signal-to-noise ratio of collected data can be improved by systems with memory in which signal correlation is exchanged for additional bit precision by prediction or by bandwidth reduction.

In practice the dynamic range of an ADC is limited by intermodulation terms due to second-order distortion, and a proper measure of an ADC's dynamic range is its spurious free dynamic range (SFDR). Today, 12 bits, 25 MHz sample rate ADCs with 100 dB SFDR are available from high-end suppliers. At lower sample rates, memory and feedback (such as in a sigma-delta converter) make possible 22 bits, 0.5 MHz sample rate ADCs with 110 dB SFDR. These parameters determine the proper insertion point of the ADC in the receiver [14]. This, of course, is a system-dependent decision, but in general the ADC is inserted in the receiver's signal flow path so that subsequent noise bandwidth reduction due to DSP lowers the quantizing noise from its input level to the SFDR of the quantizer. Insertion earlier in the flow will pull the ADC spurs up through the noise, and insertion after this point will deny the system available noise reduction opportunity.

In our IF sampling example we can theoretically sample and thus acquire the 462.5 KHz wide signal with a 16 bit converter at the 1.02 MHz rate without degradation due to aperture jitter. The actual conversion rate, of course, proceeds at the subsample rate of 200 KHz. The signal can now be translated from the quarter sample rate using a digital mixer, filtered using linear phase FIR filters, and downsampled after filtering to the reduced Nyquist rate permitted as a result of the bandwidth reduction by the filters. The problem with this scenario is that the digital system is once again emulating an analog system, and we know better than to do this. In the next section we perform the translation, filtering, and downsampling in a single process unique to the DSP world.

7. INTENTIONAL ALIASING WITH DSP

A sampled data, carrier-centered, narrowband signal can be downconverted or basebanded by emulating a conventional *I-Q* heterodyne and low-pass filter operation. An alternate scheme moves the spectrum to baseband by aliasing it as a result of a change in sample rate. Two related techniques that can alias the carrier-centered signal to baseband involve rearranging the sequence of the operations of heterodyne filtering, and resampling. In the first technique, the order of heterodyne and filtering are reversed by applying the heterodyne to the filter and to the filtered output data rather than to the input data. This relationship, called the equivalency theorem [15], is shown clearly in Eq. (5), where the argument of the heterodyne applied to the data in the convolution is factored, rearranged, and partly extracted from the summation.

$$y(n) = \sum_{k} x(n-k) e^{-j(n-k)\theta} h(k) \quad (5a)$$

$$= e^{-jn\theta} \sum_{k} x(n-k) h(k) e^{jk\theta} \quad (5b)$$

The process of imbedding the heterodyne in the filter weights is an off-line operation and requires no additional work. To this point there has been no reduction in workload, but there will be following the next step. What we have accomplished so far is to move the heterodyne through the filter operation so that it is adjacent to the resampler. The next logical step in this process

is to reorder the operations of heterodyne and resample and only apply the heterodyne to the output samples that survive the downsample operation. This is shown in Eq. (6), where we denote the output of the filter as $r(n)$. Figure 7 illustrates the sequence of reordering we have just described. At this point we have moved the resampler through the heterodyne and reduced the workload of this operation by a factor of M. We can eliminate the heterodyne entirely by a clever choice of center frequency.

$$y(n) = e^{-jn\theta} r(n), \text{where} r(n) = \sum_k x(n-k)h(k)e^{jn\theta} \tag{6a}$$

$$y(nM) = e^{-jnM\theta} r(nM) \tag{6b}$$

The digital frequency θ, with units radians/sample, experiences aliasing to $M\theta$ as a result of the M-to-1 downsampling operation. If this aliased frequency is any multiple of 2π, the aliasing accomplishes the entire heterodyne to baseband. Thus, any digital center frequency of the form $k(2\pi)/M$ (or continuous center frequency kf_S/M) can be aliased to baseband under the M-to-1 downsample.

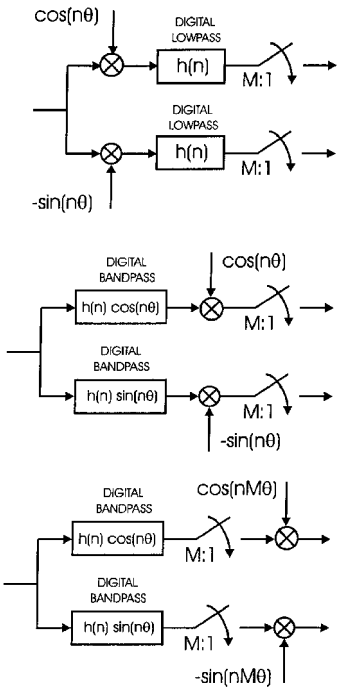

Fig. 7. Heterodyne the filter rather than the data.

Let us return for a moment to the example of the previous section, in which a signal with a 25 KHz two-sided bandwidth centered at 50 KHz and sampled at 200 KHz. If the output of the filter is downsampled 4-to-1, the quarter sample frequency aliases to baseband and the output heterodyne indicated in Eq. (6) drops out of the processing chain. For this example, the output sample rate of 50 KHz is conveniently twice the two-sided bandwidth of 25 KHz.

The final step in the processing chain is to reorder the operations of carrier-centered filtering and resampling, the configuration shown in the third option of Fig. 6. We accomplish this task by partitioning the filter into an M-path filter in anticipation of the M-to-1 downsampling. This partition is indicated in Eq. (7). For convenience of notation in Eq. (7a) we denote the heterodyned filter coefficients as $g(n)$ and in Eq. (7b) show its Z-transform. The M-path partition is shown in Eq. (7c) and the exponent representing the path index is factored from each path summation in Eq. (7d). Equation (7e) is simply a compact representation of Eq. (7d), while Eq. (7f) replaces the simple notation for the filter weights $g(n)$ with the heterodyned weights $h(n)\exp\{j(nM + p)\theta\}$. In Eq. (7g) we factor the exponent of the heterodyne and in Eq. (7h) form the polyphase form of the filter.

$$\text{denote } h(n)e^{j\theta n} = g(n) \tag{7a}$$

$$G(Z) = \sum_n g(n)z^{-n} \tag{7b}$$

$$= \sum_n g(nM)z^{-nM} + \sum_n g(nM+1)z^{-(nM+1)} + \sum_n g(nM+2)z^{-(nM+2)} + \cdots \tag{7c}$$

$$= \sum_n g(nM)z^{-nM} + z^{-1}\sum_n g(nM+1)z^{-nM} + z^{-2}\sum_n g(nM+2)z^{-nM} + \cdots \tag{7d}$$

$$= \sum_p z^{-p} \sum_n g(nM+p)z^{-nM} \tag{7e}$$

$$= \sum_p z^{-p} \sum_n h(nM+p)e^{j(nM+p)\theta} z^{-nM} \tag{7f}$$

$$= \sum_p z^{-p} e^{jp\theta} \sum_n h(nM+p) z^{-nM} e^{jnM\theta} \tag{7g}$$

$$= \sum_p z^{-p} H_p(z^M) e^{jp\theta} \tag{7h}$$

The block diagram representing the partition shown in Eq. (7) can be seen in Fig. 8. The M-path filter, the second segment of Fig. 8, represents Eq. (7g) or Eq. (7h). The third segment is obtained upon moving the resampling switch through the M-path filter stages. This is permitted when we recognize that M-delays at the input sample rate are the same as a single delay at the output sample rate. Formally, this relationship is known as the Noble Identity [16]. Note that the heterodyne terms in each state collapse to unity under the aliasing operation and that the interaction between the input delays and the resampling at the input to each filter is equivalent to a commutator delivering samples to each stage of the M-path filter.

What we have accomplished in this section is the reversal of the signal flow in a standard DSP-based narrowband receiver. The standard structure presented in the center segment of Fig. 7 orders the processing chain such that the selected signal component is first heterodyned, then filtered, and finally downsampled. In the alternate structure, the polyphase filter presented in the final segment of Fig. 8, we have reordered the processing chain so that the composite signal is first downsampled, then filtered, and the selected component is finally extracted by a phase coherent summation. Other comparisons worth noting are: In the first structure the signal is made complex as we enter the algorithm, while in the second the signal is made complex as we leave the algorithm. The low-pass filter of the first structure is repeated for both the I and Q processing paths, while the coefficients of the same (single) filter are distributed over the M-path polyphase partition. If we happen to require two or more subchannels in the architecture of the first structure, each channel would require a duplicate copy of the first process, while in the second structure every channel would share the existing M-stage polyphase partition and would only differ in their phase-coherent summations. In the limit, if all $M/2$ component channels of the input signal are desired as outputs, the collection of phase-coherent sums is a discrete Fourier transform (DFT), which of course would be implemented as a fast Fourier transform (FFT) [17]. Again, note that reversing the signal flow arrows of the resampling aliased-based receiver will lead to the signal flow graph of the resampling aliased-based transmitter.

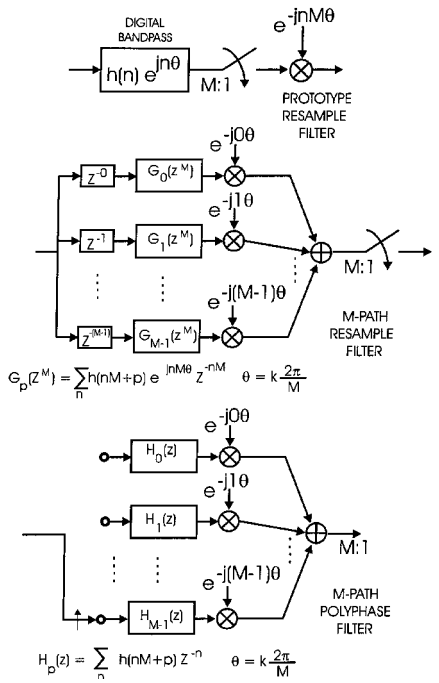

Fig. 8. Polyphase filter downsamples data.

7.1. Other Receiver Functions Implemented By DSP

A receiver performs a number of essential secondary functions required to support the primary functions of spectral translation, bandwidth reduction, matched filtering, and symbol detection. These functions include carrier recovery, timing recovery, automatic gain control, and, if required, channel equalization. In standard receivers, as suggested in Fig. 9, channel equalization is the only process performed with DSP techniques while the others are traditionally performed with analog control loops or with hybrid systems using DSP control of analog components. The control signals or data in the digital receiver section are digital and in order to interact with the analog world must be converted back to analog by a DAC and low-pass smoothing filter for each signal leaving the digital section. Returning to the analog world to effect the control of the analog functions is another example of reproducing an analog solution with a DSP emulation. The analog AGC block is the only function that must be performed in the analog domain, but we

are not obligated to supply an analog control voltage to effect the operation. We could replace the analog-controlled AGC block with a digitally controlled attenuator block and simply use a bit serial control stream for the AGC loop. Suppose we must deliver an analog control signal. As an example, we are required to control the analog AGC block with an analog control signal that spans 3 orders of magnitude (60 dB) with an accuracy of ±0.25 dB. This would require a 12-bit control word that we can generate and deliver as a serial or parallel bit stream to an external DAC and low-pass filter. Since the loop bandwidth is generally much smaller than the DSP system clock rate, we can consider the control words as being significantly oversampled. From this perspective, we can apply the digital control data words to sigma-delta DAC residing in the digital receiver and output a 1 bit data stream. This stream, after smoothing by a simple RC filter, represents the low-bandwidth full-precision analog control signal.

In a similar fashion to the analog AGC control, we have the same options for timing and carrier recovery control loops. We could replace the voltage-controlled oscillators (VCOs) with direct digital synthesizers (DDSs) and control them with digital control words, or apply the proper width (say 16 bit) control words to external DACs and analog smoothing filters, or once again requantize the low-bandwidth oversampled control words in sigma-delta modulators and output 1 bit data streams that are smoothed by simple RC filters.

We have access to a fourth option that is shown in Fig. 10. We can apply a full DSP-based solution to the carrier and timing recovery tasks. The carrier recovery can be implemented by a complex product between the complex input data and the output of a complex DDS. The only caution to exercise here is that the bandwidth of the analog anti-aliasing filters must be widened sufficiently to accommodate the input signal bandwidth plus the worst-case frequency offset and may require some coupling between the carrier loop control signal and the synthesizer of the previous translator. This increases the noise bandwidth presented to the ADCs and may permit strong adjacent channel interference to absorb part of the ADC's dynamic range, which in turn impacts the number of bits required by the ADC.

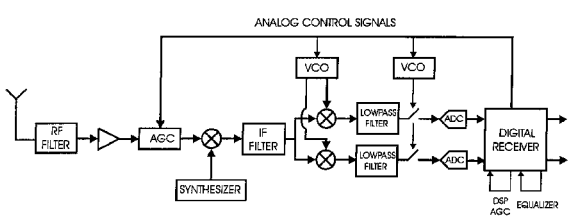

Fig. 9. Ancillary signal conditioning functions.

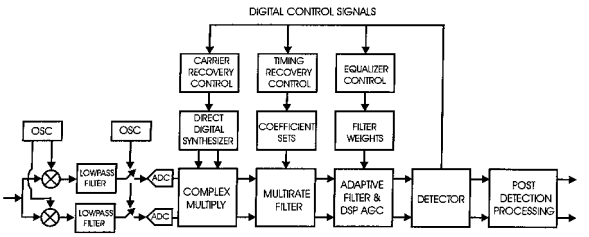

Fig. 10. DSP-based signal conditioning functions.

The timing recovery loop uses sample statistics to determine if the time position of the data samples is at the peak of the eye opening. If the samples are not phased properly, the loop changes a parameter to move the time samples to the correct position relative to the eye opening. In the conventional prototype analog solution, the sampling clock is phase shifted by controlling the voltage to a VCO. In the full DSP solution this same effect is accomplished by interpolating from the input sample times to the desired output sample times. This can be accomplished in many ways including polynomial and band-limited interpolators as well as Taylor series interpolators. These techniques are equivalent to a polyphase filter bank and differ primarily in implementation. We can imagine the interpolator as a bank of coefficients that can compute sample values over some granular time scale such as 1/16th or 1/32nd of the intersample time interval. The timing loop control mechanism controls the selection of the weights corresponding to the proper sample offset. Essentially, the filter bank raises the sample rate by a factor of M and the control loop reduces the rate by the same factor, so there is no real change in sample rate. The output samples, however, are offset relative to the input samples, and of course we do not bother computing the samples we will discard in the downsampling operation. If the input sample rate differs from the output symbol rate the control loop will periodically change the selected weights to effect the phase lock. A potential problem here is that as one clock overtakes the other we run the risk of data overflow or underflow. The clock alignment requires input and output buffering with some flow mechanism control for data transfer. A simple alternative strategy is to frequency lock the timing clock and use the resampler for phase lock.

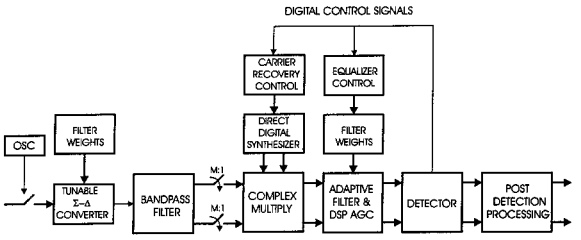

Fig. 11. Alternate DSP-based signal conditioning functions.

Figure 11 offers another configuration for a DSP-based receiver. Here the IF sampling collects data at the higher rate with the expectation that the band selection and bandwidth reduction will occur in the DSP block. If the output bandwidth is a small fraction of the collected bandwidth, we can argue that the desired signal is significantly oversampled. We can then use a tunable sigma-delta modulator to requantize the input data to a smaller number of bits (say 1, 2, or 4 depending on the oversample rate) while preserving the signal-to-quantizing noise ratio in the selected band. These requantized data are delivered to a polyphase filter bank, which will effect a bandwidth reduction, a band translation, and a sample rate change (as shown in Fig. 8). The advantage of using the sigma-delta sample rate change (as shown in Fig. 8). The advantage of using the sigma-delta modulator as a signal conditioner is that the subsequent processing can now proceed with reduced complexity hardware. For instance, if the data are requantized to 1 or 2 bits, the filtering process can be implemented without multiplication but rather with adds or with short look-up tables and adds. This is a clever way to map an algorithm into a field programmable gate array (FPGA), a processor with elementary processing elements that can perform look-up and adds very efficiently [18–20].

8. CONCLUSIONS

We have examined and reviewed a number of ways that DSP can be inserted in the signal flow path of an RC receiver or transmitter with primary emphasis on the receiver. We have shown that the DSP can be used as a substitute for specific signal processing tasks by simple substitution of the analog process by the equivalent DSP process. We suggest that this direct substitution may lead to unsatisfactory results, and in particular it is not a good idea if the ratio of sample rate to processed bandwidth is large. This can occur, for instance, when we extract a pilot tone from a larger collected bandwidth or when we pull one channel from a multichannel signal. A particularly valuable option available to the DSP designer is the resampling filter. I have this philosophy that has served me well in system design: Always operate DSP algorithms at (or near) the Nyquist rate for the signal being processed. If the signal bandwidth is being reduced, reduce the sample rate proportionally. Sample rate changes have a second interesting effect—they can alias spectra! Spectra can be aliased down with a downsample operation and can be aliased up with an upsample operation. What is remarkable is that sample rate reductions do not have to follow bandwidth reduction but can precede it, causing aliasing, which can be canceled by a phase-coherent summation at the output data rate. We also considered ancillary signal processing tasks that could be brought into the DSP realm with a change in perspective. And finally we considered the option of requantizing with a sigma-delta modulator as a signal conditioning option for a DAC (or ADC) operation and for signal conditioning when the sample rate is large compared to the bandwidth and will be reduced as part of the processing. By no means have we exhausted all the tricks and perspectives. There are many clever people addressing old problems with new ideas and there are still many more to come. I for one find delight when I discover a new DSP trick or perspective and I keep my eyes and mind open for them. I hope this overview has been useful for expanding your horizons.

REFERENCES

1. fred harris, Multirate digital signal processing in digital receivers, *Second International Symposium on DSP for Communication Systems (DSPCS-94)*, Adelaide, Australia, April 26–29, 1994.
2. fred harris, Design considerations and design tricks for digital receivers, *Ninth Kobe International Symposium on Electronics and Information Sciences*, Kobe, Japan, June 18–19, 1991.
3. fred harris and T. Wechselberger, Multirate all-digital MODEM for support of universal multiplex transport layer for digital compression, *1993 NCTA, National Cable Television Association Conference*, San Francisco, California, June 7–9, 1993. (Reprinted in *Communications Technology*, September 1993.)
4. Jim Marsh, Scott Williams, Brent Jensen, Simon Atkinson, Jonathon Strange, Dave May, Tom Bilotta, and fred harris, Quadrature demodulator and digitizer for PSK and QAM applications, *Wireless Symposium*, Santa Clara, California, February 15–18, 1994.
5. Joe Mitola, The software radio architecture, *IEEE Communications Magazine*, May 1995, pp. 26–38.
6. Rupert Baines, The DSP bottleneck, *IEEE Communications Magazine*, May 1995, pp. 46–54.
7. Jeffery Wepman, Analog-to-digital converters and their applications in radio receivers, *IEEE Communications Magazine*, May 1995, pp. 39–45.

8. Bernard Sklar, *Digital Communications: Fundamentals and Applications*, Prentice Hall, 1988.
9. fred harris, On the design of pre-equalized square-root raised cosine matched filters for DSP based digital receivers, *Twenty-Seventh Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, Oct. 31–Nov. 3, 1993.
10. fred harris, Implementing waveform shaping filters to pre-equalize gain and phase distortion of the analog signal processing path in DSP based modems, *MILCOM-94*, Ft. Monmouth, New Jersey, October 2–5, 1994.
11. fred harris, On measuring the gain and phase unbalance and DC offsets of quadrature A-to-D converters with an adaptive canceling filter, *Twenty First Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, Nov. 2–4, 1987.
12. fred harris, "Digital Filter Equalization of Analog Gain and Phase Mismatch in I-Q Receivers," *International Conference on Universal Personal Communications (ICUPC-96)*, Boston, Massachusetts, Sept. 30–Oct. 2, 1996.
13. N. S. Jaynt and Peter Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.
14. Don Steinbrecher, Some fundamental limits governing the design of digital receivers, *IEEE International Conference on Analogue-to-Digital and Digital-to-Analogue Conversion*, Swansea, Wales, Sept. 17–19, 1991.
15. John Wozencraft and Irwin Jacobs, *Principles of Communications Engineering* , John Wiley & Sons, 1967.
16. P. P. Vaidyanathan, *Multirate Filters and Filter Banks*, Prentice Hall, 1993.
17. fred harris, On the relationship between multirate polyphase filters and windowed, overlapped, FFT processing, *Twenty Third Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, Oct. 30–Nov. 1, 1989.
18. Chris Dick and fred harris, Narrow band FIR filtering with FPGAs using sigma-delta modulation encoding, *Journal of VLSI Signal Processing*, 4th quarter, 1996.
19. Chris Dick and fred harris, Multiplier-free FIR filtering with FPGAs using a tunable sigma-delta modulator, *Proceedings of the Australian Computer Architecture Workshop 96 (ACAW'96)*, Melbourne, Australia, Jan. 29–30, 1996, pp. 100–124.
20. Chris Dick and fred harris, FPGA implementation of high order FIR filters by re-quantizing the input data stream, *SPIE Conference on Phonetics East '95, Field Programmable Gate Arrays (FPGAs) for Fast Board Development and Reconfigurable Computing*, Philadelphia, Pennsylvania, October 25–26, 1995.

fred harris is a Professor in the E&CE Department at SDSU, where he holds the Signal Processing Chair and teaches courses in digital signal processing and communication systems. His particular areas of interest are multirate signal processing, fast algorithms, fast Fourier transforms, adaptive algorithms, noise feedback quantizers, and DSP algorithms applied to communication systems such as satellite and cable modems, equalizers, and other digital signal conditioning systems. He has consulted for the past 25 years on advanced DSP applications, holds a number of patents on digital receivers, and lectures around the world on DSP applications. He was formerly Chief Scientist at Tiernan Communications, where he coordinated designs of DSP algorithms and systems for advanced communication applications. He collects old toy trains and drives secretaries and editors to distraction by requesting the use of lowercase letters when spelling his name.