Frequency Modulation (FM) Tutorial
Lawrence Der, Ph.D.
Silicon Laboratories Inc.

Introduction

Edwin H. Armstrong, known as one of the founding fathers of radio technology, invented the superheterodyne radio receiver in 1918 and frequency modulation (FM) in 1933 [1]. These two concepts, along with his regenerative circuit technique developed in 1912, formed the basis of radio frequency electronics as we know it today. In the United States, FM radio stations broadcast between radio frequencies of 88 MHz to 108 MHz with a channel bandwidth of 200 kHz. FM radio was first deployed in monaural in 1940; and in 1960, FM stereo was introduced. This article presents a basic tutorial on FM with descriptions of multiplex (MPX) signaling and noise improvement techniques such as stereo-mono blending and soft mute.

FM Basics

Frequency modulation is a form of analog angle modulation in which the baseband information-carrying signal, typically called the message or information signal m(t), varies the frequency of a carrier wave. Audio signals transmitted by FM radio communications are the most common. However, FM radio can also transmit digital data with the low bandwidth digital information known as Radio Data System (RDS) in Europe and Radio Broadcast Data System (RBDS) in the U.S. The simplest approach to generating FM signals is to apply the message signal directly to a voltage-controlled oscillator (VCO) as shown in Figure 1.
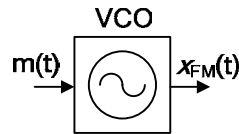

Figure 1. FM Generation with a VCO

A voltage message signal, m(t), is applied to the control voltage of the VCO, and the output signal, $x_{FM}(t)$, is a constant amplitude sinusoidal carrier wave whose frequency is ideally a linear function of its control voltage. When there is no message or the message signal is zero, the carrier wave is at its center frequency, f_c. When a message signal exists, the instantaneous frequency of the output signal varies above and below the center frequency and is expressed by

$$f_i(t) = f_c + K_{VCO}m(t)$$

where K_{VCO} is the voltage-to-frequency gain of the VCO expressed in units of Hz/V, and the quantity, $K_{VCO*}m(t)$, is the instantaneous frequency deviation. The instantaneous phase of the output signal is equal to 2π multiplied by the integral of the instantaneous frequency as shown below

$$\theta_i(t) = 2\pi f_c t + 2\pi K_{VCO} \int_0^t m(t)dt$$

where the initial condition of the phase is assumed to be zero for simplicity. Hence, the FM output signal, $x_{FM}(t)$, is given by the following equation

$$x_{FM}(t) = A_c \cos\left[2\pi f_c t + 2\pi K_{VCO} \int_0^t m(t)dt\right]$$

A few observations can be made from the FM output signal. First, the amplitude of an FM signal is constant regardless of the message signal, giving it a constant envelope property with an output power equal to $A_c^2/2$ into a 1 Ω resistor. Second, the frequency-modulated output, $x_{FM}(t)$, has a nonlinear dependence to the message signal, m(t), making it difficult to analyze the properties of an FM signal. To estimate the bandwidth of an FM signal, a single tone message signal is used as shown below

$$m(t) = A_m \cos(2\pi f_m t)$$

where A_m is the amplitude of the message signal and f_m is the frequency of the message signal. Substituting this message signal into the above formulas, we find

$$x_{FM}(t) = A_c \cos\left(2\pi f_c t + \frac{K_{VCO} A_m}{f_m} \sin(2\pi f_m t)\right)$$

$$= A_c \cos\left(2\pi f_c t + \frac{\Delta f}{f_m} \sin(2\pi f_m t)\right)$$

$$x_{FM}(t) = A_c \cos\left(2\pi f_c t + \beta \sin(2\pi f_m t)\right)$$

The quantity $\Delta f = K_{VCO} A_m$ represents the peak frequency deviation of the FM signal from the center frequency and is directly proportional to the amplitude of the message signal and the gain of the VCO. This quantity, Δf, is called the maximum instantaneous frequency deviation. The ratio of the frequency deviation, Δf, to the message signal frequency, f_m, is called the modulation index, β.

For a single tone message signal, the number of significant sidebands in the output spectrum is a function of the modulation index. This can be seen by first writing the FM output signal in terms of nth order Bessel functions of the first kind [2, 3].

$$x_{FM}(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos\left(2\pi(f_c + nf_m)t\right)$$

By taking the Fourier transform, we see a discrete FM output spectrum with magnitude coefficients as a function of β as shown in the equation below.

$$x_{FM}(f) = \frac{A_c}{2} \sum_{n=-\infty}^{\infty} J_n(\beta)\left[\delta(f - f_c - nf_m) + \delta(f + f_c + nf_m)\right]$$

The number of sidebands of an FM signal and its associated magnitude coefficient can be found with the help of Bessel function tables such as the one shown in Table 1.

β	J0	J1	J2	J3	J4	J5	J6	J7	J8
0	1								
0.25	0.98	0.12							
0.5	0.94	0.24	0.03						
1.0	0.77	0.44	0.11	0.02					
2.0	0.22	0.58	0.35	0.13	0.03				
3.0	-0.26	0.34	0.49	0.31	0.13	0.04	0.01		
4.0	-0.40	-0.07	0.36	0.43	0.28	0.13	0.05	0.02	
5.0	-0.18	-0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02

Table 1. Bessel Functions of the First Kind Rounded to Two Decimal Places.

If $A_c^2/2 = 1$, β=1, f_m = 1 kHz, and f_c = 100 kHz, then the result is the FM voltage spectrum shown in Figure 2

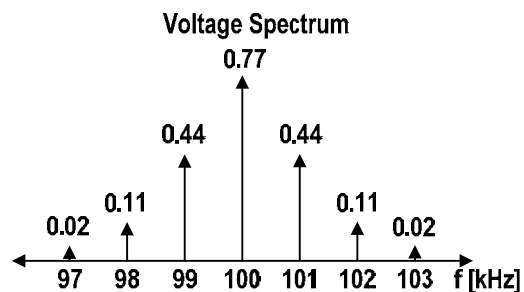Figure 2. FM Voltage Spectrum for $\dfrac{A_c^2}{2} = 1$, β=1, fm = 1 kHz, and fc = 100 kHz.

A key point of modulation index, β, is that it determines the bandwidth of the signal by determining the number of effective sidebands of an FM signal. For instance, if β=0.25, only one sideband is needed; while if β=5, eight sidebands are required. Another important point about the modulation index: it can change a lot even for a fixed frequency deviation because the message signal frequency can vary. In general, as the modulation index increases, the number of sidebands increases and the bandwidth goes up. However, the increase in modulation index due to decreasing message frequency (recall $\beta = \Delta f / f_m$) may not necessarily increase the FM bandwidth. The bandwidth is equal to the number of discrete spectral tones multiplied by the frequency spacing set by the message signal frequency f_m. For more complicated message signals, the bandwidth of an FM signal can also be approximated with Carson's rule, $BW_{FM} \approx 2(\beta+1)f_m$ [2, 3]. The empirical relation states that the number of significant spectral tones in an FM spectrum is $\approx 2(\beta+1)$, not including the carrier. For example [2], in North America, the maximum frequency deviation, Δf, is 75 kHz for commercial FM broadcasting. If the maximum message frequency is equal to 15 kHz for audio, then $\beta = 75kHz/15kHz = 5$, and the FM bandwidth is $BW_{FM} = 2(5+1)15kHz = 180kHz$. This is close to the allotted 200 kHz channel bandwidth. If the Bessel functions are used to approximate the bandwidth, the bandwidth of (2 x 8 +1)15 kHz = 255 kHz is achieved. In practice, the last few side tones may contribute negligible power, thus reducing the bandwidth to about 200 kHz (assuming the tones

that are below -10 dBc are negligible). Again, it is important to remember that these equations are derived from a single tone message signal, which is much different from real-world message signals that contain many different frequencies at the same time. In this case, the maximum frequency of the real message signal can be used as an approximation for f_m.

To recover the message signal from the FM signal, frequency demodulation must be performed. The most basic frequency demodulator consists of a frequency discriminator, which is a differentiator followed by an envelope detector, as shown in Figure 3.
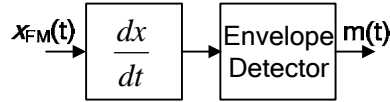

Figure 3. Ideal Frequency Discriminator

The differentiator converts the FM signal to an AM signal as shown in the following equation

$$\frac{dx_{FM}(t)}{dt} = -A_c(2\pi f_c + 2\pi K_{VCO}m(t))\sin(2\pi f_c t + 2\pi K_{VCO}\int_0^t m(t)dt)$$

and the envelope detector can be used to recover m(t) [4]. Differentiation is one of the key steps used in FM demodulation. However, an unfortunate by-product of differentiation is that it amplifies high-frequency noise and degrades the overall signal-to-noise ratio (SNR) of the recovered message signal. To compensate, FM broadcasters insert a pre-emphasis filter prior to FM transmission to amplify the high-frequency content of the message signal. All FM receivers incorporate a reciprocal de-emphasis filter after the FM receiver to attenuate high-frequency noise and interference and restore a flat message signal frequency response. Figure 4 shows the block diagrams of an FM transmitter with a pre-emphasis filter, $H_{pe}(f)$, and an FM receiver with a de-emphasis filter, $H_{de}(f)$.
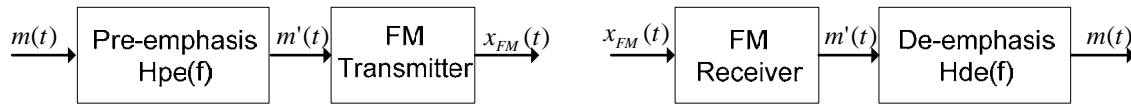

Figure 4. Pre-Emphasis and De-Emphasis in FM System

The pre-emphasis filter has a high-pass characteristic transfer function given by

$$H_{pe}(f) = 1 + j2\pi f \tau_x$$

and the de-emphasis filter has a low-pass characteristic transfer function given by

$$H_{de}(f) = \frac{1}{1 + j2\pi f \tau_x}$$

where the time constant, τ_x, is the pre-emphasis/de-emphasis time constant. The two time constants used in various regions of the world are 75 µs (in regions including the US) and 50 µs (in regions including Europe).

The SNR of an FM system for mono signals without pre-emphasis and de-emphasis is

$$SNR_{FM} = 3\beta^2(\beta+1)\ CNR$$

$$SNR_{FM} \approx 3\left(\frac{B_T}{2W}\right)^3 CNR$$

where B_T is the FM transmission bandwidth ($= BW_{FM}$), W is the message signal bandwidth ($\approx f_m$) and CNR is the carrier-to-noise ratio equal to $A_c^2/2B_T N_o$ where $N_o/2$ is the two-sided power spectral density of white noise [2]. The above SNR equation illustrates the trade-off that exists between message signal quality (SNR) and FM transmission bandwidth. With an FM transmission bandwidth of 200 kHz and a message signal bandwidth of 15 kHz ($\beta \approx 5.67$), it is reasonable to expect the SNR at the output of an FM receiver to have an FM gain of 27 dB above the CNR. However, the above SNR equations are only valid for large CNRs. As the CNR at the input of the FM discriminator is decreased, it will eventually generate impulse noise, resulting in clicks and crackling. The onset of impulse noise tells us that the FM receiver has just entered a noise threshold region known as the threshold effect. The FM threshold is defined as the minimum CNR yielding FM improvements that do not significantly deviate from the theoretical equation given for FM SNR [2]. As noted previously, the use of pre-emphasis and de-emphasis filters is one approach to improving the SNR of an FM system by attenuating high frequency noise. The actual improvement factor, I, in output SNR of an FM receiver using pre-emphasis and de-emphasis filters is

$$I = \frac{\left(W/f_x\right)^3}{3\left[\left(W/f_x\right) - \tan^{-1}\left(W/f_x\right)\right]}$$

where $f_x = 1/2\pi\tau_x$ is the 3-dB corner frequency of the pre-emphasis and de-emphasis filters [2]. With a 3-dB corner frequency of 2.1 kHz and a message signal bandwidth of 15 kHz, an improvement factor of 13 dB can be achieved from the pre-emphasis and de-emphasis filters. This improvement factor also assumes a large CNR at the input of the FM discriminator. Thus, the total SNR improvement for a mono signal above threshold from FM gain and pre-emphasis and de-emphasis filtering is $\approx$ 27 dB + 13 dB = 40 dB, assuming an FM transmission bandwidth of 200 kHz, a message signal bandwidth of 15 kHz, and a 3-dB pre-emphasis and de-emphasis corner frequency of 2.1 kHz ($\tau_x = 75\mu s$). Care must be taken when interpreting this result because the equation suggests that it is possible achieve an FM SNR of 40 dB with a carrier-to-noise ratio of 0 dB. Generally, this will not be the case because standard FM demodulators typically exhibit a threshold at 12 dB CNR [5] invalidating the above results. Moreover, the SNR improvement for stereo signals is only 17 dB above CNR [dB] for CNRs above threshold [5]. The equations below summarize the audio SNR improvements for FM

$$SNR_{MONO} = 40 + CNR \ [dB]$$

$$SNR_{STEREO} = 17 + CNR \ [dB]$$

when the CNR is above threshold [5].

Stereo FM – Multiplex Signal
Prior to 1961, monaural broadcasting of audio signals was the standard for AM, FM and TV. FM broadcasts at that time also included Subsidiary Communications Authorization (SCA) services that were multiplexed with the main monophonic channel to provide background music and other services to offices and stores. In 1961, the FCC approved the transmission of stereophonic sound, which extends the idea of multiplexing signals to generate stereo audio. One of the key requirements of the stereo multiplex signal was to be backwards compatible with the large existing base of FM monophonic receivers. To accomplish this goal, the 0 to 15 kHz baseband part of the multiplex (MPX) signal had to contain the left (L) and right (R) channel information (L+R) for monophonic reception. Stereophonic sound is achieved by amplitude modulating the (L-R) information onto a suppressed 38 kHz subcarrier in the 23 to 53 kHz region of the baseband spectrum. A 19 kHz pilot tone is added to the multiplex signal to enable FM stereo receivers to detect and decode the stereo left and right channels. The composite baseband signal format meets the backwards compatibility needed for FM mono receivers while simultaneously providing enough information for FM stereo receivers to decode the left and right stereo channel outputs. Today's MPX signal includes a 57 kHz subcarrier that carries RDS and RBDS signals [6]. Figure 5 shows a spectrum of a modern-day MPX baseband signal.
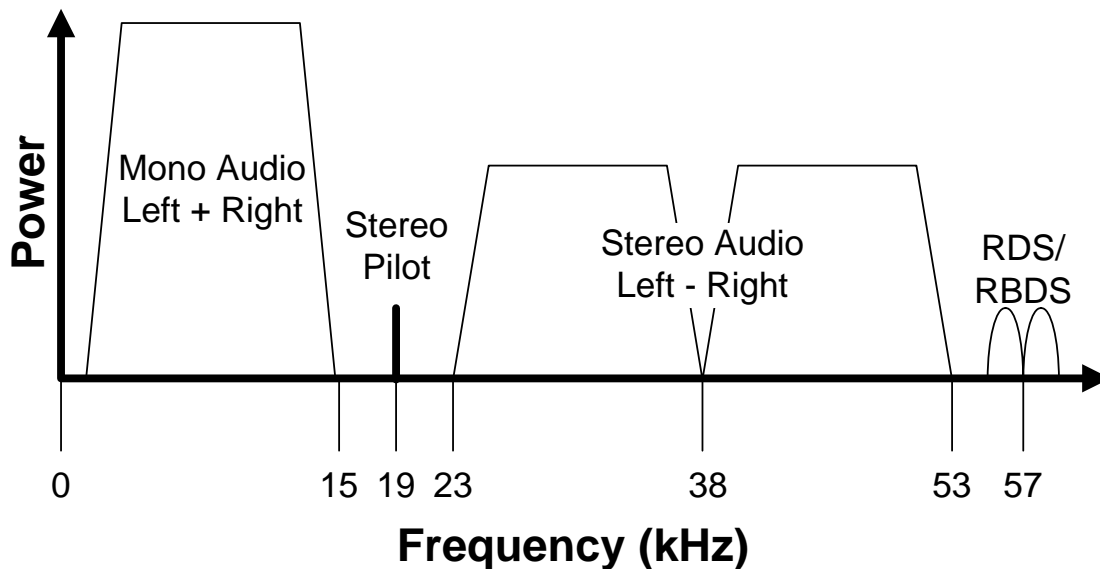

Figure 5. MPX Baseband Spectrum

The mathematical analysis presented in the previous section assumes that the message signal, m(t), is a single-tone sinusoidal signal. In reality, the message signal used in today's FM broadcasts is the MPX signal with a baseband spectrum similar to the one shown in Figure 5. The FCC has set modulation limits of 100% modulation (an instantaneous frequency deviation of 75 kHz corresponds to a 100 percent modulation) for stereophonic transmission and up to 110

percent modulation for SCA multiplex subcarriers under certain conditions [5]. Figure 6 shows an example modulation level breakdown for the various signals in a typical MPX message signal.
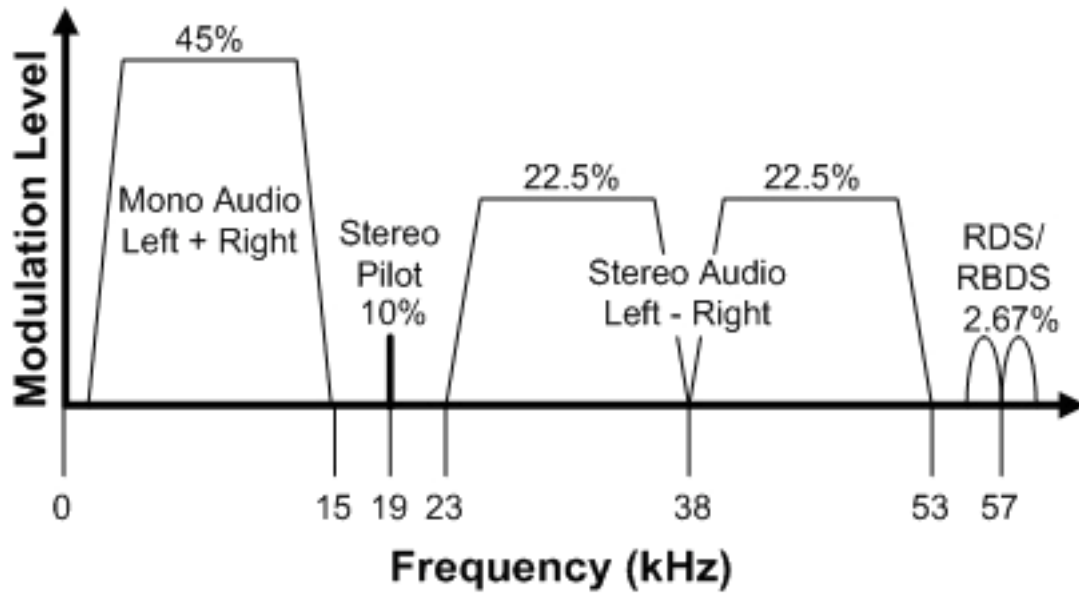

Figure 6. MPX spectrum showing modulation level

The total modulation level for the MPX signal shown in Figure 6, assuming no correlation, is the arithmetic sum of each of the subchannel levels giving 102.67 percent modulation or a peak frequency deviation of 77.0025 kHz. From the last section, the frequency deviation is related to the amplitude of the message signal by the constant, K_{VCO}, since $\Delta f = K_{VCO} A_m$. Thus, for a fixed K_{VCO}, the amplitude of all the subchannel signals within the MPX message signal must be scaled to give the appropriate total frequency deviation.
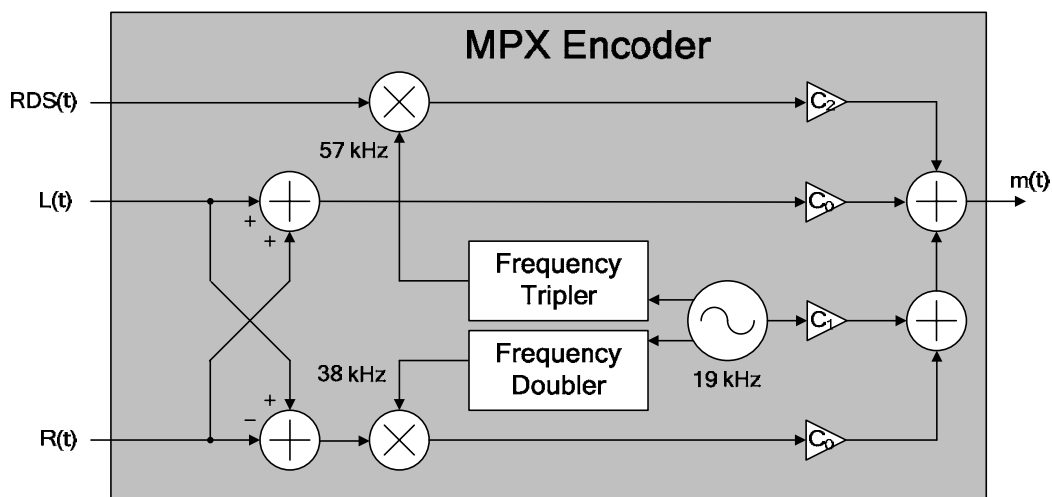

Figure 7. MPX Encoder

Figure 7 shows a conceptual block diagram of an MPX encoder used to generate the MPX signal. L(t) and R(t) denote the time domain waveforms from the left and right channels and RDS(t) denotes the time domain waveform of the RDS/RBDS signal. The MPX message signal can be expressed as

$$m(t) = C_0[L(t) + R(t)] + C_1 \cos(2\pi * 19kHz * t) + C_0[L(t) - R(t)]\cos(2\pi * 38kHz * t) + C_2 RDS(t)\cos(2\pi * 57kHz *$$

where C_0, C_1, and C_2 are gains used to scale the amplitudes of the $(L(t) \pm R(t))$ signals, the 19 kHz pilot tone, and the RDS subcarrier, respectively, to generate the appropriate modulation level.
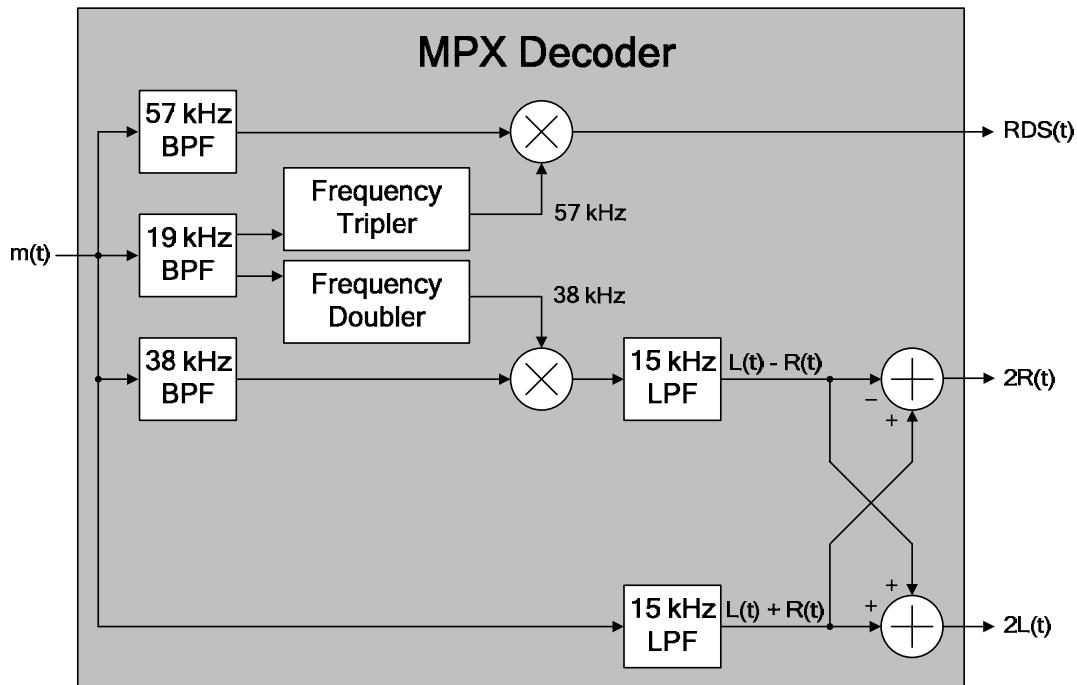

Figure 8. MPX Decoder

Figure 8 shows a conceptual block diagram of an MPX decoder used to recover the left, right and RDS signals from the MPX message signal, m(t). The message signal is applied to three bandpass filters with center frequencies at 19, 38 and 57 kHz and to a low-pass filter with a 3-dB cutoff frequency of 15 kHz. The 19 kHz bandpass filter is a high-Q filter used to extract the 19 kHz pilot tone from the MPX message signal. The recovered pilot tone is frequency-doubled and tripled to produce the required local oscillator (LO) signals needed to demodulate the (L-R) and RDS signals, respectively. By adding and subtracting the (L+R) and (L-R) signals, a scaled version of the left and right channels is recovered for stereophonic sound. RDS is brought back down by mixing with a 57 kHz LO, and the data can be recovered by sending this signal to a matched filter.

The above analysis reveals the difficulty in maintaining good stereo separation. First, if a monaural signal is applied to the input of the decoder, the pilot tone, (L-R), and RDS signals are equal to zero because they do not exist for monaural signals. The left and right outputs of the

decoder would be the same and equal to the (L+R) signal thereby recovering the mono signal. Second, any gain or phase mismatch in the generation of the MPX message signal and/or recovery of the left and right channels leads to finite stereo separation, so the left channel has some right channel information, and the right channel has some left channel information (also known as channel separation or cross talk). For example, if the 15 kHz low-pass filters of the decoder in Figure 8 have a gain mismatch of 1 percent, the stereo separation would be about -46 dB. This example illustrates that the left and right signal paths must match in both amplitude and phase to maintain good stereo separation, which can be difficult if the encoder and decoder circuits are implemented with analog circuits.

Noise Improvement Techniques

Recent implementations of FM tuners, such as Silicon Laboratories' Si4700 FM tuner and Si4701 FM tuner with RDS/RBDS, have incorporated noise improvement techniques such as stereo-mono blending and soft mute to improve the audio quality of FM radios.
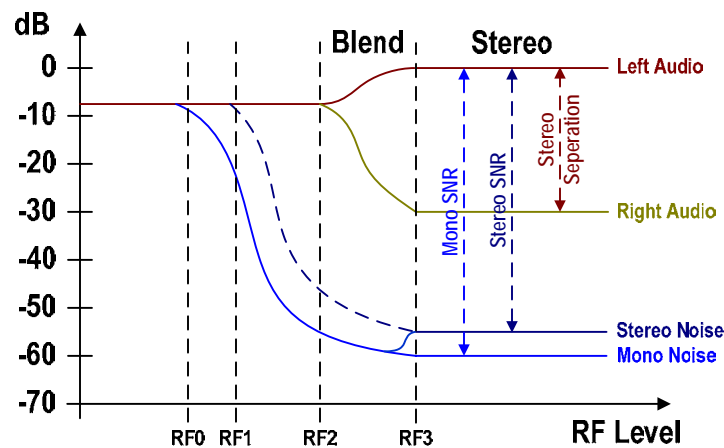

Figure 9. FM Characteristic Curve

Figure 9 shows a plot of a generic FM characteristic curve. The X axis represents the RF signal level and the Y axis represents the left audio output normalized to its maximum output level, i.e., 0 dB represents the maximum output level for the left audio output signal. The left audio, right audio, stereo noise and mono noise levels are plotted on this graph; all signals are plotted relative to the left audio output. In this example, an RF input level of RF3 and higher brings the FM tuner in full stereo mode, resulting in a stereo separation of 30 dB and a stereo SNR of 55 dB. If the FM tuner is forced in mono mode in this region, the mono SNR would be 60 dB. The larger mono SNR comes from a smaller monaural bandwidth of 15 kHz as compared to a stereo MPX signal requiring a bandwidth of 53 kHz.

In the region between RF2 to RF3, stereo-mono blending occurs as shown by the merging of the left and right audio signals. As the left and right audio signals merge, the stereo noise also merges to the mono noise and thus effectively increases the SNR of the audio signal. If blending was not implemented, the stereo noise would track the dashed dark-blue line, and the audio SNR and RF sensitivity level would be lower than an FM tuner with stereo-mono blending. In this example, RF0 could represent the sensitivity level of an FM tuner with stereo-mono blending, and RF1 could represent the sensitivity level of an FM tuner without stereo-mono blending.

Sensitivity is defined as the minimum RF input level to achieve a certain amount of audio SNR. In this fictitious example, sensitivity is defined as the RF level necessary to achieve an audio SNR of 1 dB.

In addition, when the RF input level to the FM tuner decreases, the noise level increases at a much faster rate than the rate of decrease in the audio output level. In this example, the audio output only drops by about 6 dB from its maximum output level, but the noise level can increase all the way up to the audio output level when the RF level drops below sensitivity (RF0). When this occurs, the noise and audio signals are at the same level, and this level can be relatively loud since it is only 6 dB below the maximum audio output level. One approach to minimizing the audible noise level in this weak RF region is to attenuate both the audio and the noise signals together in a technique known as "soft mute." Figure 10 shows an FM characteristic curve with soft mute. In this example, when soft mute is enabled, the audio noise and signals are attenuated by 14 dB to a level that is 20 dB below the maximum audio output level to minimize audible noise and improve the overall user experience.
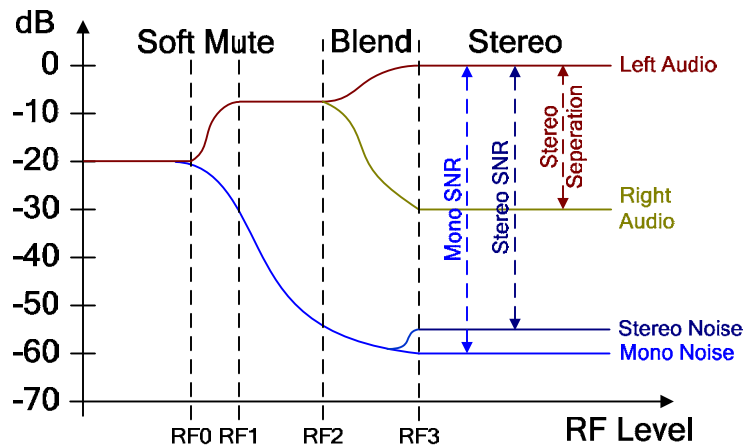
Figure 10. FM Characteristic Curve with Soft Mute

Si4700/01 FM Tuners
The Si4700 and Si4701 FM tuners are the industry's first radio tuner ICs to leverage a digital low-IF architecture and a 100 percent CMOS process technology, resulting in a completely integrated solution that requires only one external supply bypass capacitor and less than 20 mm^2 of board space. Figure 11 shows a block diagram of the Si4700 and Si4701 FM tuners. The digital low-IF architecture allows for the elimination of external components and factory adjustments due to analog process variations. This mixed-signal architecture allows digital signal processing (DSP) to perform channel selection, FM demodulation and stereo audio processing to achieve superior performance compared to traditional analog architectures.
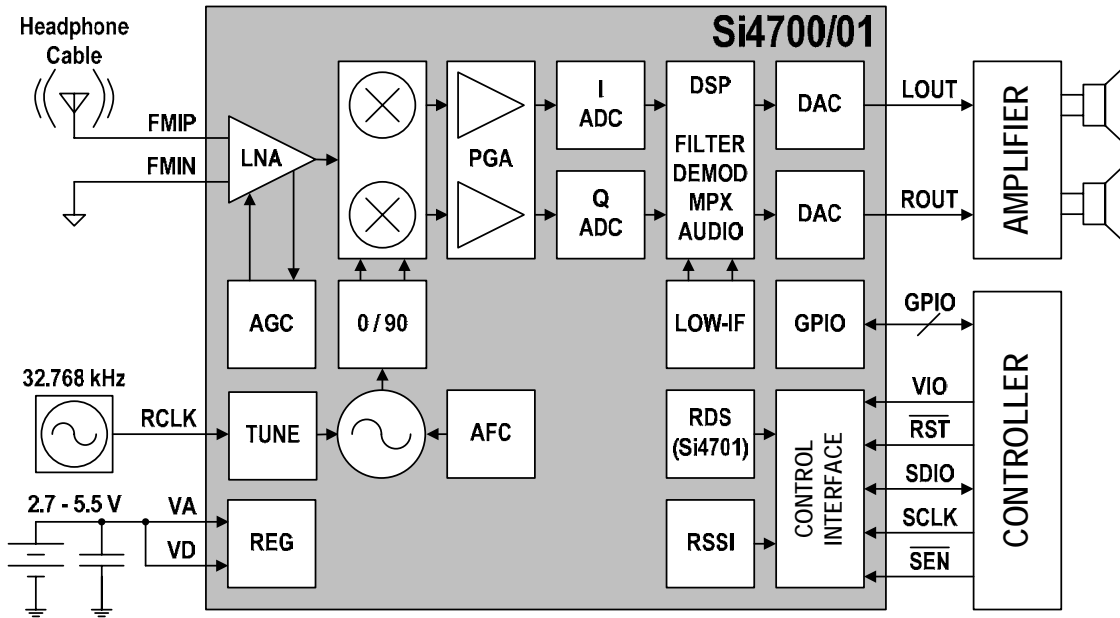
Figure 11. Block Diagram of the Si4700/01 Digital-Low IF FM Tuner

The Si4700 and Si4701 FM tuners include programmable, stereo-mono noise thresholds and soft mute parameters to allow maximum flexibility for noise improvements. DSP is utilized to provide optimum sound quality for varying signal reception conditions. This rich feature set and high levels of integration and performance are directly attributable to the digital low-IF radio architecture and the digital implementation of the FM demodulation, MPX decoding and noise improvement functions. Besides simplifying and reducing design-in time, the high integration of the digital low-IF architecture increases quality and improves manufacturability by having only one external bypass capacitor.

FM radio is one of the most prevalent forms of media communication in the world today. As listeners all over the world continue to buy and use FM radios, designers for portable devices such as MP3 players and mobile phones are increasingly including FM radio capability in their products. Understanding the basics of FM radio will assist designers in creating high performance products, be they traditional standalone radios or next-generation, multi-use devices.

References
[1] E. H. Armstrong Web Site, http://users.erols.com/oldradio/
[2] S. Haykin, Communication Systems, 3rd Edition, Wiley, 1994
[3] R. E. Ziemer, W. H. Tranter, Principles of Communications, Systems, Modulation, and Noise, Fourth Edition, Wiley, 1995
[4] B. Razavi, RF Microelectronics, Prentice Hall, 1998
[5] J. Kean, "FM Stereo and SCA Systems", National Association of Broadcasters Engineering Handbook, 9th Edition, NAB, 1999, pgs 591 – 608.
[6] S. A. Wright, "Radio Broadcast Data System (RBDS)", National Association of Broadcasters Engineering Handbook, 9th Edition, NAB, 1999, pgs. 633 – 642.