ABSTRACT
This paper presents recent trends in the area of integrated CMOS transceiver design for short distance wireless applications. This application is characterized by very low-cost and low-power solutions. Current challenges and recent trends are described and digital-oriented design opportunities for increasing integration outlined. Signal processing approaches applied to the front-end electronics find an increasing emphasis and are extremely viable.

1. INTRODUCTION
The radio frequency (RF) front-end circuits have been traditionally implemented by Gallium Arsenide (GaAs) MESFET, Si-bipolar junction transistors (BJTs), III-V heterojunction bipolar transistors (HBTs) and Silicon-Germanium (SiGe) BJTs, while the baseband digital signal processing (DSP) and analog circuits are being implemented exclusively using CMOS technologies. As the emphasis of wireless applications moves towards the personal communication systems (PCS) and wireless local area networks (WLAN) as well as the wireless entertainment electronics, light-weight, small-dimension, low-cost, low-power and a higher level of integration are becoming ever critical. These have spurred the interests in low-cost CMOS technologies. RF wireless systems using CMOS technologies are being intensively investigated mainly due to their low cost, high yield and higher level of integration including baseband circuits.

1.1 CMOS Process Technology
On the device physics side, there is a fundamental limitation of the bipolar technology that prevents from using it in low-voltage applications in favor of CMOS. Operation at 1 V power supply is difficult to achieve since the $V_{BE}$ of bipolar transistors lies around 0.7 V and the base-to-collector junction must be reverse biased. By contrast, the MOS transistor can admit a $V_{DS}$ lower than the $V_{GS}$. Therefore, if the saturation voltage is kept at a few hundred mV, there is enough room for some output dynamic range.

For MOSFET transistor application to RF circuits, there are several figure-of-merit (FOM) parameters: cutoff frequency $f_T$, maximum oscillation frequency $f_{max}$, minimum noise figure $F_{min}$, flicker noise $1/f$, power-added efficiency PAE and power gain $G_{A}$. A conventional deep-submicron CMOS technology designed for logic applications has shown to exhibit respectful RF device characteristics [1] [2]. Scaling of CMOS devices increases the $1/f$ flicker noise which is caused by carrier trapping near the thin oxide-silicon interface. On the other hand, there is no such junction in bipolar transistors. The flicker noise could be converted into close-in phase noise of voltage-controlled oscillators (VCO) and appear at the output of down-conversion mixers. Fortunately, as shown in [3], this noise frequency translation mechanism could be avoided through half-circuit topological symmetry.

Both $f_T$ and $f_{min}$ peak values already exceed 100 GHz for sub-0.1 μm deep-submicron CMOS logic devices and are predicted to double every three years [3]. Despite the continual RF performance lag behind the latest SiGe bipolar processes, at this point, their performance is adequate for wireless communication bands for up to 5 GHz. This has opened the possibilities for CMOS RF circuits meeting the stringent requirements in communication systems. Efforts have been made to show the feasibility of CMOS front-end circuits, and the performance has become comparable to those of BJT circuits. The ultimate goal is to take advantage of the CMOS process to implement a single chip radio including RF front-end, intermediate frequency (IF) modulation/demodulation circuits and baseband signal processing circuits.

1.2 CMOS Integration
There is a wide array of opportunities that integration presents. The most straightforward way would be to merge various digital sections into a single silicon die, such as DRAM or flash memory embedded into DSP or controller. More difficult would be integrating the analog baseband with the digital baseband. Care must be taken here to avoid coupling of digital noise into the high-precision analog section. In addition, the low amount of voltage headroom challenges one to find new circuit and architecture solutions.

Sensible integration of diverse sections results in a number of advantages: (1) Lower total silicon area. In a deep-submicron CMOS design, the silicon area is often bond-pad limited. Consequently, it is beneficial to merge various functions on a single silicon die to maximize the core to bond-pad ratio. (2) Lower component count and thus lower packaging cost. (3) Power reduction. There is no need to drive large external inter-chip connections. (4) Lower printed-circuit board (PCB) area, thus saving the precious "real estate."
The ultimate goal in mobile wireless integration might be a single-chip digital radio as shown in Fig. 1. The baseband controller, such as based on the ARM7 microprocessor, implements the Bluetooth protocol layer stack which is controlled by a software program stored in a non-volatile flash memory. The RF transceiver module implements the physical layer. The advanced deep-submicron CMOS process total integration leads to an extremely compact and economic implementation of this sophisticated and highly functional communication system.

1.3 Deep-submicron Integration Challenges

Deep-submicron CMOS processes present new integration opportunities on one hand, but make it extremely difficult to implement traditional analog circuits, on the other. For example, frequency control input of a low-voltage deep-submicron CMOS oscillator is an extremely challenging task due to its highly nonlinear frequency-vs.-voltage characteristics and low voltage headroom making it susceptible to the power supply and substrate noise. In such low supply voltage case, the dynamic range of the signal and thus the signal-to-noise ratio will degrade significantly. One has to look for alternative solutions, such as utilizing a voltage doubler in [4]. Furthermore, the advanced CMOS processes typically use low resistance P-substrate which is an effective means in combating latchup problems, but exacerbates substrate noise coupling into the analog circuits. This problem only gets worse with scaling down of the supply voltage. Fortunately, there is a serious effort today among major IC fabrication houses to develop CMOS processes with higher resistivity silicon substrate.

2. DIRECT MODULATION TRANSMITTER

Fig. 2 shows a general block diagram of a transmit pulse amplitude modulation (PAM) process using complex signals. It mathematically describes an arbitrary modulation process. Complex signal representation requires two physical wires that carry both real-valued parts of a complex number. Fig. 3 shows a block diagram of a PAM transmit modulation using in-phase (I) and quadrature (Q) signals that represents a natural progression towards a more physically-realizable representation.

Fig. 4 shows a block diagram of a PAM transmit modulation using a polar alternative in the form of direct amplitude and phase modulation. The direct phase modulation might be performed by modulating the oscillator tuning input in a feed-forward manner with a possible PLL loop compensation method as shown by M. Bopp [5] and B. Zhang [6]. The direct amplitude modulation might be performed by regulating the supply voltage to a constant-envelope power amplifier. This method appears to be the best choice for digital integration of mobile RF transceivers because it does not use the RF/analog-intensive up-conversion modulator. Due to this reason, the emphasis in this paper is placed on frequency synthesizers with phase/frequency modulation capability.

3. MODULATING SYNTHESIZER

RF synthesizers are used as a local oscillator (LO) in both receivers and transmitters, and remain one of the most challenging tasks in mobile RF systems because they must meet very stringent requirements of a low-cost, low-power and low-voltage monolithic implementation while meeting the phase noise and switching transient specifications. There are three major conventional frequency synthesis techniques: direct-analog mix/filter/divide, direct-digital (DDS), indirect or phase-locked loop (PLL).

3.1 PLL-based Synthesizer

A great majority of RF wireless synthesizers for mobile applications are based on the PLL structure. Under locked condition, the average output frequency of a PLL bears an exact relationship with the reference input frequency. Conventional PLL has an integer divider ratio \( N \), such that \( f_{out} = N \cdot f_{ref} \). Resolution is equal to the reference frequency which is usually selected to be the same as the channel spacing. Narrow loop bandwidths are undesirable because of long switching times, inadequate suppression of the VCO phase noise, and susceptibility to the supply and substrate noise.

Unfortunately, the PLL structure does not easily lend itself to silicon integration. Because of the spur reduction requirements, the loop filter, usually realized as a charge pump, has large external capacitors in order to achieve a low PLL bandwidth of several kHz. Realizing a monolithic capacitance on the order of a few hundred pF would require a prohibitively large area if implemented as a high-quality metal-insulator-metal (MIM) capacitor. Implementing it as a MOS capacitor would take less area, but it would likely be unacceptable because of its high leakage current and nonlinearities.
Figure 5: Periodic and deterministic phase error

3.1.1 Fractional-N Architecture

In fractional-N synthesizers, the output frequency can increment by fractions of the reference frequency, allowing the latter to be much greater than the required channel spacing. This allows wide loop filter design at the expense of fractional spurs, resulting in improved loop dynamics and attenuation of the oscillator-induced noise. The PLL bandwidth is usually set at roughly 10% of the reference frequency to avoid any significant feed-through of the reference tone and may now span several channels. In response to a change in a frequency control word, the PLL output frequency settles to the programmed value with a time-constant inversely related to the loop bandwidth.

Figure 6: Fractional-N synthesizer using a $\Sigma\Delta$ modulated clock divider

Fractional PLL can achieve arbitrarily fine time-averaged frequency division ratio of $\frac{N}{N+N+1}$ (where $N$ separates the integer and fractional parts) by modulation of the instantaneous integer division ratio of $N$ and $N+1$ (in practice, multi-bit modulus could be used as demonstrated by T. Kenny [7]). Phase detector will operate at a frequency of $f_{\text{ref}} \cdot f_{\text{off}} + \frac{N}{N+N+1} \cdot f_{\text{off}}$, the phase error of the phase detector causes VCO fractional spurs at multiple of the offset frequency $(\cdot f_{\text{off}}) \cdot f_{\text{off}}$. There are several methods to suppress the fractional spurs. A more conventional method is the analog fractional-N compensation scheme that uses an accumulator and a DAC and is based on the observation that the phase error perturbation is periodic and deterministic and could be theoretically canceled out in hardware (see Fig. 5). The second method uses a $\Sigma\Delta$ modulated clock divider [8][9] and is shown in Fig. 6. This solution is more digital in nature since it does not rely on precise analog component matching of the previous technique. It trades the reduction in fractional spurs for the increase in the noise floor.

The fractional-N frequency synthesizer architecture lends itself well to an indirect narrowband frequency modulation which could be implemented almost entirely in a digital manner. In such architectures the desired channel is selected using the digital control word. As long as the modulation data rate is lower than the PLL loop bandwidth, the average division ratio established by the digital control word could be augmented by the instantaneous value of the modulation frequency deviation. There has been some research done to increase the data rate by compensating for the PLL loop high frequency attenuation by boosting the high frequency components of the modulation signal as shown by M. Perrot [10]. After the equalized modulation signal passed through the PLL, the modulation spectrum would be restored to its original form. The digital equalizer could be embedded in the GFSK filter with little extra overhead. However, the precise loop compensation requirement makes this architecture not very practical for manufacturing.

Figure 7: Modulating wideband fractional-N synthesizer

The above problem of mismatch between the digital compensation filter and the analog PLL got addressed by W. Bax [11]. The fractional-N PLL is re-architected there to place a $\Sigma\Delta$ frequency discriminator, whose transfer function is set digitally and well controlled, in the feedback path (see Fig. 7). As a result, the only analog component left that required a substantial amount of matching is the VCO.

Fig. 8 shows the noise shaping properties of the first, second and third-order $\Sigma\Delta$ modulated clock divider. The first order of $\Sigma\Delta$ operation turns out to be equivalent to the conventional, but uncompensated, fractional-N architecture and exhibits systematic division ratio patterns that produce unacceptably large frequency tones. The left plot shows the power spectral density (PSD) of the divided clock. As shown, the third order of $\Sigma\Delta$ dithering introduces enough randomness to completely eliminate any frequency spurs that are clearly shown with the second and first order $\Sigma\Delta$ dithering. The right plot shows the PSD of the divided clock phase and it demonstrates that the second order $\Sigma\Delta$ dithering performs the high-frequency shaping of the division ratio quantization noise of 20 dB per decade, whereas the third order produces 40 dB per decade. The noise at higher frequencies then gets filtered out in the loop filter. While the phase spectrum is important to determine the noise floor, it is inconvenient to deal with frequency spurs.

3.1.2 Phase-domain PLL

A. Kajiwara addresses the chief limitation of the traditional PLL-based frequency synthesizers (see Fig. 9), namely the slow frequency switching times, which make them less desirable for the advanced portable wireless applications that use spread spectrum and frequency hopping techniques [12]. On the other hand, the direct digital synthesizers, whose switching time is extremely fast, cannot be used in a direct manner at wireless frequencies. The authors propose a DSP-based phase-domain PLL structure that features fast switching times. This architecture, however, has a major limitation of handling only integer-N division ratios.
1.3
1.4
1.5

Spectrum of Output Clock [dB] clk1(y) clk2(r) clk3(g)
Frequency: Fsamp=1GHz, (FFT: iter=50, len=409600, nfft=1048576, dlen=655613)

3rd order
2nd order
1st order

Figure 8: $\Sigma\Delta$ divided clock: clock output spectrum (left), phase spectrum (right)

Sigma-Delta Modulator; $T=80.03076904$
Sigma-Delta Modulator Divider; $T=80.03077030$

3.2 Hybrid Approach

In certain applications it is necessary to combine the two (rarely three) major synthesis techniques such that the best features from each basic method are emphasized. Most likely it is the hybrid of DDS and PLL structures that is found in certain wireless applications. Here, the wideband modulation and fast channel-hopping capability of the DDS method, that now operates at lower frequency, is combined with a frequency multiplication property of a PLL loop that up-converts it to the RF band. A. Hafez [13] describes a 900 MHz band hybrid synthesizer structure that uses a 1.10–1.85 MHz low-frequency DDS to generate a stable frequency reference to the main PLL loop. The PLL loop, instead of conventional digital frequency divider or prescaler, uses subsampled mixing to translate the RF frequency down to the $f_{ef}$. The frequency resolution of the synthesizer is established by the DDS and the PLL loop is mainly used as a frequency multiplier. Since the DDS operates at low frequency, its major limitation of high power is not a concern. Unfortunately, the subsampling process introduces an excessive noise.

4. RECEIVER

Recently, there has been a wide focus on “software radio”: a term used to describe a programmable integrated solution addressing multiple standards and modes [14, 15, 16]. Short distance wireless transceivers typically enforce stringent constraints on power dissipation and cost. Low cost solution requires the ability to integrate the transceiver with the baseband processor and the most promising technology is a standard low-cost CMOS process. Another important consideration in cost reduction is that the proposed solution must avoid off-chip components. Unfortunately, conventional design approaches requires the ability to put high-Q inductors and high-Q image-reject and channel filters in addition to low-noise integrated VCOs [14, 15]. The attempts to improve the Q, generally, result in increasing the cost of the process, thereby neutralizing the low-cost advantage of standard CMOS process. High quality RF designs are possible with alternate technologies, however, CMOS offers the advantage of cost reduction through integration and high gate density in digital circuitry. This motivates new approaches in the transceiver design such that the playing field is leveled by employing more advanced signal-processing techniques and using digital methods in low-cost CMOS processes that offer very high gate densities. An example of such a technique was presented in [17, 18] for constant envelope modulation schemes.

4.1 Two-step Conversion Receiver

Super-heterodyne type architecture which uses two-step down-conversion has been the most widely used architecture to date and offers many advantages such as reduction of unwanted energy outside the desired channel using IF filters and VGA [14]. An integrated receiver designed using this approach becomes more practical if a wide-band at RF is frequency translated to IF and followed by an IF low-pass filter before the second down-conversion stage. This is shown in Fig. 10. Channel selection is performed using a second tunable LO. This approach offers the advantages of integrating frequency synthesizers with low phase noise [14] as well as improving desensitivity to fixed and time-varying local dc-offsets due to self-mixing and other non-linearities. However, such method-
ogy requires image-reject mixer which requires dealing with gain and phase mismatches in I/Q branches. High linearity and low-noise are desirable in such architectures in addition to variable gain in the first stage. 1/f noise may be reduced by using long-channel devices. However, a larger dynamic range is required in the A/D, which implies larger power dissipation. In a two-step down conversion procedure, one could use a complex subsampling mixer as the second stage of frequency down-conversion and use digital techniques to remove I/Q mismatches. Adaptive mismatch techniques are also possible for I/Q mismatch in a direct IF sampling approach [19]. The linearity of a mixer can be improved by putting it inside a sigma-delta modulator based A/D converter loop [20] and putting a reconstruction filter in the feedback loop. A higher-power yet simple approach is to directly sample the IF frequency using an A/D and is feasible only at base stations.

4.2 Single-step Conversion Receiver

Direct conversion and very-low-IF conversion are two approaches that have recently attracted a lot of attention. In direct conversion, IF equals zero and the received signal is filtered and amplified at the baseband which simplifies the design for low-power. This is also true for very-low-IF conversion which places IF very close to zero (half to two channel spacings above zero). Direct conversion requires quadrature down-conversion with vector detection otherwise the negative frequency half-channel folds over to the positive half-channel. The channel selection is performed using a lowpass filter and no image-reject filter is required.

However, direct-conversion is plagued by spurious LO leakage in which the strong LO frequency signal in the receive path can leak through the antenna due to coupling and bounce back from nearby objects finally entering the receiver as an interferer [21, 22]. Non-linearities in the LNA and mixer also cause a dc-offset which appears with the down-converted baseband. There are two detrimental effects of the dc-offset. First, it increases the requirements on the dynamic range of subsequent components and second, it imposes severe constraints on the reverse isolation of mixer and LNA. Although capacitive coupling can cure this problem if the signal spectrum has a null at dc [18, 21, 22] or if the BER does not degrade significantly due to the dc-null placed by capacitive coupling, the more recent trend has been to estimate and remove the offset digitally [21, 23]. The detrimental effects of I/Q mismatch in quadrature down-conversion is not as severe as in image-reject architectures [22], however, even-order distortion adds to the picture where two strong interferers generate a low-frequency beat equal to the difference frequency and appear at the mixer output due to direct feed-through [22]. Converting LNA and mixers to differential topology increases the noise figure [22]. The 1/f noise of devices has a severe impact on the down-converted signal around dc and can be dealt with by increasing the size of the devices and/or by periodic offset cancellation [22].

4.2.1 Subsampling Receiver

Another approach to direct-conversion receiver is sub-sampling [24, 25]. The basic approach is to sample the RF signal at an integer fraction of carrier frequency which is greater than twice the bandwidth of the modulating signal. Spectral images of the modulating signals are repeated and down-conversion can be achieved by a low-pass filter.

A general scheme based on sub-sampling operation is presented in [24] and shown in Fig. 11 in which the RF input is sampled and-held and followed directly by discrete time analog signal processing. The baseband signal is converted using an A/D converter. Sampling the carrier frequency \( f_c \) at a rate \( f_s \) results in spectral images located at \( n f_s \pm f_c \) where \( n \) is an integer. A desired spectral image can be filtered using a discrete time analog filter. In this approach, the lowest power solution would involve a tradeoff analysis between the input rate of the A/D converter and the complexity of anti-aliasing downsampling filters in multi-stage configuration. The use of multi-stage filtering allows reuse of anti-aliasing filters [24] by appropriate selection of \( f_s \) in relation with \( f_c \). In this approach, channel select filtering, demodulation and baseband processing are done in the digital domain following the A/D converter. The final stages of the multi-stage analog filters can also be used to reduce the adjacent channel interferers, thereby reducing the dynamic range requirement and power dissipation of the A/D converter. [25] shows a subsampling mixer implemented for 1.8 GHz RF system. The mixer is implemented using a track-and-hold circuit. A differential OTA is used to transfer the sampled charge to the output in order to cancel the charge feed-through and to attain high linearity. The speed of OTA determines the maximum \( f_s \).

Another example of subsampling approach has been successfully demonstrated in [17] for binary FSK transceiver at 900 MHz. The approach used substantially reduces power dissipation by hard-limiting the filtered output of a sub-sampling mixer. The high frequency images are rejected by -60 dB using a switched capacitor analog filter. The limiter serves to act as a 1-bit A/D converter which provides an over-sampled down-converted baseband signal. The signal is decoded using a 1-bit FSK demodulator. The AGC functionality is achieved using the filter, limiter and demodulator. Although, this structure simplifies a general architecture requiring a multibit A/D converter and VGAs thereby saving power, its application cannot be directly extended to general modulation schemes.

Subsampling receiver eliminates the need for IF filters, image-reject mixers, image reject filters and analog I/Q branches thereby making these suited to integration. Further, the L0 in such a scheme operates at much lower frequency than \( f_c \). However, these trade one set of problems with another. The major issues in such receivers are noise-folding, sensitivity to clock jitter and linearity requirements. Noise folding can be reduced by using a preselect filter. The sensitivity to timing jitter and the noise figure of the system can be improved by placing N sampling switches in parallel collecting successive samples that are dumped to an output buffer simultaneously every N clock cycles [26]. The relationship of \( f_s \) with \( f_c \) can be relaxed by employing multirate subharmonic sampling which allows a larger choice of \( f_s \) while replacing the I-Q mixing with a geometric transformation.

5. SYSTEM-LEVEL LOW POWER OPERATION

Power reduction in a CMOS receiver requires system level considerations. A major factor that determines the power dissipation in the receiver is the dynamic range of the input signal. In an interference dominated environment, only one or two strong interferers are present and have the potential to overload the front-end. Hence, the receiver is designed for the worst-case scenario and the desired BER and achievable noise floor determine the sensitivity and the
required dynamic range of the receiver. If undesired spectrum can be discarded using filters, the dynamic range requirements can be relaxed for the A/D converter. In the absence of strong interferers, the entire dynamic range in the receiver is not required and under such circumstances, system level control can switch to a low-power mode which uses smaller dynamic ranges.

6. CONCLUSION
We presented several approaches which are currently used in low-power integrated CMOS transceivers for short distance wireless. Recently, digital and analog signal processing techniques have found increasing application in the generation and detection of signals in such transceivers. Many of such techniques were described and issues related to these approaches outlined. The new generation of low-power integrated RF transceivers will rely on increased role of digital approaches replacing conventional RF sub-systems.

7. REFERENCES