Tunable, Concurrent Multiband, Single Chain Radio Architecture for Low Energy 5G-RANs

T. O’Farrell, R. Singh, Q. Bai, K. L. Ford, R. Langley
Department of Electronic and Electrical Engineering
University of Sheffield
Sheffield, UK
Email: t.ofarrell@sheffield.ac.uk

M. Beach, E. Arabi, C. Gamlath, K. A. Morris
Department of Electrical and Electronic Engineering
University of Bristol
Bristol, UK
Email: m.a.beach@bristol.ac.uk

Abstract—This invited paper considers a key next step in the design of radio architectures aimed at supporting low energy consumption in 5G heterogeneous radio access networks. State-of-the-art mobile radios usually require one RF transceiver per standard, each working separately at any given time. Software defined radios, while spanning a wide range of standards and frequency bands, also work separately at any specific time. In 5G radio access networks, where continuous, multiband connectivity is envisaged, this conventional radio architecture results in high network power consumption. In this paper, we propose the novel concept of a concurrent multiband frequency-agile radio (CM-FARAD) architecture, which simultaneously supports multiple standards and frequency bands using a single, tunable transceiver. We discuss the subsystem radio design approaches for enabling the CM-FARAD architecture, including antennas, power amplifiers, low noise amplifiers and analogue to digital converters. A working prototype of a dual-band CM-FARAD testbed is also presented together with measured salient performance characteristics.

I. INTRODUCTION

The proposed fifth generation (5G) of radio access networks (5G-RAN) is expected to support a combination of a 1000 fold increase in area capacity, multiple radio access technologies (multi-RAT) and concurrent multi-connectivity using an heterogeneous radio network architecture (HetNet). While this will enable the 5G-RAN to support higher throughput connections with increased reliability for different levels of mobility, there is considerable concern that the radio network energy consumption will increase substantially [1] [2]. A major contributing factor being the multiplicity of radio transceivers used both at the base transceiver station (BTS) and the mobile station (MS), which is set to increase in 5G. In addition, the sheer difficulty of including separate transceiver chip sets in the MS for each RAT is a potential bottleneck to realising a true global 5G MS.

The majority of the hand-held MSs use multi-chain transceiver chip sets, where each chain corresponds to a specific communication standard. The BTS use separate radio units for different standards and are adopting software defined radio (SDR) units, which have the ability to work over a wide range of standards and RF bands, but only provide a single communication link at any specific time. There is a need for a major architectural change in the radio sub-systems and system-as-a-whole to equip the MS and BTS for the foreseen increase in RATs and with support for dynamic and concurrent multiband spectrum access to enhance the network capacity with a minimal amount of RF hardware for low power consumption. CM operation through a single transceiver chain is a major advancement for SDR and multi-standard chip sets.

We present the concept of a tunable, concurrent multiband, frequency-agile radio (CM-FARAD) architecture, which has the ability to concurrently work over multiple RF bands/standards using a single tunable radio transceiver. The CM-FARAD architecture is based on software defined radio (SDR) principles, but incorporates software controllable frequency-agile analogue front ends as well as a flexible digital sub-system. In this paper, the CM-FARAD architecture is presented, primarily focused on the MS, and the methods for enabling CM operation at the sub-system level, along with the challenges involved, are reviewed.

While the development of millimetre-wave systems requires more time, as a first step, the CM-FARAD architecture addresses the CM requirement over the sub 6 GHz spectrum, which is assigned to the major wireless communication standards. Additionally, this portion of spectrum is underutilized [3]. The presented radio architecture and sub-system level techniques for CM capability specifically assume system operation over the 0.4 - 6 GHz spectrum. The key performance characteristics of an CM-FARAD system are discussed through a hardware-in-the-loop test-bed using concurrent transmissions over the long term evolution (LTE) and unused digital terrestrial television (DTT) bands, which are also known as TV white spaces (TVWS).

II. MULTIBAND RADIO ARCHITECTURES

The majority of radio systems are based on either heterodyne or homodyne (direct-conversion) architectures. While the former architecture has been widely used in conventional systems, the latter, due to its simplicity and low cost, is often found in cellular radio systems. Through a homodyne architecture, tunable image-rejection filters, which operate at the carrier frequency and must have the ability to tune throughout the entire frequency range, can be avoided. Moreover, the selection of a single intermediate frequency (IF) to enable operation over various sub-GHz standards is very challenging.
Hence, the radio chip sets and SDRs in MSs tend to use a homodyne architecture.

CM reception using a single radio transceiver chain is difficult to implement in a homodyne system as the analogue front-end complexity must increase to accommodate new bands and RATs in a 5G-RAN. At the BTS, various RATs are supported by multiple radio units including SDR, whereas MSs typically use compact multi-standard transceiver chip sets. The architectural details of the majority of these chip sets are not in the public domain, whereas the ones that are, are based on the velcro approach, where each transceiver chain is realised by a homodyne radio. Deploying separate radio units at the BTS and using the velcro approach in the MS is expected to increase the complexity, size and power consumption of equipment in a 5G-RAN.

A. FARAD Architecture

Enabling tunability and CM operation in analogue and digital sub-systems of a radio are central to realizing CM-FARAD transceivers, and the design of compact, power efficient BTSs and MSs. In the analogue domain, various designs for tunable dual-band antennas [4] [5] have been proposed. However, independent tunability of bands is an issue that still needs to be addressed. Dual-band power amplifiers (PAs) [6] and low noise amplifiers (LNAs) have also been presented. Multi-band reconfigurable analogue-to-digital converters (ADCs) have also been designed to support intra-band carrier aggregation [7]. However, the power efficient and small size digitisation solutions desired in CM-FARAD, are currently unavailable.

Despite recent developments in radio sub-systems to support CM operation, the vision for a tunable CM "system-as-a-whole" remains unclear. Research on a particular subsystem typically assumes that the rest of the system already has CM capabilities, which is not the case. Thus system design challenges, including the system level testing and subsystem integration, are usually not addressed collectively. For example, subsystems are conventionally interfaced using fixed 50 Ω matching. In a system-as-a-whole design approach, optimal impedances may be specified for each interface, thereby substantially improving efficiencies.

To achieve tunable, CM operation in wireless communications systems, Mitola’s vision [8] for SDR architectures remains pertinent, whereby analogue-to-digital and digital-to-analogue conversion (ADC/DAC) should be carried out as close as possible to the antenna. The CM-FARAD system will be based on a direct RF digitization architecture, which is expected to realise a complete SDR architecture. Direct RF digitization can lead to reconfigurable yet small, power efficient and cost effective radio front ends. Fig. 1 shows the architecture of the CM-FARAD system, which comprises a multi-band tunable analogue front-end, multi-band reconfigurable ADC & DAC, digital down-converter (DDC), digital up-converter (DUC), and a baseband processing unit.

Fig. 1. Architecture of the proposed direct conversion (homodyne) System for frequency-agile applications.

III. ANTENNAS

Tunable multi-band antennas have attracted attention recently due to the increasing demand for cognitive radio (CR) and carrier aggregation (CA) in wireless communication systems. To support tunable, CM operation in a single radio transceiver, the antenna needs to support multiple frequency bands, which are independently tunable. Furthermore, in the MS the antenna needs to be compact in volume.

A frequently used solution involves either a wideband or multi-band antenna, that covers the frequency bands of interest, and tunable filters for band selectivity. The challenges in this approach are the difficulty of maintaining good antenna performance over a very wide frequency range and the complex geometry of multi-band antennas. Also, a very high capability of out-of-band noise rejection is required for the tunable filters [9]. An alternative approach is to integrate the tuning components inside the antenna to form a tunable multi-band antenna. The challenge of this solution is the difficulty of increasing the number of concurrent operating frequency bands while still maintaining the independent tunability of each band.

Various dual-band antenna designs have been reported in the literature. A planar inverted F-antenna (PIFA) was presented in [4], which demonstrated two tunable bands over the frequency range 88–2175 MHz. In [5], a dual-band folded slot antenna was designed with a wide tuning range from 1–3 GHz. However, independent tunable bands have not been sufficiently addressed in multi-band antennas using a single feed, where tuning one band may shift the resonant frequency of the other.

Slot antennas have been used in many tunable antenna designs as they have a planar structure, with an adaptable input impedance, that can be readily embedded in the ground plane of a printed circuit board (PCB), while accommodating nearby electronic components. Fig. 2(a-b) shows a novel design of an independently tunable dual-band slot antenna with a single feed. The antenna was manufactured on an FR4 PCB, where two open-ended slots are placed perpendicular to each other to reduce their mutual coupling. Two slots are adjusted to be resonant at the lower and higher frequencies, respectively, and are fed by a 50 Ω stripline printed on the back of the PCB, as shown by the reverse view in Fig. 2(a), to provide concurrent operation at two frequency bands [10]. A varactor diode per slot independently tune the two resonant frequencies.

1A dedicated hard-wired module per standard.
The measured antenna reflection coefficients under various biasing voltages are plotted in Fig. 2(c and d). The -6 dB tuning range of the lower and higher bands cover 560 – 800 MHz and 800 – 1140 MHz, respectively. The results show that each band can be tuned individually without changing the operating frequency of the other band. Also, the slot bandwidths are narrow in both bands, which provides additional front end filtering. Direct matching to an PA or LNA is achieved by adjusting the stripline feeding location.

IV. RF POWER AMPLIFIERS FOR CM-FARAD

For a CM-FARAD transmitter, the PA needs to concurrently amplify signals with different carrier frequencies and bandwidths. For 5G-RANs operating below 6 GHz, as many as five concurrent signals with bandwidths from a few MHz to over 100 MHz, distributed in bands across the range 0.4 – 6 GHz, is possible. For these bands to be efficiently utilized, the RF system should have the ability to tune between all these bands with concurrent operation and control of the bandwidth. Such specifications are extremely difficult to accommodate using current PA technologies. Two main approaches are being considered by researchers: ultra-wideband and tunable/concurrent designs. In this paper, a review of the design process of the latter approach is presented.

The field effect transistors (FET) used in the design of PAs can be modelled as a current source with internal resistance of \( R_{\text{opt}} \) in parallel with a capacitor (\( C_{ds} \)) as shown in Fig. 3. The output matching network (OMN) should transform \( Z_{\text{Antenna}} \) (typically 50 \( \Omega \)) to a complex impedance with a real part equal to \( R_{\text{opt}} \) and an imaginary part that cancels the transistor’s output capacitance. Since \( R_{\text{opt}} \) and \( C_{ds} \) can be assumed frequency-independent, the PA design reduces to the realization of an OMN that provides the complex impedance mentioned above at all frequencies of interest.

For the amplifier to operate at the intended class, the OMN is also required to present particular impedances at the second and third harmonics of the frequencies of interest. For instance, if an amplifier is to be designed for concurrent dual-band operation taking into account both the second and third harmonics, the OMN has to provide specific impedances at six different frequencies, which is challenging. For CM operation, the OMN has to provide preferred impedances not only at the fundamental and harmonics but also at the inter-modulation products, further increasing the complexity of the design [6]. The harmonics termination requirements have less effect on the performance of the amplifier and may be ignored in many cases.

The first step in the design of tunable CM PAs is to load pull the PA (in simulations or measurement) to determine the optimal terminations. The load pull test has to be performed at all the frequencies of interest and results in a contour on the Smith chart for the complex output impedance vs. frequency.
The second step in the design is to select the topology of the OMN, which should have good potential for multi-band matching, be easy to simulate, and be able to deliver the maximum output power without breaking down or becoming forward biased. The third and final step is to optimize the tunable OMN selected previously. This optimization process can be done by either physical intuition, numerical optimization [11], or direct synthesis. The performance of a CM-FARAD PA depends primarily on the tuning range and losses of the capacitors used in the OMNs.

V. BLOCKER RESILIENT LOW NOISE AMPLIFIER FOR TUNABLE MULTIBAND RECEIVERS

In 5G, the requirement to operate concurrently with up to five multiple bands across the 0.4 – 6 GHz frequency range poses a significant challenge for the LNA in terms of interference rejection and operating bandwidth. The current solution of using multiple narrowband LNAs isn’t appealing when parameters such as size, cost and power consumption are taken into account. FARAD aims to address the practical issues associated with carrier aggregation and multi-band operation using experimental circuit fabrication methods.

Since the LNA must be designed to operate in a hostile RF environment where the power in the wanted band is normally much lower than the interference (or blocker) signals, both in-band and out-of-band blocker rejection is important. The major challenges for CM operation are that of achieving high linearity (characterized by IP3) and tunable blocker rejection that can operate across a wide frequency range with minimal impact on the LNA’s noise figure (NF). Other key constraints such as size of circuit, power consumption and cost of manufacture remain as important factors for the design. LNA architectures that do away with selective blocker rejection are totally dependant on the maximum IP3 for mitigating the influence of large out-of-band blockers. The high IP3 normally comes at the expense of a higher power consumption. For example recent advances in GaN LNAs with IP3 in excess of +25dBm operating from 2GHz to 8GHz have been reported [12]. However, the power consumption is of the order of watts, which is beyond the specifications of most commercial MSs.

The implementation of blocker rejection relaxes the LNA’s IP3 requirement thus allowing power consumption and NF to be reduced. Several blocker rejection methods have been investigated. The direct filtering methods that use SAW filters were successful in early generations of wireless standards. However, the frequency limitations above 3GHz coupled with the lack of tunability has led to the push for SAW-less receivers in recent implementations. Current methods of bandpass filtering make use of frequency dependant loads at the output of the first gain stage of the LNA [13]. More recent implementations have investigated additional blocker cancellation techniques at the baseband level with the use of frequency translational filtering. An advantage of this approach comes from the blocker signal being out of phase relative to the main signal, which allows further cancellation of the blocker at the baseband [13]. An implementation of this filtering technique in a feed-forward architecture has also been demonstrated [14]. Further implementations of this technique in 40nm CMOS over the 0.4 – 6 GHz range have also been demonstrated [13].

The primary objective in extending these techniques to concurrent, multi-band implementations is to duplicate only parts of the circuit requiring the smallest areas and the lowest powers. FARAD investigates new types of blocker filtering architectures for tunable, concurrent multi-band solutions based on CMOS and BiCMOS fabrication technologies.

VI. ANALOGUE TO DIGITAL CONVERTER

The performance requirements of an ADC depend on the highest frequency input signal \(F_{\text{max}}\), the maximum information bandwidth (BW), the required sampling rate \(F_s\) and the digital resolution \(n\) bits. CM-FARAD aims to digitize multiple RF signals concurrently, within the 0.4 – 6 GHz spectrum, each with a certain BW. Hence, Nyquist sampling, which requires \(F_s > 2F_{\text{max}}\) would result in an \(F_s > 12\) GSPS, which would digitize the entire spectrum of interest. However, the power consumption of such a fast ADC would be prohibitive in the MS.

In contrast, for under-sampling or band-pass sampling (BPS) the \(F_s/2\) need only to be greater than the sum BW of the multiple bands in use, which is significantly less than \(F_{\text{max}}\). In BPS, the information at a certain carrier \(F_s\) is intentionally aliased to a desired digital intermediate frequency (IF). This relaxes the \(F_s\) requirements allowing a small form-factor, power efficient ADC to be realized. The salient challenges of using BPS are two-fold: 1) calculating the appropriate \(F_s\) that avoids spectral overlapping of the multi-band signals after sampling; and 2) avoiding excessive SNR degradation caused by aliasing or folding of noise from the input bands into the sampling bandwidth \(F_s/2\) [15].

Lin et al [16] have proposed an efficient algorithm to compute the minimum sampling frequency \(F_{s,\text{min}}\) required to down-convert multiple RF signals to multiple digital IFs. The algorithm avoids spectral overlap between the concurrent bands, which fall between DC and \(F_s/2\) after under-sampling. For example, consider a CM-FARAD system operating with three concurrent transmissions selected from four possible bands: TVWS (460 – 790 MHz), LTE(1) (0.7 – 0.96 GHz), LTE(2) (1.427 – 2.69 GHz) and WiFi (2.45 GHz).

From Lin’s algorithm, the largest \(F_{s,\text{min}}\) required for the four possible tri-band combinations are shown in Table I together with the total number of concurrent transmission combinations and the maximum aggregate bandwidth. The results show that the \(F_s\) requirements do not exceed 456.25 MHz for all possible combinations of concurrent transmissions, while the maximum \(F_{\text{max}}\) that can appear at the system input is 2.69 GHz in this case. This shows that BPS can provide a low \(F_s\) for real-time data processing in the digital domain.

VII. DUAL-BAND CM-FARAD TEST-BED

In this section, a dual-band CM-FARAD test-bed system is presented, and its capabilities in terms of tunability and inter-band interference (IBI) are measured [17] [18].
with root raised cosine filtering were analysed in a CM mode. In order to investigate the frequency agility and CM
A. Measured Results

The test-bed has the capability to concurrently operate over two separate bands within the 560 - 1100 MHz frequency
range. In order to investigate the frequency agility and CM transmission capability, various TVWS and sub-1 GHz LTE
bands changed between 2 to 137 MHz.

This invited paper has presented a highly novel architectural approach for the mobile and wireless radio systems to achieve
tunable and concurrent multi-band operation through a single transceiver chain. The benefits of the CM-FARAD architecture
are two fold. Firstly, it will provide a radio access technology for legacy as well as future wireless communications standards
through a single transceiver chain. Such a novel architecture has substantial potential to reduce the power consumption,
complexity and cost of future RANs. Such an architecture could lead to a 3-to-4 fold reduction in hardware power con-
sumption by replacing multiple radios with a single FARAD radio. Secondly, the tunable, concurrent multi-band operation
will help solve the foreseen capacity demands by efficiently utilizing the available spectrum at certain locations and times
to boost data transmission rates. The pathways to achieve tunable, concurrent multi-band operation at sub-system level
are discussed along with the challenges involved, which, the authors believe, will open new research fields in both radio
transceiver and network design. The CM-FARAD architecture is particularly suited to 5G-RANS, which will support multi-
connectivity in a HetNet scenario.

VIII. CONCLUSIONS

This work was carried out in the FARAD project (https://www.commnet.ac.uk/groups/frequency-agile-radio-farad-project/),
funded by the UK government under the EPSRC grants EP/M013723/1 and EP/M01360X/1. The work is also supported by the UK mobile virtual centre of excellence (mVCE).

REFERENCES

Fig. 5. EVM_{rms} (%) evaluation of 16-QAM and 64-QAM single carrier transmissions over the dual-band FARAD system based on CM combinations in (a), (b) and (c) and in independent transmission mode in (d).