Analog Approximate-FFT 8/16-Beam Algorithms, Architectures and CMOS Circuits for 5G Beamforming MIMO Transceivers

Viduneth Ariyarathna[®], Student Member, IEEE, Arjuna Madanayake[®], Member, IEEE, Xinyao Tang, Student Member, IEEE, Diego Coelho, Student Member, IEEE,

Renato J. Cintra[®], *Senior Member, IEEE*, Leonid Belostotski[®], *Senior Member, IEEE*, Soumyajit Mandal[®], *Senior Member, IEEE*, and Theodore S. Rappaport[®], *Fellow, IEEE*

Abstract—Emerging millimeter-wave (mmW) wireless systems require beamforming and multiple-input multipleoutput (MIMO) approaches in order to mitigate path loss, obstructions, and attenuation of the communication channel. Sharp mmW beams are essential for this purpose and must support baseband bandwidths of at least 1 GHz to facilitate higher system capacity. This paper explores a baseband multibeamforming method based on the spatial Fourier transform. Approximate computing techniques are used to propose a lowcomplexity fast algorithm with sparse factorizations that neatly map to integer W/L ratios in CMOS current mirrors. The resulting approximate fast Fourier transform (FFT) can thus be efficiently realized using CMOS analog integrated circuits to generate multiple, parallel mmW beams in both transmit and receive modes. The paper proposes both 8- and 16-point approximate-FFT algorithms together with circuit theory and design information for 65-nm CMOS implementations. Postlayout simulations of the 8-point circuit in Cadence Spectre provide well-defined mmW beam shapes, a baseband bandwidth of 2.7 GHz, a power consumption of 70 mW, and a dynamic range >42.2 dB. Preliminary experimental results confirm the basic functionality of the 8-beam circuit. Schematic-level analysis of the 16-beam I/Q version show worst-case and average side lobe levels of -10.2 dB and -12.2 dB at 1 GHz bandwidth, and -9.1 dB and -11.3 dB at 1.5 GHz bandwidth. The proposed multibeam architectures have the potential to reduce circuit area and power requirements while meeting the bandwidth requirements of emerging 5G baseband systems.

Manuscript received January 6, 2018; revised March 31, 2018; accepted April 20, 2018. Date of publication May 1, 2018; date of current version September 11, 2018. This work was supported in part by the National Science Foundation under Award 1711625, Award 1711395, and Grant 1731290 and in part by CNPq, Brazil. This paper was recommended by Guest Editor M. Alioto. (Corresponding author: Leonid Belostotski.)

- V. Ariyarathna and A. Madanayake are with the Department of Electrical and Computer Engineering, University of Akron, Akron, OH 44325 USA (e-mail: bpv1@zips.uakron.edu; arjuna@uakron.edu).
- X. Tang and S. Mandal are with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: xxt81@case.edu; xxm833@case.edu).
- D. Coelho and L. Belostotski are with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: diego.coelho2@ucalgary.ca; lbelosto@ucalgary.ca).
- R. J. Cintra is with the Signal Processing Group, Departamento de Estatística, UFPE, Recife 50740540, Brazil, and also with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: rjdsc@de.ufpe.br).
- T. S. Rappaport is with the Department of Electrical and Computer Engineering, NYU WIRELESS, New York University Tandon School of Engineering, Brooklyn, NY 11201 USA (e-mail: tsr@nyu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JETCAS.2018.2832177

Index Terms—Analog beamforming, multi-beams, 5G multi-beam arrays, low-complexity algorithms.

I. Introduction

ILLIMETER-WAVE (mmW) wireless communication promises an unprecedented change in the wireless industry. Major changes in business models and access capabilities are expected as we move from today's spectral scarcity to fifth generation (5G) networks with a glut of spectrum. The availability of spectrum above 6 GHz will motivate novel mmW systems that support a huge number of Internet of Things (IoT) devices and use several GHz of channel bandwidth, which is orders of magnitude greater than today's cellular bandwidths and also far greater than WiGig IEEE 802.11ac/ad bandwidths [1]–[5]. Thus, the copious quantities of spectral resources available in the 6-300 GHz range promise exponential increases in capacity and data rates when compared to all legacy cellular bands combined. On-chip mmW antenna array processors are required to actually realize such high data rates; i.e., in order to bring "Moore's Law" to wireless capacity [6].

A. Multi-Beam Arrays for mmW Channels

Radio propagation is much more directional at mmW compared to today's cellular bands, which impacts mmW transceiver design, power consumption, and system efficiency [7]. At mmW, most objects that are encountered by a radio wave are much larger than its wavelength. Thus the mmW bands are dominated by scattering and reflection from such large objects (vehicles, people, buildings, etc.) as shown in Fig. 1(a). The free-space path loss, which is proportional to the square of frequency as predicted by the Friis formula, can be compensated by increasing the antenna gain. Heavy attenuation also occurs due to weather (water droplets from rain) and fog/hail in the environment (see Fig. 1(b)). Such attenuation can also be compensated by increasing antenna gain; in particular, by using arrays that form sharp beams. A mmW channel in an indoor or urban environment typically consists of multiple propagation paths due to scattering, reflection, and wave-guiding effects that can be exploited for MIMO communications even in the presence of obstructions in the direct path or when communicating around corners. Thus, the ability to form multiple sharp steerable beams that are adapted under algorithmic control is absolutely essential

2156-3357 © 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

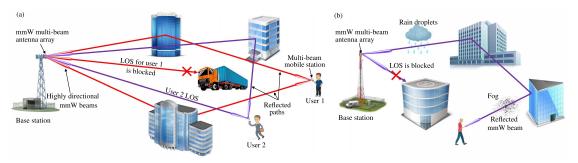


Fig. 1. Examples of highly directional beam-like propagation in mmW wireless chennels, which can suffer from obstructions in dynamically changing mobile environments.

for taking advantage of real-world mmW channels [5]–[9]. Specifically, mmW access points will need to create a large number of sharp wide-bandwidth beams [10]–[12] operating in both transmit and receive modes in order to achieve both capacity and multiple access goals for real-world channels. Thus, wideband multi-beamforming is necessary in order to achieve the orders-of-magnitude increases in capacity, data rate, and geographical penetration demanded by the explosive growth in wireless applications [6], [13]–[17].

B. Channel Models and Multi-Beam Access Points

Highly directional propagation in mmW wireless channels can obstruct sharp beams in dynamically changing mobile environments. For example, human and vehicle motion can result in a rapidly changing channel with blocking and unblocking happening rapidly in an unpredictable manner. An obstacle such as a person or motor vehicle in the direct line-of-sight (LOS) of the wave can cause 40 dB or more signal attenuation [6]. Fortunately, the ray-like propagation characteristics at mmW enable reflection-based connections to be maintained using highly directional antenna arrays even when the LOS path is blocked. In the best case, the LOS plus several non-LOS reflected paths between the transmitter and receiver can increase system capacity for MIMO communications. Electronically steerable multi-beam-capable arrays are required to utilize these reflected paths in highly dynamic mobile environments. Such arrays are thus of great importance in order for base stations to provide simultaneous, independent, and wideband mmW MIMO links to many mobile transceivers.

C. Organization

Here we propose analog current-mode circuits that generate multiple beams by using approximate discrete Fourier transforms (DFTs) for spatial filtering. The paper is organized into seven sections. RF system considerations involved in 5G systems are briefly discussed in Section II. Section III then provides a basic introduction to spatial FFT-based multi-beam architectures. Section IV describes the mathematical approach for finding approximate DFT transforms and introduces the 8-and 16-point approximate transforms and their sparse factorizations. Details of current-mode circuit design for achieving 8- and 16-beams are discussed in Section V. Detailed analysis and preliminary experimental results of a circuit realization of the 8-point version are presented in Section VI. These results are compared with a baseline digital implementation in Section VII. Section VIII concludes the paper.

II. REVIEW OF RF SYSTEM CONSIDERATIONS

Multi-beam systems are envisioned for 28-GHz 5G wireless network base stations, mobile stations, micro base stations, pico cells, and user equipment. For mobile 5G systems, compact and energy-efficient highly integrated multi-beam solutions are desirable to improve battery life and reduce heat-dissipation problems while enabling directional agility. The wide bandwidths of 5G systems present new challenges in realizing fully integrated transceivers and make some currently favored transceiver architectures, such as passive mixer-based receivers, not directly applicable [18].

As an example of such a solution, a recent work by IBM [19] demonstrates fully integrated dual-polarization 16-element arrays for 28-GHz 5G applications. The transceiver architecture presented in [19] is a 2-step sliding-IF half-duplex architecture as shown in Fig. 3.

Unlike steerable beamformers using tunable delays at mmW bands, the proposed multi-beam architecture operates in the baseband while supporting up to 1.5 GHz of bandwidth per beam. It utilizes the spatial frequency distribution of directed mmW energy: each mmW beam has a unique spatial frequency that remains intact through up-down conversion in the mixers. Thus, the proposed baseband analog-FFT can realize multiple far-field mmW beams for a variety of transceiver architectures without requiring tunable mmW delay lines.

Aside from the beamforming method, the key RF components in such a system are the low-noise amplifier (LNA) and the power amplifier (PA). LNA design for antenna arrays has been studied earlier [20]. These works have shown that array sensitivity as a whole is impacted significantly by antenna mutual coupling. For a given beamformer, an array active reflection coefficient, Γ_{act} , can be calculated, and the LNA is designed such that it is noise-matched to Γ_{act} . However, when beamformer coefficients are not known at the time of the design or when the same LNAs are simultaneously used for forming more than one beam, [21] has found that noise matching the LNA to the loaded reflection coefficient of the array is more appropriate and results in lower sensitivity on average than if designed for a specific value of Γ_{act} .

Antenna mutual coupling affects the transmitter as well. Some of the transmitter power is lost in the coupled paths [22]. Similar to LNAs, the output network of PAs has to be designed such that the inefficiency due to coupling loss and the stand-alone PA efficiency are optimized. This optimization is independent of beamformer operation and is performed based on both electromagnetic properties of the antenna array, which are obtained from array simulations or measurements,

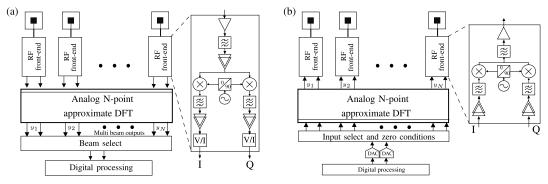


Fig. 2. (a) Receive-mode multi-beam system with down-converted analog basband beamforming; (b) transmit-mode multi-beam system using baseband analog beamforming.

and the specific performance of the selected PA output transistors. These system-level issues are not discussed further in this paper. The remaining sections focus on a baseband implementation of the FFT-based multi-beam beamformer where issues associated with frequency dependent beam squint is insignificant [23].

III. SPATIAL FFT-BASED MULTI-BEAM ARCHITECTURES

The DFT is described by an $N \times N$ linear transform (LT) that splits N-samples of a signal into its frequency components. Performing the DFT operation across the signals obtained from a uniformly spaced linear antenna array produces an orthogonal set of beams where each bin corresponds to a particular direction, thus realizing a multi-beam beamformer. Multiple broadband transmit/receive beams that are orthogonal to each other in space are a critical need for emerging wireless systems as well as defense applications in radar and electronic warfare; as they achieve greater capacity by enabling spatial diversity for MIMO [7]. The Fast Fourier Transforms (FFTs) [24] are fast algorithms for efficiently computing the DFT. They are essential for a tremendous number of applications, such as wireless communications [25], [26], networking [27], sensor networks [28], [29], cognitive radio [30], radar [31], imaging [32], [33], filtering [34], correlation [35], and radio-astronomy [36]. Any type of FFT implementation has computational error due to the use of finite precision arithmetic to realize irrational matrix coefficients. Thus, it may be possible to compute the spatial DFT operation faster without losing much performance if we approximate the DFT itself. Given this possibility, if we can find the simplest possible approximate LT to replace a DFT with tolerable error, we would be able to improve upon traditional FFTs in terms of speed and hardware complexity. In particular, while the DFT is mostly implemented digitally, the proposed simple transforms may have efficient analog implementations. These are likely to have power and layout area advantages over digital implementations for broadband applications.

We propose to achieve such FFT-like performance by LT approximation and fast algorithm factorizations that compromise on accuracy [37]. We start with a maximum error tolerance that must be met, and iterate over all the possible matrix representations of the transform, mark those that are within the desired error bound, and use the one that has the simplest analog implementation. In particular, transforms with small integer coefficients are desirable because they enable straightforward current-mode implementations by changing

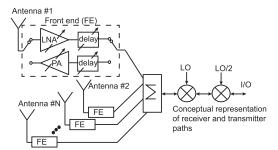


Fig. 3. A representative 28-GHz 5G half-duplex transceiver architecture based on delay-and-sum beamforming [19].

the W/L (width/length) ratios of CMOS transistors to implement multiplication by the coefficients. Such current-mode implementations are intrinsically fast because their bandwidth is only limited by the poles of the current mirrors and not by the maximum clock rate as for digital counterparts [38].

We propose approximate DFTs in which complex phasing networks corresponding to twiddle factors in conventional analog-FFT implementations [39] are replaced by simple weights $\pm 1, 0, \pm \frac{1}{2}$ that can be easily realized in both analog and digital hardware. For example, while a digital circuit that implements the conventional twiddle factors with a certain precision using fixed-point arithmetic would need $\mathcal{O}(N \log(N))$ complex multipliers, the simplified weights can be realized using only bit shifters. Similarly, current-mode, current-mirror-based integrated circuit (IC) implementations of the approximate algorithms would need only 2-4 wellmatched transistors for realizing each coefficient. The area and power consumption of both these signal processing approaches (digital and analog) scale with accuracy requirements [40]. This paper will show that analog implementations of the proposed approximate DFT algorithms offer significantly lower power and area consumption than digital ones for accuracy levels and operating bandwidths that are appropriate for mmW multi-beam 5G transceivers. Fig. 2 shows system configurations for both transmit and receive modes that can produce N simultaneous beams by using the proposed analog N-point approximate-DFT (ADFT), where $N \in \{8, 16\}$.

IV. MATHEMATICAL BACKGROUND

A. Discrete Fourier Transform

The *N*-point DFT is characterized by a linear orthogonal transformation matrix \mathbf{F}_N whose entries are [34]:

$$[\mathbf{F}_N]_{k,n} = \omega_N^{nk},$$

where $\omega_N = \exp(-2\pi j/N)$ is the Nth root of unity and $j = \sqrt{-1}$. If N is even, the following relationship holds true:

$$[\mathbf{F}_N]_{k+\frac{N}{2},n} = (-1)^n [\mathbf{F}_N]_{k,n},$$
 (1)

for k = 0, 1, ..., N/2 - 1 and n = 0, 1, ..., N, and

$$\left[\mathbf{F}_{N}\right]_{k,n+\frac{N}{2}}=\left(-1\right)^{k}\left[\mathbf{F}_{N}\right]_{k,n},$$

for k = 0, 1, ..., N and n = 0, 1, ..., N/2 - 1. Therefore, the DFT matrix can be written as follows [39]:

$$\mathbf{F}_N = \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{A}_{0,1} \\ \mathbf{A}_{1,0} & \mathbf{A}_{1,1} \end{bmatrix}. \tag{2}$$

Here the four blocks have size $N/2 \times N/2$ and the entries of the sub-matrices $\mathbf{A}_{0,1}$, $\mathbf{A}_{1,0}$, and $\mathbf{A}_{1,1}$ are identical to the entries of $\mathbf{A}_{0,0}$ except for sign changes. If N is a multiple of four, further symmetries are present in the terms of the block $\mathbf{A}_{0,0}$. The entries of $\mathbf{A}_{0,0}$ satisfy:

$$\left[\mathbf{A}_{0,0}\right]_{k+\frac{N}{4},n} = (-j)^n \left[\mathbf{A}_{0,0}\right]_{k,n},\tag{3}$$

for k = 0, 1, ..., N/4 - 1 and n = 0, 1, ..., N, and

$$\left[\mathbf{A}_{0,0}\right]_{k,n+\frac{N}{4}} = (-j)^k \left[\mathbf{A}_{0,0}\right]_{k,n},\tag{4}$$

for k = 0, 1, ..., N and n = 0, 1, ..., N/4-1. Therefore, the sub-matrix \mathbf{A}_{00} can also be split into sub-matrices in a similar way as in (2).

The symmetries in (1)-(4) imply that the DFT matrix \mathbf{F}_N contains only N/4 different complex numbers making analog hardware implementation comparatively simple.

B. DFT Approximations

Because the DFT and its approximations are matrices, we define the latter through a matrix mapping [41]–[44]:

$$f: \mathbb{C}^{N/4-1} \to \mathbb{C}^N \times \mathbb{C}^N$$
$$\mathbf{a} \mapsto \hat{\mathbf{F}}_N,$$

where $\mathbf{a} = \begin{bmatrix} a_1 \ a_2 \dots a_{N/4-1} \end{bmatrix}^{\mathsf{T}}$ is an N/4-point complex parameter vector and the entries of $\hat{\mathbf{F}}_N$ are given by:

$$\left[\hat{\mathbf{F}}_N\right]_{k,n} = (-1)^p (-j)^t a_{nk \bmod N/4},$$

where $a_0 = 1$, $p = n \mod N/2 + k \mod N/2$ and $t = n \mod N/4 + k \mod N/4$. Such mapping ascribes the DFT symmetries in (3) and (4) to the resulting approximated DFT matrices [39], [44].

Furthermore, the entries of parameter vector **a** are expected to be simple to ensure the low-complexity cost of the resulting approximations [43]. In fact, the set of dyadic rationals is a suitable choice for the numerical domain of the parameter vector **a** [44]. Thus, we require that $a_n \in \mathcal{P}^2$, where $\mathcal{P} = \{0, \pm 1, \pm 2, \pm 1/2\}$ is the selected low-complexity number set [42], [43], [45]. Such low-complexity complex entries can dramatically reduce the complexity of the resulting approximations. This is because the dyadic elements represent trivial multiplications [39]; therefore, the derived approximation is multiplierless by construction.

A good DFT approximation is expected to exhibit similarity and proximity to the exact DFT matrix in some sense. As a proximity measure, we adopt the following figure of merit:

$$d(\mathbf{a}) = \left\| \hat{\mathbf{F}}_N - \mathbf{F}_N \right\|_{\mathrm{F}}^2,$$

where $\|\cdot\|_{F}$ denotes the Frobenius norm [44].

Thus, by minimizing the above proximity measure, meaningful DFT approximations can be derived. However, besides matrix symmetries, DFT approximations are also expected to satisfy the orthogonality property. Strict orthogonality can also be relaxed in favor of near-orthogonality [46]. We can ensure this property for the proposed class of DFT approximations by searching for DFT approximations whose product $\hat{\mathbf{F}}_N \hat{\mathbf{F}}_N^{\dagger}$ is close to a diagonal. Thus, orthogonality and near-orthogonality can be quantified according to the deviation from diagonal measure [42], [43], [46]:

$$\phi(\hat{\mathbf{F}}_N) = \frac{\left\| \operatorname{diag} \left(\hat{\mathbf{F}}_N \hat{\mathbf{F}}_N^{\mathsf{H}} \right) \right\|_{\mathsf{F}}}{\left\| \hat{\mathbf{F}}_N \hat{\mathbf{F}}_N^{\mathsf{H}} \right\|_{\mathsf{F}}},\tag{5}$$

where the superscript H denotes the Hermitian conjugation. As a criterion, candidate approximations, whose deviation from orthogonality exceed 0.2, are considered unsuitable. The value 0.2 stems from [47] as a reference model for good approximations [42], [43], [45].

Additionally the inverse of $\hat{\mathbf{F}}_N$ is required to be well-defined to allow perfect reconstruction. For such, it suffices that the determinant is nonzero or far from zero.

Mathematically, this requires solving the following optimization problem:

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} d(\mathbf{a}),\tag{6}$$

subject to the following constraints:

- 1) the entries of **a** satisfy $a_n \in \{x + jy | x, y \in \mathcal{P}\}.$
- 2) the inverse transformation is well-defined;
- 3) the inverse matrix must also be of low-complexity;
- 4) orthogonality or near-orthogonality must be satisfied (cf. (5)).

The resulting approximation is furnished by $f(\mathbf{a}^*) = \hat{\mathbf{F}}_N^*$. Although the search space of (6) grows exponentially in terms of N, we focus on the solution of (6) for particular small blocklength transforms, namely $N \in \{8, 16\}$. For the 8- and 16-point DFT, the corresponding parameter vector contains only one and three elements, respectively. Thus for these cases the optimization problem can be solved by means of exhaustive search. By solving the optimization problem (6) for the 8- and 16-point approximations, the following parameters are obtained, respectively: [1-j] and $[1-j/2,1/2-j/2,1/2-j]^{\top}$.

C. Fast Algorithms for Optimal DFT Approximations

Generally, fast algorithm techniques for discrete time transforms aim at reducing the number of multiplication operations [39]. This is fundamentally because the multiplication operations are prone to consume more hardware resources for execution than addition and bit-shifting operations.

In the case of ADFTs, the approximate transforms are natively multiplierless; thus their associated fast algorithms focus on reducing the number of additions and/or bit-shifting operations.

1) 8-Point DFT Approximation: The 8-point DFT approximation matrix resulting from the optimal solution in (6) is f([1-j]), given by:

$$\hat{\mathbf{F_8}} = \frac{1}{2} \times \begin{bmatrix} 2 & 1-j & -2j & -1-j & 2 & -1+j & 2j & 1+j \\ 2 & 1-j & -2j & -1-j & -2 & -1+j & 2j & 1+j \\ 2 & -2j & -2 & 2j & 2 & -2j & -2 & 2j \\ 2 & -1+j & 2j & 1-j & -2 & 1+j & -2j & -1+j \\ 2 & -2 & 2 & -2 & 2 & -2 & 2 & -2 \\ 2 & -1+j & -2j & 1+j & -2 & 1-j & 2j & -1-j \\ 2 & 2j & -2 & -2j & 2 & 2j & -2 & -2j \\ 2 & 1+j & 2j & -1+j & -2 & -1-j & -2j & 1-j \end{bmatrix}.$$
 (7)

Considering the factorization methods described in [34], [39], [42], [43], the following matrix factorization is derived:

$$\widehat{\mathbf{F}} = \mathbf{P} \cdot \mathbf{A}_4 \cdot \mathbf{D} \cdot \mathbf{A}_3 \cdot \mathbf{A}_2 \cdot \mathbf{A}_1 \tag{8}$$

where

D = diag([1, 1, 1, j, 1, j, j, 1]), and **P** = $\begin{bmatrix} \mathbf{e}_0 & \mathbf{e}_4 & \mathbf{e}_2 & \mathbf{e}_5 & \mathbf{e}_1 & \mathbf{e}_7 & \mathbf{e}_3 & \mathbf{e}_6 \end{bmatrix}^{\mathsf{T}}$ is a permutation matrix whose columns \mathbf{e}_i , for $i = 0, 1, \dots, 7$ are vectors with null entries expect for the *i*th component that is unity. This particular factorization requires only 26 analog additions as the total computational cost.

2) 16-Point DFT Approximation: The resulting 16-point DFT approximation is $f([1-j/2 \ 1/2-j/2 \ 1/2-j]^{\mathsf{T}})$ and can be written as follows:

$$\hat{\mathbf{F}}_{16} = \frac{1}{2} \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{A}_{0,1} \\ \mathbf{A}_{1,0} & \mathbf{A}_{1,1} \end{bmatrix}, \tag{9}$$

where the equation can be derived, as shown at the top of the next page.

The DFT approximation matrix $\hat{\mathbf{F}}_{16}$ can be factorized as [39] and [42]:

$$\hat{\mathbf{F}}_{16} = \mathbf{B}_1 \cdot \mathbf{D} \cdot \mathbf{B}_2 \cdot \mathbf{B}_3 \cdot \mathbf{B}_4 \cdot \mathbf{B}_5, \tag{10}$$

and $\mathbf{D} = 1/2 \operatorname{diag}(1, 1, 1, 1, 1, 1, 1, 1, 1, j, j, j, j, j, j)$.

D. Adoption of DFT Approximates for Beamforming Systems

The most important concern for the use of the approximated DFT algorithms in multi-beam systems is the increase in side-lobe levels that occurs as an artifact of the approximation process. As long as these increases are within an acceptable level, the approximate algorithms become good candidates for beamforming applications. Fig. 4 compares numerically simu-

$$\mathbf{A}_{0,0} = \begin{bmatrix} \frac{2}{2} & \frac{2}{2-1i} & \frac{2}{1-1i} & \frac{2}{1-2i} & \frac{2}{1-1i} & \frac{2}{2-2} & \frac{2}{1-2i} & \frac{2}{1-2i} & \frac{2}{1-1i} & \frac{2}{2-2} & \frac{2}{1-2i} & \frac{2}{1-2i} & \frac{2}{1-1i} & \frac{2}{2-2} & \frac{2}{1-2i} & \frac{2}{1-2i}$$

lated beams for the exact DFT and the ADFT for the 8- and 16point scenarios against the normalized spatial frequency. The maximum magnitude error of 8-point approximated transform beam response (Fig. 4 (b)) with respect to exact DFT responses (Fig. 4 (a)) was calculated as 3.0% where as for 16-point case it was calculate to be 2.9%. In addition, the main-lobes point in the same directions as the exact transforms, which make the approximate transforms well suited for multi-beamforming applications. However, note from Fig. 4 (b) that the shapes of four side lobes differ from those of the other four. A similar result can also be observed in Fig. 4(d). Since the matrix entries of the approximate transform are constrained to a limited choice of integer values and orthogonality was relaxed in favor of near-orthogonality, the filter banks associated to the approximate and exact DFT are different giving rise to an error in the side lobes. Nevertheless, such a minor distortion does not affect the targeted application in the sense that it is smaller than the secondary lobe of the exact DFT.

The method given in (6) can also be applied for larger matrix sizes such as N = 1024, 2048, The resulting approximation matrices possess better properties such as nearorthogonality and spectral behavior that is close to the exact DFT. The main problem in deriving such larger approximate DFT matrices is the necessity of creating associated fast algorithms: the larger the matrix, the harder it becomes to derive such algorithms. At the time of writing this manuscript, we are able to generate up to 64 beams using multiplierless approaches, and up to 1024 beams using $\mathcal{O}(N)$ complexity for multipliers. A possible alternative for generating such large approximations is the use of scaling methods. Essentially, larger ADFTs can be generated by re-using smaller ADFTs. This avoids the hassle of deriving fast algorithms for large matrices, which the authors find to be impractical for matrices larger than 32×32 , while still furnishing approximations with $\mathcal{O}(N)$ complexity for multipliers.

V. CIRCUIT TOPOLOGIES, BEAMFORMING ARCHITECTURES

Realization of the analog-DFT is key to implementing the mmW receiver and transmitter architectures shown in Fig. 2. Early attempts to implement analog DFT processors used op-amp circuits to realize the weights of the DFT matrix [48]. This approach is slow and difficult to scale to larger arrays because the twiddle factors become closer to each other as the FFT size increases, making them harder to realize accurately. More recently, a 0.13 μ m CMOS 8-point Cooley-Tuckey FFT processor for orthogonal frequency division

multiplexing (OFDM) applications was reported [49], [50]. The processor uses a time-interleaving bank of sample-andholds and discrete time analog multipliers, and has been tested with 1 GS/s OFDM inputs. However, dedicated input signals are used to represent the FFT coefficients, which makes scaling difficult. 2-D rectangular LC lattices implemented on CMOS have also been proposed for computing analog DFTs of spatial input signals [51]. The method has been verified using numerical simulations, and bandwidths of > 10 GHz are possible for on-chip implementations. However, such large bandwidths require small inductor and capacitor values, which are difficult to realize accurately, and unwanted mutual coupling between the inductors is also an issue. The analog FFT processor in [52] uses a current-mirror-based architecture to scale the input current by the twiddle factor weights. However, the authors had to approximate the weights to the first decimal place for ease of implementation, resulting in degraded beam shapes. Finally, the work in [53] describes a 16-point analog domain FFT using a charge-reuse analog Fourier transform (CRAFT) engine. The circuit uses charge reuse to achieve an input bandwidth of 5 GHz. However, the design requires RF samplers in the front-end, and inaccuracies in the capacitance network lead to twiddle factor errors that make scaling difficult.

A critical issue faced by all previous approaches for realizing analog DFTs has been that accurate twiddle factor values are difficult to generate on-chip. The level of difficulty grows as the FFT size increases since the factors become closer to each other, which results in performance degradation. Hence it makes sense to allow some error margin and implement approximate transforms with integer twiddle factors. The resulting implementations are more scalable since the transform coefficients are now constrained to a small set of Gaussian integers. The approximate transforms in (7) and (9) satisfy this special property, i.e., are limited to small integer coefficients $\mathcal{P} \in \{0, \pm 1, \pm 2\}$. This property enables high-bandwidth current-mode analog ICs in which well-controlled geometric parameters (namely, the W/L ratios of current mirror transistors) determine the integer coefficients.

A. Analog Current-Mode ADFT Designs

Let the current-mode signals captured by the N Nyquist-spaced antennas of an N-beam multi-beamforming system be $x_{\text{in}} = [x_1, x_2, \dots, x_N]^T$. The beam outputs $y = \hat{F_N} \cdot x_{\text{in}}$ where $y = [y_1, y_2, \dots, y_N]^T$ correspond to unique directions of arrival given by $\psi_i = \sin^{-1}(\frac{2k}{N})$ where

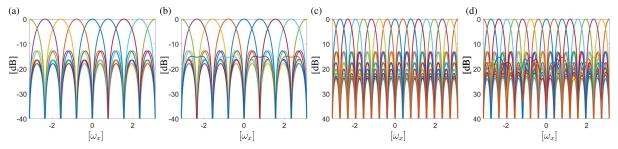


Fig. 4. (a) 8-point exact DFT beams and the corresponding (b) approximated-DFT beams; (c) 16-point exact DFT beams and the corresponding (d) approximated-DFT beams as a function of the spatial frequency.

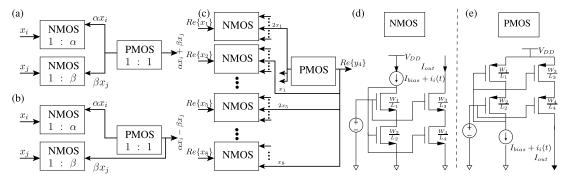


Fig. 5. Current-mode implementation of (a) addition and (b) subtraction operations, which are the primary functions for implementing the analog ADFT circuit shown in Fig. 2 to realize mmW receivers and transmitters; (c) example showing the implementation of row 4 of $Re\{\hat{F}_8\}$; (d) NMOS and (e) PMOS current mirros designed using a low-voltage cascode topology.

$$k = \begin{cases} i; & 1 \le i \le N/2 \\ -i; & N/2 < i \le N \end{cases}$$
 and i is the output bin number. In current mode, the output current at each output bin y_i requires implementation of $\sum_{k=0}^{N-1} p_{ik} x_k$ where p_{ik} denotes a matrix coefficient. The addition and subtraction arithmetic

In current mode, the output current at each output bin y_i requires implementation of $\sum_{k=0}^{N-1} p_{ik}x_k$ where p_{ik} denotes a matrix coefficient. The addition and subtraction arithmetic required for this calculation can be implemented directly using NMOS and PMOS current mirrors as shown in Fig. 5(a). Here α and β are the weights by which the input current needs to be scaled. Thus the transforms in (7) and (9) can both be realized using analog current-mode CMOS. The $Re\{\hat{F}_8\}$ and $Im\{\hat{F}_8\}$ transforms can be implemented separately to realize the full transformation. For example, implementation of the 4th row of $Re\{\hat{F}_8\}$ is shown in Fig. 5(c).

In section V-B, we discuss an 8-point current-mode design that follows this approach. Such an approach would in general require $\mathcal{O}(N^2)$ current mirrors for generating N-beams. A digital architecture, which implements a fast algorithm (e.g. Fourier, discrete sine/cosine), involves the implementation of butterfly matrices. Use of such sparse factorization matrices reduces the hardware complexity of the transform. Given that a sparse factorization exists for the transform of interest, the same principle can be applied to analog implementations. Therefore we take advantage of a sparse factorization that reduces the arithmetic complexity of the approximate transforms given in (7) and (9). It is realized by implementing each factorization individually and then combining them.

After observing the factorization matrices of (8) and (10), it can be seen that each row and column of the factorized matrix consists of a maximum of two elements. This implies that i) each output of the factorization stage requires an addition of the form $pa_i + qa_j$ where $p, q \in \mathcal{P}$ and a_i and a_j are i^{th} , j^{th} $0 < i, j \leq N$ inputs to the factorized matrix; and ii) each input a_i has to be copied a maximum of two times.

Thus an implementation of each stage of the factorization given in (8) and (10) comprises of N NMOS current copiers producing two current copies and N PMOS mirrors attached to the outputs of each stage. The full realization of the analog circuit requires implementation of all the factorization stages and separate copies of the hardware for the real and imaginary parts as shown in Fig. 6. The analog implementation of the last factorization stage B_1 in the real signal path requires small changes in the circuit compared to the implementation (denoted by B_1) used in the imaginary signal path. Therefore it is denoted as B_1' . In the design of the B_1' block, the polarity of the signals entering from inputs 10-16 have been negated to account for the -1 generated by multiplication of the outputs 10-16 of the imaginary component at the B_2 stage by j.

B. 8-Point ADFT Implementation in 65 nm CMOS

Given the effort required to design different stages, the factorized version of the transformation is not beneficial for realizing analog transforms with smaller values of N. Therefore, the 8-point ADFT was designed using the final matrix in (7). Real and imaginary parts of F_8 were separately realized using 65-nm general-purpose (GP) CMOS technology cells and BSIM4 RF transistor models using the approach shown in Fig. 5(c). Both types of current mirrors were realized using low-voltage cascode topology to reduce systematic errors due to finite output impedance while simultaneously obtaining enough voltage headroom to accommodate four transistors in series (see Figs. 5(d) and (e)). The W/L values used for the circuit that implements the NMOS and PMOS mirrors in Fig. 5 are tabulated in Table I. For 1:1 NMOS mirrors, the output branch transistor sizes were set as W_3/L_3 = W_1/L_1 , $W_4/L_4 = W_2/L_2$, while for 1:2 NMOS mirrors the

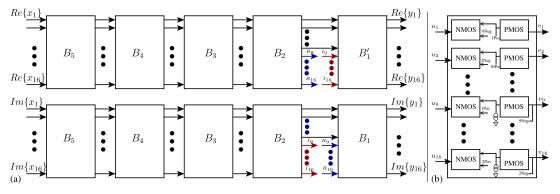


Fig. 6. (a) System architecture of the 16-point analog ADFT; (b) realization of the B₅ factorization stage using current mirros.

sizes were set to $W_3/L_3 = 2W_1/L_1$, $W_4/L_4 = 2W_2/L_2$. The design was simulated using Cadence Spectre and initial results were reported in [45]. It was then laid out; post-layout simulation results and noise analysis are shown in Section VI.

Fig. 7 shows simulated beam patterns for sinusoidal inputs at 2 and 4 GHz. These are in good agreement with the exact DFT responses (maximum magnitude error of 6% and 10% for 2 GHz and 4 GHz respectively), but are distributed unevenly versus angle. Specifically, the beams are more concentrated around 0° than around -90° and 90° . This is due to the nature of the DFT, which samples the frequency domain linearly, i.e. $\omega_x = \frac{2\pi k}{N}$ where k is the bin number and ω_x corresponds to the spatial frequency. As a result, $\omega_x = -\omega_{ct} \sin \psi$ where ω_{ct} is the normalized temporal frequency and ψ is the spatial angle. Thus each output bin of the DFT corresponds to a different angle which has a \sin^{-1} relationship with k. In particular, the beam directions are given by $\sin^{-1}\left(\frac{2k}{N}\right)$ for a Nyquist-sampled array, resulting in an uneven distribution.

C. 16-Point ADFT Implementation in 65 nm CMOS

The reduced hardware complexity of the factorized version makes it attractive for higher values of N. Specifically, the number of current copies needed for realizing an N-point ADFT transform in non-factorized form has a maximum value of $4N^2$ (asymptotically $\mathcal{O}(N^2)$), whereas it is $4S_NN$ in the factorized form where S_N is the number of factorization stages (realized using NMOS mirrors). Since $S_N \ll N$, the latter number asymptotically converges to $\mathcal{O}(N)$. The factorized circuit also needs $\mathcal{O}(N)$ PMOS mirrors for performing current addition/subtraction, so the total number of mirrors required remains $\mathcal{O}(N)$ as compared to $\mathcal{O}(N^2)$ for the direct approach. For example, 308 current copies are needed for implementing the factorized version of the 16-point ADFT in current mode, whereas a direct realization would require 768 copies. Although the number of PMOS mirrors required for addition/subtraction of currents is higher for the factorized implementation (160 versus 96), the overall complexity associated with the factorized implementation is still significantly lower. Moreover, this performance benefit becomes larger as N increases. Thus the individual factorization stages given by equations (11) to (15) were implemented using 65-nm CMOS. The real and imaginary components of the designs were implemented separately and the top-level architecture is shown in Fig. 6. The inputs to the design were assumed to be $5-\mu$ A peak-to-peak RF signals superimposed on a $100-\mu$ A DC

TABLE I
CURRENT MIRROR SIZING FOR THE 8-POINT ADFT CIRCUIT

	W_1	L_1	W_2	L_2
1:1 NMOS	3.6 μm	$0.15 \ \mu \text{m}$	2.8 μm	$0.15 \ \mu \text{m}$
1:1 PMOS	29.6 μm	0.15 μm	32 μm	$0.15~\mu{\rm m}$

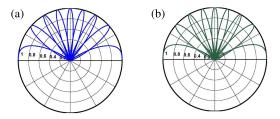


Fig. 7. Simulated beam patterns generated by the analog 8-point ADFT design at (a) 2 GHz and (b) 4 GHz (maximum magnitude beam errors of 6% and 10%, respectively).

bias current. The factorized stages were designed and cascaded starting from B_5 to B_1 to generate the 16-point transformation as shown in Fig. 6. The diagonal matrix D is implemented as a cross-connection of wires. All NMOS and PMOS mirrors used are low-voltage cascode current mirrors, as in the case of the 8-point design. The output bias currents of the stages were equalized by using PMOS mirrors with appropriate bias currents. The transistor sizes (named as shown in Fig. 5 (d) and (e)) used for the basic NMOS and PMOS mirrors in each individual stage are shown in Table II. The transistor sizes of the output branch were set depending on the magnitude of the matrix coefficient that was realized.

D. Simulated Beams

The circuit was simulated in Cadence Spectre using noiseless input signals generated by MATLAB. For a given direction of arrival (DOA) ψ , 16 spatially Nyquist-sampled sinusoidal signals, which emulate downconverted plane waves, were generated. The simulation frequencies of the signals were chosen to be within the baseband bandwidth of interest, which is smaller than the bandwidth of the circuit being simulated [13]. These inputs were fed to the circuit and the simulated output waveforms were recorded. To obtain the array factor for each bin of the ADFT, waveforms were generated for different values of ψ in the range -90° to 90° . The output waveforms were then exported to MATLAB in order to compute the beamformed signals as a function of ψ .

W/L VALUES USED TO REALIZE CURRENT MIRRORS AT EACH FACTORIZED STAGE IN THE 16-POINT ADFT CIRCUIT

B5 B4 B3 B2 B1

NMOS L PMOS L

TABLE II

	B_5		B_4		B_3		B_2		B_1	
	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS
W_1	0.8 μm	3.7 μm	0.8 μm	3.7 μm	1.8 μm	3.7 μm	1.2 μm	3.7 μm	1.2 μm	$3.7~\mu\mathrm{m}$
L_1	$0.15 \ \mu m$	0.24 μm	$0.15 \ \mu m$	$0.24 \ \mu m$	$0.15 \ \mu m$	$0.12 \ \mu m$	0.06 μm	$0.12 \ \mu m$	0.06 μm	$0.12 \ \mu m$
W_2	0.6 μm	4 μm	0.6 μm	4 μm	$1.4~\mu\mathrm{m}$	4 μm	$1.2~\mu\mathrm{m}$	4 μm	1.2 μm	4 μm
L_2	0.09 μm	0.24 μm	$0.09 \ \mu m$	$0.24~\mu{\rm m}$	$0.06~\mu{\rm m}$	$0.12 \ \mu m$	0.06 μm	$0.12 \ \mu m$	0.06 μm	$0.12 \ \mu m$

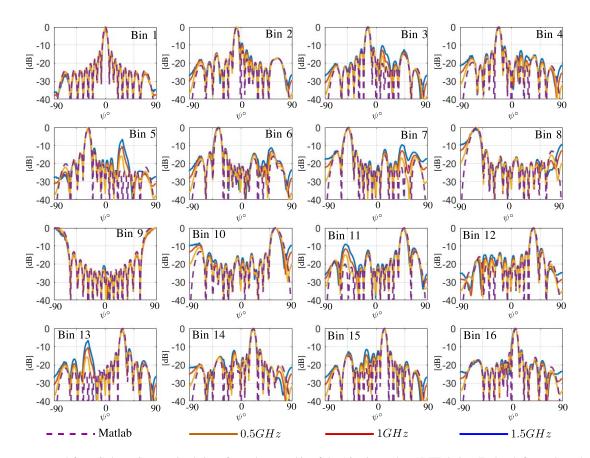


Fig. 8. Beams generated from Cadence Spectre simulations for each output bin of the 16-point analog ADFT design. Each sub-figure shows beam patterns for different IF bandwidths. The patterns for an ideal accurate DFT were simulated using MATLAB and are also shown (dashed magnenta lines). Simulated baseband channel bandwidths in this work correspond to 1 to 2 GHz, as required for wideband 5G channels operating at carrier frequencies of 28 or 38 GHz [13].

Fig. 8 shows the power patterns of each beam at different frequencies.

In general, the beam shapes obtained from Cadence simulations closely follow theoretical ADFT responses. The side lobe levels of the beam patterns at 500 MHz remain consistent with theoretical responses except for bins 5, 7, 11, and 13, which have significantly higher side-lobe levels in the stop band than expected. This effect arises due to slight deviations between the theoretical coefficients and those realized by the cascaded current-mode architecture. Since the bias currents grow in magnitude as we traverse from input to the output through the factorization stages, the absolute values of the current matching errors also tend to grow. Thus, the realized coefficient values deviate from the required ones, resulting in deviations in the beam patterns. Nevertheless, these changes are in the stop band and are lower than the maximum side lobe level. Thus, they are of minor importance for beamforming

applications. However, they become more significant as the frequency increases. At a baseband frequency of 500 MHz the average and worst case peak lobe levels were -12.9 dB and -11.9 dB, respectively. At 1 GHz and 1.5 GHz these numbers increased to -12.2 dB and -11.3 dB (average) and -10.2 dB and -9.1 dB (worst case), respectively.

VI. POST-LAYOUT ANALYSIS AND EXPERIMENTS

The 8-point current-mode ADFT circuit was laid out and fabricated in the UMC 65-nm RF-CMOS process. The chip was designed for a 1.8-V power supply, and the transistor sizes were chosen to ensure a current mirror bandwidth of at least 1 GHz for a DC bias current of 100 μ A. As shown in Fig. 2, the chip is placed between the RF front-end and digital backend of a mmW transmitter or receiver array, both of which are expected to have voltage-mode interfaces. Thus, V-I and I-V

TABLE III

SIMULATED CHANGES IN BEAM POWER LEVELS DUE TO PROCESS VARIATIONS AND MISMATCH FOR THE 8-POINT ADFT CIRCUIT AT 200 MHz. THE ANGLES CORRESPOND TO THE DOA VALUES FOR MAXIMUM RESPONSES FOR EACH BEAM

Beam angle	Mean (µ)	Standard deviation (σ)
0°	28.6%	4.6%
-14.5°	10.4%	2.8%
−30°	20.7%	2.9%
-48.5°	10.7%	2.9%
$\pm 90^{\circ}$	21.1%	3.1%
14.5°	10.7%	3.0%
30°	20.7%	3.1%
48.5°	10.6%	2.8%

converter circuits have to be added to the current-mode ADFT core. Such circuits were designed, and the performance of the overall implementation was then analyzed.

A. Effect of Mismatch on the Beams

The effect of transistor mismatch on the beam patterns were studied using Monte-Carlo simulations. In each simulation, transistor sizes were randomly chosen based on probability distributions provided by the foundry that include both process variation and mismatch. Table III shows the simulated power deviations of the eight beams due to process variation and mismatch for an input frequency of 200 MHz. Means and standard deviations of the probability distributions, which are nearly Gaussian, are shown; these have maximum values of 28.6% and 4.6% respectively. The deviations in beam width and pointing direction were small (<1% in both cases) and are not shown. Finally, the average and worst case (3σ) peak side lobe levels were also degraded by relatively small amounts (0.2 dB and 1.8 dB, respectively).

B. Design of V-I/I-V Converter Circuits

As shown in Fig. 2, V-I/I-V converter circuits are necessary for interfacing the current-mode ADFT core with external circuits. Thus, such converters were added to provide 50 Ω impedance at each input and output port of the ADFT. The noise and linearity of the V-I converter dominate the dynamic range (DR) of the multi-beamformer. Our design (see Fig. 9(d)) uses a common-gate input stage for impedance matching. The bias current is adjusted via V_n to set the desired input impedance $Z_{in} \approx 1/g_s$, while the AC current is mirrored to create the output current I_{out} . The DC value of I_{out} , which sets the power consumption and bandwidth of the current-mode core, can be independently adjusted via V_p .

Fig. 9 shows simulation results for the V-I converter for a NMOS bias current (set by V_n) of 1.6 mA, a PMOS bias current (set by V_p) of 1.1 mA, and an output bias current of (1.6-1.1)=0.5 mA. The power consumption is 2.38 mW. A Bode plot of the effective small-signal transconductance G_m (see Fig. 9(c)) shows a -3 dB bandwidth of 2.7 GHz. The lower cut-in frequency is set to $f_c \approx g_s/(2\pi C_{dc})$ by the value C_{dc} of an input DC blocking capacitor (not shown here). The circuit is well-matched over the useful frequency range: $|S_{11}|$ is approximately -16 dB from 10 MHz

to 4 GHz as shown in Fig. 9(b). The input-referred noise power spectral density (PSD), including noise from the source resistance, is $\approx 1 \text{ nV/Hz}^{1/2}$ as shown in Fig. 9(f). The resulting noise figure (NF) is 6 to 8 dB over the operating bandwidth, which is adequate since in practice this baseband circuit will be preceded by an RF receiver chain. The total integrated (1 MHz to 10 GHz) output current noise is $i_{out,n} = 1.0 \ \mu A_{rms}$.

The simulated total harmonic distortion (THD) versus input amplitude at two frequencies is shown in Fig. 9(e). As expected, THD levels increases with frequency. The maximum input amplitudes for THD < 5% are $V_{in,max} = 30$ mV and 13 mV at 100 MHz and 1 GHz, respectively. Assuming no further low-pass filtering to limit the output bandwidth, the dynamic range (DR) of the circuit is

$$DR = 20\log_{10}\left(\frac{V_{in,max}G_m/\sqrt{2}}{i_{out,n}}\right). \tag{16}$$

The resulting values are 49.9 dB and 42.2 dB at 100 MHz and 1 GHz, resulting in an effective number of bits (ENOB) of 8.0 and 6.7 bits, respectively. Similarly, the signal to noise and distortion ratio (*SNDR*) of the ADFT circuit is

$$SNDR = 20 \log_{10} \left(\frac{V_{in} / \sqrt{2}}{\sqrt{v_{in,n}^2 + \alpha^2 V_{in}^2 / 2}} \right), \tag{17}$$

where $v_{in,n} = i_{out,n}/G_m$ is the total input-referred noise voltage, and α is the THD at an input amplitude of V_{in} . For a 1 GHz input, the maximum value of $SNDR \approx 32 \text{ dB}$ occurs for an input amplitude of $V_{in} = 7$ mV, for which THD = 2.5%. This results in a significantly lower ENOB of 5.0 bits. Note that the capacity of the wireless system depends on both THD and SNDR; the relative importance of these specifications thus needs further study. Given that the noise and linearity of the current-mode core and output I-V converter do not limit system performance (the latter is simply a 50 Ω resistor), the ENOB of the final analog outputs is limited either by THD or SNDR to the values stated above. For the remainder of the paper, we will make the conservative assumption that the overall ENOB is limited by SNDR to 5.0 bits. This level of precision is sufficient for most mmW communications applications.

The finite bandwidth of the current mirrors in the ADFT core slightly reduces the output bandwidth compared to that shown in Fig. 9(c). The -3 dB bandwidth of a single mirror is given by $BW \approx g_m/C_{tot}$, where g_m is the transconductance of the input transistor and $C_{tot} \approx 2C_{gs}$ is the total parasitic capacitance where C_{gs} is the gate-source capacitance of transistors. BW can be improved at the cost of current matching accuracy (and eventually beam shape fidelity) by decreasing the transistor area WL, since $C_{tot} \propto WL$ and threshold-voltage mismatch $\sigma_{\Delta V_{th}} \propto 1/\sqrt{WL}$. Alternatively, it can also be improved at the cost of power consumption by increasing $g_m \propto \sqrt{I_{ds}}$. In our case the value of g_m (and thus BW) can be adjusted through the bias voltage V_p in the V-I converters. The actual circuit uses N=3 cascaded mirrors (one NMOS, two PMOS) in the signal path. Assuming that these are identical, the bandwidth is further reduced to $BW \times \sqrt{2^{1/N} - 1} \approx BW/2$.

Fig. 9(a) shows a labeled die photograph of the fabricated chip. The active area is 190 μ m \times 270 μ m. A printed circuit

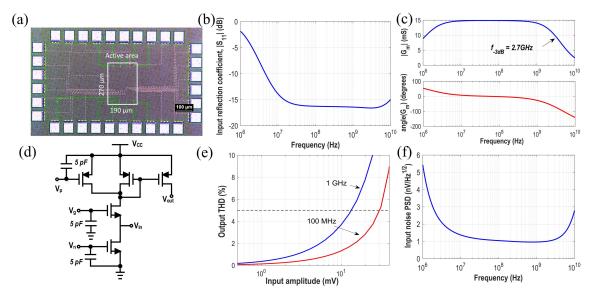


Fig. 9. (a) Die photograph of the chip in the UMC 65 nm 1P/8M CMOS process. (b) Simulated input reflection coefficient |S₁₁|. (c) Simulated Bode plot of the small-signal transconductance. (d) The proposed V-I converter circuit. (e) Simulated total harmonic distortion (THD) versus input amplitude. (f) Simulated input-referred noise power spectral density (PSD) over the frequency range from 1 MHz to 10 GHz. Bias conditions for the V-I converter were as described in the text.

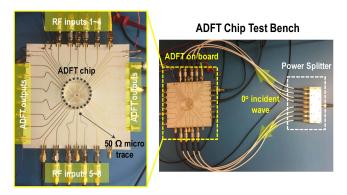


Fig. 10. Experimental setup for testing the 8-point ADFT IC.

board (PCB) for testing the chip has been designed and tested as described in the next section. The die is attached to the PCB using a chip-on-board technique to minimize parasitic capacitance and inductance from the package.

C. Preliminary Experimental Results

A 10-mil-thick PCB using Rogers 4350B substrate material was implemented to create 50 Ω microstrip traces for the RF inputs and outputs of the ADFT chip, as shown in Fig. 10. The bare die was wire-bonded on the PCB. An off-the-shelf one-to-eight 0° phase-shift power splitter with approximately 0.5 MHz to 1.5 GHz bandwidth was used to simulate the outputs of an eight-element uniform linear antenna array with RF signals impinging on them at a DOA of 0°. The bias current of the V-I converter was adjusted by setting the NMOS bias current using $V_n = 0.67 \text{ V}$ and the PMOS bias current using $V_p = 1.34$ V. Moreover, the gate voltage V_g was set to 1.45 V to ensure that all transistors in the V-I converter remain saturated. An external DC blocker (approximately 0.5 MHz to 8.0 GHz bandwidth) was used at each input to isolate the DC level of the RF inputs from the V-I converters, thus ensuring proper biasing of the chip.

Initial experiments were focused on verifying the basic functionality of the chip at low input frequencies (≤ 20 MHz). Fig. 11 compares the experimental and simulated THD of the ADFT chip as a function of the input signal amplitude at two input frequencies: 2 MHz and 10 MHz. Assuming that the maximum allowable THD = 5%, the largest allowable input amplitude based on the experiments are $V_{ein,max} = 13.5$ mV and 9.1 mV at 2 MHz and 10 MHz, respectively. These numbers are in reasonable agreement with the simulated values of $V_{sin,max} = 21.4$ mV and 11.2 mV at 2 MHz and 10 MHz, respectively. In addition, by comparing the THD curves in Fig. 9(e) and Fig. 11, we observe that the linear range of the entire receiver chain (Fig. 11) is significantly degraded compared to that of a single LNA (Fig. 9(e)). This is because of the limited linear range of the ADFT current-mirror matrix.

The measured outputs of the ADFT chip in response to narrowband input signals at 10 MHz and 0° DOA are shown (after normalization) as the red dots in Fig. 12. The output of the 0° beam is much larger than that of the other beams, as expected. However, these have significantly larger power levels (on average, $\sim 10\%$ of the 0° beam) compared to the circuit simulation results shown as blue dots (on average, $\sim 0.16\%$ of the 0° beam). In addition, both these results are worse than the theoretical ADFT outputs to a 0° DOA input, which are 1 for the 0° beam and 0 for all other beams. The observed non-zero outputs of the other beams, which correspond to degradation of beam orthogonality, are due to finite current mirror output impedances (systematic errors) and transistor mismatches (random errors). These effects result in errors in the ADFT matrix coefficients and thus the beam shapes. Note that the circuit simulations result in more accurate beam shapes than the experiments because they include systematic errors but not random ones. The latter can be modeled using Monte-Carlo simulations, as shown in Table III, and can be reduced by i) using optimized layouts to reduce mismatch, such as common centroid geometries; and ii) increasing transistor area

	Di	gital	Analog		
	8-point	16-point	8-point	16-point	
T_{cpd}	1032 ps	1067ps	-	-	
$f_{s,max}$	969 MHz	937 MHz	-	-	
BW	485 MHz	468 MHz	1 GHz	1 GHz	
Power	106 mW ⁽¹⁾	215 mW ⁽¹⁾	70.4 mW	162.4 mW	
Area	$34894 \ \mu \text{m}^2 \ ^{(2)}$	66556 μ m ² (2)	51300 μm^2	(layout not done)	
ADC power	19.2 mW	38.4 mW	19.2 mW	38.4 mW	
V to I power	-	-	38 mw	76.1 mW	
Total power (≈1 GHz BW)	231.2 mW ⁽³⁾	468.4 mW ⁽³⁾	127.6 mW	276.9 mW	
Total area (≈1 GHz BW)	69788 µm ² (4)	133112 µm ² (4)	51300 µm ²	<u>-</u>	

TABLE IV

COMPARISON OF 8- AND 16-POINT ANALOG ADFTS WITH DIGITAL IMPLEMENTATIONS IN THE 45 nm FREEPDK LIBRARY

(1) Power of digital ADFT cores only (no ADC); (2) Area of a single digital core; (3) Total power of two digital cores; (4) Total area of two digital cores.

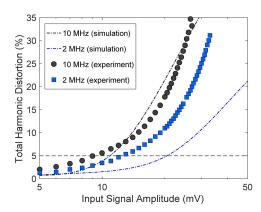


Fig. 11. Measured THD of the 8-point ADFT IC to narrowband inputs at two different input frequencies.

WL and gate overdrive voltage to further reduce mismatch at the cost of increased capacitance and reduced bandwidth.

VII. COMPARISON WITH A BASELINE DIGITAL IMPLEMENTATION

Equivalent digital designs for the proposed 8-point and 16-point ADFTs were implemented in VHDL to allow their bandwidth, power, and area requirements to be compared with the proposed analog implementations. Both digital cores were synthesized using the NCSU 45 nm FreePDK library [54]. Note that this is a more advanced technology than the 65-nm process used for the analog designs. The input word length for the digital synthesis was set to 6 bits to be comparable with the SNDR-limited analog ENOB of 5.0 bits.

The figures of merit for the both digital and analog imple-

mentations are listed in Table IV for comparison. The number of digital cores N_c needed for processing a bandwidth B (assuming polyphase sampling) is given by $N_c = \left(\frac{2B}{f_{s,max}}\right)$. Thus to handle a bandwidth of 1 GHz, $N_c \approx 2$. The total power consumption of the digital equivalent implementation includes that of the digital cores as well as the analog-to-digital converters (ADCs). For quadrature receivers, the digital implementation requires two ADCs per antenna (in total $2 \times N$ for an N-point transform), while the analog implementation only requires two ADCs per sampled beam (real and imaginary outputs). Therefore, when implementing an N-point transform in analog, the number of ADCs required is 2M where $1 \le M \le N$. The numbers in Table IV assume the worst case, i.e. that all N beams are sampled, resulting in M = N. ADC

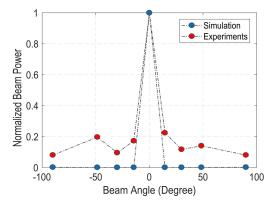


Fig. 12. Measured output responses of the 8-point ADFT chip to narrowband input signals corresponding to a DOA of 0°: simulation (blue) and experiments (red), respectively. The input frequency was 10 MHz.

power was estimated by assuming a converter with suitable specifications (≥ 1 Gs/s, ENOB = 5 to 6 bits) and the lowest possible Walden figure of merit (FoM). We searched Dr. Boris Murmann's ADC survey [55] for this purpose. As of writing, the lowest reported FoM is 28.7 fJ/conversion for ENOB = 5.5 bits at 1 Gs/s [56]. As shown in Table IV, since the bandwidth of the digital cores is ≈ 460 MHz, the ADC power is 2×0.46 GHz $\times 2^{5.5} \times 28.7$ fJ = 1.2 mW per channel.

According to the table, the total power consumption of the proposed analog implementations are $\approx 45\%$ and $\approx 41\%$ less than that of the digital implementations for the 8-point and 16-point cases, respectively. This is despite the fact that the digital results were obtained using a more advanced process (45 nm versus 65 nm). The area metric also indicates a $\approx 27\%$ reduction in area for the 8-point analog circuit even though the corresponding digital core was implemented in 45 nm. Therefore, our results suggest that the proposed analog beamforming solution is significantly more power- and area-efficient than digital implementations with similar performance.

There is another significant advantage of the analog implementation. In a fully digital implementation all antenna outputs have to be amplified to span the full-scale input range of the ADCs. For example, the design in [56] has a input single-ended range of 300 mV. Amplifying all *N* input signals to such large levels prior to digitization is not trivial; the amplifiers have to be linear and so are power hungry. This is even more difficult if directional blockers are present, since these blockers are not rejected until after the beamformer and so the amplifiers and the ADCs have to deal with them.

On the other hand, the proposed analog multi-beamformer can be placed right after a baseband mixer, so only M amplifiers (one per beam) are needed. Moreover, since these amplifiers operate after the beamformer, their linearity requirements are relaxed since the beamformer can greatly suppress blockers.

VIII. CONCLUSION AND FUTURE WORK

At mmW frequencies, wireless signals suffer from heavy attenuation due to obstructions, weather, and other environmental conditions. Such attenuation can be compensated by using array processing to improve transmitter and receiver directivity. Moreover, mmW channels typically have multiple propagation paths due to scattering and reflection that provide spatial diversity in the presence of changing environmental conditions. Thus, the ability to form multiple sharp steerable beams is of great importance in mmW communication systems, and mmW access points need many beams in both transmit and receive modes. Broadband multi-beam analog architectures have been discussed to address this need. ADFT algorithms with small integer coefficients that closely match FFT-based beam-patterns have been proposed for beamforming, and analog CMOS architectures for 8 and 16 simultaneous beams using these transforms have been discussed. The 8-point circuit was realized directly, i.e. by mapping the proposed 8point ADFT matrix to analog current mirrors. This approach has a hardware complexity (number of mirrors) of $\mathcal{O}(N^2)$, making it difficult to realize higher values of N for 5G systems with massive MIMO front-ends. Thus, a more scalable approach was used for the 16-point ADFT. Instead of directly implementing the matrix, individual sparse factorization stages were mapped to current mirrors. This approach reduces the number of current mirrors to $\mathcal{O}(N)$, resulting in lower hardware complexity and circuit area. However, realizing even larger values of N remains challenging.

Beamforming circuits were designed in 65-nm GP CMOS technology. The designs were simulated using Cadence Spectre to obtain the multi-beam array factors. The 8-point version was laid-out and fabricated in UMC 65 nm RF-CMOS, and preliminary experimental results confirm basic functionality of the chip. Moreover, simulation results for the 16-point version show high beam fidelity up to 1.5 GHz of baseband bandwidth, which is sufficient for proposed 5G communications standards.

Future work may include the investigation of larger ADFTs. The investigation of ADFTs are problematic as *N* increases given that good ADFTs are found solving the nonlinear discrete optimization problem defined in Section IV-B. A possible solution to investigate larger ADFTs is to *re-use* smaller optimal approximations. A similar methodology has been used for approximating the discrete cosine transform in the context of image compression [57] and could be applied in the search of ADFTs. This could be derived from fast algorithms that express a large DFT as a function of smaller DFTs, such as the Cooley-Tukey fast algorithm.

The proposed V-I converter circuit topology is not particularly power efficient, which significantly increases total power consumption. Future circuit design efforts will thus be focused on designing a more power-efficient V-I topology. The resulting 8- and 16-point ADFT circuits will be laid out, fabricated, and then tested using uniform linear antenna arrays for different values of DOA and signal bandwidth.

ACKNOWLEDGMENT

The authors would like to thank A. Nikoofard and J. Liang for help with integrated circuit layout.

REFERENCES

- [1] K. Haneda *et al.*, "5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–7. [Online]. Available: http://arxiv.org/abs/1602.07533.
- [2] K. Haneda et al., "Indoor 5G 3GPP-like channel models for office and shopping mall environments," in Proc. IEEE Int. Conf. Commun. Workshops (ICC), May 2016, pp. 694–699.
- [3] F. Boccardi *et al.*, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [4] R. Mudumbai et al., "Distributed transmit beamforming: Challenges and recent progress," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 102–110, Feb. 2009.
- [5] G. MacCartney, Jr., et al., "Millimeter-wave human blockage at 73 GHz with a simple double knife-edge diffraction model and extension for directional antennas," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–6.
- [6] T. S. Rappaport et al., Millimeter Wave Wireless Communications (Prentice Hall Communications Engineering and Emerging Technologies Series from Ted Rappaport). Englewood Cliffs, NJ, USA: Prentice-Hall, 2015.
- [7] S. Sun et al., "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.
- [8] R. W. Heath et al., "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [9] R. Méndez-Rial et al., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.
- [10] K.-J. Koh and G. M. Rebeiz, "0.13-μm CMOS phase shifters for X-, Ku-, and K-band phased arrays," *IEEE J. Solid-State Circuits*, vol. 42, no. 11, pp. 2535–2546, Nov. 2007.
- [11] K.-J. Koh et al., "A millimeter-wave (40–45 GHz) 16-element phased-array transmitter in 0.18-μm SiGe BiCMOS technology," IEEE J. Solid-State Circuits, vol. 44, no. 5, pp. 1498–1509, May 2009.
- [12] H. Hashemi et al., "A fully integrated 24 GHz 8-path phased-array receiver in silicon," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, vol. 1. Feb. 2004, pp. 390–534.
- [13] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: It will work!" IEEE Access, vol. 1, pp. 335–349, May 2013.
- [14] Aalto University, BUPT, CMCC, Nokia, NTT DOCOMO, New York University, Ericsson, Qualcomm, Huawei, Samsung, Intel, University of Bristol, KT Corporation, and University of Southern California, "5G channel model for bands up to 100 GHz," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–56.
- [15] S. Rangan et al., "Millimeter-wave cellular wireless networks: Potentials and challenges," Proc. IEEE, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [16] G. R. MacCartney and T. S. Rappaport, "Rural macrocell path loss models for millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1663–1677, Jul. 2017.
- [17] T. S. Rappaport *et al.*, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design (invited paper)," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [18] H. Krishnaswamy and L. Zhang, "Analog and RF interference mitigation for integrated MIMO receiver arrays," *Proc. IEEE*, vol. 104, no. 3, pp. 561–575, Mar. 2016.
- [19] B. Sadhu et al., "A 28 GHz 32-element phased-array transceiver IC with concurrent dual polarized beams and 1.4 degree beam-steering resolution for 5G communication," in Proc. IEEE Int. Solid-State Circuits Conf., Feb. 2017, pp. 128–129.
- [20] K. F. Warnick et al., "Minimizing the noise penalty due to mutual coupling for a receiving array," *IEEE Trans. Antennas Propag.*, vol. 57, no. 6, pp. 1634–1644, Jun. 2009.
- [21] L. Belostotski et al., "Low-noise amplifier design considerations for use in antenna arrays," *IEEE Trans. Antennas Propag.*, vol. 63, no. 6, pp. 2508–2520, Jun. 2015.

- [22] H.-S. Lui et al., "A note on the mutual-coupling problems in transmitting and receiving antenna arrays," *IEEE Antennas Propag. Mag.*, vol. 51, no. 5, pp. 171–176, Oct. 2009.
- [23] S. M. Perera *et al.*, "Wideband N-beam arrays using low-complexity algorithms and mixed-signal integrated circuits," *IEEE J. Sel. Topics Signal Process.*, to be published.
- [24] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, Apr. 1965.
- [25] X. Zhang et al., "DFT spread generalized multi-carrier scheme for broadband mobile communications," in Proc. Int. Symp. Pers., Indoor Mobile Radio Commun., Sep. 2006, pp. 1–5.
- [26] P. Xia et al., "DFT structured codebook design with finite alphabet for high speed wireless communication," in Proc. 6th IEEE Consum. Commun. Netw. Conf. (CCNC), Jan. 2009, pp. 1–5.
- [27] G. Berardinelli et al., "On the potential of zero-tail DFT-spread-OFDM in 5G networks," in Proc. 80th IEEE Veh. Technol. Conf. (VTC Fall), Sep. 2014, pp. 1–6.
- [28] S. Gupta, "An adaptive and efficient data delivery scheme for DFT-MSNs (delay and disruption tolerant mobile sensor networks)," in *Proc. Int. Conf. Adv. Eng., Sci. Manage. (ICAESM)*, Mar. 2012, pp. 99–104.
- [29] Y. Wang *et al.*, "Protocol design and optimization for delay/fault-tolerant mobile sensor networks," in *Proc. 27th Int. Conf. Distrib. Comput. Syst.* (*ICDCS*), Jun. 2007, p. 7.
- [30] T. Hunziker et al., "Spectrum sensing in cognitive radios: Design of DFT filter banks achieving maximal time-frequency resolution," in Proc. 8th Int. Conf. Inf., Commun. Signal Process. (ICICS), Dec. 2011, pp. 1–5.
- [31] Y. Wang et al., "Generalized DFT waveforms for MIMO radar," in Proc. 7th IEEE Sensor Array Multichannel Signal Process. Workshop (SAM), Jun. 2012, pp. 301–304.
- [32] M. Sezgin et al., "A novel DFT/RDFT based subband representation for the fusion of remote sensing images," in Proc. 2nd Int. Conf. Recent Adv. Space Technol. (RAST), Jun. 2005, pp. 611–616.
- [33] J. Li *et al.*, "Robust multiple watermarks for medical image based on DWT and DFT," in *Proc. Int. Conf. Comput. Sci. Converg. Inf. Technol.* (*ICCIT*), Nov./Dec. 2011, pp. 895–899.
- [34] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2009.
- [35] A. F. Molisch et al., "DFT-based hybrid antenna selection schemes for spatially correlated MIMO channels," in Proc. 14th IEEE Pers., Indoor Mobile Radio Commun. (PIMRC), vol. 2. Sep. 2003, pp. 1119–1123.
- [36] L. Dong et al., "The research for effects of window functions in radio astronomy," in Proc. 3rd Int. Congr. Image Signal Process. (CISP), vol. 7. Oct. 2010, pp. 3064–3073.
- [37] K. Lengwehasatit and A. Ortega, "Scalable variable complexity approximate forward DCT," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 11, pp. 1236–1248, Nov. 2004.
- [38] S. M. McDonnell *et al.*, "Compensation and calibration techniques for current-steering DACs," *IEEE Circuits Syst. Mag.*, vol. 17, no. 2, pp. 4–26, 2nd Quart., 2017.
- [39] R. E. Blahut, Fast Algorithms for Digital Signal Processing. Cambridge, U.K.: Cambridge Univ. Press, 2010.
- [40] R. Sarpeshkar, "Analog versus digital: Extrapolating from electronics to neurobiology," *Neural Comput.*, vol. 10, no. 7, pp. 1601–1638, 1998.

- [41] N. J. Higham, Functions of Matrices: Theory and Computation (Other Titles in Applied Mathematics). Philadelphia, PA, USA: SIAM, 2008, ch. 1, pp. 1–34.
- [42] D. Suarez et al., "Multi-beam RF aperture using multiplierless FFT approximation," Electron. Lett., vol. 50, no. 24, pp. 1788–1790, 2014.
- [43] C. J. Tablada et al., "A class of DCT approximations based on the Feig-Winograd algorithm," Signal Process., vol. 113, pp. 38–51, 2015.
- [44] V. Britanak et al., Discrete Cosine and Sine Transforms. San Francisco, CA, USA: Academic, 2007.
- [45] V. Ariyarathna et al., "Multi-beam 4 GHz microwave apertures using current-mode DFT approximation on 65 nm CMOS," in Proc. IEEE Int. Microw. Symp. (IMS), May 2015, pp. 1–4.
- [46] B. N. Flury and W. Gautschi, "An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form," SIAM J. Sci. Stat. Comput., vol. 7, no. 1, pp. 169–184, Jan. 1986.
- [47] T. I. Haweel, "A new square wave transform based on the DCT," Signal Process., vol. 81, no. 11, pp. 2309–2319, Nov. 2001.
- [48] An Analog FFT Beamformer for Acoustic Applications, Office Naval Res., Arlington, VA, USA, Mar. 1978.
- [49] M. Lehne and S. Raman, "An analog/mixed-signal FFT processor for wideband OFDM systems," in *Proc. IEEE Sarnoff Symp.*, Mar. 2006, pp. 1–4.
- [50] M. Lehne and S. Raman, "A prototype analog/mixed-signal fast Fourier transform processor IC for OFDM receivers," in *Proc. IEEE Radio Wireless Symp.*, Jan. 2008, pp. 803–806.
- [51] E. Afshari et al., "Ultrafast analog Fourier transform using 2-D LC lattice," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 8, pp. 2332–2343, Sep. 2008.
- [52] N. Sadeghi et al., "Analog DFT processors for OFDM receivers: Circuit mismatch and system performance analysis," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 9, pp. 2123–2131, Sep. 2009.
- [53] A. Farahmand and M. R. Zahabi, "An energy efficient, high speed analog FFT processor for MB-OFDM UWB receivers," in *Proc. Int. Congr. Technol.*, Commun. Knowl. (ICTCK), Nov. 2014, pp. 1–6.
- [54] FreePDK45:Contents. Accessed: Mar. 21, 2018. [Online]. Available: https://www.eda.ncsu.edu/wiki/FreePDK45:Contents
- [55] ADC Performance Survey 1997–2017 (ISSCC & VLSI Symposium). Accessed: Mar. 21, 2018. [Online]. Available: https://web.stanford.edu/murmann/adcsurvey.html
- [56] K. D. Choo et al., "Area-efficient 1GS/s 6b SAR ADC with charge-injection-cell-based DAC," in IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, vol. 59. Jan./Feb. 2016, pp. 460–461.
- [57] M. Jridi et al., "A generalized algorithm and reconfigurable architecture for efficient and scalable orthogonal approximation of DCT," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 2, pp. 449–457, Feb. 2015.

Authors' photos and biographies not available at time of publication.