Generator-Based Broadband Analog Baseband Receivers for Massive MIMO Arrays

Ethan Chou

Electrical Engineering and Computer Sciences
University of California, Berkeley

December 1, 2021

Generator-Based Broadband Analog Baseband Receivers for Massive MIMO Arrays

by

Ethan Chou

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Science, Plan II

in

Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ali M. Niknejad, Chair
Professor Borivoje Nikolic

Fall 2020

Generator-Based Broadband Analog Baseband Receivers for Massive MIMO Arrays

by Ethan Chou

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

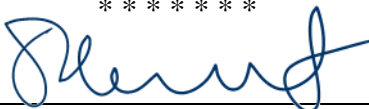Approval for the Report and Comprehensive Examination:

Committee:

Professor Ali M. Niknejad
Research Advisor

12/16/2020

(Date)

* * * * * * *

Professor Borivoje Nikolic
Second Reader

12/18/2020

(Date)

Abstract

Generator-Based Broadband Analog Baseband Receivers for Massive MIMO Arrays

by

Ethan Chou

Master of Science, Plan II in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Ali M. Niknejad, Chair

Massive multiple-input multiple-output (MIMO) wireless communications systems operating at millimeter-Wave (mm-Wave) promise to be an enabling technology for high-capacity, next-generation mobile networks. This work describes the design of an integrated high-bandwidth analog baseband section receiver for the receive chain of such systems, and utilizes several circuit techniques to address the challenges and issues encountered at such bandwidths. By leveraging the Berkeley Analog Generator (BAG) framework, the layout design process for the analog baseband can be captured for efficient portability to different CMOS technologies. To demonstrate the concept of the generator-based design, iterations in 28nm bulk CMOS and 16nm FinFET have been produced, while maintaining performance comparable to the reported state-of-the-art. A test chip of the 28nm iteration has been fabricated, projected to achieve variable gain from 3 to 39 dB and bandwidth of 2.5 GHz set by fourth-order filtering while consuming 16 mW from a 1 V supply, according to layout extracted simulations.

To Mom and Dad

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Objective

To address the rapidly growing demand for mobile data consumption with new wireless applications like autonomous driving and virtual reality on the rise, wireless channel capacity needs to increase dramatically. As a promising solution to this issue, the massive MIMO paradigm, in which a large number of channel elements spatially multiplex user data to increase spectral efficiency, has gained much attention over the years as the enabling technology of the next-generation mobile wireless networks [17]. Moreover, implementations of massive MIMO systems at mm-Wave E-band (60-90 GHz) carrier frequencies allows for large arrays of transceivers to have a compact form factor for deployment in a base station, owing to the reduced size of mm-Wave antennas. The widely available spectrum at E-band, which allows the use of high channel bandwidths to maximize network capacity, further increases the advantage of mm-Wave implementations [21].

A large-scale effort at UC Berkeley to demonstrate such a massive MIMO system with a 128-antenna array serving 16 simultaneous users in the E-band, termed "Hydra", is currently in progress at the time of this work. The application-specific integrated circuits (ASIC) that make up the custom uplink front-end include the analog baseband (ABB) circuitry, traditionally comprised of the variable gain amplifier (VGA) and low pass filter (LPF), whose main purpose is signal conditioning to maximize the analog-digital converter's (ADC) dynamic range. This work focuses on the design of the ABB for the Hydra receive chain, but can also be generalized to use in other similar receive chains. Such ABB's for mm-Wave receive arrays highlight unique design challenges compared those for lower frequency sub-6GHz single channel receivers.

- **Bandwidth, Noise, and Linearity:** Though dependent on the technology transit frequency f_T, obtaining broadband gain at continuously higher bandwidths can quickly consume enormous amounts of power, while clever design choices can alleviate this power increase. Note that this refers to broadband gain, where the bandwidth is

determined by an RC product and thus ideally extends to DC, as opposed to passband gain, where the bandwidth is determined by a resonant LC tank and is narrowband around the resonant frequency.

Noise requirements are more difficult to meet at higher bandwidths, as the integrated output noise must aggregate a greater amount of the noise spectrum within the signal bandwidth. High linearity is also typically more difficult to achieve as higher frequency signals excite more nonlinear distortion, and open loop structures are generally preferred which are inherently more nonlinear.

- **Power, Area, and Variability:** As with any per-channel element, the power and area of a single analog baseband element is crucial as the total power and area consumption will be that of a single element multiplied by the number of channels, magnifying any design inefficiencies. However, additional challenges arise when considering integration with the RF front-end blocks like the low noise amplifier (LNA) or downconversion mixers, which must operate at mm-Wave with low noise and high linearity requirements. These blocks also liberally use area-intensive on-chip inductors, transformers, and transmission lines for high frequency and bandwidth operation. Thus, they consume a large fraction of the overall receiver power and area budget, leaving limited amounts for the analog baseband circuits.

  Simply having a larger number of array elements also heightens the importance of variability from process, voltage, and temperature (PVT) as there are more elements and thus more sources of variation. PVT-sensitive circuit parameters of the analog baseband such as DC offset and cutoff frequency must include proper calibration circuitry to tune these parameters to nominally desired values.

For an analog design at these frequencies, layout parasitics can have significant impact on the performance, resulting in multiple passes of layout modifications and extraction. As CMOS technology scales, the design rules become more complex, and thus the design effort during this process can increases significantly. By the same reasons, porting an existing design from one technology can also take significant effort. BAG aims to address this issue by providing a framework integrated with Cadence EDA tools to script technology-agnostic schematic and layout generators [2]. The generator script captures the design process in a generic fashion such that the same design can be generated in other technologies, ensured to be compliant with design rules, with a set of specifications as inputs. This circuit generator-based approach with BAG is extensively used in the design of the ABB. This brings fast layout modifications to access extracted performance, internalizes the complex design rules away from the designer, and provides relative ease of porting the design to another technology.

The objective of this work is to demonstrate a low power yet high performance ABB imple-

mentation for mm-Wave massive MIMO receivers, whose layout design can be captured in a generator for rapid and efficient porting in various deep sub-micron CMOS technologies. Along with the generators for the ADC and digital signal processing, a fully technology-agnostic and multi-channel baseband receive ASIC can then be generated in multiple technologies with different specifications, and be capable of targeting various applications.

1.2 Thesis Organization

This work is organized as follows: In Chapter 2, the architectural design decisions of the Hydra massive MIMO receive module is briefly discussed, including the mm-Wave front-end, baseband data conversion and processing, and digital back-end. Chapter 3 derives the specifications for the ABB, and analyzes several high-bandwidth amplification and filtering techniques, many of which are incorporated into the design. Several linearity enhancement techniques for the open-loop structures found in high frequency baseband circuits are also briefly addressed. The circuit design is discussed in detail in Chapter 4, culminating in a complete prototype ABB design iteration in 28nm CMOS. The layout extracted simulations concerning bandwidth, gain, noise, and linearity are also presented. Chapter 5 shifts the focus from circuit design to modular layout design of the ABB that can be captured by a generator, specifically BAG, for rapid migration of the layout from one technology to another. To demonstrate the practicality of this concept, a design iteration in 16nm FinFET is generated and integrated into a complete 16-channel baseband ASIC. Finally, Chapter 6 compares the two ABB design iterations to the state-of-the-art and concludes this work.

Chapter 2

Massive MIMO Receive Chain

2.1 Massive MIMO Architecture

Generally speaking, massive MIMO systems employ a number of access point transceivers (M) much greater than the number of users (K). When $M \gg K$, simple linear beamforming algorithms can be used to achieve nearly optimal user tracking and inter-user interference cancellation. To accomplish this, a centralized architecture can be used, whose power consumption is directly influenced by interconnect bandwidth and computational requirements. The distributed architecture used for the Hydra system is shown in Fig. 2.1, where C modules perform maximal-ratio combining (MRC) beam-forming across their respective P antennas, reducing data order to K. Data is accumulated among neighboring modules in sequence until the final distributed module sends data to a central element for post-processing. Data interconnect scales with K, and dimension of matrix multiplication operations scales with the greater of K or P. As previously mentioned, in a massive MIMO system, $M \gg K$ and so P may be made arbitrarily smaller than M based on the number of modules. This relaxes interconnect and computational requirements with significant power savings [16].

A discrete implementation of Hydra reported in [16] demonstrates a prototype array using a distributed architecture consisting of 20 base station antennas operating at 75 GHz with 250 MHz of channel bandwidth, validating a full MIMO uplink signal processing chain. Implemented with custom radio boards, off-the shelf mm-Wave radio components, and FPGA's, the array is large and power hungry.

The Hydra Head and Hydra spine ASIC's are components of a low-power integrated solution that can support higher bandwidths of 2.5 GHz. The Head ASIC contains the RF front-end and baseband beamforming, while the Spine ASIC contains the ADC's, digital baseband signal processing, and wireline links. Specifically, the Spine module is intended to act as a distributed beamforming interconnect that provides data converter interfaces to multiple analog sub-arrays. This allows the sub-arrays to be combined in the digital domain to

realize larger effective arrays. The interface between the Head and Spine modules consists of 32 analog differential inputs and 32 analog differential outputs, providing 16 channels of I/Q baseband for receive. Together, both modules as a receive chain can support full aperture hybrid beamforming or solely digital beamforming. For full aperture hybrid beamforming, the mm-Wave front-ends perform an initial stage of beamforming, so that each analog interface between the Head and Spine modules corresponds to a different beam. Multiple spines can then be daisy-chained to combine the beams from multiple front-ends, expanding the aperture of the array. For digital beamforming, the Head module simply performs downconversion to baseband and the Spine module performs a stage of beamforming in the digital domain before being daisy-chained with other Head/Spine sub-arrays to realize larger arrays. In either mode, a Tail module then performs centralized baseband processing and user interference cancellation on the aggregate outputs of all the daisy-chained sub-arrays. The chip boundaries of the system are shown in Fig. 2.1.
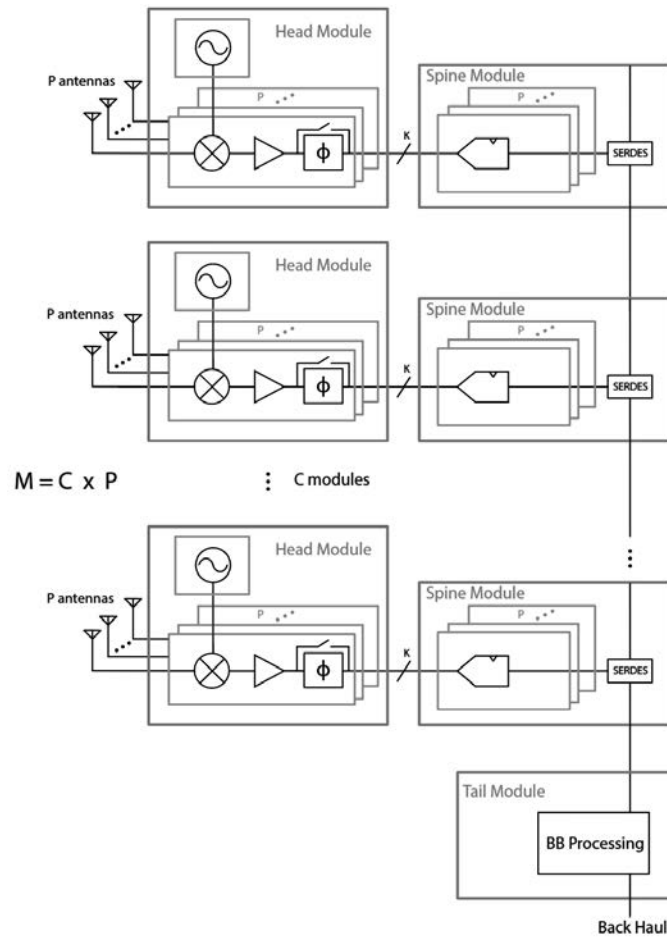

Figure 2.1: Distributed architecture of the Hydra massive MIMO uplink system.

2.2 Mm-Wave Front-End

The mm-Wave front-end receiver (RX), intended for operation over E-band, poses several design challenges for massive MIMO applications and CMOS implementations. First, the use of RF phase shifting and combining to filter out-of-band (OOB) interferers before down-conversion for high out-of-band linearity, as commonly seen in single-beam phased arrays, would require $M \times K$ mm-Wave phase shifters. This would add a significant power and area overhead, so instead baseband phase shifting and combining with analog or digital Cartesian beamformers is more suitable, as shown in Fig. 2.2(a). Second, correlated noise averaging across the RX array provides a $10log(M)$ improvement in the signal-to-noise ratio (SNR) relative to a single-element RX, which implies a relaxed noise figure (NF) requirement. However, the number of RX array elements M also mandates that the power consumption and area of each element be minimized.

Conventional mm-Wave RX's employ multi-stage LNA's that are capacitively or magnetically coupled for wideband and low-NF operation. However, these interstage passive networks occupy significant area, as well as introduce high loss at mm-Wave, which requires greater power to compensate for these signal losses to achieve a certain amount of gain. Because of these area and power penalties, and the fact that a relaxed NF can be tolerated, a direct-conversion, passive mixer-first RX is chosen, which minimizes the number of mm-Wave stages in the signal path. Mixer-first RX's, while commonly used in sub-6GHz applications, have their own design challenges, especially at mm-Wave. The most significant is that to minimize NF and maximize out-of-band blocker rejection, the on-resistance of the mixer switches R_{sw} needs to be much less than the matched source resistance R_s, or $R_{sw} \ll R_s$. This implies the use of wide mixer switches that require significant power in the LO buffers to provide sufficient swing.

The proposed mixer-first RX uses two techniques to provide wideband matching to the 50Ω antenna while simultaneously presenting a small capacitive load to the LO buffer. First, the mixer switches are purposely downsized to present a high input impedance $Z_{in,mix} > R_s$, which can be matched to the antenna using a wideband impedance transformation network consisting of an L-match and input shunt resonator. The shunt resonator neutralizes the pad capacitance and improves the matching bandwidth. Another benefit is that the matching network also provides passive voltage gain that amplifies the signal before entering the noisy passive mixer, which compensates for the higher noise due to the increased R_{sw}. Second, the mixer switches can be further downsized without increasing $Z_{in,mix}$ by introducing frequency-translational negative feedback between the open-loop baseband amplifier outputs and the RF input, as shown in Fig. 2.2(b). If the auxiliary feedback mixer switches are sized to obtain a loop gain of ≈ 1 (which also eliminates stability concerns), $Z_{in,mix}$ can be decreased by a factor of 2 within the loop bandwidth. The gain of the baseband amplifier is also part of the loop, so the auxiliary feedback switches can be considerably downsized compared to the mixer, and thus contribute to only a small capacitance overhead. By leveraging these

two techniques, a prototpye in 28nm CMOS achieves an $S11 < 10$ dB and $NF < 10$ dB, over the entire mm-wave E-band, while consuming only 12 mW and occupying 0.085 mm². The complete analysis, circuit implementation, and measurement results of the front-end can be found in [14], [15].
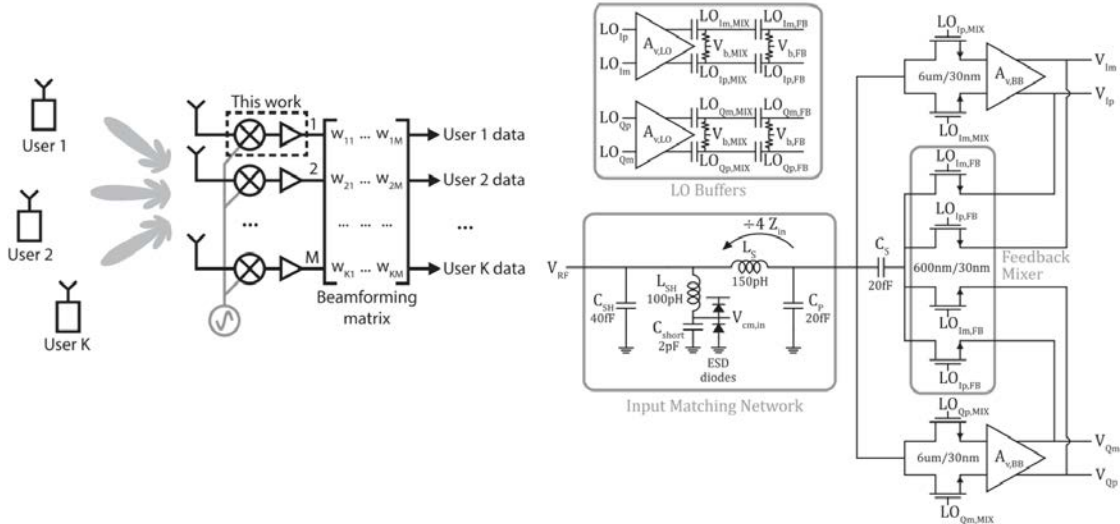
Figure 2.2: (a) Massive MIMO RX front-end system with baseband beamforming [15] and (b) RX front-end schematic [14].

2.3 Analog Baseband

The analog baseband circuitry of the RX chain is shared between the Head and Spine modules. The Head module contains the analog beamforming processing (with a bypass option if only digital beamforming is desired) consisting of phase shifters and summations, while the Spine module contains the traditional signal conditioning circuitry to properly drive the ADC. Both basebands of each module contain some amount of variable-gain amplification and filtering to mitigate interference. The ADC consists of a 6-bit ENOB, 5 GS/s, 8-slice time-interleaved SAR (TISAR) ADC. The analog baseband circuitry of the Spine ASIC, shown in Fig. 2.3, is the focus of this work.

2.4 Digital Baseband

The digital baseband for each channel consists of a root-raised-cosine FIR filter to prevent aliasing of of out-of-band blockers and noise. This is followed by DC offset correction and
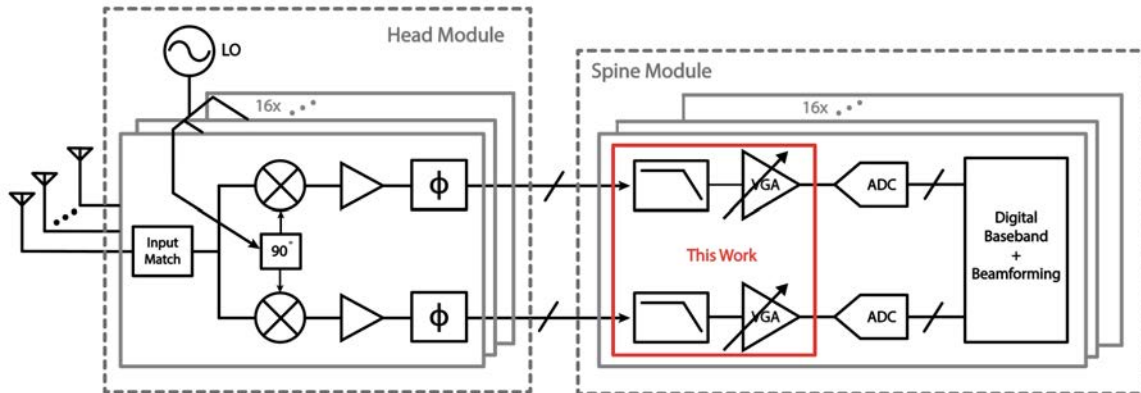
Figure 2.3: Block diagram of one module of the Head/Spine RX chain.

I/Q correction, intended to remove errors from the base station antenna that they follow. The sub-array conjugate beam-forming coefficients are determined by a Golay correlator that extracts magnitude, timing, and phase information about the effective channel by finding the peak cursor in the time domain [16]. Finally, for each beam, timing deskew, decimation, and distribution are performed before the Spine module output is serialized. The serializer/deserializer (SERDES) circuitry supports 16 streams at 8 bit samples (4 bit I, 4 bit Q). These outputs can then be daisy-chained with neighboring Head/Spine sub-arrays.

Chapter 3

Broadband Analog Baseband Sections

3.1 Design Considerations

To derive the specifications of the analog baseband, the entire RX chain's desired performance must be partitioned with the front-end and ADC performance. As stated in Section 2.3, the 6-bit ENOB, 5 GS/s TISAR ADC (dynamic range DR $\approx$ 32 to 38 dB) has a full scale voltage of V_{FS}=400 mV differential peak-to-peak. As mentioned in section 2.2, the front-end achieves NF=10 dB.

- **Gain:** The maximum gain of the RX chain can then be calculated as

$$A_{v,max} = \frac{\overline{v_{IRN,50\Omega}}}{\overline{v_{n,ADC}}} = \frac{-174dBm + 10log_{10}(BW) + NF}{V_{FS}/DR_{5bits}} \approx 36dB \qquad (3.1)$$

  Accounting for safety margins concerning higher NF and to get the RX noise 6dB above the ADC noise floor increases the total gain requirement to 48 dB. Since the front-end provides 24 dB of gain, the maximum gain of the ABB is set to be 24 dB. The minimum gain is determined by the largest swing that the front-end may present to the analog baseband, and in this case is set to 15 dB.

- **Bandwidth:** The desired nominal baseband bandwidth is 2.5 GHz, set by a second-order pole roll-off. Of course, a higher-order pole is more desirable to achieve greater attenuation of out-of-band blockers. By using switched passives, the bandwidth should also be programmable for different data rate applications and calibration over PVT. The sampling frequency of the Nyquist-rate ADC is consistent with the desired baseband bandwidth of 2.5 GHz.

- **Noise:** Because of the high gain in the front-end, the noise performance of the RX is dominated by the front-end. Recall that the cascaded NF, assuming all the stages are impedance matched, is given by

$$NF = NF_1 + \frac{NF_2 - 1}{A_1} + \frac{NF_3 - 1}{A_1 A_2} + \frac{NF_4 - 1}{A_1 A_2 A_3} + ... \qquad (3.2)$$

where all the NF terms are in linear noise *factor* units. Because the interface between the front-end and ABB is matched, it can be shown with the following that the ABB noise performance is rather insignificant to the total RX NF.

$$NF_{RX} = NF_{FE} + \frac{NF_{ABB} - 1}{A_{v,FE}} \tag{3.3}$$

Thus, the noise specifications for the ABB will be constrained in the scenario of driving the ADC in a standalone setup. To ease the noise requirements in this mode, the ADC is intended to operate with a ENOB of 5 instead of 6 bits, which entails a DR of approximately 32 dB (given by $DR = 6.02 \cdot ENOB + 1.76$ dB, assuming a full-scale input). In that case, the maximum output integrated noise should be

$$DR = \frac{(\frac{1}{2}\frac{1}{\sqrt{2}}V_{FS})^2}{\overline{v_{o,n}}^2} \Rightarrow \overline{v_{o,n}}^2 = 2 \times 10^{-5}V^2 \tag{3.4}$$

Assuming the ADC noise dominates, this corresponds to a NF given by

$$NF = 1 + \frac{\overline{v_{in,amp}}^2}{\overline{v_{Rs}}^2} = 1 + \frac{\overline{v_{o,n}}^2/A_v{}^2}{4kTR_s \cdot BW} \tag{3.5}$$

which yields an NF of 14.1 to 23.0 dB for the maximum and minimum gain settings, respectively.

- **1 dB Compression Point:** If driving the ADC in a standalone setup, the output 1-dB compression point (OCP1dB) must exceed $V_{FS} \approx -4$ dBm (for a 50Ω system) to ensure the ADC operates with maximum dynamic range. Neglecting second-order interaction, recall the cascaded output compression point is given by

$$\frac{1}{P1dB} = \frac{1}{P1dB_1 A_2 A_3} + \frac{1}{P1dB_2 A_3} + \frac{1}{P1dB_3} + ... \tag{3.6}$$

where the compression terms should be in linear voltage units and the input compression point for any stage can be obtained by simply dividing by its respective gain. If in cascade with the front-end, the input 1-dB compression point (ICP1dB) can be calculated as

$$\frac{1}{V_{FS}} = \frac{1}{ICP1dB_{FE} \cdot A_{v,FE} \cdot A_{v,BB}} + \frac{1}{ICP1dB_{BB} \cdot A_{v,BB}} \tag{3.7}$$

which yields a $OCP1dB = A_{v,ABB} \cdot ICP1dB_{ABB} \approx V_{FS}$, indicating that the in-band ABB linearity dominates the overall RX in-band linearity, as expected.

- **DC Offset:** With the relatively high gain present in ABB's, DC offset can easily saturate the RX chain. Moreover, AC coupling is usually not tolerable in direct-conversion receivers due to the signal content present at low frequencies (down to

10kHz for this receiver). Several designs use servo feedback loops with a LPF to extract the DC offset [26], [25], [1]. However, this solution still presents issues with a high-pass pole attenuating signal content and settling time.

In this design, the signal path is DC-coupled to prevent any loss of information in the signal band down to approximately 10kHz, and also eliminates the need for large area-intensive passives and the impact of capacitor bottom plate parasitics on the signal path. Furthermore, a high-pass corner cutoff frequency of 10kHz would potentially require impractically large on-chip resistor or capacitor values. However, the elimination of AC-coupling means DC offset can accumulate along the signal path and eventually saturate the ABB, leading to significant compression, clipping, and degradation of the ADC dynamic range. In this design, static digital calibration is used to correct the DC offset at various points along the signal path, targeting < 2 LSB's (≈ 6 mV) of offset at the output for negligible impact on the ADC DR.

- **Power Consumption and Area:** One of the challenges faced by designs for massive MIMO applications is the critical need to minimize the power consumption and area, as these quantities will be multiplied by the number of array elements. This does not lend well to area-intensive peaking inductors or large coupling capacitors. To first order, the power consumption of amplifiers will be determined by the need for bandwidth or low noise, while linearity and gain are generally more dependent on topology, technology parameters, channel length, or bias points. As a simplified first-pass starting point, the minimum bias currents for bandwidth and noise limited differential pair common source amplifiers can be derived as

$$I_{bw} = \frac{1}{2}\frac{A_v\omega_{bw}V_{ov}C_L}{1 - A_v\omega_{bw}/(\omega_T/\eta)}, I_{noise} = \frac{8kT(1 + \gamma A_v)f_{bw}SNR_{min}}{A_vV_{sig}}\frac{V_{i,max}}{V_{sig}} \qquad (3.8)$$

where $\eta = C_{dd}/C_{gg}$, γ is the MOS channel noise factor, V_{ov} is the overdrive voltage, ω_T is the unity gain frequency, V_{sig} is the input signal, and $V_{i,max}$ is the largest tolerable input signal for SNR purposes.

Based on the top-level floorplan and power consumption budgeting for the Hydra Spine chip, the ABB is targeted to consume at most 20 mW from a 0.9 - 1 V supply and occupy 0.05 mm^2.

Number of Stages

To obtain higher gain-bandwidth (GBW) than what is reasonably obtainable with a single stage amplifier, an n-stage cascaded architecture is commonly used. Consider a cascade of n identical gain cells each with a bandwidth BW_c. It is derived in [18] that the overall bandwidth of the n-stage cascaded amplifier is given by

$$BW_{tot} = BW_c \sqrt[m]{2^{1/n} - 1} \qquad (3.9)$$

DC Gain	15 - 24	dB
3-dB Bandwidth	2.5	GHz
Noise Figure	14.1 - 23.0	dB
Output $CP1dB$	-4	dBm
Max Power	20	mW
Area	0.05	mm^2

Table 3.1: Performance specifications

where m is equal to 2 for a first-order gain cell and m is equal to 4 for a second-order gain cell. Thus, the overall bandwidth decreases by a factor of $\sqrt[m]{2^{1/n} - 1}$. For an overall gain of A_{tot}, the gain of each cell needs to be $A_{tot}^{1/n}$. Since $GBW_c = A_{tot}^{1/n}BW_c$ and $GBW_{tot} = A_{tot}BW_{tot}$, the required GBW of each cell can then be derived as

$$GBW_c = \frac{GBW_{tot}}{A_{tot}^{1-1/n}\sqrt[m]{2^{1/n} - 1}} \tag{3.10}$$

showing that for a fixed, required total GBW, as the number of stages n increases, the GBW required for the cell decreases, relaxing the power consumption or design difficulty of each stage. This trend in shown in Fig. 3.1 for A_{tot} = 30 dB and BW_{tot} =2.5 GHz. This makes intuitive sense because while the overall bandwidth decreases by a factor of $\frac{\sqrt[m]{2^{(1/(n+1)-1}}}{\sqrt[m]{2^{(1/n)-1}}} < 1$ with each additional stage, the overall gain increases by a factor of $A_{tot}^{1/n} > 1$, so the overall GBW experiences a net increase. However, because each stage has lower gain, the noise of each stage accumulates more rapidly, placing an upper bound on the number of stages for reasonable noise performance, below approximately 5 [26].

3.2   Broadband Amplifier Techniques

Cherry-Hooper Amplifier

The Cherry-Hooper amplifier [3], shown in Fig. 3.2(a), is widely used for wideband VGA designs ([26], [1]). At its core, the Cherry-Hooper amplifier consists of a transconductor G_{m1} (*series-series* feedback), followed by a transimpedance amplifier (TIA) consisting of another transconductor G_{m2} with R_f in feedback (*shunt-shunt* feedback). At first glance, by leveraging the method of open-circuit time constants and feedback analysis on input and output resistances, the poles at nodes 1 and 2 can be roughly estimated as

$$\omega_1 \approx \frac{1}{R_{in,2}C_1} = \frac{1}{(\frac{R_f+R_{o2}}{1+G_{m2}R_{o2}})C_1} \approx G_{m2}/C_1 \tag{3.11}$$
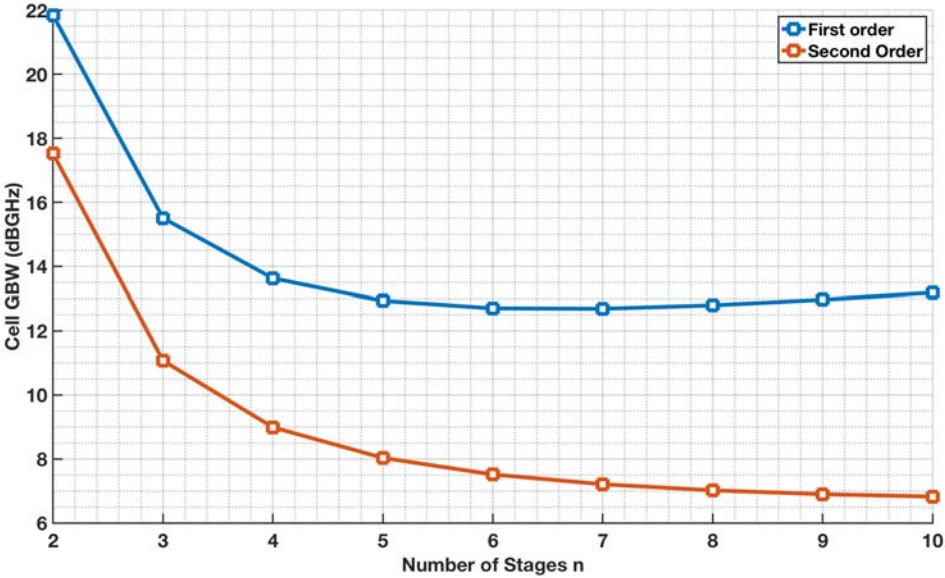
Figure 3.1: Cell GBW vs. number of stages for $A_{tot} = 30$ dB and $BW_{tot} = 2.5$ GHz.
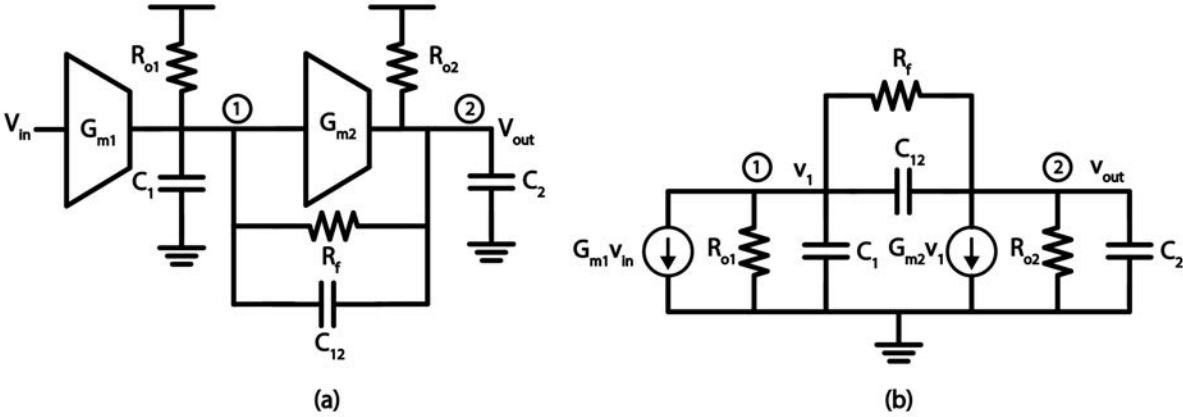

Figure 3.2: (a) Generalized schematic of the Cherry-Hooper amplifier and b) equivalent small-signal model used to calculate the transfer function.

and

$$\omega_2 \approx \frac{1}{R_{out,2}C_2} = \frac{1}{(\frac{R_{o2}}{1+G_{m2}R_{o2}})C_2} \approx G_{m2}/C_2 \tag{3.12}$$

where C_1 and C_2 are the total capacitances at nodes 1 and 2, respectively, and it assumed $R_f \ll R_{o2}$. Both of these poles are typically at higher frequency than the output pole of a common source amplifier. However, notice that at high frequencies C_2 shunts the output of G_{m2}, which decreases the loop gain of the TIA and affects the closed loop resistances, thus changing the pole locations. To accurately determine the bandwidth, the equivalent small signal model, shown in Fig. 3.2(b) can be used to derive the transfer function [26]:

$$A_v(s) = A_{v0}\frac{1 - s(C_{12}/G_{m2})}{s^2B(R_f/G_{m2}) + sD + 1} \tag{3.13}$$

$$B = C_1C_{1,2} + C_1C_2 + C_{1,2}C_2 \tag{3.14}$$

$$D = C_{1,2}R_f + C_1\frac{R_f + R_2}{G_{m2}R_2} + \frac{C_2}{G_{m2}} \tag{3.15}$$

where $A_{v0} = G_{m1}R_f$ is the total DC gain if $G_{m2}R_f \gg 1$, $G_{m2}R_2 \gg 1$, $R_2 \gg R_f$, and $R_1 \gg 1/G_{m2}$. To first order, the dominant pole is $\approx 1/D$. If $C_{1,2} \ll C_1, C_2$ and $R_2 \gg R_f$, the pole locations reduce to the preliminary estimates of $\omega_1 = G_{m2}/C_1$ and $\omega_2 = G_{m2}/C_2$. Thus, the Cherry-Hooper amplifier has the advantage of greater bandwidth than common source amplifiers by moving the poles to higher frequencies without a significant decrease in gain. This implies that the gain and bandwidth are decoupled and can be maximized separately to first order, in contrast to the typical gain-bandwidth tradeoff of a single pole amplifier. The disadvantage is that because the Cherry-Hooper amplifier is typically in a stacked configuration as shown in Fig. 3.3(a), headroom issues quickly become pronounced with low supplies, especially if the voltage drop across the R_f is significant. To circumvent this issue, the differential pairs can equivalently be split into two explicit stages as shown in 3.3(b), at the penalty of greater C_1 from parasitics and typically less signal current transfer from the Gm to the TIA. Common-mode feedback (CMFB) can also be employed to ensure little DC current and thus voltage drop across R_f.

It is interesting to note that an optimal R_f value exists to maximize the GBW of the Cherry-Hooper cell. This is because the G_m stage is loaded by the Miller capacitor of the C_{gd} of the TIA stage input devices, an equivalent capacitance of $C_{Miller} = G_{m,TIA}R_f \cdot C_{gd}$. Thus, gain increases with R_f while bandwidth decreases, resulting in an intermediate value of R_f that corresponds to a maxima in the gain-bandwidth.

Cascoding

Cascoding, as shown in Fig. 3.4(a), can mitigate the Miller effect on C_{gd} by introducing a low impedance node at the drain of of the active device M_1. The impedance at the intermediate node x is $\approx \frac{1}{g_{m2}}$, and thus the C_{gd} of M_1 undergoes a Miller effect transformation of $C_{gd}(1+$
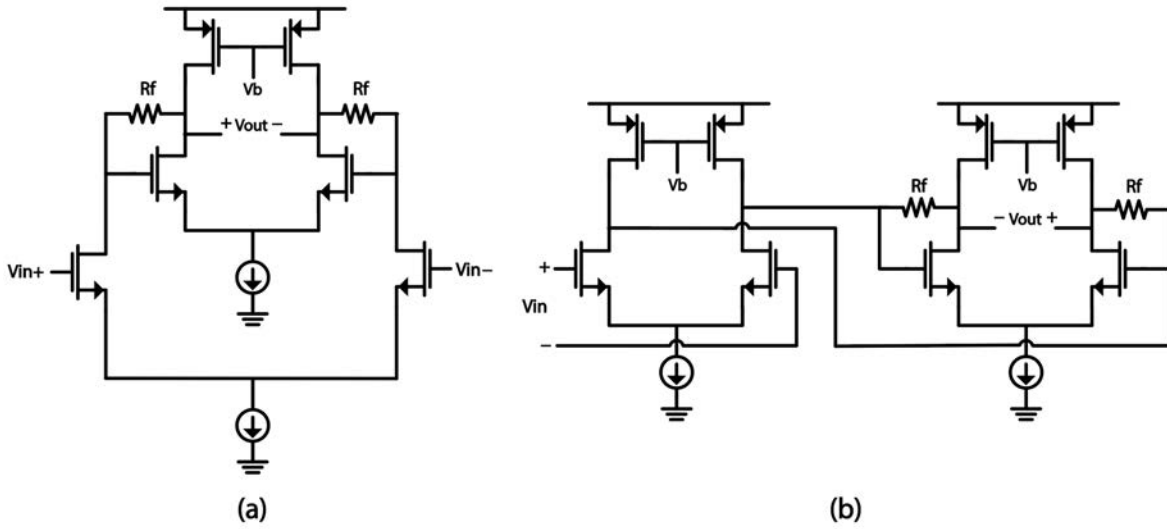
Figure 3.3: (a) Typical implementation of the Cherry-Hooper amplifier that suffers from headroom limitations and (b) implementation more compatible with low supply voltages.
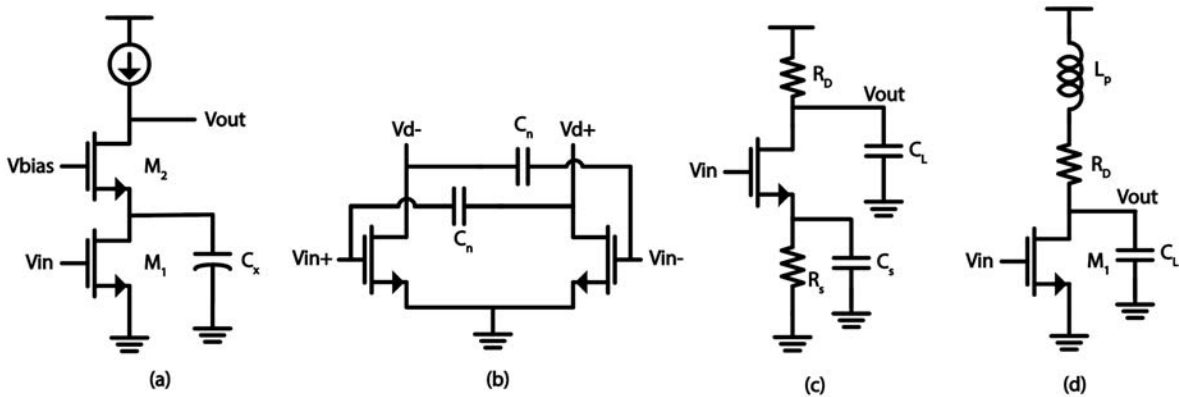

Figure 3.4: (a) Cascoding, (b) capacitive neutralization, (c) capacitive degeneration, and (d) inductive shunt peaking.

$\frac{g_{m1}}{g_{m2}}$) as opposed to $C_{gd}(1 + g_{m1}R_D)$ without cascoding. This can substantially increase the input pole of the amplifier, or equivalently increase the output pole of the preceding stage. An added benefit is that the output resistance is increased to $\approx g_{m2}g_{m1}(r_{o1}||\frac{1}{sC_x})$, which is desirable for a transconductance stage. It also can be shown in [22] that the output noise contribution of the cascode device is negligible within frequencies where the impedance seen looking into the drain of M$_1$, $r_{o1}||\frac{1}{sC_x}$, is much less than the impedance seen looking into the source of M$_2$. However, cascoding presents two drawbacks: (1) the parasitic capacitance C_x can begin shunting the signal current to ground as its impedance decreases and (2) the cascode device takes up valuable headroom in low-supply designs. To limit these drawbacks, special care should be taken in layout to minimize C_x and the cascode device can be biased with a relatively low overdrive voltage.

Capacitive Neutralization

The Miller effect on C_{gd} can also be mitigated with capacitive neutralization, in which cross-coupled capacitors C_n are connected between the non-inverting gate and drain of a differential pair, as shown in Fig. 3.4(b). This results in a Miller capacitance of $C_n(1 - |A_v|) = -C_n(|A_v| - 1)$, where A_v is the *positive* voltage gain between the non-inverting nodes. Thus, the Miller capacitance of $C_{gd}(1 + |A_v|)$ can be effectively cancelled out or reduced by this neutralization capacitor if $A_v > 1$, moving the input pole higher in frequency and increasing the bandwidth. In practice, C_n is implemented by the parasitic capacitance of metal routing, explicit MOM capacitors, or MOS capacitors (depending on the acceptable tracking over PVT and the desired quality factor of C_n) so accurate matching of a nonlinear, bias-dependent capacitance C_{gd} is difficult to achieve. If $C_n > C_{gd}$, this may cause peaking in the frequency response so a conservative value of C_n is typically used. One drawback of this technique is that the output incurs an extra capacitance of $C_n(1 - 1/|A_v|)$, which will decrease the output pole frequency. It is also worth mentioning that in common-mode, C_n simply appears in parallel with C_{gd}, effectively increasing the input-to-output feedforward capacitance and potentially introducing common-mode stability issues.

Capacitive Degeneration

Other methods to achieving greater bandwidth involve introducing additional zeros in the frequency response to cancel the dominant pole, as seen with capacitive degeneration, shown in Fig. 3.4(c). Introducing a capacitance at the active device's source increases the effective G_m at high frequency by shunting out the source resistance that decreases G_m, introducing a zero that can be used to compensate for the gain roll off due to the output pole $\omega_{p1} = 1/(R_DC_L)$. The frequency response of G_m can be expressed as

$$G_m(s) = \frac{g_m(sR_sC_s + 1)}{sR_sC_s + g_mR_s/2 + 1} \tag{3.16}$$

where $w_z = 1/(R_s C_s)$ and $w_{p2} = (1 + g_m \frac{R_s}{2})/(R_s C_s)$. If ω_z cancels ω_{p1}, the new dominant pole is at higher frequency at ω_{p2}. However, imperfect cancellation will result in peaking and distortion in the frequency response, so w_z needs to be carefully placed to ensure the level of peaking is tolerable.

Inductive Shunt Peaking

Inductive shunt peaking also broadbands the frequency response by placing an inductor in series with R_D, or in *shunt* with the output load, shown in Fig. 3.4(d). Intuitively, the impedance of the inductor rises as the impedance of the load capacitor decreases, introducing a zero that can compensate the effect of the output pole. The impedance of the RLC network can be expressed as

$$Z(s) = (sL_p + R_D) || \frac{1}{sC_L} = \frac{R_D[s(L_p/R_D) + 1]}{s^2 L_p C_L + s R_D C_L + 1} \tag{3.17}$$

where the ratio of the zero to original output pole, $m = \frac{R_D C_L}{L_p/R_D} = \frac{R_D{}^2 C_L}{L_p}$, determines the amount of bandwidth extension and peaking. $m = 2$ is commonly chosen as desirable compromise with a bandwidth extension of 80% and a modest normalized peak frequency response of 1.03 [18]. In practice, the implementation of L_p with on chip spiral inductors introduces two drawbacks: (1) the added parasitic capacitance of the inductor and routing can severely limit the bandwidth extension benefit and (2) the large area of inductors typically seen for low-GHz designs leads to an area-intensive layout that can complicate routing and make integration into the entire RX chain layout difficult due to the large footprint. Active inductors have been used to circumvent these drawbacks ([23], [27]) but have their own associated penalties with headroom, noise, transistor parasitics, and the need for above-V_{DD} biasing.
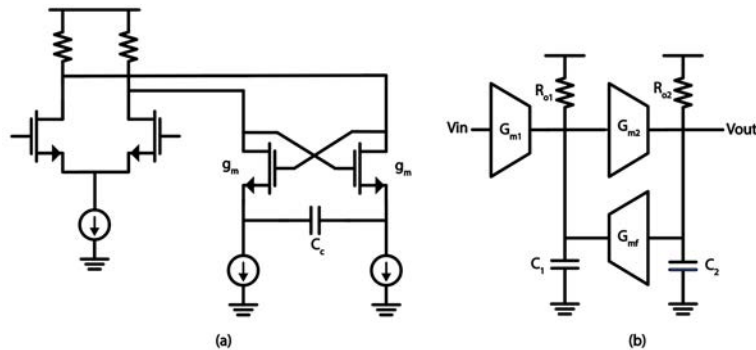

Figure 3.5: (a) Negative impedance conversion applied to a differential pair amplifier and (b) active feedback.

Negative Impedance Conversion

While the previously mentioned broadband techniques required no additional power, active devices consuming additional current can also be used. Negative impedance conversion (NIC) creates a negative capacitance that can be used to cancel the parasitic capacitance at the amplifier output and extend the bandwidth. This is done by using a cross-coupled pair as shown in Fig. 3.5(a). If the C_{gs} of the cross-coupled devices are considered, the admittance presented by the NIC can be expressed as

$$Y_{NIC} = \frac{1 - sC_{gs}/g_m}{(\frac{C_{gs}}{C_c} + 2)\frac{1}{g_m} + \frac{1}{sC_c}} \tag{3.18}$$

and for frequencies well below f_T ($s \ll g_m/C_{gs}$), an impedance of $Z_{NIC} = -(C_{gs}/C_c + 2)/g_m - 1/(sC_c)$ is obtained, which consists of a negative resistance in series with a negative capacitance. The negative series resistance lowers the Q of the negative capacitance and can also increase the gain of the amplifier (as it is in parallel with the amplifier's output resistance).

Active Feedback

The use of negative feedback with an active device to extend the bandwidth is used in [1], [26], [11] and consists of a feedback transconductance cell G_{mf} in feedback around a second stage transconductance cell G_{m2} to feed a portion of the output of G_{m2} back to its input, as shown in Fig. 3.5(b). The transfer function of this two stage amplifier can be expressed as

$$A_v(s) = A_{v0}\frac{1}{s^2A + sB + 1} \tag{3.19}$$

where

$$A_{v0} = \frac{1 + G_{m1}g_{m2}R_{o1}R_{o2}}{G_{mf}G_{m2}R_{o1}R_{o2}}, A = \frac{R_{o1}R_{o2}C_1C_2}{1 + G_{mf}G_{m2}R_{o1}R_{o2}}, B = \frac{R_{o1}C_1 + R_{o2}C_2}{1 + G_{mf}G_{m2}R_{o1}R_{o2}} \tag{3.20}$$

and the 3-dB bandwidth is given by

$$\omega_{-3dB,activeFB} = \sqrt{\frac{1 + G_{mf}G_{m2}R_{o1}R_{o2}}{R_{o1}R_{o2}C_1C_2}} \tag{3.21}$$

Compared to a resistively loaded common source amplifier with $\omega_{-3dB,CS} = 1/(R_DC)$, the bandwidth of the amplifier with active feedback increases by a factor of

$$\frac{\omega_{-3dB,activeFB}}{\omega_{-3dB,CS}} = \sqrt{1 + G_{mf}G_{m2}R_D{}^2} \tag{3.22}$$

assuming $R_{o1} = R_{o2} = R_D$, $C_1 = C_2 = C$ and that the active feedback adds no capacitance to the output nodes of each stage. Like any negative feedback loop with multiple poles, stability may be difficult to ensure especially at high frequency. The feedback loop also adversely loads the signal path.

3.3 High-Frequency Filtering
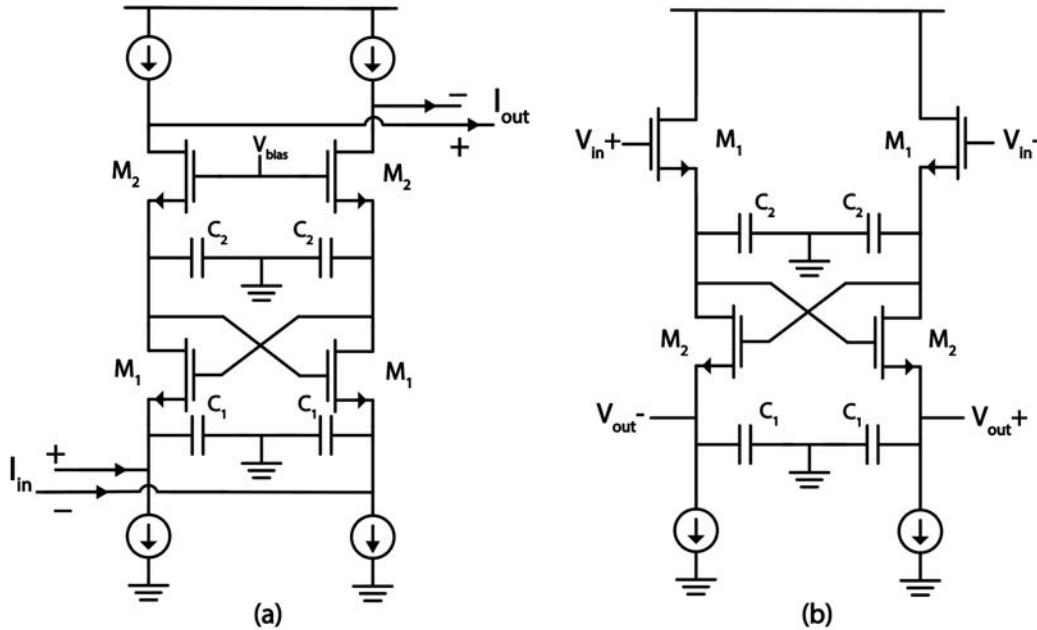
Figure 3.6: (a) Common-gate-based (CGB) current-mode filter and (b) source-follower-based (SFB) voltage-mode filter.

For high frequency baseband circuits, Gm-C filters are the predominant filters of choice, which trade off lower power consumption for compromised linearity due to their open-loop structure. In this work, Gm-C voltage-mode source-follower-based (SFB) and current-mode common-gate-based (CGB) biquads, shown in Fig. 3.6, are chosen as the Gm-C filter of choice due to their unique properties of simultaneous lower power consumption and improved linearity, while finding common application in high frequency analog basebands ([25], [4], [6]).

To analyze this type of filter, first consider the first-order current-mode gm-C filter in Fig. 3.7, where the single pole is located at $\omega_p \approx g_m/C$. It displays a unique noise-shaping property in that the output current noise spectrum is high-pass shaped by the in-band zero (located at $w_z = \omega_p/(g_m R_s)$) in the transfer function, due to the fact at low frequency the high impedance of capacitor C forces the noise of M$_1$ to circulate within the transistor, and at high frequency the low impedance of C allows all the noise to reach the output. Notice that the input current source creates an in-band degeneration for M$_1$, which improves the noise and linearity performance. As stated in [20], this type of filter can be viewed as a "pipe". In the passband, the filter functions like a lossless pipe in which the input current is equal to the output current, and thus no noise or intermodulation distortion components can be added to it. In the stop-band, however, a current "leakage" path can allow noise
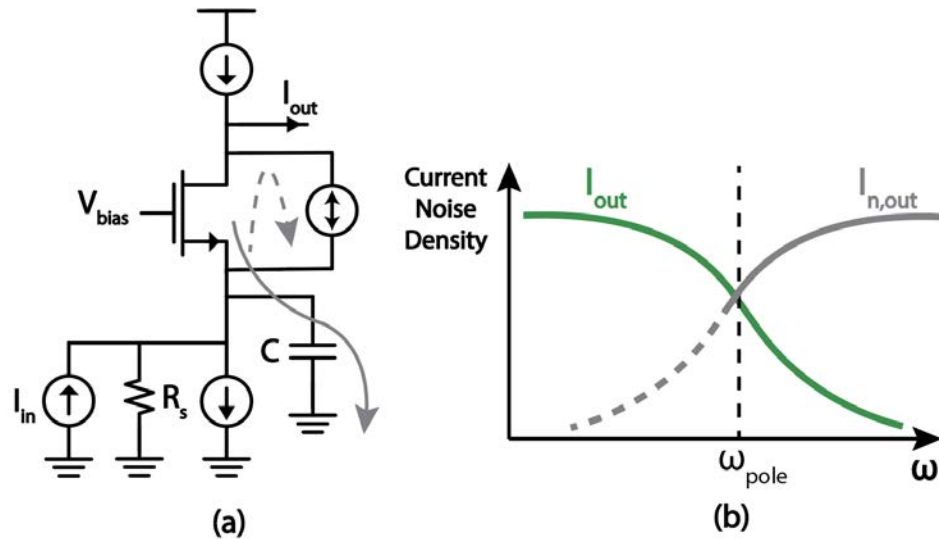
Figure 3.7: Schematic of a "pipe" filter and its noise behavior for low frequencies (dashed grey) and high frequencies (solid grey) relative to the pole.

and distortion to enter the pipe and reach the output. For example, this path is created by the low impedance in the stop-band by capacitor C. To truly realize such a filter, the condition of a perfectly unitary transfer function must be satisfied; in practice, finite output resistances of the current sources and input source prevent this.
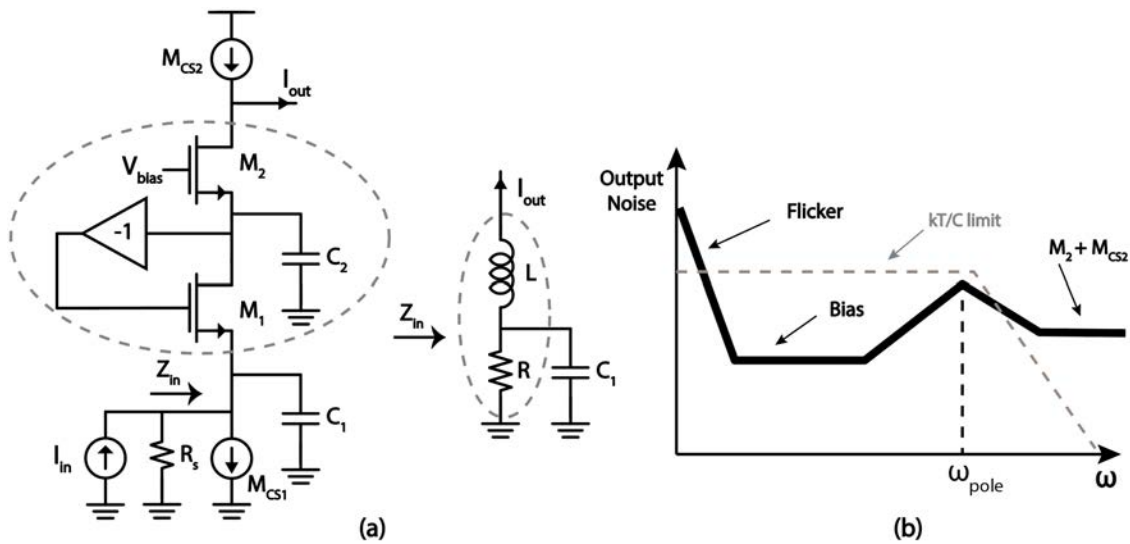
Figure 3.8: (a) Biquad realization with active inductor (b) and output noise PSD.

To extend this pipe filter to a biquad with two complex conjugate poles, an RLC network can be introduced with the use of active inductors, realized through M_2, C_2, and feedback around M_1, as shown in Fig. 3.8(a). At low frequency, the inductor presents a low impedance to the source and passes all the input current to the output. As the frequency increases, the inductor impedance increases and presents a less desirable path for the input current to flow, while capacitor C_1 increasingly shunts more signal current to ground. Assuming $g_{m1} = g_{m2} = g_m$, the impedance of the active inductor can be derived to be

$$Z_{AI} = \frac{sC_2}{g_m{}^2} \frac{1}{1 + sC_2/g_m} \tag{3.23}$$

and corresponds to an inductance $L = C_2/g_m{}^2$ in parallel with a resistance $R = 1/g_m$. Incorporating this synthesized RLC network into the CGB filter pipe filter, the second-order, low-pass transfer function is expressed as:

$$H(s) = \frac{i_{out}}{i_{in}} = \frac{1}{s^2 \frac{C_1 C_2}{g_{m1} g_{m2}} + s(\frac{C_2 - C_1}{g_{m2}} + \frac{C_1}{g_{m1}}) + 1}, g_{m1} = g_{m2} = g_m \Rightarrow \frac{g_m{}^2/(C_1 C_2)}{s^2 + s(g_m/C_1) + g_m{}^2/(C_1 C_2)} \tag{3.24}$$

Notice that the current gain is unity, the desirable condition for a lossless pipe in which no external or additional current is injected into the signal path. The complex conjugate pole frequency is given by

$$\omega_0 = \sqrt{\frac{g_{m1} g_{m2}}{C_1 C_2}}, g_{m1} = g_{m2} = g_m \Rightarrow \omega_0 = \frac{g_m}{\sqrt{C_1 C_2}} \tag{3.25}$$

and the quality factor Q is given by

$$Q = \frac{\sqrt{\frac{g_{m1}}{g_{m2}}} \sqrt{\frac{C_1}{C_2}}}{\frac{C_1}{C_2} + \frac{g_{m1}}{g_{m2}}(1 - \frac{C_1}{C_2})}, g_{m1} = g_{m2} \Rightarrow Q = \sqrt{\frac{C_1}{C_2}} \tag{3.26}$$

The input impedance corresponds to that of an LRC resonator and so has a band-pass characteristic, peaking at ω_0 with an impedance equal to the shunt loss of the inductor, $1/g_m$.

$$Z_{in} = \frac{s/C_1}{s^2 + s(g_m/C_1) + g_m{}^2/(C_1 C_2)} \tag{3.27}$$

The input-referred current noise can be derived as

$$IRN^2 = \overline{i_{n,cs1}{}^2} + |Q\frac{s}{\omega_0} + 1/(g_m R_s)|^2 \overline{i_{n,1}{}^2} + |\frac{s}{\omega_0}\frac{1 - Q^2}{Q} - (\frac{\omega}{\omega_0})^2|^2 \overline{i_{n,2}{}^2} + |\frac{\omega}{\omega_0 Q} + 1 - (\frac{\omega}{\omega_0})^2|^2 \overline{i_{n,cs2}^2} \tag{3.28}$$

The output noise PSD is illustrated in Fig. 3.8(b). Notice that the noise transfer function to the output for M_1 displays a bandpass property due to the fact C_2 filters both the signal and the noise injected by M_1. On the other hand, the noise transfer function to the output

for M_2 displays a high-pass property due to the fact for frequencies higher than the poles, C_2 begins to become a short and all the noise injected by M_2 can make its way to the output. For the current sources, the noise of M_{cs1} is simply processed as part of the input signal, and the noise of M_{cs2} is injected at the output and entirely makes its way to the output. At low frequencies the noise is dominated by flicker noise. As the frequency approaches ω_0, M_{cs1}, M_1, and M2 all contribute increasing amounts of output noise that peaks at ω_0. Beyond ω_0, the only noise that is not filtered out is contributed by M_2 and M_{cs2}. Thus, in-band high-pass noise shaping is achieved, compared to traditional filters where most of the noise is in the filter passband. This noise high-pass property within the passband potentially leads to a low in-band noise design. However, the out-of-band noise can be folded back into the signal band of interest by the subsequent sampling of the ADC. Therefore, additional filtering, for example by a shunt capacitor at the load to create another pole, may be needed when converting the signal back to a voltage to feed the ADC.

The CGB filter can also achieve high out-of-band linearity due to two mechanisms unique to this topology: (1) similar to the high-pass noise shaping property, no intermodulation distortion products can be generated in the filter passband as long as the lossless pipe condition is satisfied (independent of the location of the blockers with respect to the passband) and 2) capacitor C1 at the input filters and absorbs out-of-band blockers before entering the nonlinear devices and modulating the input transistors' gate-source voltages. Thus, we can expect the intermodulation distortion products to decrease as the blockers are moved further from ω_0. Note that this passive blocker attenuation at the filter input is not present in other topologies; for example, in an op-amp RC filter, a blocker current signal injected in the virtual ground is not filtered, forcing the op-amp to sink/source this current, independently of the frequency of the blocker signal with respect to the filter passband. Of course, the input impedance still impacts the linearity, so it should be expected that the worst distortion occurs at ω_0, according to Equation 3.27. A more quantitative analysis of the advantageous noise and linearity mechanisms can be found in [20].

The SFB filter is essentially the voltage-mode equivalent of CGB filter (the source follower ideally provides unity voltage gain while the common gate amplifier provides unity current gain). The cross coupled devices M_2 once again form an active inductor to create two complex conjugate poles, and the transfer function, pole frequencies, and Q are identical to equations 3.24, 3.25, and 3.26, respectively. Similarly, the noise due to the current source and M_2 are shaped by a high-pass transfer function, moving their noise in the output spectrum to outside the passband. Unlike the CGB filter, however, the linearity is no longer mainly determined by the bandpass input impedance. Instead, as with any feedback structure, the linearity improves with a larger loop gain ($g_{m1}Z_{out}$, where Z_{out} is the output impedance of the current source). Thus, a larger g_m improves linearity, and a larger g_m implies minimizing the overdrive voltage, V_{ov}. A lower V_{ov} implies a more current efficient transconductor (lower current to achieve a certain g_m and lower power consumption). This trend is opposite to conventional Gm-C filters which require a higher input transconductor V_{ov} and thus greater

power consumption to achieve higher linearity. This describes the main advantage of the SFB filter. However, both g_m and R_{out} decrease with higher frequency beyond the passband due to parasitic capacitances, which decreases the loop gain and therefore the linearity. Therefore, the SFB filter has lower out-of-band linearity compared to the CGB filter.

3.4 Linearity

While favorable for lower frequency applications, the use of traditional op-amp closed-loop structures ([5], [8], [10]) is typically avoided at high baseband frequencies due to the large power consumption required for high unity-gain frequency op-amps, as well as difficulties regarding stability and compensation for sufficient phase margin as the effects of capacitive parasitics become more prevalent. Therefore, the inherent benefits of high input-limited linearity from large loop gain in closed-loop topologies are not efficiently realizable at the higher frequencies of interest. To maintain sufficient linearity in the baseband chain, other design techniques for enhanced input-limited and output-limited linearity must be explored. Because of the differential implementation of baseband signal processing, the focus of the preceding discussion focuses on third-order nonlinearity ($IIP3, ICP_{1dB}$).

Source Degeneration

Recall that input-limited linearity originates from nonlinear transconductance g_m, or the conversion of linear input voltage to nonlinear output drain current. For ideally linear V-I conversion, this implies that the overall transconductance G_m should be constant with respect to the input voltage, where $I_{out} = G_m \Delta V_{in}$. Source degeneration, shown in Fig. 3.9(a), widens the differential input voltage range that provides a constant G_m by series feedback that reduces the signal swing applied between the gate and source, most easily observed by noticing that $G_m = \frac{g_m}{1+g_m R_s} \approx \frac{1}{R_s}$ when $g_m R_s \gg 1$, indicating that G_m is now independent of the input voltage.

Source degeneration presents two drawbacks in that it lowers the DC gain and the resistor voltage drop lowers available headroom by $I_{ss}R_s/2$. To alleviate the second issue, the degeneration resistors can be equivalently reconfigured as in Fig. 3.9(a). However, it suffers from greater noise and offset voltage due to the mismatch between the two split tail current sources [22].

Complementary Input

A complementary input amplifier consists of both NMOS and PMOS devices being used as the transconductance device with their gates tied together, as shown in Fig. 3.9(b). This results in the sum of the both devices' small signal drain currents and the net $G_m = g_{m,n} + g_{m,p}$. Historically, PMOS devices have approximately half the mobility as NMOS devices, and so sizing them larger to achieve similar g_m as the NMOS device results in twice
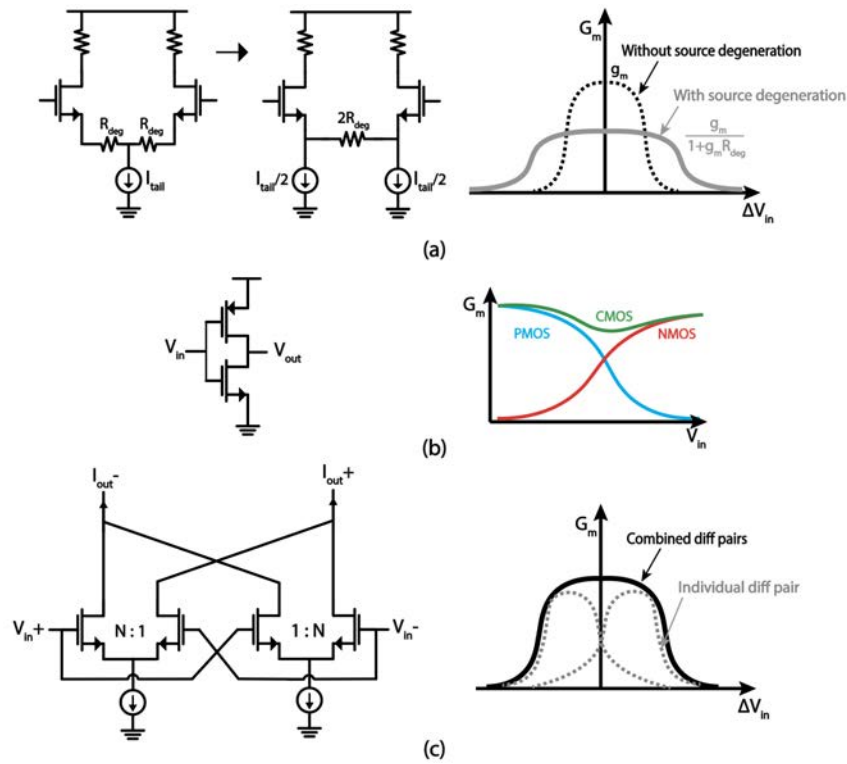
Figure 3.9: Various V-I linearity enhancement techniques such as (a) source degeneration, (b) complementary input and (c) parallel differential pairs.

the G_m but three times the input capacitance. However, in deep-submicron (DSM) CMOS technologies, the gap between minimum channel length NMOS and PMOS device mobility is very small and in fact almost identical in 16nm and 28nm nodes, thus making them viable amplifier choices.

Several linearity advantages can be obtained with complementary inputs if the devices are sized for similar g_m. Even if driven single-ended, the second-order nonlinearity of the NMOS and PMOS devices cancel to first order, which leads to a high IIP2 [7]. More importantly, the inherent push-pull operation of the NMOS and PMOS devices helps maintain a constant G_m over a wide input range, as the devices compensate for one another across a wide input range, with the NMOS handling the higher large-signal excursions of the input swing and the PMOS handling the lower ones. The net G_m is thus relatively linear over a wider range even though each device experiences substantial distortion.

Parallel Differential Pairs

Shown in Fig. 3.9(c), parallel differential pairs is based on the concept that the input differential voltage range in which linear G_m is obtained can also be widened by introducing horizontal shifts in the opposite directions in the G_m versus Δv_{in} curve of two differential pairs and then summing their current outputs together. The horizontal shifts can be performed by intentionally creating an input offset by mismatching the differential pair device sizing by a ratio N, and the output currents can be summed by simply shorting their drains together. A qualitative explanation is that due to the more gradual slopes in the I-V curves of the constituent differential pairs, the overall linear transconductance range is widened. Given the DC parameters of the device, an optimal ratio N for extended G_m input range exists. This technique is also known as the 'multi-tanh' technique [12].

Current-mode Operation

While the above enhanced linearity techniques improve the input-limited linearity, processing the signal in the current domain can improve the output-limited linearity, as current-mode operation typically use low impedances to guide the signal current. Thus, the voltage swings are minimized and the devices experience smaller drain-source voltage swings that do not deviate as far from their designed operating point. In other words, nonlinearities are less heavily excited, producing less nonlinear distortion. Recall that output-limited linearity originates from nonlinear output conductance g_{ds}, whose effect becomes pronounced under large drain-source voltage swing and low drain-source voltage, or when the device operates near triode [28]. The output resistance of MOS transistors in DSM technologies is highly nonlinear as many different physical effects like channel length modulation and drain induced barrier lowering (DIBL) contribute varying amounts of nonlinearity at different bias points. Thus, the current-mode CGB filter is expected to outperform the voltage-mode SFB filter in terms of output-limited linearity, considering all else equal.

Chapter 4

Circuit Implementation

4.1   Architecture

The inputs to the ABB are AC-coupled to isolate the DC common-mode from that of the RF front-end. As stated previously, the entire subsequent signal path is then DC-coupled. One issue with DC coupling is that the output common-mode level of the previous stage directly sets the input common-mode level of the next stage, which adds design complexity in ensuring one shared common mode properly biases both stages. A differential 100Ω physical resistor is used to terminate the input. To keep the ABB linear when interfacing to a mm-Wave front-end design that already provides significant gain, a programmable attenuator is placed at the input. The rest of the ABB consists of a certain number of Cherry-Hooper (CH) amplification cells to achieve high bandwidth, and current-mode filters to achieve higher linearity. The filters can be conveniently placed between the two stages of the Cherry-Hooper gain cells (with proper input and output resistances), as the signal is in the current domain at that point. As will be discussed later, the desired number of stages is determined to be two. Thus, the ABB core will be comprised of a first-stage Cherry-Hooper amplifier, a second-stage Cherry-Hooper amplifier which contains the current-mode biquad filter in between the Gm stage and TIA, and a buffer stage intended to drive the ADC. Additionally, each Cherry-Hooper amplifier implements a real pole and shares gain programmability. The proposed architecture is shown in Fig. 4.1. A CMFB loop maintains proper common-mode DC level between at the interfaces between the Gm, filter, and TIA of the second stage CH, while the first-stage CH is self biased. Gain tuning is implemented with programmable feedback resistors in both CH stages. A programmable feedback capacitor in both stages sets the two real poles, while the filter provides an additional two complex conjugate poles for a fourth-order overall frequency response. DC offset calibration is corrected at various points along the signal path, in this case at the second-stage CH Gm and TIA.

It is unusual that the filter is placed in the second Cherry-Hooper stage, as any out-of-band blockers are desired to be attenuated as early as possible in the signal chain. However,

the CGB filter presents a complicated and stringent design tradeoff amongst noise perfor-
mance and shaping, signal loss, and parasitic capacitance that can degrade its pole charac-
teristics, all of which will necessitate greater power consumption in the filter. To alleviate
this issue, the filter can be placed in the second stage while the first-stage CH is designed
to be a low-noise amplifier with first-order filtering to provide preliminary out-of-band fil-
tering and relax the noise performances of the filter, second-stage CH, and buffer. However,
the gain of the first stage necessitates greater linearity performance for these latter stages,
again highlighting the need for filtering in the current-domain and linearity enhancement
techniques.
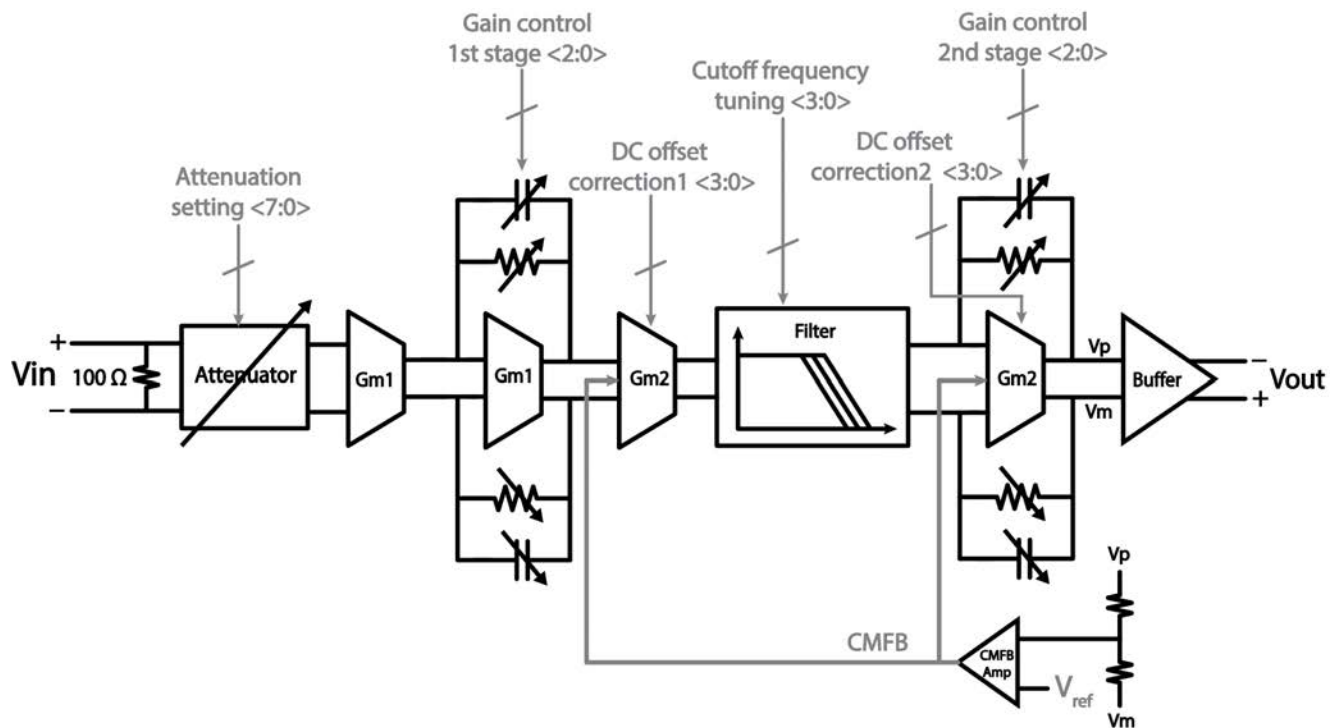
Figure 4.1: Block-level schematic of the ABB.

4.2 Design Approach

The design begins with determining the number of stages to maximize the achievable gain-
bandwidth within a reasonable power consumption. As explained in Section 3.1, increasing
the number of stages to meet a certain overall GBW allows the designer to relax the GBW
of each stage, which is desirable from a power consumption perspective. To roughly gain
perspective, the gain-bandwidth for a differential pair with a dominant output pole, active

load, and fanout $= 1$ can be expressed as $GBW = g_m/C_L = g_m/(2C_{dd}+C_{gg}) = f_T/n$ where n is usually between 1.2 and 1.8, depending on the technology. In 28nm technology, the layout extracted f_T up through multiple metal layers can be well over 300 GHz. Given the gain specifications, the required GBW for each cell is well below the theoretical limit for even one stage. However, the required cell GBW for one stage is more than twice that for two stages, so two stages are chosen to potentially reduce power consumption and relax the design for GBW. Note that since there will be multiple filtering poles setting the ABB bandwidth, the approximation (from [18]) $BW_{tot} \approx \frac{0.833\omega_p}{\sqrt{n_{poles}}}$ yields that the four poles need to be at roughly 6 GHz for an overall bandwidth of 2.5 GHz, assuming all poles are at the same frequency.

The desired small-signal frequency response can first be obtained. Because both Cherry-Hooper stages and buffer each drive capacitive loads, the overall transfer function is simply the cascaded product of each individual transfer function. However, placing the current-mode LPF in between the Gm and TIA stage of a Cherry-Hooper amplifier presents several modifications to the Cherry-Hooper transfer function in Equation 3.13 to represent $H_{CH2}(s)$, due to the fact the insertion of the filter introduces an additional intermediate node. By representing the small-signal model of the filter by its transfer function $H_{LPF}(s)$, its corresponding input and output resistances and capacitances, and the transfer function of a conventional Cherry-Hooper amplifier $H_{CH}(s)$, equations 3.13 and 3.24 can be leveraged to form

$$H_{CH2}(s) = H_{CH}(s)H_{LPF}(s)\frac{1}{1 + sC_{int}R_{in,filter}} \tag{4.1}$$

where C_{int} is the total capacitance at the node of the Gm output and filter input, and $R_{in,filter}$ is the input resistance of the filter, or the $1/g_m$ of the common-gate input stage in this design. An approximation is made by grouping the Gm and LPF as a transconductor with an intermediate pole $\omega_{int} = 1/(C_{int}R_{in,filter})$ that produces a filtered output current. Thus, in the above, C_1 and R_{o1} of $H_{CH}(s)$ from Equation 3.13 should now refer to the total capacitance at the node of the filter output and TIA input, and output resistance of the filter, respectively. The overall transfer function of the ABB is given by

$$H_{ABB}(s) = H_{CH1}(s)H_{CH2}(s)H_{Buffer}(s) \tag{4.2}$$

Next, quantitative description of the linearity and noise performance is provided. As will be explained in detail in Section 4.3, the implementation topology of each block or stage must be mentioned to describe their linearity and noise performance: The first CH stage is composed of inverter-based amplifiers as described in Section 3.4, the LPF is implemented as the CGB biquad as described in Section 4.3, the second stage CH is composed of source degenerated NMOS-input differential pairs as described in Section 3.4, and the buffer is implemented with parallel differential pairs as described in Section 3.4.

The overall OCP1dB, in terms of ICP1dB (expressed as P1dB below for compactness) for each stage for the ABB can be expressed as

$$\frac{1}{OCP1dB_{ABB}} = \frac{1}{P1dB_{CH1} \cdot A_{v,CH1} \cdot G_{m2} \cdot A_{i,filter} \cdot Z_{TIA2} \cdot A_{v,buffer}} +$$
$$\frac{1}{P1dB_{Gm2} \cdot G_{m2} \cdot A_{i,filter} \cdot Z_{TIA2} \cdot A_{v,buffer}} +$$
$$\frac{1}{P1dB_{filter} \cdot A_{i,filter} \cdot Z_{TIA2} \cdot A_{v,buffer}} + \frac{1}{P1dB_{TIA2} \cdot Z_{TIA2} \cdot A_{v,buffer}} +$$
$$\frac{1}{P1dB_{Buffer} \cdot A_{v,buffer}}$$

(4.3)

where $A_{i,filter}$ is the DC current gain of the LPF and Z_{TIA2} is the transimpedance of the TIA. Note that the ICP1dB of each block is given as the noise quantity identical to it's input signal quantity, so ICP1dB of the LPF and TIA are given as currents while the ICP1dB of the voltage gain and Gm stages are given as voltages.

To express the equations for memory-less linearity, recall the approximation for a large-signal output current to input voltage relation can be approximated as a power series as

$$i_{out} \approx g_1 v_i + g_2 v_i^2 + g_3 v_i^3 + ... \tag{4.4}$$

The derivation of highly accurate equations for the power series coefficients are outside the scope of this work. Instead, the IIP3 contributions of each term are approximated by the following equations and needed coefficients can be obtained through simulation. The equations assume that the linearity is input-limited and that the only odd order distortion products are due to third-order nonlinearity. At low frequencies, the main nonlinearity contributions of each stage are calculated as

$$ICP1dB_{CH1} = \sqrt{0.11}\sqrt{\frac{1}{ICP1dB_{Gm1} \cdot G_{m1} \cdot Z_{TIA1}} + \frac{1}{ICP1dB_{TIA1} \cdot Z_{TIA1}}}$$
$$ICP1dB_{Gm1} = \sqrt{0.11}\sqrt{\frac{4}{3}\frac{g_{1,nmos} + g_{1,pmos}}{g_{3,nmos} + g_{3,pmos}}}$$

(4.5)

where the addition of the power series coefficients for the NMOS and PMOS devices will enhance the first-order response and mitigate the third-order nonlinearity by expanding the linear input range, as explained in 3.4.

$$ICP1dB_{Gm2} = \sqrt{0.11}\frac{1}{3}I_{tail}\sqrt{g_m(R_{deg} + 2/g_m)^3} \tag{4.6}$$

From [20],

$$ICP1dB_{Filter} = \sqrt{0.11}\frac{8g_m^5 R_s \omega_1^2}{3\omega_p^3 |jg_m g_3 \omega_1 + g_2^2 \omega_p|\sqrt{I_1}} \tag{4.7}$$

where ω_1 and I_1 are the frequency and magnitude of the input test tone(s), respectively, ω_p is the cutoff frequency, and g_m is the transconductance of the active common-gate device. From [9],

$$ICP1dB_{TIA2} = \sqrt{0.11}\sqrt{\frac{4}{3}\frac{z_1^4}{z_3}\frac{g_m}{(1+R_f/R_s)}\frac{1}{R_f}} \tag{4.8}$$

where g_m is the transconductance of the input pair, R_f is the value of the feedback resistor, R_s is the output resistance of the preceding current-output stage, and z_1 and z_3 are the power series coefficients for a transimpedance, or output voltage to input current, relation. The source degeneration is assumed to be disabled.

$$ICP1dB_{Buffer} = \sqrt{0.11}\sqrt{\frac{4}{3}\frac{g_1(1+N)}{g_3(1+1/\sqrt{N})}}, g_1 = \sqrt{\mu_n C_{ox}\frac{W}{L}I_{tail}}, g_3 = -(\mu_n C_{ox})^3\frac{1}{8\sqrt{I_{tail}}} \tag{4.9}$$

where N is the multiplicity ratio, and the power series coefficients are for the referenced device sizing, derived in [22]. Finally, it should be noted that after linearity-enhancing circuit topologies and techniques are used and their component values (the source degeneration resistor, for example) are constrained, linearity is determined mainly by biasing, such as the V_{ov} of a differential pair or the nominal V_{DS} allocated to devices that experience large drain swings.

Noise can then be similarly analyzed. Note that the input-referred noise (IRN) of each block is given as the noise quantity identical to it's input signal quantity, so IRN of the LPF and TIA are given as current noise while the IRN of the voltage gain and Gm stages are given as voltage noise.

$$IRN_{ABB}^2 = IRN_{CH1}^2 + \frac{IRN_{Gm2}}{A_{v,CH1}^2} + \frac{IRN_{filter}}{A_{v,CH1}^2 G_{m2}^2} +$$
$$\frac{IRN_{TIA2}}{A_{v,CH1}^2 G_{m2}^2 A_{i,filter}^2} + \frac{IRN_{buffer}}{A_{v,CH1}^2 G_{m2}^2 A_{i,filter}^2 Z_{TIA2}^2} \tag{4.10}$$

Due to the large signal bandwidth, thermal noise becomes a significant noise source relative to flicker noise. At low frequency, the input-referred thermal noise spectral density for each term is given by

$$IRN_{CH1}^2 = \frac{8kT\gamma}{(g_{m,n}+g_{m,p})G_m} + \frac{8kT}{(g_{m,n}+g_{m,p})G_m^2}[\frac{\gamma/(g_{m,n}+g_{m,p})_{TIA}}{(\frac{R_f}{1+sR_fC_i})^2} + \frac{1}{R_f}] \tag{4.11}$$

$$IRN_{Gm2}^2 = \frac{8kT\gamma(g_{m,in}+g_{m,load})}{(\frac{g_{m,in}}{1+g_{m,in}R_{deg}/2})^2} + 4kTR_{deg/2} \tag{4.12}$$

$$IRN_{TIA2}^2 = \frac{8kT}{R_f} + \frac{IRN_{Gm2}^2}{(\frac{R_f}{1+sR_fC_i})^2} \tag{4.13}$$

where C_i is the input capacitance of the respective TIA stages, and it is assumed Gm2 is sized identically as TIA2.

$$IRN_{buffer}{}^2 = 8kT\gamma(\frac{1}{g_{m,in}} + \frac{g_{m,load}}{g_{m,in}{}^2}) \tag{4.14}$$

Finally, IRN_{filter} is given by Equation 3.28. Flicker noise terms, whose noise spectral density is given by $\overline{i_n{}^2}/\Delta f = \frac{KI_D}{L^2 C_{ox} f}$ (K is a process-dependent parameter and I_D is the drain current), are omitted here for brevity, but can easily be included as an additive noise current term with the same transfer function to the output as the thermal noise for the same MOS device. The total integrated output noise of the ABB can be calculated by

$$\overline{v_{o,n}^2} = \int_0^\infty IRN_{ABB}{}^2|H_{ABB}(s)|^2 df \tag{4.15}$$

and then converted to NF in the same way as in Equation 3.5.

Finally, the calibration circuitry can be designed to cancel DC offset at certain locations within the signal path to prevent any stage from saturating. Similar to noise performance, the first gain stages within the signal path are the most critical, as their DC offsets are effectively amplified by subsequent gain stages. The input referred DC offset of the ABB is given by

$$V_{OS,ABB} = V_{OS,CH1} + \frac{V_{OS,CH2}}{A_{v,CH1}} + \frac{V_{OS,buffer}}{A_{v,CH1}A_{v,CH2}} \tag{4.16}$$

If the DC offset exceeds the range of the calibration, the gate area of the input pairs in the first stage needs to be increased to lower the random offset.

4.3   Circuit Design

Termination and Attenuator

A large AC short capacitor is placed on the center tap of the 100Ω differential termination resistor to provide termination of the common mode for high frequencies. The attenuator is implemented as a differential string resistor ladder as shown in Fig 4.2 , where the output voltage switches tapped at each resistor segment are thermometer encoded to select one of eight attenuation levels, for a maximum attenuation of -18 dB with linear-scale intermediate attenuation increments of $n/8$, where $n = 1, 2...8$. Because the resistive divider is placed in shunt with the termination resistor, a large total resistor value of $8R_{att}$ for the attenuator should be used to maintain an input resistance of approximately 50Ω. On the other hand, R_{att} also influences the input pole to the first-stage Cherry-Hooper amplifier which is located at

$$\omega_{in} = 1/(C_{in}nR_{att}||[(8-n)R_{att} + 50\Omega]) \tag{4.17}$$

which should be kept higher than the desired bandwidth. A value of $R_{att}= 150\ \Omega$ sets the bandwidth greater than 5GHz while maintaining $S11 < -10$ dB. Note that the attenuation level in dB adds directly to the noise figure.
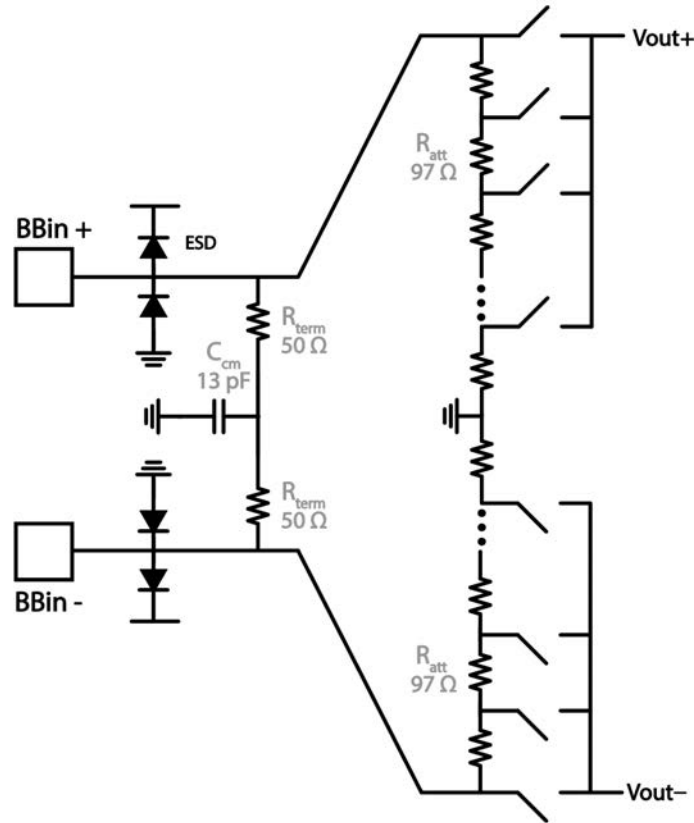

Figure 4.2: Schematic of the input termination and attenuator.

Cherry-Hooper Amplifiers

As the first stage of the Cherry-Hooper amplifier in the ABB, the design of the Gm stage is mainly concerned with low noise and current efficient transconductance that will partly set the overall DC gain and noise figure. Input-referred noise and DC gain for the CH amplifier are both improved by increasing g_m. As explained in Section 3.4, the use of inverter-based transconductors results in increased G_m for the same bias current for efficient transconductance. One drawback is that the NMOS device contributes relatively high flicker noise. However, because the operation frequency is relatively low compared to the process f_T, the channel length can be increased from minimum length to reduce the flicker noise without significantly degrading the frequency response. To reduce the input capacitance presented to

the previous stage, cascode devices are employed to mitigate the Miller effect. The schematic of the first-stage CH amplifier is shown in Fig. 4.3. The cascoded inverters are left as pseudo-differential structures as there is not enough headroom remaining for stacking more devices as current sources at the common-source node. While this means there is no common-mode rejection in the first stage CH, the front-end that drives the ABB is assumed to have high common-mode rejection. It may be worth noting that pseudo-differential amplifiers display higher input linearity characteristics than their differential counterparts with tail current sources [22].

The cascoded inverter for the G_m stage is self-biased to mid-rail by placing a large feed-back resistance between the input and output nodes. In order to keep the input impedance large and therefore not load the attenuator or input matching, the feedback resistance should be as large as possible as it is given by

$$R_{in} = \frac{R_{fb}}{1 + (g_{m,n} + g_{m,p})(g_{m,n}r_{o,n}{}^2||g_{m,p}r_{o,p}{}^2)} \tag{4.18}$$

Thus, a long channel MOS device biased in triode is used to realize a feedback resistance of $> 50k\Omega$. In its highest gain setting, the first-stage CH amplifier achieves a gain of 24 dB, a ICP1dB of -20.1 dBm, and an output-referred noise (ORN) of 61.9 nV/$\sqrt{Hz}$. In its lowest gain setting, it has a gain of 7 dB, an ICP1dB of -4.1 dBm, and a ORN of 7.9 nV/$\sqrt{Hz}$. The bandwidth of 3-5 GHz is set by the synthesized first-order pole from the TIA feedback passives. The first-stage CH amplifier consumes 7.1 mA of current.

The second stage Cherry-Hooper amplifier relaxes the noise and gain performance and shifts the focus to high linearity to handle the larger signal swings. To maximize the output swing headroom, ultra-low threshold NMOS differential pairs with active loads are employed, which have the tail current source for common-mode rejection. To extend the input linear range, the differential pairs are source degenerated. A degeneration enable switch is included for adjusting flexibility in trading of between linearity and DC gain depending on the desired ABB gain setting. To mitigate the Miller effect and improve high frequency performance, the differential pairs are also capacitively neutralized. The PMOS active loads' gate bias is provided by the same CMFB loop. The error amplifier is a high-gain, single-stage folded cascode op-amp, shown in Fig. 4.5. A 5.5 pF compensation capacitor sets the loop band-width at 5 MHz with a phase margin of 73 degrees. The schematic of the second-stage CH amplifier is shown in Fig. 4.4; the output RC filter simultaneously provides compensation and filters the noise on the output voltage. Assuming the filter that processes the signal in the intermediate current domain is bypassed and source degeneration is not enabled, in its highest gain setting, the second-stage CH achieves a gain of 18 dB, a ICP1dB of -17.5 dBm, and a ORN of 37.5 nV/$\sqrt{Hz}$. In its lowest gain setting, it has a gain of 3 dB, a ICP1dB of -7.8 dBm, and a ORN of 5.2 nV/$\sqrt{Hz}$. The bandwidth of 3-5 GHz is set by the synthesized first-order pole from the TIA feedback passives. With source degeneration enabled, the gain decreases by 4 dB while the input compression point roughly increases by 6 dB. The Gm cell
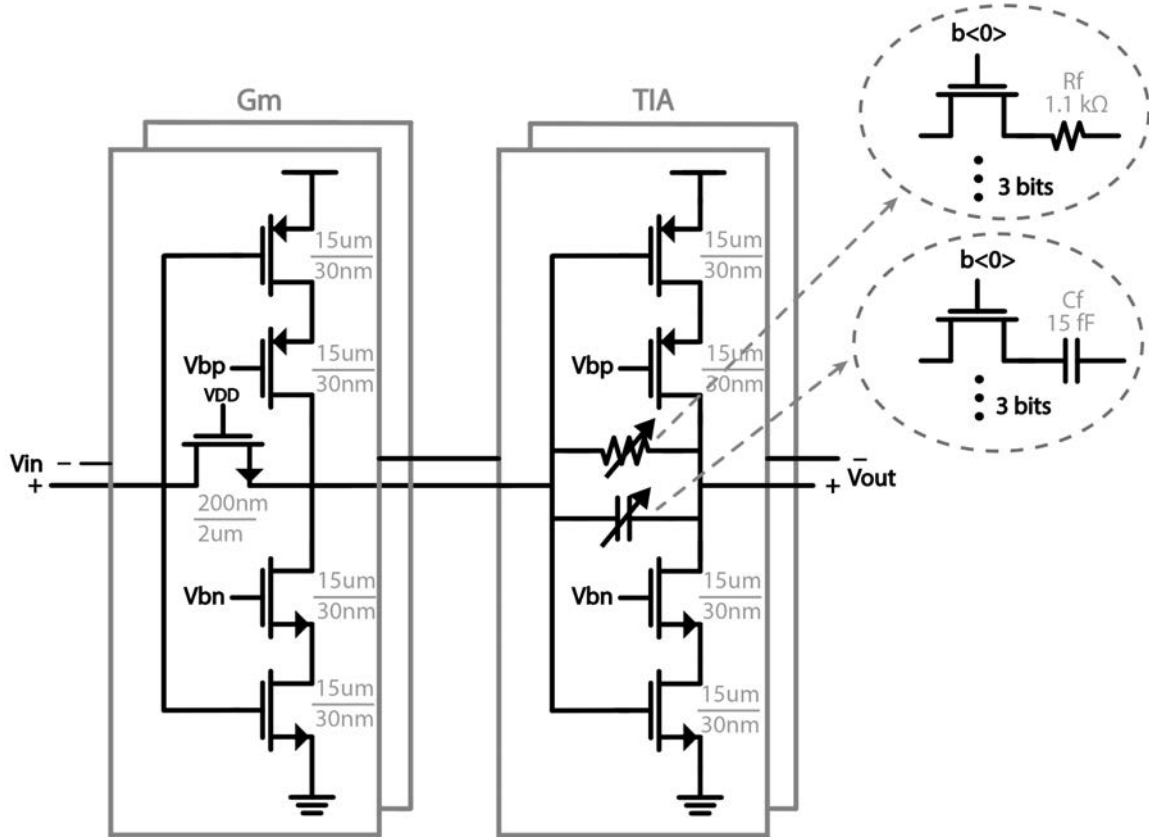
Figure 4.3: Schematic of the first-stage Cherry-Hooper amplifier.

and TIA of the second-stage CH amplifier consume 2.2 mA of current. As the second stage, it is also designed for finer gain tuning within a smaller range of gain as the NF is mainly determined by the noise performance of the first stage.

Variable gain that is controlled digitally is performed by implementing the resistive feedback of each TIA with a programmable array of resistors. Additionally, a real pole can be conveniently synthesized with the parallel connection of the feedback resistor and a capacitor. For each gain setting, the parallel capacitance value is chosen to maintain a fixed pole frequency $f_{p,TIA} = \frac{1}{2\pi R_f(i)C_f(i)}$, where i is determined by the number of bits. Each element of the array can be selected via switches, which are placed on the virtual ground side of the TIA, where the low voltage swing will cause less modulation of the nonlinear switch on-resistance and thus produce lower high frequency distortion. Ultra-low threshold NMOS devices are used to minimize the on-resistance of the switches despite the TIA's self-biased common mode being near mid-rail, to avoid the additional diffusion capacitance overhead from using a complementary pass gate. Being in series with the feedback capacitors, the
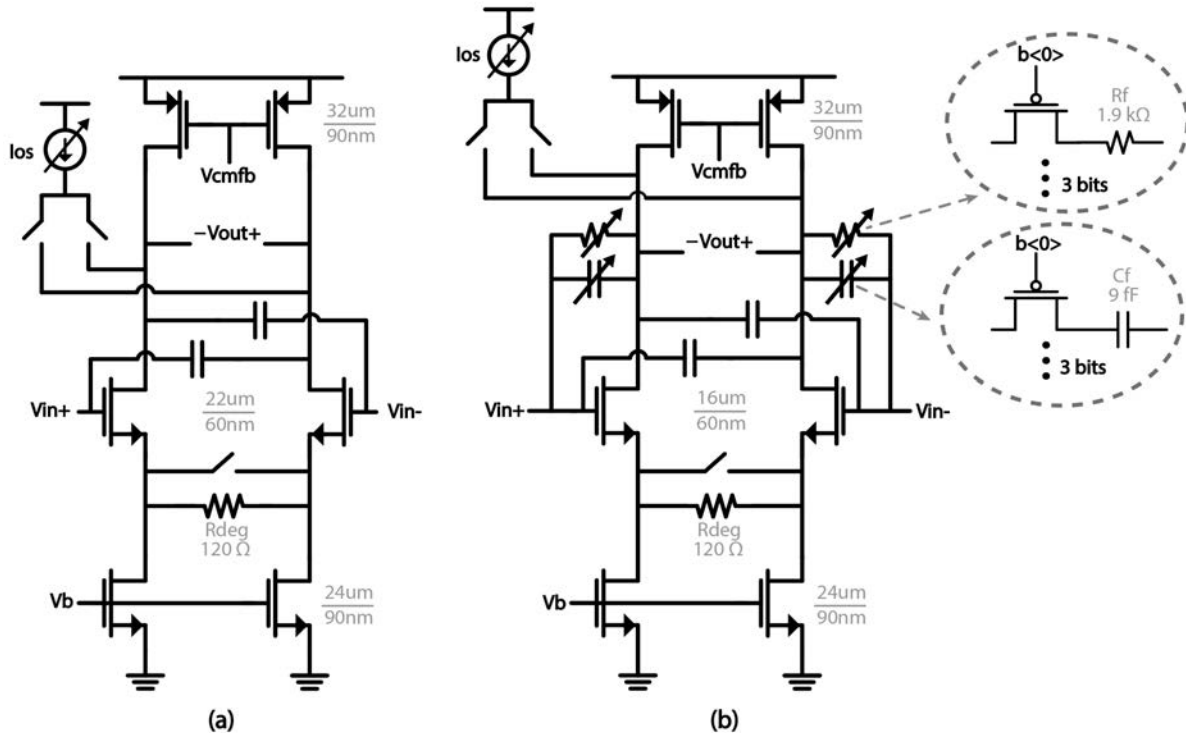
Figure 4.4: Schematic of (a) Gm cell and (b) TIA of the second-stage Cherry-Hooper amplifier.

switches' on resistance introduces a parasitic zero at $1/(R_{on}C_f)$, so the switch on-resistance must be low enough to push this zero well beyond the signal band. Both passives are implemented as a multiple of a common unit passive for better matching, not only between the differential signal paths but also between the I/Q ABB paths.

Current-mode LPF

The current-mode LPF adopts the CGB-filter design explained in Section 4.3. The biquad is designed for a Butterworth response to ensure minimal in-band ripple and steep roll-off, while a slower roll-off is not as critical due to the presence of other filtering poles in the ABB. The quality factor and thus Butterworth ($Q = 1/\sqrt{2}$) response of the filter is set by the ratio of capacitors $Q = \sqrt{\frac{C_1}{C_2}}$, while the cutoff frequency is set by the capacitor values and g_m of the cross coupled devices, yielding $\omega_0 = \frac{g_m}{2\sqrt{C_1 C_2}}$. The maximum input impedance of the LPF that occurs at ω_0 is given by $1/g_m$, and is set by the tolerable in-band attenuation and limited by power consumption, determined to be 125Ω in this design. These three
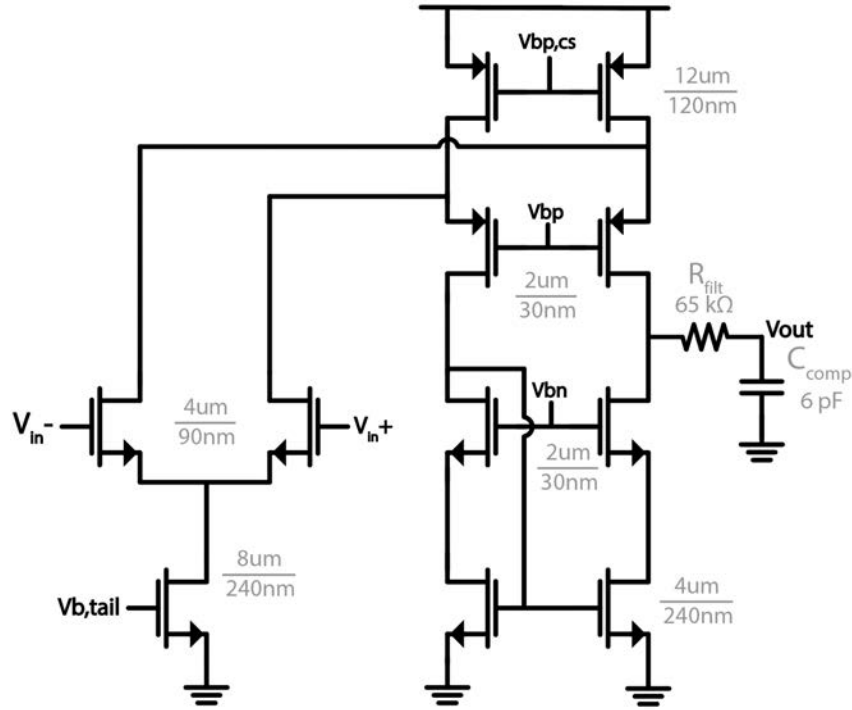
Figure 4.5: Schematic of the CMFB error amplifier used in both the LPF and the second-stage CH amp.

constraints determine the three design variable of C_1, C_2, and g_m. In this design, w_0 is set to be between 3 and 7.2 GHz, and the in-band gain is -3.4 dB. The frequency response of the LPF is shown in Fig. 4.6. The schematic of the filter design is shown in Fig. 4.7.

According to simulation, the current sources in the pipe filtering branch are the dominant noise contributors to the filter integrated output noise, contributing both significant thermal and flicker noise. To address this, the current sources use long channel devices and are biased with a very high V_{ov}. To further reduce the noise, the cutoff frequency of the LPF can be set slightly greater than the intended bandwidth to push the noise "bump" in the output noise spectrum mentioned in outside of the ABB bandwidth. The output noise spectrum is shown in Fig. 4.6(b).

It is determined from simulation that the main limitation to the filter linearity is hard distortions, which describes the case when the signal current becomes comparable to the bias current. In this case, the LPF linearity can be improved by simply increasing the bias current (within a reasonable power consumption budget) and thus the capability to handle large signal current without significant compression. The ICP1dB for this design is 6.28 mA

differential peak-to-peak.

To provide cutoff frequency tuning to account for PVT variation, the capacitor C_2 is implemented as a variable capacitor bank. Making C_2 variable instead of C_1 results in less variation in other characteristics of the LPF, such as linearity (in which C_1 plays an important role in filtering out-of-band blockers) and quality factor (because the nominal Q is less than 1 and C_2 is in the denominator of the quality factor expression). Note that in order to maintain constant Q ,which is typically desirable, both C_1 and C_2 capacitors should be made variable and change by the same amount. However, it was determined that doing so complicated the layout enormously, and because the in-band flatness was the most important aspect of the LPF as opposed to group delay or roll-off, only C_2 is used for tuning. As a consequence, the nominally designed Butterworth LPF does exhibit more of a Bessel characteristic for lower cutoff frequency settings as C_2 increases and the Q decreases. Similar to the programmable capacitor bank of an LC voltage-controlled oscillator (VCO), the switch is placed on the axis of symmetry between two series capacitors that are twice as large as the desired capacitance value. This prevents the parasitic capacitances of the switch from directly loading the signal path nodes and capacitively divides any potentially large signal swings (from a large blocker, for example) across the switch, making it more linear.
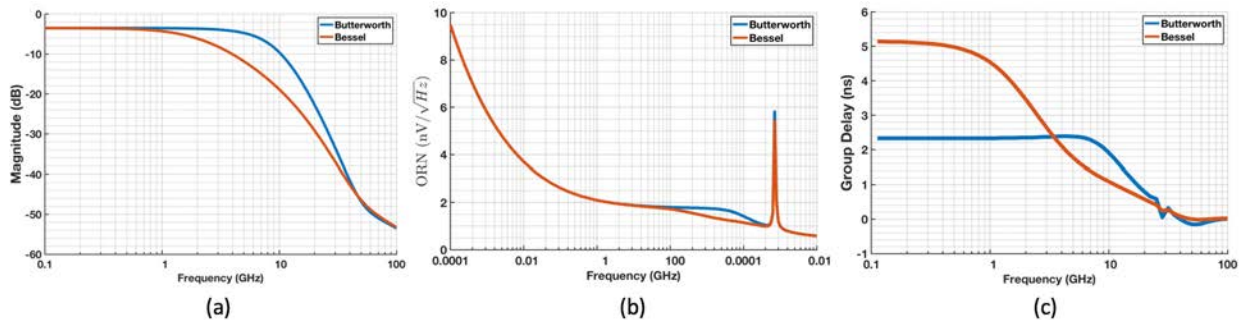

Figure 4.6: (a) Magnitude response, (b) output-referred noise PSD, and (c) group delay of the LPF with nominal cutoff frequency (Butterworth) and low frequency cutoff (Bessel).

To minimize the effect of parasitics on the input node of the filter, a low voltage swing input PMOS common-gate current-buffer is added in between the Gm stage and filter. As with any current-mode circuit, the input should be a low impedance of $1/g_m$ in this case to ensure most of the signal current produced by the preceding Gm stage flows into the filter. Similarly, the output stage employs a folded common-gate buffer with cascode loads to maintain low impedance on the drains of M_2 (and $M_{CS,p}$) and high output impedance to drive the subsequent TIA.

Finally, a CMFB loop sets the output common mode the filter output, using the same error amplifier as shown in Fig. 4.5. A 6 pF compensation capacitor sets the loop bandwidth at 5 MHz with a phase margin of 68 degrees. The common-mode bandwidth is designed to be low as it is intended to cover slow-varying on-chip variations in temperature or bias. The LPF consumes 4.4 mA of current, including the input and output buffers and CMFB error amplifier. It is worth noting that the positive feedback of the cross-coupled differential pair used to implement the active inductor should be checked for stability. By breaking the loop at the gates of these devices, the loop gain can be derived as

$$LG(s) = \frac{\frac{1}{g_m r_{o,cs}} + s\frac{Q}{\omega_0}}{\frac{s^2}{\omega_0^2} + s\frac{Q+1/Q}{\omega_0} + 1} \tag{4.19}$$

The loop gain for this design has over 75 degrees of phase margin.
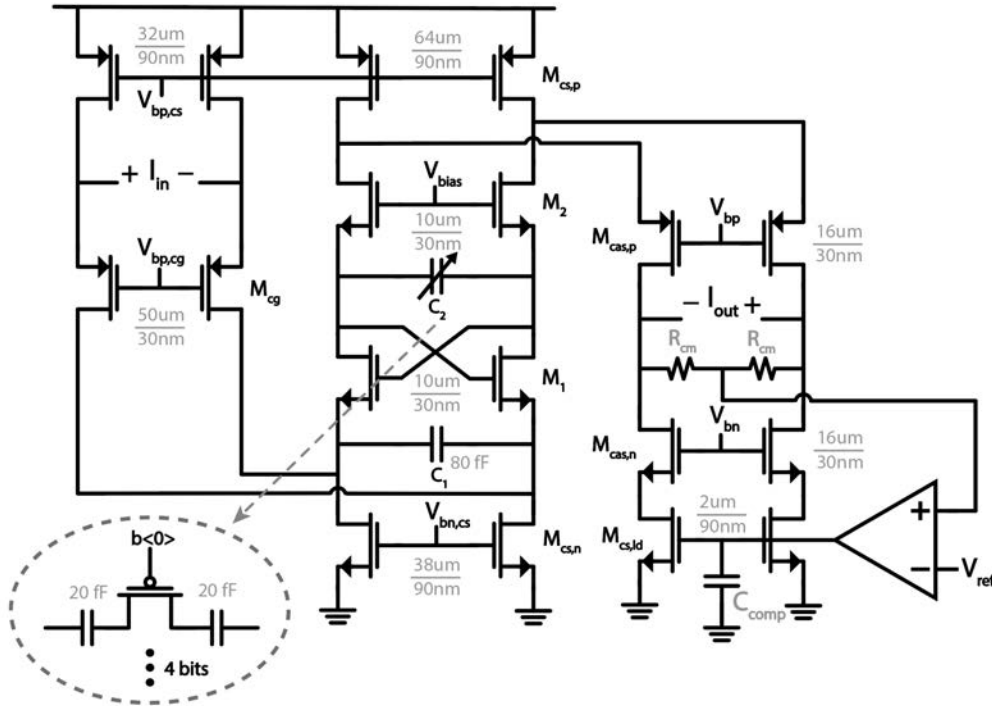

Figure 4.7: Schematic of the current-mode biquad filter with input/output stages and internal CMFB loop.

Output Buffer

Following the two Cherry-Hooper stages, the output buffer requires high linearity performance to handle a large input signal swing, as well as a low output resistance to drive the

large capacitive load presented by the ADC. Using source degeneration for sufficient linearity and a low load resistance results in signal attenuation, so the linearity-enhancing technique of parallel differential pairs explained in Section 3.4 is employed with a ratio N=2. Although this is not close to the theoretical ratio N for optimal linearity, $N = 2$ provides sufficient linearity without adding significant complexity and parasitics in the layout and routing as higher ratios would incur. PMOS diode-connected loads are used to create a low output resistance of $1/g_{m,p}$, which also define the output common-mode and removes the power overhead for a CMFB loop. Another important design consideration is bandwidth which is dictated by the output settling behavior within the sampling clock cycle of the ADC, or more precisely, the tolerable dynamic error. For a 6-bit ENOB ADC, by choosing a settling time of 4τ (where $\tau = \frac{1}{\omega_{bw}}$) a dynamic error of 1.8% has insignificant impact on the ADC performance. Of course, this assumes the buffer does not enter the nonlinear slewing regime from large input swings, which will nominally be at most the ADC full-scale voltage. This can be accomplished by sizing the input pair for a large enough overdrive V_{ov}. From analog circuit concepts,

$$V_{ov} > \omega_{BW} \cdot V_{o,max}/GBW_{buffer} \tag{4.20}$$

The buffer needs to have a 3dB bandwidth greater than 3.2 GHz while driving the differential 100 fF capacitive load presented by the ADC, where the required 3dB bandwidth is given by

$$f_{3dB,buffer} \geq \frac{N_\tau}{2\pi T_{samp}/2} \tag{4.21}$$

where N_τ is the number of time constants and T_{samp} is the ADC sampling clock period. The schematic of the final design is shown in Fig. 4.8. The buffer has a DC gain of -2 dB and bandwidth of 9 GHz (corresponding to more than 11τ for settling), while displaying a ICP1dB of +6.4 dBm and ORN of 2.3 nV/$\sqrt{Hz}$ and consumes 2.3 mA.

DC Offset Cancellation

DC offset is cancelled by using a 4-bit current digital-analog converter (DAC) with a full-scale current of 650 μA to introduce varying amounts of current imbalance through the drains in the differential pairs of the second-stage CH amplifier to ensure that at no point along the signal path the DC offset saturates the ABB. By controlling the polarity of the current-steering switches, the output current of the DAC, I_{OS}, can be steered to the appropriate branch to compensate for the offset that has accumulated along the signal path up to this point. The amount of offset that can be corrected is given by

$$V_{offset} = \pm k\Delta I_{offset}R_{out}, k = 0, 1, ..15 \tag{4.22}$$

which indicates that it is favorable to choose a high resistance node to minimize the amount of offset correction DC current needed. The offset correction circuitry is designed to compensate up to 430 mV of offset at the ABB output for the highest gain setting and 90 mV
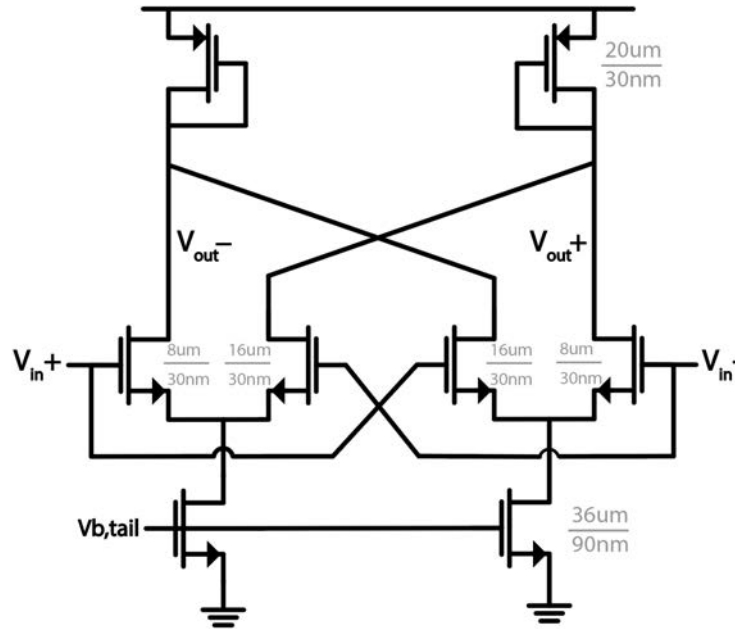
Figure 4.8: Schematic of the buffer to drive the ADC.

for the lowest gain setting, with 4 bits of step resolution.

4.4 Simulated Performance

All simulations results are with the extracted layout, with the ABB driving a 100fF capacitive load. This represents the input capacitance of the time-interleaved ADC and the significant parasitic capacitance of the H-tree routing to reach each time-interleaved slice.

Frequency Response

The ABB is first characterized by its AC small-signal performance. The small-signal frequency response of the ABB is shown in Fig. 4.9 for various gain settings, displaying the variable gain and low-pass filtering capability. The DC gain can be programmed between 3 and 39 dB, with a 3-dB bandwidth of 2.25 and 2.5 GHz, respectively. At 10 GHz, the attenuation is greater than 30 dB, highlighting the out-of-band rejection functionality provided by the LPF and TIA's. Fig. 4.10 shows the DC gain programmability with 48 digital gain control codes that are combinations of the three gain control bits of each stage's TIA. Of course, the combinations that disable all the feedback resistors for either stage are not
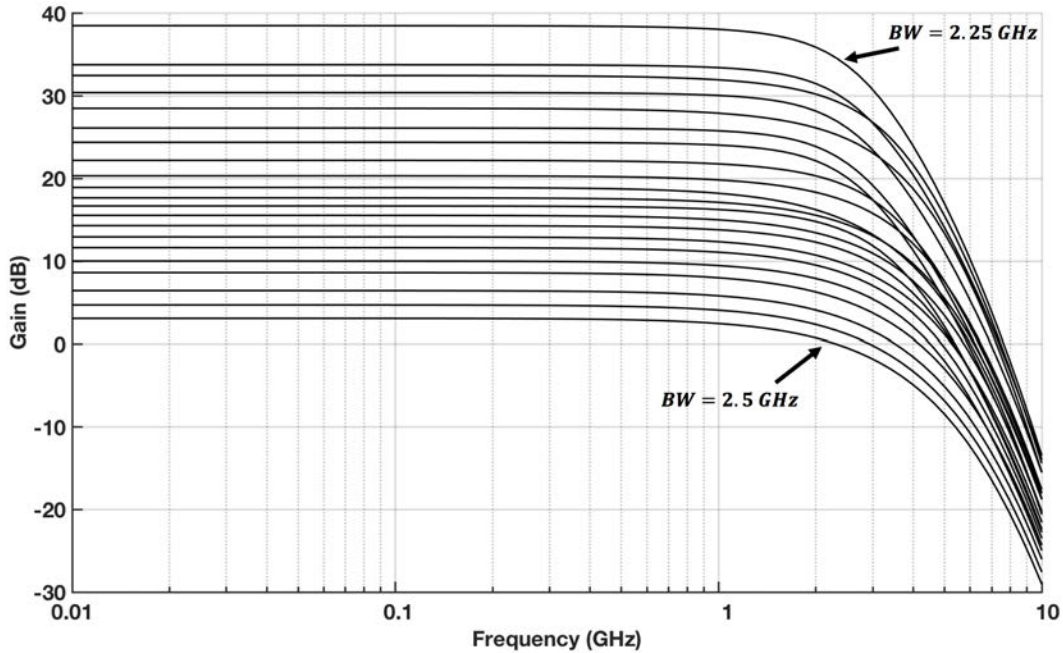
Figure 4.9: Frequency response of the ABB for various gain settings.

valid. While linear-in-dB gain characteristic was not necessary for the ABB in its intended application, the gain characteristic is roughly linear up until the highest few gain settings.

The small-signal bandwidth can be adjusted as shown in Fig. 4.11, for the lowest and highest gain settings. While not shown in the figure for clarity, the lower cutoff frequency extends close to DC due to the DC coupling of the system, and so the lower cutoff frequency is thus determined by the off-chip AC coupling which can use much larger capacitors that would not be feasible on-chip. Fig. 4.12 shows the 3-dB bandwidth of the ABB versus the 16 digital gain control codes that tune the filter cutoff frequency, for the lowest and highest gain settings. In the highest gain setting, the bandwidth can be tuned from 1.7 GHz to 2.6 GHz, while in the lowest gain setting the bandwidth can be tuned up to 3.4 GHz. Based on process variability Monte Carlo simulations, the tuning range should be sufficient to cover the $\approx 20\%$ bandwidth variation, which is asymmetrically skewed toward the lower frequency distribution. If needed, additional bandwidth tuning can be performed by adjusting the bias current of the filter to modify its g_m.

Linearity

Input and output compression are simulated with a 100 MHz input tone. The gain compression curves of gain versus input power for various gain settings are shown in Fig. 4.13,
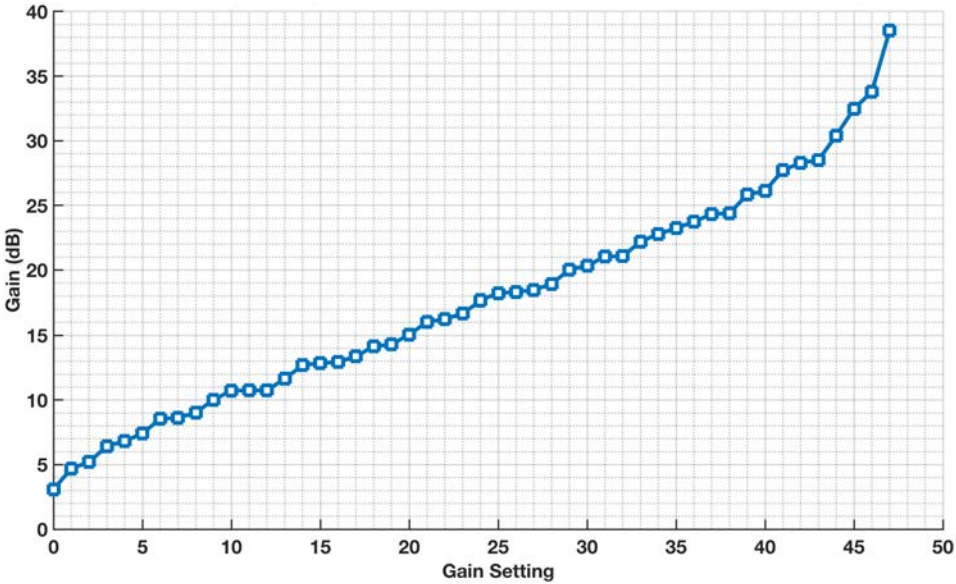
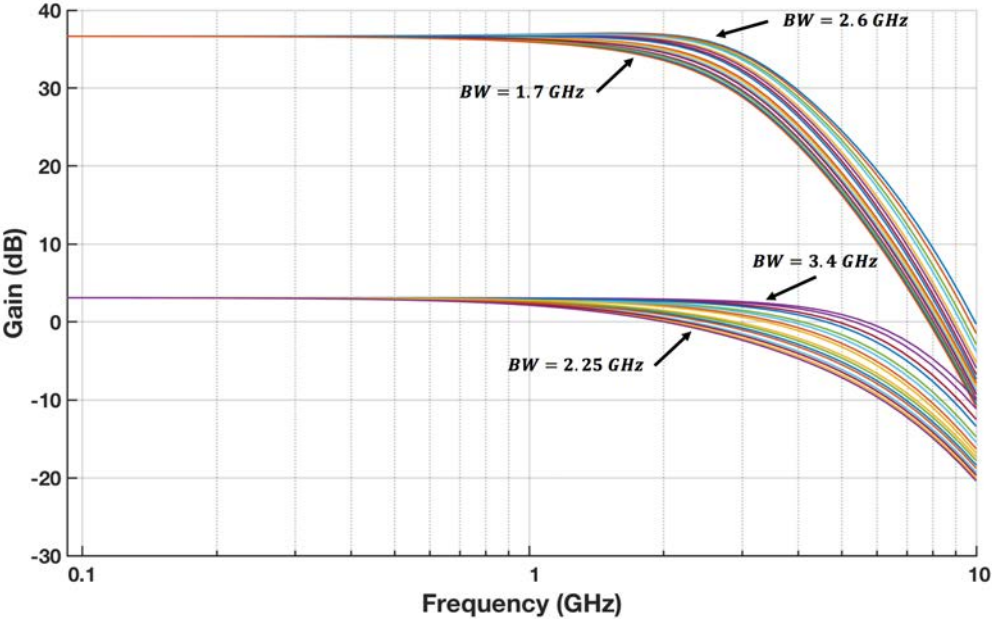Figure 4.10: DC gain vs. gain control code.

Figure 4.11: Bandwidth response with varying filter cutoff frequency settings, for highest and lowest gain settings.
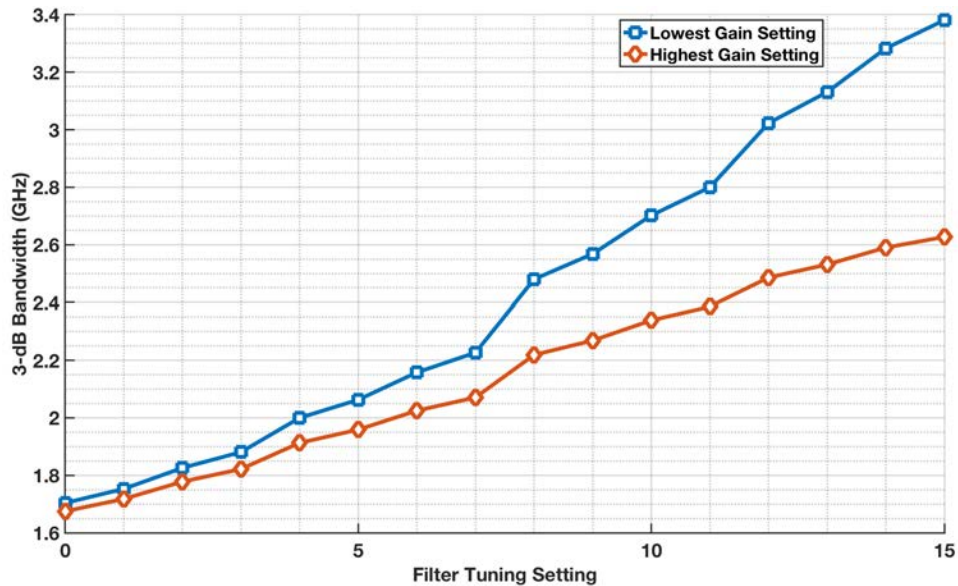
Figure 4.12: 3-dB bandwidth of the ABB (for the lowest and highest gain settings) vs. filter cutoff frequency tuning code.

displaying ICP1dB of -40.1 dBm and -15.9 dBm for the highest and lowest gain settings, respectively.

To characterize the linearity of the ABB, a 100 MHz input tone is used to simulate the input and output gain compression points. The ICP1dB and OCP1dB versus gain setting, with higher codes having higher gain, are shown in Fig. 4.14. Observing the overall trend, the ICP1dB decreases as the VGA gain increases, and the OP1dB increases as the VGA gain increases, both of which are expected trends for a typical VGA. However, it is seen that certain gain settings are not consistent with this trend and display degraded linearity performance. These settings are those in which the first stage CH amplifier has very high gain, while the second stage CH amplifier has very low gain. The high gain of the first stage imposes a high input linearity on the second stage Gm as it receives a large input signal. To exacerbate the issue, the preceding inverter based TIA sets the input common mode to mid rail, limiting the V_{gs}, and hence overdrive, available to the differential pair, making it difficult to remain linear for a large input range with a small overdrive voltage. The low gain of the second stage implies the feedback resistance of the TIA is low, which increases the output loading on the TIA amplifier and thus decreases its loop gain. This increases the impedance seen looking into the TIA and results in greater voltage swings at the TIA input, further hurting the linearity. These settings can be avoided in use if fine gain control is not needed. Nonetheless, all gain settings achieve an OCP1dB greater than -4 dBm, with the majority of settings having a OCP1dB much greater than 0 dBm.
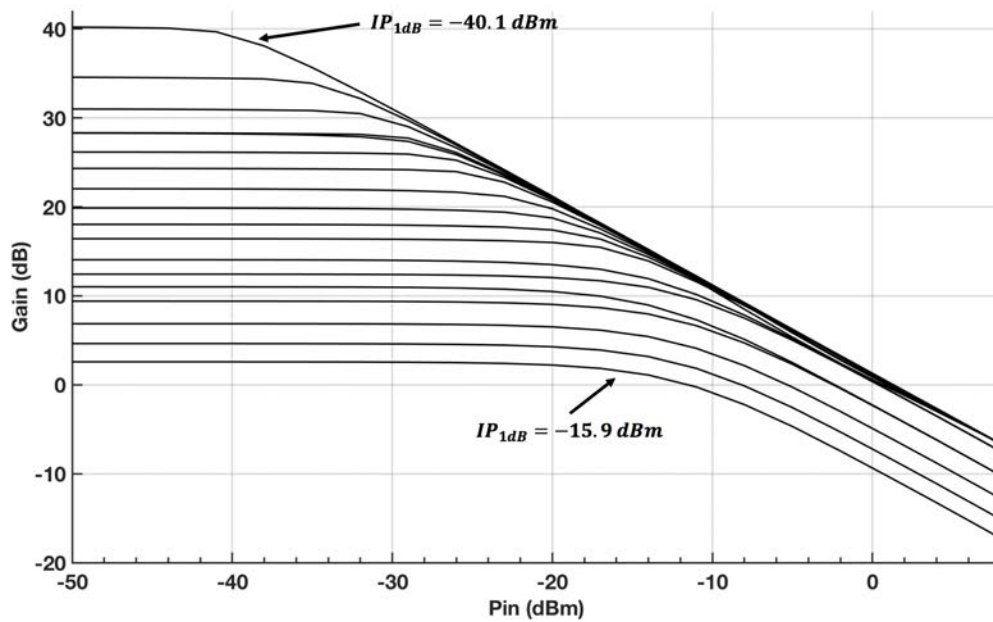
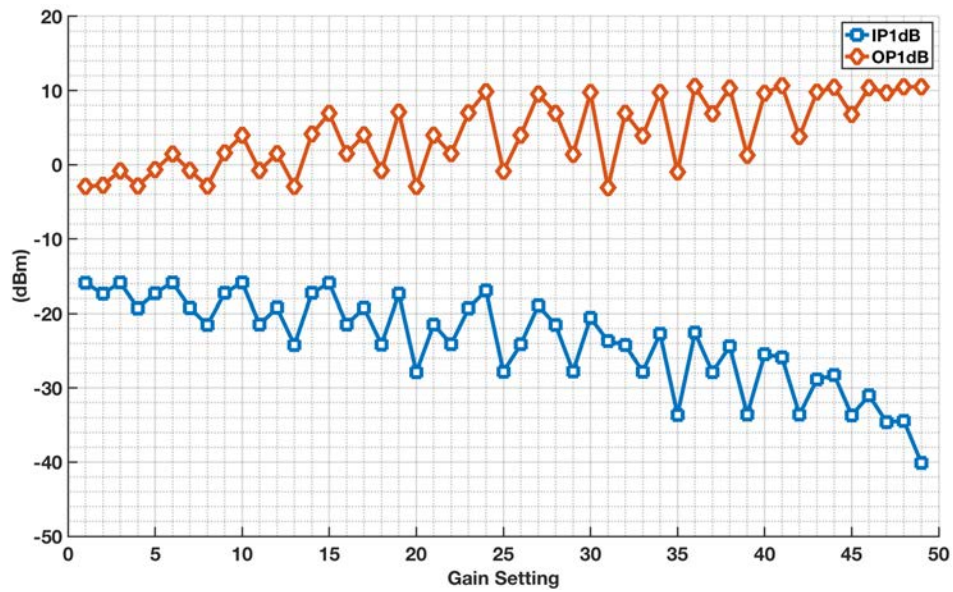Figure 4.13: Gain compression curves for various gain settings.

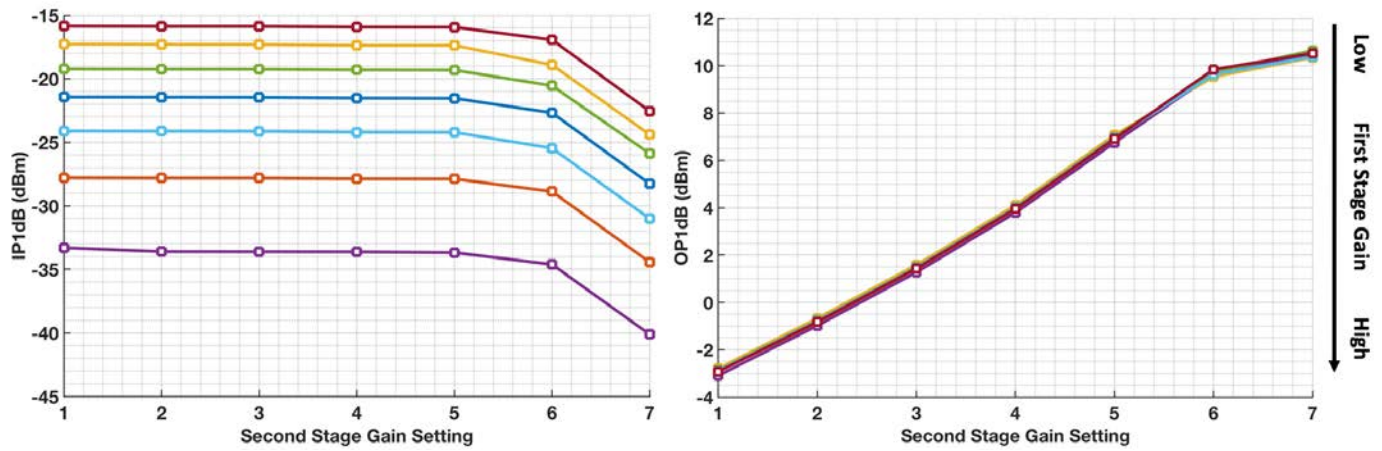Figure 4.14: ICP1dB and OCP1dB vs. gain setting.

Figure 4.15: ICP1dB (left) and OCP1dB (right) vs. second-stage gain setting for various first-stage gain settings.

Fig. 4.15 shows the ICP1dB and OCP1dB versus the second-stage gain settings, where a higher setting corresponds to higher gain; each curve is plotted for a constant first-stage CH amplifier gain setting. It is observed that the ICP1dB is mainly determined by the gain of the first stage, implying that the second stage is input limited. The OCP1dB is nearly identical for all first-stage gain settings.

Noise

Next, the noise performance of the ABB is characterized. The NF for various gain settings is shown in Fig. 4.16. At 1 GHz, approximately the middle of the bandwidth, a NF of 9.8 dB for the maximum gain setting and 15.4 dB for the minimum gain setting are obtained. As expected, the NF is almost entirely dependent on the gain setting of the first CH amplifier stage. The main noise contributors are thermal noise from extracted resistive components and the flicker noise from the NMOS devices, both in the first stage. Fig. 4.17 shows NF versus the second-stage gain settings, where a higher setting corresponds to higher gain; each curve is plotted for a constant first-stage gain setting. As expected, the NF is mainly determined by the first-stage gain setting. The lesser impact on NF by the second-stage justifies its role for fine gain tuning with robust NF performance.

DC Offset

The DC offset of the ABB must also be verified to ensure its expected levels are compatible with the calibration circuitry for cancellation. Using a Monte Carlo simulation with mismatch, the output-referred DC offset at the maximum gain setting is determined to have
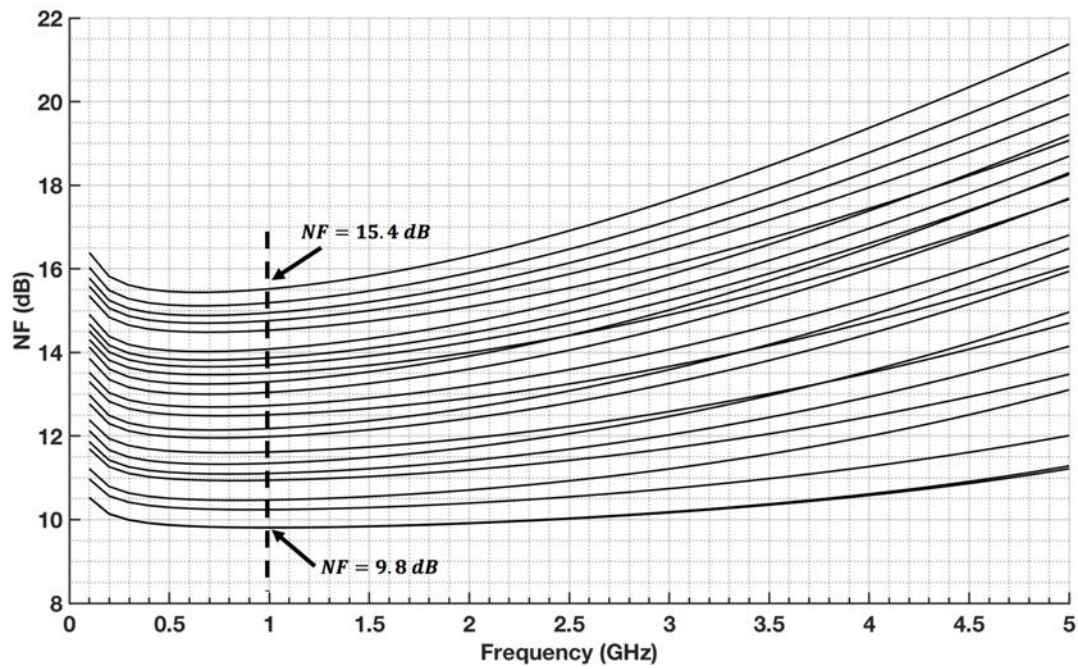
Figure 4.16: NF for various gain settings.
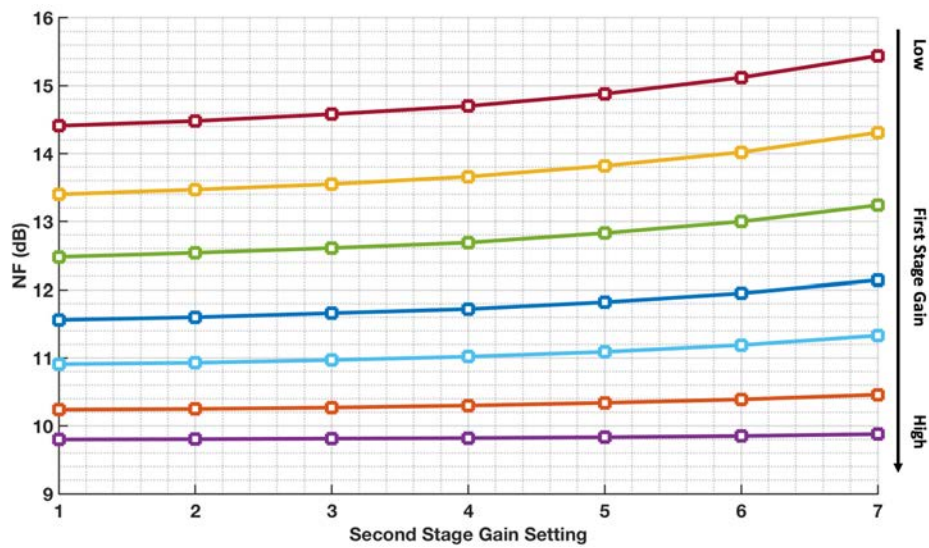
Figure 4.17: NF vs. second-stage gain setting for various first-stage gain settings.

Technology	28nm CMOS		Power Consumption
DC Gain (dB)	3 - 39	1st-stage CH	7.1 mW
Bandwidth (GHz)	2.5	2nd-stage CH (Gm+TIA)	2.2 mW
Noise Figure (dB)	9.8 - 15.4	Filter	4.4 mW
ICP1dB (dBm)	-15.9 - -40.1	Buffer	2.3 mW
Supply (V)	1	**ABB**	**16 mW**

Table 4.1: Performance and power consumption distribution of the ABB in 28nm.

be 3σ=550 mV. For the minimum gain setting, the output-referred DC offset is 3σ=20 mV. Thus, the DC offset correction design mentioned previously in this chapter has sufficient dynamic range and resolution within the $\pm\ 3\sigma$ range to calibrate the residual output-referred DC offset to below 2 LSB's of the ADC.

The simulated performance and power consumption distribution is summarized in Table 4.1.

Chapter 5

Schematic and Layout Generator Design

As CMOS technology continues to scale, analog design becomes more time consuming due to the increasing complexity of the layout design rules, and modifications to the layout design can thus be costly. Modifications to the layout are inevitable because post-layout effects can significantly impact the initial design's performance, especially at high frequency. In particular, for the ABB layout can impact the performance in the following ways:

- **Bandwidth**: Due to the short channel lengths of the devices in a continuous row of fingers, the parasitic capacitance between gate, drain, and source metallization stacks from M1 to higher level metals is substantial. This capacitance can be between the terminals for one finger or between the terminals for two adjacent fingers. Overall, the bandwidth can be significantly decreased.

- **Matching**: Any asymmetries in layout between the devices and routing of the differential signals will impact the matching. DC offset will be affected by random mismatches in V_T if an insufficient number of edge dummy devices are included or due to process gradients, and by systematic mismatches if high-current routing are different lengths because that will lead to different amounts of DC IR drop. Beyond DC, asymmetries will lead to nonzero CM-DM gain ($A_{v,CM-DM}$) which is highly undesirable as unwanted CM components that are expected to be rejected are erroneously being converted and included in the DM signal. Moreover, this implies second-order distortion products will remain in the differential signal and degrade second-order nonlinearity metrics like $IIP2$ and $HD2$.

- **Noise**: While the noise models of the active devices are dependent on various device layout parameters, parasitic resistance introduced by routing and interconnects introduces additional sources of thermal noise. For example, bias voltages on high resistance routing can introduce thermal noise to the signal path if appropriate filtering is not included. Layout also introduces new coupling paths of noise from nearby circuits to

sensitive signal lines, such as through parasitic capacitive coupling from routing or through the substrate. Thus, shielding and isolation with ground planes and guard rings should be considered.

- **Common-Mode Rejection**: The parasitic capacitance also decreases the CM rejection of unwanted CM noise and interference coupled from the supply or substrate, by decreasing the output impedance of tail current sources.

In addition, porting the design to a new technology can also be very time consuming even if the circuit topology remains completely fixed, as a new set of layout rules must be followed. In particular, a layout ported between 16nm FF and 28nm bulk has to account for many differences in the front-end-of-line (FEOL) design rules due to the inherent device processing differences between FinFET and bulk planar CMOS. In addition, differences in the back-end-of-line (BEOL) design rules must be accounted for as well, such as different number of metal layers, minimum metal spacings, and electromigration width constraints. Therefore, layout and schematic generators in BAG can significantly speed up this iterative layout process by quickly generating DRC/LVS clean layouts and schematics when design values are changed.

5.1   Layout Generator Design

It must be emphasized that contrary to popular belief, BAG *does not* automatically generate layouts based on some internal algorithm or flow. Instead, the layout generator aims to capture the layout as a generic arrangement of transistors, passives, and building blocks higher in the layout hierarchy, as well as routing between all these components. Thus, the adaptability and robustness of the generator to produce DRC/LVS clean layouts that are also optimized for analog design is solely dependent on the ability of the designer writing the generator to account for both proper analog layout techniques (e.g. common centroid, inter-digitation, dummies, shielding) and robustness to ensure the layout is dynamic enough to change as expected for different input parameters or even technologies.

In order to make the layout generator technology-agnostic and flexible for dynamic input parameters, BAG provides an extensive framework that encodes the layout design space in a quantized grid that references "tracks", which becomes the generic unit used for referencing spacing in both cell placement and routing. This value is typically some factor of the technology's minimum metal spacing, and all spacing, routing, and placement in the layout is constrained to an integer or half-integer number of tracks. It is also helpful to highlight the distinction between scripts based on the two distinct classes of *AnalogBase*, which is meant for FEOL layouts with MOS devices and the first few metal layers, and *TemplateBase* scripts, which is meant for BEOL floorplan and routing between lower hierarchy layouts. The BAG layout framework is described in detail in [2].

For example, Fig. 5.1 shows the floorplan for the second-stage Cherry-Hooper amplifier Gm cell, arranged in such a way that each row only contains devices that would both be changed together by the same amount. That way, if the number of fingers for the devices is changed as a dynamic input parameter, the devices can simply expand to the sides in the direction indicated by the arrows while maintaining a symmetrical layout. However, notice that placement of the MOM capacitors could be moved into the same row as the switch in order to decrease the height and routing length of the layout, if the possibility of the switch device having an increased number of fingers was not a possible case to consider. In general, a layout encoded in a generator will use regular, repetitive, and low-complexity floorplans in order to make it easier for the layout to be adaptable as parameters change. This usually leads to longer and less efficient routing and thus the generated layout usually is not optimal for the specific set of input parameters. This is exacerbated by the fact that BAG restricts any two consecutive metal layers from routing in the same horizontal or vertical direction. Another issue is that a better layout generator accounts for additional possible cases. For example, as the number of fingers for the input pair is increased, the layout may exceed a desired maximum width or aspect ratio. In that case, the generator needs to be flexible enough to detect this case and split the transistor into multiple rows.
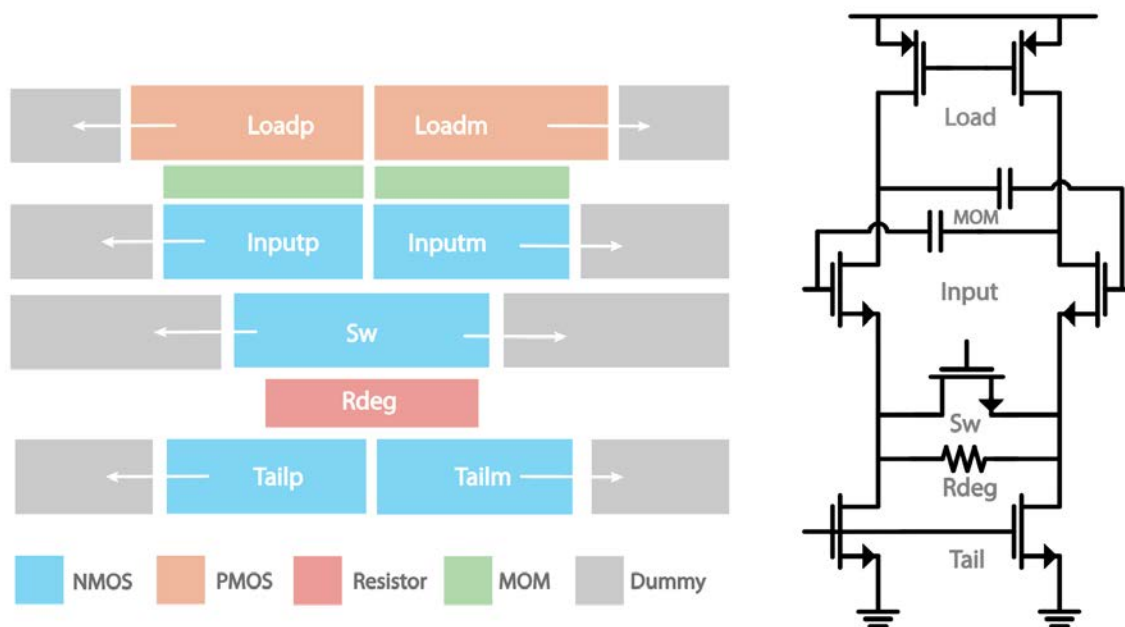

Figure 5.1: Example layout floorplan of the second stage transconductance cell.

An extensive amount of layout and schematic generators were written for this work, starting with ubiquitous sub-blocks in *AnalogBase* like differential pairs, current sources, active loads, cascodes, and switches. Parameters common to all these blocks include channel

length, finger width, number of fingers, threshold flavor, and whether to add a guard ring. Thus, this encompasses parameters such as current mirror ratios and drive strength ratios. Passives like resistors and MOM capacitors also have their own respective classes and parameters such as width, length, and number of metal layers to specify their values. From there, generators for various blocks of the ABB can be written in *TemplateBase* using the aforementioned sub-blocks as fundamental building blocks:

- **Programmable String Resistor Ladder**: Used for the input attenuator.
  Parameters: Resistor segment size, number of segments, aspect ratio of the resistor ladder

- **Differential Cascoded Inverter**: Used for the first-stage CH amplifier transconductance cells, utilizing a shared junction cascode unit cell layout, as shown in Fig. 5.2(b). This type of layout reduces the parasitic capacitances at the intermediate cascode node X by sharing the diffusion for the active device's drain and cascode device's source. Because this node is usually internal, the diffusion-metal contacts can also be removed, further reducing the capacitance. The lower capacitance at this node improves the high frequency characteristics of the cascode structure, as explained in section 3.2.
  Parameters: whether to use a triode MOS resistor in feedback or a programmable bank of feedback passives, number of bits for feedback resistors and capacitors and their unit sizes

- **Neutralized Differential Pair**: Used for the second-stage CH amplifier transconductance cells and output buffer. The neutralization capacitance is first obtained implicitly, i.e. with intentionally large capacitive routing parasitics between same-polarity gate and drain lines using a unit cell layout shown in Fig. 5.2(a). If more capacitance is needed, MOM capacitors are additionally used.
  Parameters: whether to include source degeneration, value of the degeneration resistor, neutralization capacitor size

- **Current-mode Filter**: Used for the current-mode filter in the second-stage CH amplifier and includes its input/output stages.
  Parameters: number of bits for the frequency tuning capacitor bank, unit capacitor size

- **Folded Cascode Op amp**: Used for the CMFB error amplifier for both the filter and second-stage CH amplifier stages.
  Parameters: filtering resistor and compensation capacitor sizes, whether to use NMOS or PMOS input.

- **Top Level**: The top-level generator that performs the inter-stage signal path routing on all the aforementioned blocks, as well as the routing for the CMFB loop, biasing, and digital calibration signals. Current mirror bias and supply decoupling capacitors are also added where there is empty space.

Parameters: Number of stages, amount and placement of decoupling capacitors, width and spacing of overlaid top-metal power grid for supply/ground.

As of the time of this work, BAG is not optimized yet for high-frequency layouts, and so does not support built-in functions to create shared junction cascode and implicitly neutralized differential pair layouts, with unit cells as shown in Fig. 5.2. Therefore, in order to prevent this aspect of the layout from becoming a limiting factor, custom generators to support such layouts were written. Parameters such as the spacing between gate and drain routing lines (brought up the top thin-metal layer) and width of the lines were included.
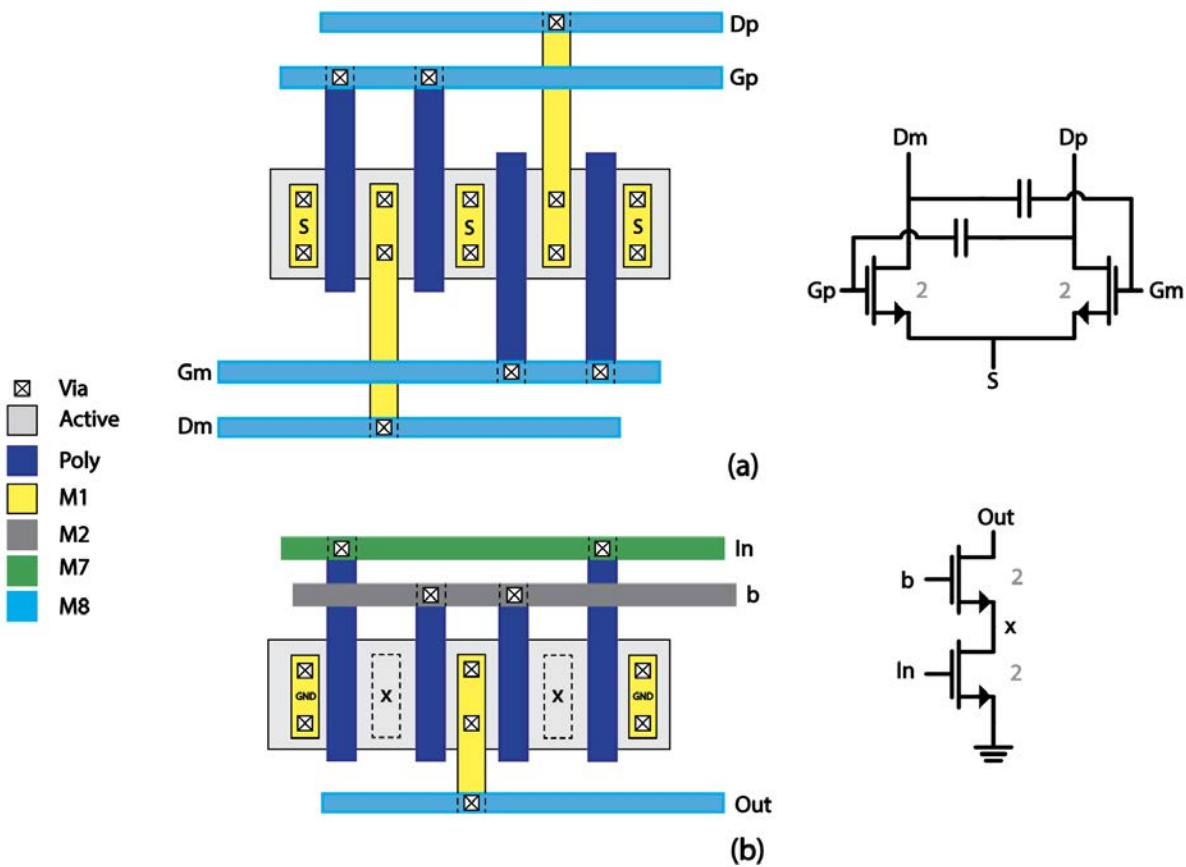

Figure 5.2: (a) Implicitly neutralized differential pair unit cell layout and (b) shared junction cascode unit cell layout.

5.2 Schematic Generator Design

After the layout is generated, a schematic generator can then map component properties of the circuit layout to a corresponding schematic. This is done after the layout is generated because dynamically-calculated information specific to the layout, such as the number of dummies for matching and alignment, is needed. Once this information is obtained, it is passed alongside the input parameters to the schematic generator to assign component values of devices to a schematic template. The schematic template is simply a human-readable netlist of a certain circuit topology with generic wrapper components whose properties and values will need to be assigned by the schematic generator. The schematic generator is also flexible in that it can also reconnect different nets and array or delete component instances.

5.3 28nm Test Chip

The design was taped out for standalone testing on a 1.2 x 2.8 mm die in 28nm bulk CMOS technology on a flip-chip I/O run. The layout of the test chip is shown in Fig. 5.3(a). Both input and output pads are differential. All the bias currents are generated from an on-chip 5-bit current DAC that uses a precise off-chip current reference. The two CMFB reference voltages are generated from an on-chip 6-bit resistor ladder voltage DAC. As mentioned previously, the input is capacitively coupled off-chip and is also resistively terminated with a large decoupling capacitor connected at the termination 50Ω resistor common-mode node to provide a clean AC short to ground. The 50Ω terminated output driver [1], shown in Fig. 5.3(c), is inductively peaked with a spiral inductor to support a 5 GHz bandwidth and source degenerated for higher linearity, and consumes 4 mA. All the digital bits for bias currents and voltages, gain and cutoff frequency tuning, and offset cancellation can be programmed with an on-chip scan chain.

5.4 16nm Design

By leveraging the layout and schematic generators, a similar ABB design was generated and in 16nm FF, targeting a more relaxed set of specifications for gain and bandwidth for the same application, of only 24 dB of gain and 1 GHz of bandwidth. The amount of time typically spent porting a layout design from one process to another was reduced significantly by using the generators once the schematic design was resized. The two generated ABB core layouts in 28nm bulk and 16nm FF are shown in Fig. 5.5.

To better optimize the performance for a given set of specifications and technology parameters, several minor topology modifications are added on top of the existing generator design to augment the design flexibility. Due to the lower gain-bandwidth requirements and

---

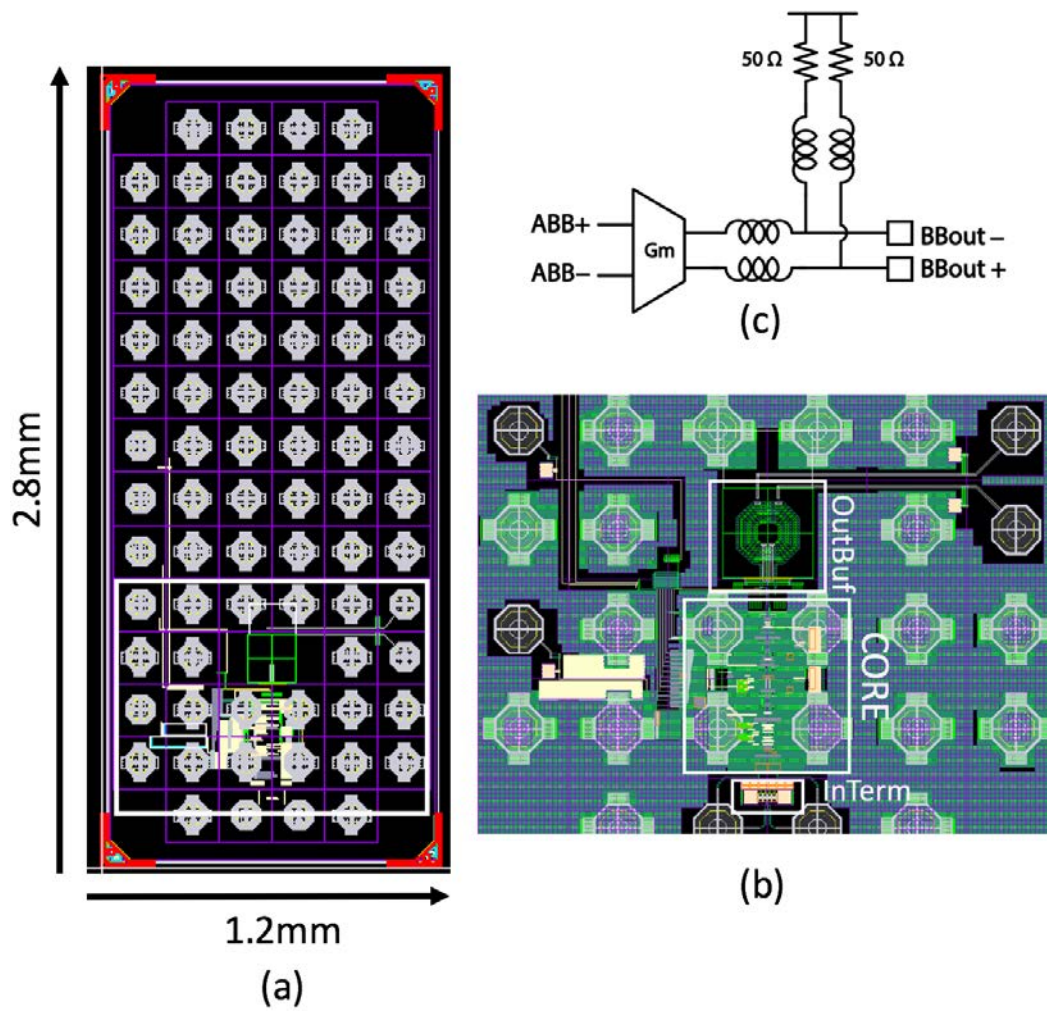[1]Thanks to Lorenzo Iotti for the output driver used in this tapeout.

Figure 5.3: (a) Test chip layout, (b) expanded view of active area highlighted in white on the test chip layout, and (c) schematic of output buffer.

the higher intrinsic gain of FinFET devices, a single-stage design is used to minimize power, shown in Fig. 5.4. Moreover, the Gm stage forgoes a pseudo-differential cascode inverter for a current-starved inverter design that now provides common-mode rejection. The bias voltages for its current sources must be provided by common-mode feedback. The buffer has an input pair mismatch of ratio N = 1 (which collapses into a conventional differential pair), with capacitive degeneration to provide peaking in the frequency response to extend the bandwidth. The attenuator, filter, and TIA Gm topologies remain exactly identical as in the 28nm iteration.
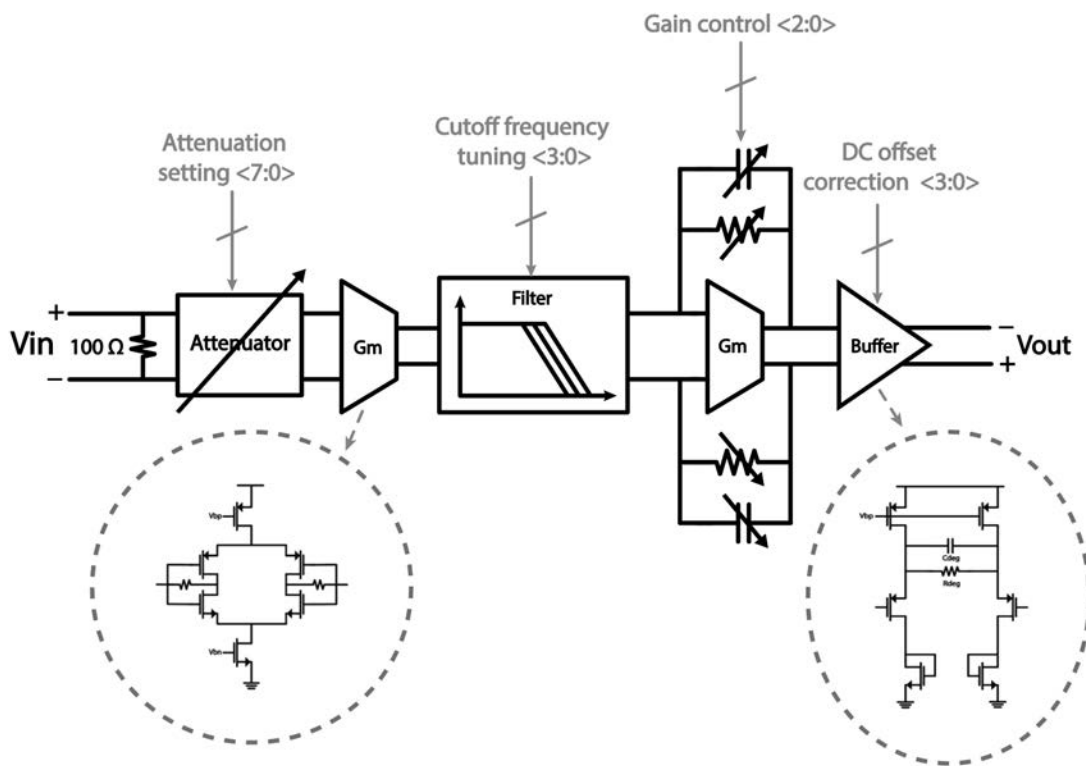
Figure 5.4: Block-level schematic of the ABB 16nm iteration, highlighting certain circuit level implementations.

Due to licensing issues with TSMC 16nm FF at the time of this work, simulated performance plots for this design are unavailable. However, the simulated performance of the extracted layout and power consumption distribution is listed in Table 5.1. This ABB in 16nm is integrated into a complete 16-channel baseband Hydra Spine ASIC, which includes ADC's, digital signal processing, and clock distribution.
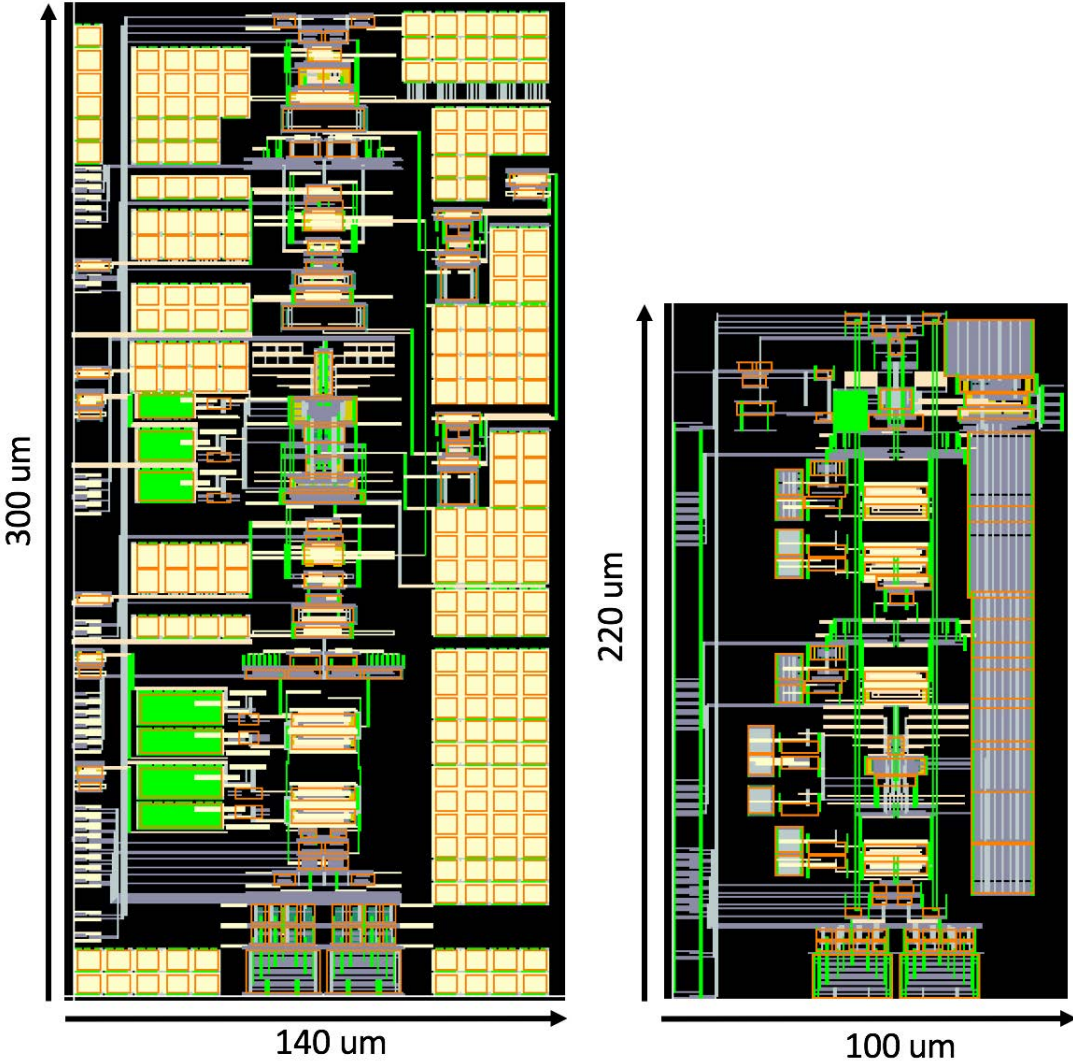
Figure 5.5: Layout of the ABB core in 28nm (left) and in 16nm FF (right). The overlaying power grid comprised of the top two metal layers is not shown.

Technology	16nm FF		Power Consumption
DC Gain (dB)	8 - 30	Gm	2.2 mW
Bandwidth (GHz)	1	Filter	2.0 mW
Noise Figure (dB)	10.1 - 17.4	TIA	2.4 mW
ICP1dB (dBm)	-15.1 - -34.1	Buffer	1.2 mW
Supply (V)	0.9	**ABB**	**7.8 mW**

Table 5.1: Performance and power consumption distribution of the ABB in 16nm.

Chapter 6

Conclusions

The challenges and design solutions as well as methodologies of high bandwidth analog baseband sections for mm-Wave receivers have been presented here. The proposed design incorporates several bandwidth, filtering, and linearity enhancement techniques to address challenges encountered in high gain-bandwidth designs. Furthermore, the adaptation of the layout design into an analog generator enables greater design space exploration as important post-layout effects are more quickly obtained, alongside greater reusability regarding porting the design across different technologies. Two designs have been generated in both 28nm CMOS and 16nm FF, and a comparison with the state-of-the-art in Table 6.1 indicates competitive performance with recently reported analog baseband sections for mm-Wave receivers. The 28nm CMOS prototype has been fabricated in a standalone testing setup and is pending measurements.

There are several future improvements that can be applied to this work:

1. **Sizing Script**: While the generators for creating corresponding layouts and schematics have been implemented for the circuits presented in this work, the device sizing that is inputted to these generators is still determined in a traditional analog design manner - a mix of simulation result interpretations and local first-pass optimizations using external scripting tools. A sizing script can capture this iterative process by leveraging python scripting in BAG with automated simulation testbenches to verify and update the sizing values. This would close the design loop and enable full design automation from specifications to generated schematic and layout that meets these specifications.

2. **More Dynamic Generator Script**: The layout generator script can be improved to be more dynamic. It is inevitable that certain aspects of the layout will be assumed to be fixed on first-pass generator designs, and refined to be dynamic or parameterized on upon porting the design to different technologies, where the assumptions will show themselves to be problematic. Examples include the spacing of certain blocks relative to unrelated features in the layout, the power grid structure, and the routing path of certain signals that need to traverse long distances. Analog layout is very particular to

	This Work (sim)		[25]	[26]	[6]	[19]	[24]	[13]
Technology	**28nm**	**16nm**	65nm	90nm	90nm	40nm	40nm	65nm
DC Gain (dB)	**3 - 39**	**8 - 30**	3 - 31	-10 - 50	0.1 - 19.6	0 - 40	10.6 - 30	2 - 32
Bandwidth (GHz)	**2.5**	**1**	0.98 - 1	2.2	0.95	1	0.97 - 1.1	0.9
Noise Figure (dB)	**9.8 - 15**	**10 - 17**	6 - 21	17 - 30	20 - 35	N/A	N/A	N/A
ICP1dB (dBm)	**-16 - -40**	**-15 - -34**	-4 - -31	-13 - -55	-4.4$^{\dagger\sharp}$	N/A	N/A	-18 - -34
Supply (V)	**1**	**0.9**	1.1	1	1	1.1	1.1	1.2
Power (mW)	**16**	**7.8**	32 - 48*	2.5	10.8	18*	30 - 42*	24.8*
Core Area (mm^2)	**0.042**	**0.015**	0.2*	0.01	0.15	0.16*	0.54*	0.24*

Table 6.1: Performance Summary and Comparison

* I + Q channels
† ICP1dB is theoretically calculated from IIP3
♯ @ 0dB gain

the design, and capturing this unpredictability for different incarnations of the design in methodological way will always have room for improvement to create a truly dynamic generator for simpler future reuse.

Bibliography

[1] D. Cai et al. "Design of Ultra-Low-Power 60-GHz Direct-Conversion Receivers in 65-nm CMOS". In: *IEEE Transactions on Microwave Theory and Techniques* (2013).

[2] E. Chang et al. "BAG2: A Process-Portable Framework for Generator-Based AMS Circuit Design". In: *IEEE Custom Integrated Circuits Conference* (2018).

[3] E.M. Cherry and D.E. Hooper. "The design of wide-band transistor feedback amplifiers". In: *Proceedings of the Institution of Electrical Engineers* 110.2 (1963), pp. 375–389.

[4] S. D'Amico et al. "A 9.5mW analog baseband RX section for 60GHz communications in 90nm CMOS". In: *Radio Frequency Integrated Circuits Symposium* (2011).

[5] S. D'Amico et al. "A CMOS 5 nV/$\sqrt{Hz}$ Hz 74-dB-Gain-Range 82-dB-DR Multistandard Baseband Chain for Bluetooth, UMTS, and WLAN". In: *IEEE Journal of Sold-State Circuits* 43.7 (2008), pp. 1534–1541.

[6] S. D'Amico et al. "A Low-Power Analog Baseband Section for 60-GHz Receivers in 90-nm CMOS". In: *IEEE Transactions on Microwave Theory and Techniques* 62.8 (2014), pp. 1724–1735.

[7] Hooman Darabi. *Radio Frequency Integrated Circuits and Systems*. Cambridge University Press, 2015.

[8] H. Elwan, A. Tekin, and K. Pedrotti. "A Low-Noise Analog Baseband in 65nm CMOS". In: *IEEE Custom Integrated Circuits Conference* (2010).

[9] M. Ghanad, C. Dehollain, and M. Green. "TIA Linearity Analysis for Current Mode Receivers". In: *IEEE International New Circuits and Systems Conference* (2018).

[10] N. Ghittori et al. "Analog baseband channel for GSM/UMTS/WLAN/Bluetooth reconfigurable multistandard terminals". In: *IEEE International Symposium on Circuits and Systems* (2006).

[11] S. Gilal and B. Razavi. "10-Gb/s Limiting Amplifier and Laser/Modulator Driver in 0.18-um CMOS Technology". In: *IEEE Journal of Sold-State Circuits* 8.12 (2003), pp. 2138–2146.

[12] B. Gilbert. "The multi-tanh principle: a tutorial overview". In: *IEEE Journal of Solid-State Circuits* 33.1 (1998), pp. 2–17.

[13] Masahiro Hosoya, Toshiya Mitomo, and Osamu Watanabe. "A 900-MHz bandwidth analog baseband circuit with 1-dB step and 30-dB gain dynamic range". In: *ESSCIRC* 55 (2010).

[14] L. Iotti, S. Krishnamurthy, and A.M. Niknejad. "A Low-Power 70–100-GHz Mixer-First RX Leveraging Frequency-Translational Feedback". In: *IEEE Journal of Sold-State Circuitss* 45.9 (2020), pp. 1770–1780.

[15] L. Iotti, G. LaCaille, and A. M. Niknejad. "A 12mW 70-to-100GHz Mixer-First Receiver Front-End for mm-Wave Massive-MIMO Arrays in 28nm CMOS". In: *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)* (2018).

[16] G. LaCaille et al. "Design and Demonstration of a Scalable Massive MIMO Uplink at E-Band". In: *IEEE Conference on Communications* (2020).

[17] E. G. Larsson et al. "Massive MIMO for next generation wireless systems". In: *IEEE Commun. Mag.,* 52.2 (2014), pp. 186–195.

[18] Thomas H. Lee. *The Design of CMOS Radio-Frequency Integrated Circuits.* Cambridge University Press, 2003.

[19] M. Miyahara et al. "An 84 mW 0.36 mm2 Analog Baseband Circuits for 60 GHz Wireless Transceiver in 40 nm CMOS". In: *Radio Frequency Integrated Circuits Symposium* (2012).

[20] A. Pirola, A. Liscidini, and R. Castello. "Current-Mode, WCDMA Channel Filter With In-Band Noise Shaping". In: *IEEE Journal of Sold-State Circuitss* 45.9 (2010), pp. 1770–1780.

[21] A. Puglielli et al. "Design of Energy- and Cost-Efficient Massive MIMO Arrays". In: *Proc. IEEE* 104.3 (2016), pp. 586–606.

[22] Behzad Razavi. *Design of Analog CMOS Integrated Circuits.* McGraw-Hill, 2016.

[23] E. Säckinger and W. C. Fischer. "A 3-GHz 32-dB CMOS Limiting Amplifier for SONET OC-48 Receivers". In: *IEEE Journal of Sold-State Circuits* 35.12 (2000), pp. 1884–1888.

[24] V. Szortyka et al. "A 42mW Wideband Baseband Receiver Section with Beamforming Functionality for 60GHz Applications in 40nm Low-Power CMOSS". In: *Radio Frequency Integrated Circuits Symposium* (2012).

[25] Y. Wang et al. "A Linear-in-dB Analog Baseband Circuit for Low Power 60GHz Receiver in Standard 65nm CMOS". In: *Radio Frequency Integrated Circuits Symposium* (2013).

[26] Y. Wang et al. "Design of a Low Power, Inductorless Wideband Variable-Gain Amplifier for High-Speed Receiver Systems". In: *IEEE Transaction on Circuits and Systems—I: Regular Papers* 59.4 (2012), pp. 696–707.

[27] C. Wu, J. Liao, and S. Liu. "A 1V 4.2mW Fully Integrated 2.5Gb/s CMOS Limiting Amplifier using Folded Active Inductors". In: *Proceedings of IEEE Symposium on Circuits and Systems* (2004), pp. 1044–1047.

[28] H. Zhang and E. Sánchez-Sinencio. "Linearization Techniques for CMOS Low Noise Amplifiers: A Tutorial". In: *IEEE Transaction on Circuits and Systems—I: Regular Papers* 58.1 (2011), pp. 22–36.