A 56-GHz Fractional-\(N\) PLL With 110-fs Jitter

Yu Zhao*, Member, IEEE, Onur Memioglu, Member, IEEE, Long Kong*, Member, IEEE, and Behzad Razavi*, Fellow, IEEE

Abstract—A fractional-\(N\) phase-locked loop (PLL) architecture incorporates a switched-current finite impulse response (FIR) filter to suppress the \(\Delta \Sigma\) modulator (\(\Delta \Sigma M\)) noise. Using a compact, low-power divide-by-8 circuit and realized in 28-nm CMOS technology, the PLL exhibits a phase noise of \(-98\) dBc/Hz at 1-MHz offset in the fractional-\(N\) mode while consuming 23 mW and occupying an active area of 0.1 mm\(^2\).

Index Terms—\(\Delta \Sigma\) noise, fractional-\(N\) synthesis, master-slave sampling filter, noise folding, non-linearity, phase-locked loop (PLL).

I. INTRODUCTION

The problem of clock generation with low jitter becomes more challenging as communication systems target higher performance. For example, PAM4 transmitters operating at 112 Gb/s or 224 Gb/s can incorporate a 56-GHz phase-locked loop (PLL) for multiplexing. Such an application poses three conditions on the design. First, the PLL jitter must remain far below the symbol period, e.g., about 100 fs, for a data rate of 224 Gb/s. Second, the PLL should preferably be realized as a fractional-\(N\) loop so as to operate with different crystal frequencies and possibly correct for crystal inaccuracies. Third, multi-lane systems make it desirable to use a low-power, compact PLL design per lane rather than distribute a 56-GHz clock across lanes and over long interconnects.

Prior fractional-\(N\) designs in this frequency range have achieved rms jitters ranging from 200 to 500 fs while consuming between 31 and 46 mW and requiring chip areas from 0.38 to 0.55 mm\(^2\) [1], [2], [3]. This article proposes a fractional-\(N\) PLL architecture and a number of circuit techniques that achieve an rms jitter of 110 fs with a power of 23 mW. Fabricated in 28-nm CMOS technology, the experimental prototype occupies an active area of 0.1 mm\(^2\).

Section II provides the background for this work. Section III introduces the proposed finite impulse response (FIR) filter and its properties and Section IV describes the PLL architecture. Sections V–VIII deal with the design of the building blocks, and Section VIII presents the experimental results.

II. BACKGROUND

A. PLL Trade-Offs

Fractional-\(N\) PLLs face a trade-off among three noise components: 1) the voltage-controlled oscillator (VCO) phase noise; 2) the \(\Delta \Sigma\) modulator (\(\Delta \Sigma M\)) quantization noise (q-noise); and 3) the reference phase noise. As the loop bandwidth (BW) decreases, so do the second and third contributions but at the cost of raising the first. For output jitter values around 100 fs\(_{\text{rms}}\), the reference phase noise is negligible if the BW is less than 10 MHz, making the trade-off between the first two the prominent issue. This motivates us to reduce the \(\Delta \Sigma\) q-noise by additional methods.

Numerous techniques have been developed to address this point [4], [5], [6], [7], [8], [9], [10], [11], [12]. For example, a digital-to-time converter (DTC) can produce in the reference path phase jumps equal to those created by the feedback divider, thereby presenting a nominally constant error to the phase detector (PD) [4], [9], [12]. This approach, however, requires DTC gain calibration and, more importantly, assumes that the DTC integral nonlinearity is small enough to negligibly fold down the high-pass \(\Delta \Sigma\) noise. This in turn demands tight matching among the DTC unit delays.

B. Use of FIR Filters

Another method filters the \(\Delta \Sigma\) noise before it reaches the VCO. In [13], for example, an FIR filter is placed after the feedback divider so that delayed copies of the divider output are combined with proper weighting (see Fig. 1). As shown in Fig. 2(a), the divider output phase jumps due to \(\Delta \Sigma\) modulation can be viewed as discrete-time samples of a function. If properly delayed, scaled, and summed, these samples yield a much “quieter” output. For the FIR filter to provide attenuation beyond, say, \(f_{\text{REF}}/10\), its unit delay must be comparable to \(T_{\text{REF}}\), a condition afforded by a chain of

Fig. 1. FIR filtering techniques.
the FIR filter. BW can reach floop filter and the VCO. It is demonstrated in [13] that the loop φnegligibly affecting the loop stability [13].

delay line are clocked by the prescaler output (at 7 GHz), the flipflops (FFs) [see Fig. 2(b)] [13]. Since all of the FFs in the delay line are clocked by the prescaler output (at 7 GHz), the excess delay introduced by the FIR filter is less than 50 ps, negligibly affecting the loop stability [13].

This scheme creates delayed copies of the divider output at φ2, φ3, etc., but scaling and summing phase quantities is not straightforward. For this reason, one can first find the phase difference between these signals and the reference, represent the differences in the voltage domain, and then perform scaling and summation (see Fig. 3) [13]. The resulting feedback signal, Vout, exhibits much less fluctuation and drives both the loop filter and the VCO. It is demonstrated in [13] that the loop BW can reach fREF/4 with negligible ΔΣ noise contribution.

Although attractive, the foregoing approach does suffer from some nonlinearity as it generates a voltage for the VCO. This point is illustrated in Fig. 4, where two XOR gates form a section of the FIR filter, and the output summation is performed by R1, R2, and C1 [13]. Feedback signal φ2 is a delayed copy of the divider output, φ1. We note that φ1 experiences phase fluctuations Δta, Δtb, etc., due to ΔΣ modulation, and φ2 reproduces these jumps after one reference period, TREF. The output switch samples Vout under the command of the reference, VREF. We show that the output samples at V5 have a nonlinear dependence on the phase fluctuations.

Suppose C1 in Fig. 4 is initially charged to VDD. At t = t1, φ2 rises and C1 begins to discharge through R2. At t = t2, φ1 goes high, causing C1 to discharge further. With the aid of superposition for V1 and V2, we obtain the sampled output as

\[V_S(t_S) = V_{out}(t_S) = V_{DD} \left[ \frac{R_1}{R_1 + R_2} \exp\left(\frac{-t_S - \Delta t_a}{\tau}\right) + \frac{R_2}{R_1 + R_2} \exp\left(\frac{-t_S - \Delta t_b}{\tau}\right) \right] (1)\]

where \(\tau = R_1 R_2 C_1 / (R_1 + R_2)\). If \(\tau \gg t_S - t_a\) and \(t_S - t_b\), then

\[V_S(t_S) \approx V_{DD} \left[ 1 - \frac{t_S}{\tau} + \frac{\Delta t_a}{R_2 C_1} + \frac{\Delta t_b}{R_1 C_1} \right] (2)\]

revealing that \(\Delta t_a\) and \(\Delta t_b\) are scaled and linearly combined— as expected of an FIR filter. This becomes clearer if we view Δt0 (the divider output phase) as a function, \(x(t)\) and note that Δt0 is equivalent to \(x(t - T_{REF})\). In practice, however, \(\tau\) cannot be arbitrarily large because it would lead to low phase detection gain and hence high phase noise contribution from the FIR filter and the following stages.

The key observation here is that the exponential action in (1) makes \(V_S\) a nonlinear function of phase jumps \(\Delta t_a, \Delta t_b\), etc. Aiming before the output summation occurs, this phenomenon folds high-pass \(\Delta \Sigma\) noise to the baseband. We remark that the nonlinearity originates from two effects. First, the charge delivered to C1 is nonlinear with respect to the phase jumps because the current flowing through the resistors changes with the output voltage. Second, the branches are never “tristated,” i.e., each resistor charges or discharges C1 at all times. The significance of these two points becomes clearer in Section III.

It can be shown that a \(K\)-tap FIR filter produces a sampled output equal to

\[V_S(t_S) = V_{DD} \sum_{k=1}^{K} R_{||} \frac{R_{||} \exp\left(-t_S - \Delta t_k\right)}{R_k} \frac{1}{\tau} \] (3)

where \(R_{||} = R_1||R_2||\cdots||R_K\) and \(\tau = R_1||C_1\). Moreover, the approximation leading to (2) can be applied here as well to obtain

\[V_S(t_S) \approx V_{DD} \left[ 1 - \frac{t_S}{\tau} + \sum_{k=1}^{K} \frac{\Delta t_k}{R_k C_1} \right] (4)\]

Fig. 5 plots the simulated output phase noise of a PLL employing such an FIR filter with \(K = 22\) and a third-order \(\Delta \Sigma\)M. All other blocks are noiseless. We observe that

---

**Fig. 2.** (a) Output of a divider driven by a ΔΣM. (b) Delay elements in the FIR filter.

**Fig. 3.** FIR scaling and summation.

**Fig. 4.** Resistor-based two-tap FIR filter.
Fig. 5. Simulated $\Delta \Sigma M$ phase noise with a 22-tap resistor-based FIR filter ($f_{\text{REF}} = 250$ MHz, $f_{\text{out}} = 7$ GHz).

Fig. 6. Simulated $\Delta \Sigma M$ phase noise with a 22-tap CP-based FIR filter ($f_{\text{REF}} = 250$ MHz, $f_{\text{out}} = 7$ GHz).

doubling $\tau$ lowers the plateau by 6 dB and the integrated jitter from 920 to 460 fs$_{\text{rms}}$. This can be predicted by writing the Taylor series for the exponential terms in (3).

An important advantage of the proposed FIR filter over charge-pump (CP) PLLs is that it avoids q-noise folding due to the up/down current mismatch. Such a mismatch manifests itself even in multi-input CPs acting as an FIR filter [14], [15]. The CP up/down current mismatch introduces non-linearity and q-noise folding [16] because the magnitude of the output current depends on the input phase error. According to simulations, a 22-tap CP-based loop with 1% mismatch suffers from considerable folding (see Fig. 6), exhibiting a rise in the integrated jitter from 4 to 81 fs$_{\text{rms}}$. Another advantage of the proposed approach is that it obviates the need for multiple feedback dividers.

III. PROPOSED FIR FILTER

A. Basic Idea

We propose a “switched-current” FIR topology that considerably reduces the nonlinearity described above. We begin with the two-tap structure shown in Fig. 7, where nominally equal current sources $I_1$ and $I_2$ are controlled by $\phi_1$ and $\phi_2$, respectively. Also depicted are the $\phi_1$ and $\phi_2$ waveforms along with the output. Assuming $V_{\text{out}}$ begins at zero and the current sources are ideal, we observe that

$$V_{\text{out}}(t) = -I_1 \Delta t_b - \frac{I_2}{C_1} \Delta t_a + \frac{I_1 + I_2}{C_1} t. \quad (5)$$

In analogy with (1), we recognize a two-tap FIR filter response here, with coefficients $\alpha_1 = -I_1/(I_1 + I_2)$ and $\alpha_2 = -I_2/(I_1 + I_2)$. Note that $V_{\text{out}}$ is sampled by $V_{\text{REF}}$ so as to perform phase comparison.

The remarkable result here is that $V_{\text{out}}(t)$ in (5) is a linear function of $\Delta t_a$ and $\Delta t_b$, thereby minimizing noise folding. The linearity is obtained fundamentally because the current sources are tristated and have a relatively high output impedance, thus allowing $C_1$ to store an amount of charge representing the phase difference between the reference and each feedback clock. By contrast, the resistive circuit of Fig. 4 continuously perturbs the output. Using the notation introduced in Section II, we write

$$V_S = \frac{I_1 + I_2}{C_1} [\alpha_1 x(t) + \alpha_2 x(t - T_{\text{REF}}) + t]. \quad (6)$$

Departures of $\alpha_1$ and $\alpha_2$ from their nominal values slightly alters the FIR transfer function but does not cause nonlinearity. For a $K$-tap realization, we have

$$V_S(t_S) = \frac{1}{C_1} \sum_{j=1}^{K} I_j \sum_{j=1}^{K} \alpha_j (t_S - \Delta t_j) \quad (7)$$

where $\alpha_j = I_j / \sum_{m=1}^{K} I_m$.

The design of the proposed switched-current FIR filter must deal with a number of questions: 1) how should $N$ and the FIR response be chosen? 2) how should the total current and $C_1$ be chosen? 3) how does the finite output impedance of the current sources affect the performance? 4) how much is the tolerable noise of the current sources? and 5) how much is the tolerable mismatch among the current sources? We address these questions in Sections III-B and V.

B. FIR Filter Implementation

The FIR filter is realized as shown in Fig. 8(a), where the feedback divider output, $\phi_1$, is delayed by a chain to produce $\phi_2, \ldots, \phi_{22}$. 

Authorized licensed use limited to: UCLA Library. Downloaded on January 05,2023 at 15:24:46 UTC from IEEE Xplore. Restrictions apply.
To extract the phase information from these signals, we perform AND functions with the reference. Thus, the pulse widths at the AND gate outputs are equal to the phase differences. Current source $j$ then converts its input pulsewidth to charge, establishing the necessary FIR coefficient, $k_j$, and $C_1$ sums the results. This capacitor begins with a zero initial condition so as to avoid intersample interference.

With the proper choice of $k_j$, the arrangement shown in Fig. 8(a) subjects the divider phase jumps to a low-pass response. It also performs phase comparisons with the reference. We thus expect $V_{out}$ to contain little q-noise.

The unit delays in Fig. 8(a) merit two remarks. First, they are clocked by $f_{VCO}/8 = 7$ GHz, where $f_{VCO}$ denotes the VCO frequency. This choice is justified in Section IV-B. Second, their delay value is selected according to the $\Delta \Sigma$ fractional code word (FCW) to be either $T_1 = 28 \times (8T_{VCO})$ or $T_2 = 29 \times (8T_{VCO})$, where $T_{VCO} = 1/f_{VCO}$. The reason for this “binary delay” is explained in Section VI.

Fig. 8(b) depicts the unit current source implementation. The cascode structure employs a timing scheme that halves the power consumption and yet achieves high linearity. Initially, both $M_1$ and $M_2$ are OFF. At the rising edge of $f_{REF}$, $M_1$ turns on, bringing $V_A$ down to a desired value. Then, on the rising edge of $\phi_j$, $M_1$ turns off, $M_2$ turns on, and $C_1$ begins to charge.

In this design, we ensure an overlap of 8 ps between $V_Y$ and $V_X$ so that $V_A$ does not take off toward $V_{DD}$ during this transition.

As shown in Fig. 8(b), the voltage developed across $C_1$ is applied to another capacitor, $C_F$, when $C_{K2}$ goes high. This action prohibits the transient on $C_1$ from reaching the VCO and causing jitter. With $C_F \approx 0.25C_1$, some charge sharing occurs, but it can be shown that this effect does not introduce nonlinearity. After $C_{K2}$ falls, $C_F$ reset $C_1$.

**IV. PROPOSED PLL ARCHITECTURE**

The proposed fractional-$N$ PLL architecture is shown in Fig. 9. A $LC$ VCO is followed by a low-power, compact $\div 8$ circuit acting as a prescaler and a multimodulus divider (MMD). The MMD output travels through a switched-current FIR filter, a PD, and a sampler, returning a voltage to a $Gm$ stage, which drives the loop filter. The FFs comprising the FIR delay units are clocked by the $\div 8$ circuit output. The third-order $\Delta \Sigma M$ is realized by a MASH 1-1-1 topology. With a word length of 20 bits, this modulator provides a frequency resolution of 2 kHz at 56 GHz, i.e., about 0.04 ppm. This is well below typical crystal frequency tolerances.

**A. Design Considerations**

The proposed architecture entails a number of considerations. First, since an MMD operating at 56 GHz would consume substantial power and require inductive peaking, we lower the clock frequency before applying it to the MMD. The trade-offs in this choice are quantified in Section IV-B.

Second, despite the limited speed of the 28-nm CMOS devices, the PLL incorporates only one inductor (in the VCO) so as to occupy a small footprint. This is made possible by a new $\div 2$ circuit topology used in the prescaler (see Section VII-A). Third, the output of the FIR filter is sampled on capacitor $C_F$ before it is reset to zero according to Fig. 8(a). Fourth, the Gm stage in Fig. 9 exhibits a low-frequency voltage gain of 30 dB, thus relaxing the voltage compliance at the FIR filter output. This greatly relaxes the design of the unit current sources, but the noise of the Gm circuit must be managed. Fifth, the loop BW is chosen equal to 4 MHz so as to minimize
the total integrated phase noise due to the VCO and the FIR-filtered contribution of the $\Delta \Sigma M$. Sixth, to save power, the VCO and divider chain employ no buffer. Thus, the VCO must absorb the input capacitance of the prescaler.

The PLL of Fig. 9 incorporates a sampling PD, generating a voltage, $V_X$, proportional to the phase error. To compute the PD gain, $K_{PD}$, we assume integer-$N$ operation and note that the FIR filter current, $I_{tot}$, flows through $C_1$ in Fig. 8(a). With a slope of $I_{tot}/C_1$, $V_X$ provides a change of $\Delta t I_{tot}/C_1$ for a phase error of $\Delta t$ seconds. The PD gain in V/rad is thus equal to

$$K_{PD} = \frac{I_{tot}}{2\pi f_{REF} C_1}. \quad (8)$$

The proposed PD can be approximately modeled by the following transfer function [17]:

$$H_{PD}(j\omega) = K_{PD} \cdot \frac{1}{1 + \frac{C_2}{C_1 f_{REF}} j\omega} e^{-j\omega T_{REF}/2} \sin(\omega T_{REF}/2). \quad (9)$$

### B. Choice of Prescaler Modulus

The prescaler divide ratio $M$ in Fig. 9 is preferably equal to a power of 2. A greater value favors the design of the MMD but at the cost of larger phase jumps in $\phi_1$. For example, with $M = 8$ and hence $f_P = 7$ GHz, the maximum phase step in $\phi_1$ is equal to $\pm 286$ ps for a MASH 1-1-1 modulator. With $M = 4$, on the other hand, this jump is only $\pm 143$ ps. The PLL output phase noise is the absence of the FIR filter exhibits a bandpass shape and is equal to

$$S_{\phi, \Delta \Sigma} = \frac{4\pi^2 M^2}{12 f_{REF}} \left(2\sin(\frac{\pi}{f_{REF}})^2 \right)^2 |G(f)|^2 \quad (10)$$

where a MASH 1-1-1 architecture is assumed and $G(f)$ denote the PLL transfer function [18]. To counter the effect of $M^2$, the FIR filter length must be greater so as to provide a sharper roll-off, translating to a larger number of FFs. The total number of FFs is given by $(K - 1) f_P/f_{REF}$, where $K$ is the number of taps and $f_P$ the prescaler output frequency.

In summary, as $M$ increases: 1) the prescaler draws more power; b) the MMD draws less; and c) the FIR filter runs at lower clock frequency but requires a larger number of taps. Since most of the power consumption by the filter occurs in its clock path, we can write

$$P_{FIR} = \frac{f_P}{f_{REF}} \frac{(K - 1) f_P}{f_{REF}} C_{FF} V_D^2 \quad (11)$$

where $C_{FF}$ denotes the clock input capacitance of each FF. A higher $M$ yields a lower $f_P$ while demanding a higher $K$. To determine the optimum $M$, we assume a certain $\Delta \Sigma$-induced jitter, e.g., 40 fs rms, and a loop BW of 4 MHz, and quantify these three effects.

Plotted in Fig. 10(a) are the PLL output spectra due to the $\Delta \Sigma M$ q-noise for the two cases, both exhibiting an rms jitter of 40 fs. We now use transistor-level simulations to compute the power drawn by the prescaler, the MMD, and the FIR filter, obtaining the values shown in Fig. 10(b). It follows that $M = 8$ is preferable. While a more aggressive design could consider $M = 16$, it necessitates $K = 56$ for the FIR filter and hence a very large number of FFs.

### V. FIR Filter Design

We have presented the FIR filter topology in Section IV but also have raised a number of questions about its attributes in Section III-A. In this section, we deal with the questions.

#### A. Filter Response and Length

The frequency response of the filter is determined by the number of its taps, $K$, and its coefficients, $k_1, k_2, \ldots, k_k$. In this work, we employ a Chebyshev response as it does not require a high resolution for the unit current sources. Specifically, we have $k_1 = k_{21} = 10$ units, $k_2 = k_{21} = 3$ units, \ldots, and $k_{11} = k_{12} = 5$ units. The number of taps, $K = 22$, is chosen as a compromise between the filter power consumption and the $\Delta \Sigma$ q-noise suppression. The efficacy of the proposed FIR architecture can be assessed by several metrics: 1) the $\Delta \Sigma$ q-noise spectrum is reduced by 18 dB at 10 MHz; 2) the integrated $\Delta \Sigma$ noise is suppressed by 12 dB; and 3) the probability density function of the phase error is narrowed from $\pm 2 \Delta t_{div}$ at the MMD output to (equivalently) $\pm 0.3 \Delta t_{div}$ at the FIR output (see Fig. 11).

#### B. Mismatch and Noise of Current Sources

As mentioned in Section III, random mismatches among the current sources in Fig. 8(a) do not introduce nonlinearity but alter the response. In a typical current-source array, PMOS mismatches can be readily maintained below 10%. We thus perform Monte Carlo simulations to determine the variation of the FIR response and hence the PLL output jitter with this amount of mismatch. Plotted in Fig. 12, the tight distribution reveals the robustness of the design.
where \( w(t) \) denotes a square pulse extending from 0 to \( \Delta t \) with a height of unity. This noise is sampled at a rate of \( f_{\text{REF}} \), yielding an output spectrum of [19]

\[
S_{\phi}(f) = \frac{1}{C_1} |W(f)|^2 S_\phi(f)
\]  

(13)

where \( W(f) \) is the Fourier transform of the square pulse and \( S_\phi(f) \) the spectrum of \( I_n \). It is also proved in [19] that window-integrated, sampled white noise has a white spectrum if \( \Delta t < T_{\text{REF}} \). That is

\[
S_{\phi}(f) = \frac{1}{(C_1 + C_F)^2} \frac{\Delta t}{f_{\text{REF}}} S_\phi(f)
\]  

(14)

where the charge sharing between \( C_1 \) and \( C_F \) is included. Dividing this spectrum by \( K_{\text{PD}}^2 \) yields the input-referred phase noise arising from the FIR filter

\[
S_{\phi, \text{white}} = 4\pi^2 \frac{\Delta t}{T_{\text{REF}} (C_1 + C_F)^2} \frac{C_1^2}{I_{\text{tot}}^2} S_\phi(f)
\]  

(15)

In a similar manner, the effect of flicker noise current, \( S_{1/f}(f) \), can be formulated as

\[
S_{\phi, 1/f} = 4\pi^2 \frac{\Delta t^2}{T_{\text{REF}}^2 (C_1 + C_F)^2} \frac{C_1^2}{I_{\text{tot}}^2} S_{1/f}(f)
\]  

(16)

where noise aliasing is neglected. As a worst case estimate, we assume the maximum \( \Delta t \) that occurs in the fractional-N mode. For a MASH 1-1-1 modulator, and a prescaler divide ratio of 8, \( \Delta t \) reaches \( \pm 16 \) \( T_{\text{VCO}} \approx 280 \) ps. We note that doubling \( I_{\text{tot}} \) and \( C_1 \) reduces both thermal noise and flicker noise by 3 dB.\(^1\) With \( I_{\text{tot}} = 7 \) mA, \( C_1 = 6 \) pF and \( C_F = 1.5 \) pF,\(^2\) we obtain an input-referred phase noise of \(-157 \) dBc/Hz at 1-MHz offset. This value translates to \(-110 \) dBc/Hz at the PLL output. With fractional-N operation, we expect similar results. According to simulations, the output jitter rises by 9 fs when the noise of the FIR current sources is included.

### C. Output Resistance of Current Sources

As mentioned in Section III, the proposed switched-current FIR filter relies on two properties to avoid noise folding: tristate action of each branch, and a current provided by each branch that is relatively independent of the output voltage. The latter premise must be reexamined in view of the finite output resistance of the current sources.

Consider the simplified two-tap circuit shown in Fig. 14, where \( R_1 \) and \( R_2 \) denote the output resistance of \( I_1 \) and \( I_2 \), respectively. We repeat the calculations of Section III for this case, obtaining

\[
V_{\text{out}}(t) = (V_{\text{DD}} + I_1 R_1)
\]

\[
\left[ 1 - \exp \left( \frac{R_1}{R_2} \Delta t_a + \frac{R_2}{R_1} \Delta t_b - t \right) \right] \frac{1}{R_1 C_1} \Delta t.
\]

(17)

Interestingly, the FIR action (scaling and summation of \( \Delta t_a \) and \( \Delta t_b \)) occurs here before nonlinearity takes over.

\(^1\)We keep \( I_{\text{tot}}/C_1 \) to 1.1 \( \text{GVs} \) so that \( \Delta t \) is large enough to accommodate the \( \Delta \Sigma \) phase error.

\(^2\)The FIR current sources charges \( C_1 \) from zero to 0.5 V in 430 ps. This charging time is less than half of the reference period.
Thus, if FIR filtering suppresses the high-frequency noise, the exponential introduces negligible folding.

Of course, if we decrease $R_1$ and $R_2$ indefinitely, folding becomes significant. This points to the need for the cascode topology in Fig. 8(b). Plotted in Fig. 15 are the PLL output phase noise profiles due to the $M$ for simple and cascode current sources. The latter is close to that shown in Fig. 10(a) for $f_{P} = 7$ GHz, a 22-Tap FIR, and ideal current sources. Nevertheless, the high output resistance is afforded for only a maximum output voltage of about 600 mV, thereby requiring the voltage gain provided by the $Gm$ stage (see Section IV).

It is possible to improve the linearity of the FIR filter by increasing $C_1$ so that the exponential in (17) can be approximated by its linear terms. However, this increases $\Delta t$ in (15) and (16), raising the noise contribution of the current sources.

To appreciate the efficacy of the FIR filter, we simulate the fractional-$N$ PLL without it, obtaining an integrated jitter of 190 fs$_{\text{rms}}$, 70% of which arises from the $\Delta \Sigma$ noise. If the loop filter’s BW is reduced by a factor of 2, the overall jitter falls to a minimum of 140 fs$_{\text{rms}}$, still far above our overall target. According to simulations, the PLL output q-noise floor changes negligibly as we reduce FCW from $2^{-4}$ to $2^{-16}$ (see Fig. 16). Thus, the FIR concept proves effective for small FCWs as well.

VI. DELAY LINE DESIGN

The unit delays in the FIR filter of Fig. 8(a) consist of FFs and are clocked with a period of $T_{\text{div}} = 8T_{\text{VCO}}$. With this clocking method, $\phi_2-\phi_{22}$ carry the feedback information for the PLL to lock [13]. However, since $T_{\text{div}}$ and $T_{\text{REF}}$ do not bear an integer ratio, the phase difference sensed by the filter accumulates with time, eventually reaching large values [see Fig. 17(a)]. This in turn produces an excessive voltage change on $C_1$ and causes the current sources(s) to collapse. To resolve this issue, the delay elements can assume either of two values, namely, $T_1 = 28T_{\text{div}}$ or $T_1 = 29T_{\text{div}}$, so as to create a tight bound for this error. Programmed individually in conjunction with the $\Delta \Sigma M$ FCW, $\alpha$, the delay of Stage $j$ is set according to the following rules. If the accumulated error from Stage 1 to Stage $j$ is less than $T_{\text{div}}$, then $T_1 = 28T_{\text{div}}$ is selected. Otherwise, $T_2 = 29T_{\text{div}}$. As depicted by the waveform in Fig. 17(b), the delay from $\phi_2$ to $\phi_3$ is compensated by one more $T_{\text{div}}$, limiting the phase error in the last tap, $\phi_{22}$, to about $T_{\text{div}}$.

The FFs employ a true single-phase clock (TSPC) structure (see Fig. 18). The total extracted capacitance of the clock input is 1.2 FF. This work employs 600 FFs, consuming $P = fCV_{DD}^2 = 6$ mW at 7 GHz.

VII. DIVIDER DESIGN

The power and area consumption of dividers can become significant at the frequencies of interest here. For this reason, it is desirable to avoid current-mode logic and inductive peaking. This section deals with both the prescaler and the MMD.
A. Prescaler

The $\div 8$ circuit consists of three cascaded $\div 2$ stages. The first must present a low capacitance to the VCO while providing sufficiently large voltage swings to drive the second. We propose a low-power, compact topology. As shown in Fig. 19(a), the circuit is based on two dynamic latches and a third inverter in the feedback path for proper toggling.

The performance is dramatically improved by introducing an unclocked feedforward path from A to B so that the signal arrives at the latter before $S_2$ turns on. This “predictive” path therefore initiates a desirable change in the state at B while $S_2$ is still OFF. Proper scaling of $\text{Inv}_4$ with respect to those in the main path extends the upper end of the lock range while imposing some limitation on the lower end. That is, at sufficiently low clock frequencies, the feedforward path overwhelms the main path, causing failure.

Switches $S_1$ and $S_2$ in Fig. 19(a) are realized by NMOS transistors that are 2 $\mu$m wide, with their gates residing at a dc level of $V_{DD}$. As the gate voltages of $S_1$ and $S_2$ rise above $V_{DD}$, their $V_{gs}$ reach 1.28 V, slightly stressing these devices. Using the model described in [20], we compute the estimated lifetime at 0.01% cumulative failure rate to be 11.3 years at 140 °C.

The capacitively-coupled clocks then allow a low resistance for these devices. Fig. 19(b) shows the simulated input frequency range of the proposed $\div 2$ stage over process corners. We observe that feedforward raises the maximum speed from 55 to 68 GHz and limits the lower end to 43 GHz. The circuit draws 1.8 mW at 56 GHz and presents an input capacitance of 10 fF to the VCO.

B. Multimodulus Divider

The MMD is implemented by a cascade of $\div 2/3$ stages [21] and its modulus can range from 16 to 63. The first two stages are implemented by TSPC logic and the rest by CMOS logic [22]. The MMD draws 1 mW at 7 GHz.

VIII. JITTER CONTRIBUTIONS

Fig. 20 plots the simulated PLL output phase noise components arising from the reference, the $\Delta \Sigma M$, the FIR filter, the Gm stage, the VCO, and the feedback divider chain for a loop BW of 4 MHz. According to simulations, the $\Delta \Sigma M$ q-noise translates to an rms jitter of 160 fs without the proposed FIR filter.

The VCO is based on a complementary LC topology with $W/L = 15 \mu m/40$ nm for all four transistors. The 45-pH differential inductor is realized as a parallel stack of metal-9 and metal-8 spirals so as to reduce the resistance. Although this inductor’s $Q$ is about 20 at 56 GHz, the addition of programmable capacitors drops the $Q$ to approximately 13. The VCO achieves a tuning range of 52.3–56.8 GHz with a phase noise of $-94 \text{ dBc/Hz}$ at 1-MHz offset while consuming 7.2 mW.

Table I summarizes the corresponding jitter contributions, demonstrating the efficacy of the FIR filter in suppressing the $\Delta \Sigma M$ q-noise. The VCO and the Gm stage are the principal contributors.
IX. EXPERIMENTAL RESULTS

The proposed PLL has been fabricated in TSMC’s 28-nm CMOS technology. Fig. 21 shows a photograph of the die, whose active area measures approximately 0.1 mm$^2$.\(^3\) Operating with a 1-V supply, the prototype consumes 23 mW. The external 250-MHz reference is supplied by Crystek’s CRBSCS-01-250 250-MHz crystal oscillator.\(^4\) Its measured phase noise is plotted in Fig. 22. For ease of measurement, the output of the $\div 8$ prescaler in Fig. 9 is monitored for characterization of the PLL.

Fig. 23 shows the measured spectrum at this point with FCW = 28.004. The fractional spur at 1-MHz offset has a level of $-65.7$ dBc, which translates to $-47.7$ dBc at the VCO.\(^3\)

\(^3\)The area of the loop filter is dominated by the series capacitor to create a zero at 800 kHz.

\(^4\)The supply voltage of the crystal oscillator is 5.5 V.

Fig. 24 plots the fractional spur levels as FCW varies from 28.0015 to 28.06 and hence the fractional spur offset frequency from 0.4 to 15 MHz. We should make two remarks. First, the measured spur levels reported here are about 3 dB lower than those in [23]. This has been achieved by separating the supply lines of the delay line, the Gm stage, and the FIR filter and refabricating the chip. Second, wireline receivers typically allow a BW of tens of megahertz [24], [25] or even above 100 MHz [26], [27] in their clock and data recovery (CDR) loop. As the CDR ensures the recovered clock phase tracks that of the received data, fractional spurs are rejected within the CDR loop BW. For example, a $-50$-dBc spur at 2-MHz offset negligibly affects the receiver performance even output.
though it translates to a deterministic jitter of 12.7 fs rms at the transmitter output.

Due to our phase noise analyzer limitations, the ±8 output is applied to an off-chip divide-by-2 circuit for phase noise measurements. Fig. 25(a) plots the measured phase noise in the fractional-N mode with an offset and the integrated jitter of 99/fs. Fig. 25(b) shows the measured phase noise in the fractional-N mode with an offset and the integrated jitter of 99/fs. The integrated jitter is computed in two different BWs. First, for a fair comparison with the prior art, the offset ranges from 10 kHz to 40 MHz, yielding a total of 110 fs rms. Second, the offset ranges from 40 MHz to the Nyquist frequency of 3.5 GHz, revealing another 31 fs. (Due to the equipment limitation, this measurement reads the phase noise values directly from the spectrum of the ±8 output.) Thus, the total jitter from 10 kHz to 3.5 GHz is 114 fs rms.

In order to study the advantage of the binary delay line, we set the FCW to 0.127 and measure the PLL phase noise. We first set a single delay value for the entire FIR chain and then enable the binary delay option. Plotted in Fig. 26 are the phase noise profiles for the two cases. Enabling the binary delay reduces the phase noise by 1 to 1.5 dB below 3-MHz offset and the integrated jitter by 10 fs.

Table II summarizes the measured performance of our prototype and compares it to that of prior-art 60- and 30-GHz fractional-N PLLs. We observe a nearly twofold reduction in jitter, an 8.3-dB improvement in the figure of merit (FoM), and more than a threefold reduction in area. We also list FoM2, proposed in [28], since the reference frequency in our prototype is higher than other PLLs in Table II. The reference spur level is -50 dBc and translates to a deterministic jitter of 12.7 fs rms in the transmitted data.

X. Conclusion
This article proposes a new fractional-N PLL architecture that suppresses the ΔM noise and lends itself to a compact low-power design. A new ±2 circuit is also described.

References


Yu Zhao (Member, IEEE) received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2013, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Los Angeles, Los Angeles, CA, USA, in 2015 and 2022, respectively. From 2015 to 2018, he was with Ubitinx Technology, Inc., San Jose, CA, USA, where he was worked on frequency synthesizers for Bluetooth and WiFi 6. He is currently with the RFIC Design Team, Hisilicon, Shanghai. His research interests include low-jitter frequency synthesizers for wireless and wireline transceivers.

Onur Memioglu (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2015 and 2018, respectively, and the Ph.D. degree from the Communication Circuits Laboratory, University of California at Los Angeles, Los Angeles, CA, USA, in 2021. He was an RFIC Design Engineer with Mikro-Tasarim, Ankara, from 2015 to 2018. He is currently with the RFIC Design Team, Apple, Cupertino, CA, USA.

Long Kong (Member, IEEE) received the B.E. degree in microelectronics from Shanghai Jiao Tong University, Shanghai, China, in 2011, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Los Angeles, Los Angeles, CA, USA, in 2013 and 2016, respectively. In 2016, he joined Oracle, Santa Clara, CA, USA, as a Senior Hardware Engineer, where he was working in high-speed SerDes transceivers. He then joined Apple Inc., Cupertino, CA, USA, in 2017, as an RFIC Design Engineer. He has been working in Bluetooth and ultrawideband radios for wearable and mobile devices. His research interests include frequency synthesizers, wireless and wireline transceivers, radars, and clock recovery for data communication systems.


Dr. Razavi is a member of the U.S. National Academy of Engineering and a fellow of the U.S. National Academy of Inventors. He received the Beatrice Winner Award for Editorial Excellence at the 1994 ISSCC, the Best Paper Award at the 1994 European Solid-State Circuits Conference, the Best Panel Award at the 1995 and 1997 ISSCC, the TRW Innovative Teaching Award in 1997, the Best Paper Award at the IEEE Custom Integrated Circuits Conference in 1998, and the McGraw-Hill First Edition of the Year Award in 2001. He was the co-recipient of both the Jack Kilby Outstanding Student Paper Award and the Beatrice Winner Award for Editorial Excellence at the 2001 ISSCC. He received the Lockheed Martin Excellence in Teaching Award in 2006, the UCLA Faculty Senate Teaching Award in 2007, and the CICC Best Invited Paper Award in 2009 and 2012. He was a co-recipient of the 2012 and 2015 VLSI Circuits Symposium Best Student Paper Awards and the 2013 CICC Best Paper Award. He was also recognized as one of the top ten authors in the 50-year history of ISSCC. He received the 2012 Donald Pederson Award in Solid-State Circuits. He was also a recipient of the American Society for Engineering Education PSpW Teaching Award in 2014 and the 2017 IEEE CAS John Choma Education Award. He has served as an IEEE Distinguished Lecturer. He served on the Technical Program Committees of the International Solid-State Circuits Conference (ISSCC) from 1993 to 2002 and the VLSI Circuits Symposium from 1998 to 2002. He has also served as a Guest Editor and an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and the International Journal of High Speed Electronics. He served as the Founding Editor-in-Chief for the IEEE SOLID-STATE CIRCUITS LETTERS.