Delft University of Technology

High-Power Digital Transmitters for Wireless Networks

Bootsman, R.J.

DOI

Publication date

2025

Document Version

Final published version

Citation (APA)

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

High-Power Digital Transmitters
for Wireless Networks

Robert Jan Bootsman

High-Power Digital Transmitters for Wireless Networks

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op donderdag 18 december 2025 om 17.30 uur

door

Robert Jan BOOTSMAN

Elektrotechnisch ingenieur,
Technische Universiteit Delft, Nederland,
geboren te Nieuw-Vennep, Nederland.

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie:

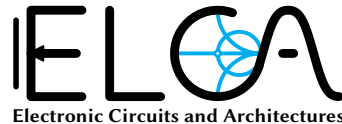|	
Rector Magnificus,	voorzitter
Prof. dr. ing. L.C.N. de Vreede,	Technische Universiteit Delft, promotor
Dr. S.M. Alavi,	Technische Universiteit Delft, promotor

Onafhankelijke leden:

Prof. dr. ir. W.D. van Driel,	Technische Universiteit Delft
Prof. dr. ir. A.B. Smolders,	Technische Universiteit Eindhoven
Prof. dr. ir. B. Nauta,	Technische Universiteit Twente
Prof. dr. P. Wambacq,	imec Leuven en Vrije Universiteit Brussel, België
Prof. dr. N. Llombart,	Technische Universiteit Delft, reservelid

Overig lid:

Dr. ir. F. van Rijs,	Ampleon Netherlands B.V.

Keywords:	Digital RF Transmitters, Power Amplifiers, Harmonic Matching, High-Speed Digital, CMOS, LDMOS, Heterogeneous Integration, Flip-Chip
Cover:	Worlds first high-power fully digital transmitter, by R.J. Bootsman

The author set this dissertation in LaTeX.

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Contents

List of Figures

List of Tables

Glossary

5G fifth generation

ACLR adjacent channel leakage ratio

ACW amplitude code word

ADC analog-to-digital converter

BJT bipolar junction transistor

BW bandwidth

CCDF complementary cumulative distribution function

CDF cumulative distribution function

CG common gate

CMOS complementary MOS

cmWave centimeter wave

CS common source

CW continuous wave

DAC digital-to-analog converter

DCO digitally controlled oscillator

DDTX Doherty DTX

DFE digital front-end

DNL differential nonlinearity

DPD digital pre-distortion

DSP digital signal processing

DTX digital transmitter

DUC digital up-conversion

DUT device under test

EER envelope elimination and restoration

EM electromagnetic

ENOB effective number of bits

ESD electrostatic discharge

ET envelope tracking

EVM error-vector magnitude

FBW fractional bandwidth

FEM finite element method

FET field-effect transistor

HEMT high-electron-mobility transistor

IC integrated circuit

INL integral nonlinearity

inshin integrated shunt inductor

IO input/output

LMBA load-modulating balanced amplifier

LO local oscillator

LSB least significant bit

LUT look-up table

MIM metal–insulator–metal

MIMO multiple-input multiple-output

MLCC multi-layer ceramic capacitor

MMIC monolithic microwave IC

mMIMO massive MIMO

mmWave millimeter wave

MOM metal–oxide–metal

MoM method of moments

MOS metal–oxide–semiconductor

MSB most significant bit

PA power amplifier

PAPR peak-to-average power ratio

PCB printed circuit board

PD-LMBA pseudo-Doherty LMBA

PDF probability density function

PLL phase-locked loop

RBW radio bandwidth

RF radio frequency

RFDAC radio frequency digital-to-analog converter

RMS root-mean-square

RX receiver

SCPA switched-capacitor power amplifier

SMD surface-mounted device

SMPA switch-mode power amplifier

SNR signal-to-noise ratio

SOI silicon on insulator

SPI serial peripheral interface

TX transmitter

UBM under bump metallization

VBW video bandwidth

VNA vector network analyzer

VSA vector signal analyzer

xMIMO extreme MIMO

ZVS zero-voltage switching

Symbols

α Conduction angle of analog transconductance power amplifiers, or generic constant possibly provided with a subscript

A Multi-phase activation phasors of a complex modulated signal that align with the I and Q axes of the complex plane

β Phase propagation constant

B Susceptance in siemens (S)

B Multi-phase activation phasors of a complex modulated signal that are 45° offset from the I and Q axes of the complex plane

BV_{DSS} Drain–source breakdown voltage (e.g., avalanche) under the assumption the device is OFF, where, typically, the gate is connected to the source

C Capacitance in farad (F)

C_L Load capacitance, in this dissertation typically for a digital driver

γ Self-loading factor of CMOS inverters, defined in Eq. (4.7) as $\gamma = C_o/C_i$

γ Complex propagation constant, $\gamma = \alpha + \mathrm{j}\beta$

d duty cycle of a rectangular pulse, defined at the 50 % point of the rising edge to the 50 % point of the falling edge

D D-parameters, the mixed-signal equivalent to analog S-parameters. Most notably D_{21}, the normalized digital forward transfer, as defined in (e.g.) Eq. (5.18)

da Digital input, normalized to be compatible with traditional linear power waves, as defined in Eq. (5.7), or Eq. (5.16) when only considering the transfer to the fundamental frequency

Δ Discrete difference or change

η Efficiency, most notably η_D the drain efficiency, and η_S the system efficiency

f Frequency in hertz (Hz)

f_0 Design frequency or center frequency of a resonant circuit

f_c Frequency of the RF carrier or RF upconverting clock

$f(\cdot)$ Generic mathematical function

f Fan-out of a digital logic gate, being the ratio of its load capacitance over its input capacitance

F Overall effective fan-out of a tapered buffer chain

f_s Sampling frequency of a digital or mixed signal system

$\mathscr{F}_t$ Fourier transform as defined in Eq. (A.1)

F_{up} Current or power utilization factor dependent on the upconversion architecture

G Conductance in siemens (S)

g_m Transconductance in siemens (S), typically a specification for a MOSFET per unit gate width

I_{XY} Current flowing from node X to node Y in ampere (A), e.g., I_{DS} for drain–source current

I In-phase component of a complex modulated signal

$\Im$ Imaginary component of a complex number

IQ Complex plane to visualize a complex baseband envelope or constellation

j Imaginary unit of a complex number, i.e., $\mathrm{j}^2 = -1$

L Inductance in henry (H)

M Capacitance multiplication factor, being the ratio of the total (tapered) driver chain capacitance over the load capacitance, as defined in Eq. (4.10)

M Mutual inductance in henry (H)

N_b Number of bits, a measure of resolution

ω Angular frequency in radian per second ($\mathrm{rad\,s^{-1}}$)

ω_0 Angular design frequency or angular center frequency of a resonant circuit, i.e., $\omega_0 = 2\pi f_0$

ω_c Angular frequency of the RF carrier or RF upconverting clock, i.e., $\omega_c = 2\pi f_c$

P Power in watt (W), e.g., P_{RFout} the RF output power into a load

ϕ Modulation phase, or vector angle in baseband

P_{DD} Power consumed or converted in a power supply domain, most notably $P_{DD,\mathrm{RF}}$ the power supply for the RF power output stage, $P_{DD,\mathrm{dr}}$ the power supply for the digital drivers driving the output stage segments, and $P_{DD,\mathrm{core}}$ the power supply used for the CMOS core devices (digital, or all thin oxide devices)

Q Quality factor of a component or resonant circuit

Q Quadrature component of a complex modulated signal

R Resistance in ohm (Ω)

ρ Envelope magnitude

$\Re$ Real component of a complex number

R_L Load resistance

ς Empirical technology constant that relates the propagation delay of a logic gate driven by a step input to that of one driven by a preceding gate with a realistic propagation delay

θ Geometrical angle

t_p Propagation delay of a digital logic gate, most notably t_{p0} the intrinsic propagation delay of an inverter driven by a step input, t_{pLH} the propagation delay from logic low to logic high and t_{pHL} vice versa, and $t_{p,\mathrm{int}}$ the self-loaded propagation delay of an inverter driven by a step input

t_{rf} Rise and fall time of a rectangular pulse, assuming rise time and fall time are identical, i.e., $t_r = t_f$. This quantity is either provided in seconds (s) or as a percentage of the pulse's (RF) period, and—unless specified otherwise—assumes a linear slope from $0\,\%$ to $100\,\%$

V_{DD} Supply voltage in volt (V), most notably $V_{DD,\mathrm{RF}}$ the supply voltage for the RF power output stage, $V_{DD,\mathrm{dr}}$ the supply voltage for the digital drivers driving the output stage segments, and $V_{DD,\mathrm{core}}$ the supply voltage used for the CMOS core devices (digital, or all thin oxide devices)

V_T Threshold voltage in volt (V)

V_{XY} Voltage between nodes X and Y in volt (V), e.g., V_{GS} for gate–source voltage and V_{DS} for drain–source voltage. When $X = Y$, it is a bias voltage referenced to ground, see for example V_{DD}

W_G Transistor gate width in meter (m)

X Reactance in ohm (Ω)

Y Admittance in siemens (S), $Y = G + \mathrm{j}B$

Z Impedance in ohm (Ω), $Z = R + jX$

Z_c Characteristic impedance of a transmission line

$\hat{\ }$ Normalized value, e.g., a current $\hat{I}$ per unit length, or a value mapped to the range $[0, 1]$

Summary

Mobile data demand and capacity have grown exponentially for decades. This trend is expected to continue in the coming years, and new techniques and communication standards are being adopted to accommodate this growing demand. The energy consumption of the mobile networks associated with this growth is of specific concern, with an estimated 3.6 % of the global electrical energy being consumed by 2030. The exponentially growing data capacity is enabled by the tremendous technological development in integrated circuits (ICs), specifically in the domain of digital-oriented CMOS technologies. With digital logic becoming ever faster, their switching performance provides new opportunities for RF transmitters. Over the last decade, this has led to enormous progress in digital transmitters (DTXs).

However, the typical supply voltages of digital-oriented CMOS technologies are too low to reach the power levels required for mMIMO base stations. The market for high-power RF applications optimizes their technologies for minimized losses, and increased power density and gain. This results in a performance gap between what digital CMOS can provide today and what is required for next-generation base stations. Benefiting from the increased functionality and power savings from the developments in digital CMOS while maintaining the power levels provided by technologies—such as LDMOS or GaN—is taking the best of both worlds. This leads to the research objective of this dissertation:

> "How can digital-oriented low-power CMOS technology be combined with high-power RF technology such that energy-efficient operation of next-generation sub-7 GHz base stations can be achieved?"

To answer this question, several demonstrators have been designed to pioneer combining CMOS technologies with high-power RF technologies.

The first demonstrators serve as the proof-of-concept for high-power DTXs. Its digital controller is fabricated in 40 nm CMOS and bond-wire-connected to an LDMOS power die with a segmented output stage. The first demonstrator targets class-BE operation at 2.1 GHz to show the feasibility of high-power DTX with high energy efficiency. Its measurements show this is the case, reaching a peak RF power of 18.5 W with drain and system efficiencies of 67 % and 60 %, respectively. Complex modulated signals are also measured using QAM, yielding an ACLR of −46.1 dBc for a bandwidth of 10 MHz. The second demonstrator investigates suitable operation classes for segmented power output stages in DTX architectures, introducing digital class-C operation. It targets higher RF bandwidths and is designed for 1.0 GHz operation using the same hardware. The used output matching network is comparable to that of analog class-B, with a resistive match for the fundamental and all harmonics shorted. The digital class-C operation uses digital current scaling with a reduced RF duty cycle, yielding a linear DTX transfer up to compression (i.e., no gain expansion as in analog class-C). A peak RF output power of 25.9 W was measured with drain and system efficiencies of 76 % and 73 %, respectively. Modulated signals result in an ACLR of −48.3 dBc for a bandwidth of 8.8 MHz. To enhance the efficiency

of DTX in back-off, the third demonstrator uses the digital class-C operation in a 2-way Doherty configuration. It targets wideband operation at 2.0 GHz, measuring a −1 dB-power bandwidth of > 430 MHz and a −10 %-drain efficiency bandwidth of > 370 MHz. It measures a peak RF output power of 39.1 W with a drain efficiency of 57 %, while only consuming 0.19 W in standby. More importantly, using a 7 MHz 256-QAM modulated signal with a PAPR of 5.5 dB, the average drain and system efficiency were measured at 49 % and 46 %, respectively. It also achieved an ACLR of −53 dBc at that same bandwidth. All these demonstrators have shown degraded performance due to parallel resonances in the driver's supply path. Still, all these demonstrators have shown much higher output powers than available in the state-of-the-art, while reaching comparable linearity performance. More importantly, these demonstrators contribute experience and a pathway to more advanced high-power DTX designs.

The next demonstrator targets a higher operating frequency, the 3.5 GHz 5G band. A higher (effective) DTX resolution is required to support larger modulation bandwidths, for which a new CMOS controller and segmented LDMOS power die are designed. This time, a high-density flip-chip approach is pioneered, resulting in an assembly with over 4000 interconnects. High-speed stacked thin-oxide drivers are implemented in the CMOS controller, and the LDMOS process is modified to support a lowered threshold voltage. Again, digital class-C operation is chosen for its high efficiency, linear operation, and wideband potential. This results in operation with a measured smooth transfer, proving that the high resolution is achieved. It achieves above 10 W output power at 3.5 GHz, with drain and system efficiencies of 45 % and 40 %, respectively. The second demonstrator using this high-density flip-chip hardware aims to show modulated signals at the 1.8 GHz band, using a push-pull configuration matched again for digital class-C operation. With the output balun deembedded, this demonstrator reaches 20 W of RF output power with drain and system efficiencies of 68 % and 63 %, respectively. Modulated performance results in an ACLR of −43.9 dBc using a 13 MHz 256-QAM signal, and, at a larger bandwidth of 53 MHz 64-QAM, an ACLR of −37.0 dBc. These ACLR levels are retained over a more than 10 dB power back-off range, illustrating the resolution of this demonstrator.

The knowledge gained from designing these demonstrators is presented in the early chapters of this dissertation, providing the reader with important aspects of designing high-power DTXs. This ranges from practical aspects of the heterogeneous integration used, such as electrical compatibility and packaging, to designing high-speed drivers and the high-level modeling of DTXs. A mathematical definition of a DTX's transfer is proposed, which relates its numerical baseband input to the output power at RF. Further, a power model capable of estimating DTX performance in terms of power and efficiency is proposed. This power model combines the theory, presented in the chapters before, into a handful of equations that describe the power relations in a DTX by first-order approximation, which are useful for hand calculations and can help conceptual understanding of the underlying relations. These background chapters guide the reader in implementing future high-power DTXs, and the power relations can be used to optimize these future designs from both the digital CMOS and power technology perspectives.

Samenvatting

De vraag naar mobiele data en de beschikbare capaciteit ervan groeien al decennia exponentieel. De verwachting is dat deze trend doorzet in de komende jaren, waarbij nieuwe technieken en communicatiestandaarden worden ontwikkeld om de groei bij te kunnen houden. Het groeiende energieverbruik van mobiele netwerken is verontrustend, omdat verwacht wordt dat het wereldwijde elektriciteitsverbruik van mobiele netwerken in 2030 is verviervoudigd ten opzichte van 2025. De exponentieel groeiende datacapaciteit wordt mogelijk gemaakt door de overweldigende technologische vooruitgang in geïntegreerde schakelingen (IC). Dit geldt specifiek voor CMOS-technologieën, die steeds snellere en zuinige digitale schakelingen mogelijk maken. Deze verbeteringen brengen ook nieuwe mogelijkheden voor radiozenders (RF-zenders) met zich mee, wat de laatste tien jaar tot enorme vooruitgang in digitale radiozenders (DTX) heeft geleid.

De voedingsspanning van de digitale CMOS is echter typisch te laag om de vermogens te genereren die nodig zijn voor mMIMO-zendmasten. Dat de marktpartijen voor radiozendertoepassingen met een hoog vermogen hun technologieën optimaliseren voor geminimaliseerd verlies en verhoogde vermogensdichtheid en versterkingsfactor, leidt tot een prestatiekloof tussen wat digitale CMOS kan bieden en wat nodig is voor de volgende generatie zendmasten. Wanneer men profiteert van de toegenomen functionaliteit en de energiebesparing die de ontwikkeling van digitale CMOS biedt – terwijl de zendvermogens van LDMOS- of GaN-technologieën behouden blijven – pakt men het beste van beide werelden. Dat leidt tot het onderzoeksdoel van dit proefschrift:

> "Hoe kunnen digitale CMOS-technologieën zodanig gecombineerd worden met radiozendertechnologieën voor hoogvermogenstoepassingen, dat de volgende generatie sub-7 GHz zendmasten op een energie-efficiënte wijze kunnen werken?"

Om deze vraag te beantwoorden zijn er diverse prototypes ontworpen om de combinatie van CMOS-technologieën met radiozendertechnologieën voor hoogvermogenstoepassingen te pionieren.

De eerste prototypes dienen als haalbaarheidsbewijs van hoogvermogens-DTXen, waarvan de digitale regelaar gefabriceerd is in een 40 nm CMOS-proces. Deze digitale regelaar is middels draadbindingen verbonden met een LDMOS-vermogens-chip, die een gesegmenteerde uitgangstrap heeft. Het eerste prototype richt zich op klasse BE werking op 2,1 GHz om de haalbaarheid van hoogvermogen-DTX met een goede energie-efficiëntie aan te tonen. Metingen hiervan laten zien dat dit het geval is, waarbij een piek RF-uitgangsvermogen van 18,5 W wordt bereikt met een *drain-* en systeemefficiëntie van respectievelijk 67 % en 60 %. Complex gemoduleerde signalen zijn ook gemeten gebruikmakend van QAM, wat een zijbandvermogensratio (ACLR) van –46,1 dBc heeft bij een bandbreedte van 10 MHz. Het tweede prototype onderzoekt geschikte werkingsklassen voor gesegmenteerde vermogenstrappen, waarbij digitale klasse C geïntroduceerd wordt. Het richt zich op grotere RF-bandbreedtes en is ontworpen voor werking op 1,0 GHz gebruikmakend van dezelfde

hardware als in het eerste prototype. Het impedantieaanpassingsnetwerk aan de uitgang is vergelijkbaar met dat van analoog klasse B, met een ohmse weerstand voor de fundamentele toon en alle harmonischen kortgesloten. De digitale klasse-C-werking gebruikt digitale stroomschaling met een verlaagde RF-arbeidscyclus, wat leidt tot een lineaire DTX-overdracht tot aan compressie (d.w.z. geen overdrachtsgroei zoals in analoge klasse C). Een piek RF-uitgangsvermogen van 25,9 W is gemeten met een *drain*- en systeemefficiëntie van respectievelijk 76 % en 73 %. gemoduleerde signalen resulteren in een ACLR van −48,3 dBc voor een bandbreedte van 8,8 MHz. Om de efficiëntie van de DTX voor verlaagde uitgangsvermogens te verbeteren gebruikt het derde prototype digitale klasse-B-werking in een tweeweg Doherty-opstelling. Het richt zich op werking op 2,0 GHz, waarbij een −1 dB-vermogensbandbreedte van meer dan 430 MHz gemeten is en een −10 %-efficiëntiebandbreedte van meer dan 370 MHz. Er is een piek RF-uitgangsvermogen van 39,1 W gemeten met een *drain*efficiëntie van 57 % terwijl er slechts 0,19 W in de wachtstand verbruikt wordt. Belangrijker nog, gebruikmakend van een 7 MHz 256-QAM signaal met een piek-tot-gemiddeld-vermogensverhouding (PAPR) van 5,5 dB, de gemiddelde gemeten *drain*- en systeemefficiëntie zijn respectievelijk 49 % en 46 %. Er wordt verder een ACLR van −53 dBc bereikt bij diezelfde bandbreedte. Al deze prototypes toonden verminderde prestaties door parallelresonanties in de stuurtrapvoeding. Dan nog tonen deze prototypes veel hogere uitgangsvermogens dan tot nu toe beschikbaar in toonaangevend onderzoek, terwijl de lineariteit vergelijkbaar is. En nog belangrijker: de ervaring opgedaan met deze prototypes baant een weg naar geavanceerdere DTX-ontwerpen.

Het volgende prototype richt zich op hogere zendfrequenties, namelijk de 3,5 GHz 5G-band. Een hogere (effectieve) DTX-resolutie is nodig om grotere modulatiebandbreedtes te ondersteunen, waarvoor een nieuwe CMOS-regelaar en gesegmenteerde LDMOS-vermogens-chip zijn ontworpen. Ditmaal is een flip-chip-aanpak met een hoge dichtheid gepionierd, wat leidt tot een assemblage met meer dan 4000 verbindingen. Snelle gestapelde transistoren met een dunne oxidelaag zijn gebruikt in de stuurtrap en het LDMOS-proces is aangepast om een lagere drempelspanning te verwezenlijken. Opnieuw is digitaal klasse C gekozen vanwege zijn hoge efficiëntie, lineaire overdracht en breedbandige potentieel. Hieruit volgt werking met een gemeten gladde overdracht, wat bewijst dat de hoge resolutie behaald is. Het haalt meer dan 10 W uitgangsvermogen op 3,5 GHz, met een *drain*- en systeemefficiëntie van respectievelijk 45 % en 40 %. Het tweede prototype dat gebruikmaakt van de hogedichtheid-flip-chip-hardware richt zich op het vertonen van gemoduleerde signalen in de 1,8 GHz band, waarbij gebruikgemaakt wordt van twee schakelblokken in een balansschakeling en wederom een impedantieaanpassingsnetwerk voor digitale klasse-C-werking. Met de verliezen van de *balun* ge-*de-embed* haalt dit prototype 20 W RF-uitgangsvermogen met een *drain*- en systeemefficiëntie van respectievelijk 68 % en 63 %. Met gemoduleerde signalen levert dit een ACLR van −43,9 dBc op bij een 13 MHz 256-QAM signaal en, bij een grotere bandbreedte van 53 MHz 64-QAM, een ACLR van −37,0 dBc. Deze ACLR-niveaus blijven behouden over meer dan 10 dB vermogensterugregeling, wat toe te schrijven is aan de goede resolutie van dit prototype.

De kennis die is opgedaan tijdens het ontwerpen van deze prototypes wordt gepresenteerd in de hoofdstukken aan het begin van dit proefschrift. De lezer wordt daar voorzien van belangrijke informatie over aspecten van het ontwerp van hoogvermogens-DTXen, variërend van de praktische kant van de gebruikte heterogene integratie, zoals elektrische combineerbaarheid en het verpakken, tot het ontwerpen van snelle digitale stuurtrappen

en het modelleren van een DTX op een hoog niveau. Een wiskundige definitie van de DTX-overdracht wordt voorgesteld, welke het RF-uitgangsvermogen in verband brengt met de numerieke invoer. Verder wordt een vermogensmodel voorgesteld dat in staat is om de DTX-prestaties op het gebied van vermogen en efficiëntie in te schatten. Dit vermogensmodel combineert de theorie, zoals gepresenteerd in de hoofdstukken ervoor, in een handvol vergelijkingen die de DTX-vermogensverhoudingen beschrijven bij eerste orde benadering. Deze vergelijkingen zijn nuttig voor snelle handmatige berekeningen en het conceptuele begrip van de onderliggende verhoudingen. Deze achtergrondhoofdstukken begeleiden de lezer in het ontwerp van toekomstige hoogvermogens-DTXen en geven vermogensverhoudingen die gebruikt kunnen worden om deze toekomstige ontwerpen te optimaliseren, aan de kant van zowel de digitale CMOS als van de vermogenstechnologie.

1

Introduction

$\mathbf{T}$HE world is more interconnected today than ever before. Not only in terms of physical transportation options, such as global trading, but also through digital communication. The ocean floors are lined with optical fiber connecting the continents, data centers sprinkled across the globe, and communication satellites flying through space. Closer to home, landlines and cable TV were introduced in the early 20th century. In the 21st century, this infrastructure is increasingly being replaced by fiber-optic communication networks and, more and more often, fiber-to-the-home.

Another important means of communication is by cellular (mobile) phone. Provided there is reception, one can access the internet anywhere, at any time. Even though cellular technology is wireless, a lot of infrastructure is required to make it work reliably. Data centers are required by, among others, mobile service providers, but more importantly, an extensive network of cellular base stations is required. These base stations consist of a (digital) network connection, electronics to process the in and outgoing information, digital-to-analog converters (DACs), upconverters, amplifiers, and antenna(s) that make the transmit (TX) chain and, vice versa, low noise amplifiers, downconverters, and analog-to-digital converters (ADCs) to implement the receive (RX) chain. Both are needed to exchange information wirelessly using radio frequency (RF) signals. Moreover, good wireless cell coverage is achieved by placing sufficient base stations.

1.1 Exponential Growth of Data Capacity

Mobile data demand and capacity have grown exponentially for decades. This trend is expected to continue in the coming years, not only in high-income countries. Developing countries often 'leapfrog' past old landline technology directly toward modern mobile technology [1]. Figure 1.1 shows the predicted mobile traffic worldwide, reaching 280 EB[1]/month by 2028, an almost 1200× increase compared to 2011 [2]. Even when compensating for the growth in mobile subscriptions in this period, this amounts to an increase of 750× the traffic *per user*.

Cellular networks evolve using new techniques while adopting new communication standards to accommodate this growing demand. The standard currently being implemented

---

[1]One exabyte (EB) is equal to 1 000 000 000 000 000 000 (10^{18}) bytes, or 1 000 000 000 gigabytes (GB).

1

FWA

151.1

5G

225.0

4G
or older

54.6

0.234

2010   2020   2030
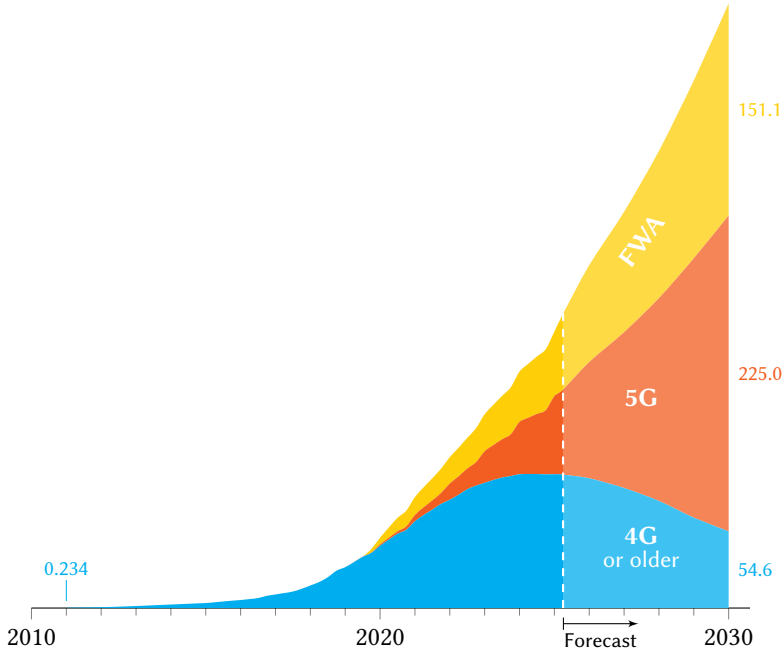
Forecast

Figure 1.1: Monthly mobile network traffic in EB/month, historic data and forecast [2]. While 4G traffic is expected to have reached its peak, 5G and fixed wireless access (FWA) will take over, and the total traffic keeps growing exponentially.

is the fifth generation (5G) of mobile networks. Its developing partnership (5G-PPP) states as the first key challenge that 5G should provide a "1000 times higher wireless (...) capacity (for a given area) and more varied service capabilities compared to 2010" [3]. These varied service capabilities are implemented by three "5G" variants. Two variants have different applications than 4G has already provided: 'ultra-reliable low-latency communications' and 'massive machine type communications,' requiring changes mainly on the control plane and protocol level. This dissertation focuses on the electronics that must enable the total capacity for 5G, with a special focus on their energy efficiency, integration level, and the bandwidth they can provide. In addition, it will provide an outlook even beyond 5G, towards 6G, as demand will most likely keep increasing.

1.1.1 Technological Advancement

The exponentially growing data capacity is enabled by the tremendous technological development in integrated circuits (ICs). This development trend was observed in the early 1960s, resulting in a self-fulfilling prophecy colloquially known as Moore's Law, which initially stated that the economic optimum number of transistors on an IC would double yearly [4]. The semiconductor manufacturing industry collectively aimed to shrink transistor dimensions to keep this prophecy going, as illustrated by Fig. 1.2, eventually halving the transistor area every two years. Since the 1980s, the main subject of this technology scaling became digital-oriented CMOS technology, with major IC manufacturers launching new CMOS nodes every year.

Figure 1.2: The number of integrated transistors per microchip, illustrating Moore's Law over time. Source: [4–6]

The power density did not change in first-order approximation while the transistor count increased. This so-called Dennard scaling was formulated in 1974, additionally stating that a circuit's delay time would decrease by 1.41× with every technology generation (halving of transistor area) [7]. This scaling law meant that the computing power for the same amount of electrical power would not double every two years but every 1.33 years instead. Without this immense growth in technology, we wouldn't be able to store and process the huge amounts of data we require today.

1.1.2 Network Energy Requirements

In 2018, mobile networks consumed an estimated total energy of 122 TW h globally [8]. Following current trends, as shown in Fig. 1.3, this is projected to grow over tenfold to 1300 TW h by 2030. Up to 3.6 % of the global electricity supply is then expected to power mobile networks, leading to large CO_2 emissions since most electricity is still generated from fossil fuels [9].

Aside from environmental concerns, it also brings about economic concerns. With increasing energy consumption, the costs of operating cellular networks also increase. This implies a higher subscription cost per user, ignoring current geopolitical tensions driving up energy prices. Users may not want—or simply be unable—to pay this price, even though they would get the 750× increase in capacity in return. Evidently, the expected power consumption does not scale proportionally to the expected capacity, showing the overall benefit of technological progress.

1

Figure 1.3: Predicted electric power consumption of networks [8].

1.2 Trends in Next-Generation Base Stations

Several techniques can be applied to increase data capacity, thus reaching the promise of increasing it by 1000 times. These can be divided into two rough categories: higher frequency bands or more transmitters with lower output power.

First, the use of higher frequency bands is briefly discussed, even though it is shown that this solution is not the be-all and end-all for future base stations. More focus is placed on the use of more transmitters as its consequences are more relevant to this dissertation's scope. Using more transmitters comes as small cells or as multiple-input multiple-output (MIMO) systems.

1.2.1 Higher Carrier Frequencies

More bandwidth becomes available when shifting to higher carrier frequencies. These higher frequency bands, also referred to as centimeter and millimeter wave (cmWave and mmWave) bands, are currently used (among others) for satellite communications and radar, for example, in automotive sensors, such as adaptive cruise control and parking sensors, or in speed cameras. Current use of the mmWave bands still provides enough available bandwidth, so there's little competition for (new) 5G and 6G band allocations.

These mmWaves tend to be attenuated very strongly by almost everything in our environment, including our bodies. Therefore, the range of these base stations is limited. This means that more of them can be placed without mutual interference, increasing the total potential capacity. However, this is simultaneously the major drawback for mmWave. A mmWave base station requires so-called line-of-sight to work properly: unobscured by buildings, windows, trees, or our hands, and it requires a correct phone orientation. Hence, the sub-7 GHz frequency range currently in use by 4G remains necessary for reliable cell coverage.

1.2.2 Small Cells, Massive MIMO, and Their Energy Consumption

Another domain to increase the available bandwidth is limiting the cell coverage in 2D space by decreasing the range of a base station such that fewer users are connected per base station, a technique known as small cells. Having fewer users connected means each user can use a larger portion of the available capacity while staying within the sub-7 GHz bands. There is a trend of decreasing cell size to handle the increased data traffic. A smaller cell also means more cells, but less RF power is required per cell. However, at best, the total transmitted RF power is the same to reach full coverage[2].

Another spatial method to increase capacity is focusing the RF beam toward the intended user, called beamforming. This beamforming can be achieved using multiple antennas and electronically adjusting their signals' relative amplitude and phase. Increasing the number of antennas gives more control over the beams, such as focusing the radiated power into narrower beams, which enhances the SNR for a given user. This enhanced SNR increases the data capacity of the channel, or one can choose to keep the SNR the same by scaling down the radiated output power. A 5G system using an array of 8×8 antennas could be considered massive MIMO (mMIMO), with increasing antenna elements for 5G-Advanced (5.5G) toward 16×16. For 6G, this will increase further to 32×32, which can be considered extreme MIMO (xMIMO) [11, 12].

Unfortunately, using many more cells or antennas significantly increases system complexity, overhead, and energy consumption. Due to the increased number of transmitters (TXs) and receivers (RXs) compared to 4G, the power consumption of the small-signal parts is duplicated, while the total transmitted RF power remains comparable [13]. The current state-of-the-art GaN-based base station power amplifiers (PAs) show up to 49 % average drain efficiency for typical signals, as illustrated in Fig. 1.4 [12]. This final PA stage accounts for 59 % of the total TX power consumption at this typical power output [13]. The drain efficiency lowers at lower output powers, for example, in low-traffic situations or with increasing modulation complexity and signal peak-to-average power ratio (PAPR). The remaining analog signal generation, small-signal parts, and PA quiescent current have a more constant power consumption that account for 46 % of the TX power consumption at very low output power, causing the significant energy consumption when duplicating these blocks.

1.3 Technology Scaling and Digital-Intensive RF Transmitters

The incredible progress in digitally oriented CMOS performance and economics also benefited classical analog and RF designs to some extent. More specialized CMOS processes fabricated on insulating substrates, called silicon on insulator (SOI), improve RF performance by minimizing parasitic capacitance and substrate losses. With digital logic becoming ever faster, their switching performance provides new opportunities for RF transmitters. Namely, by integrating more transistors, more functionality and reconfigurability can be provided. As such, more parts of the transmit chain have been digitized, which includes digital

---

[2]When the range (radius) r of a cell is decreased, the transmitted RF power required scales quadratic (r^2) to keep received signal-to-noise ratio (SNR) the same [10]. The covered area by one transmitter also scales by r^2. These perfectly cancel, so by first order approximation the number of transmitters to cover an identical area has to increase by the same amount as the individual transmit power is decreased.

1

(a)

(b)
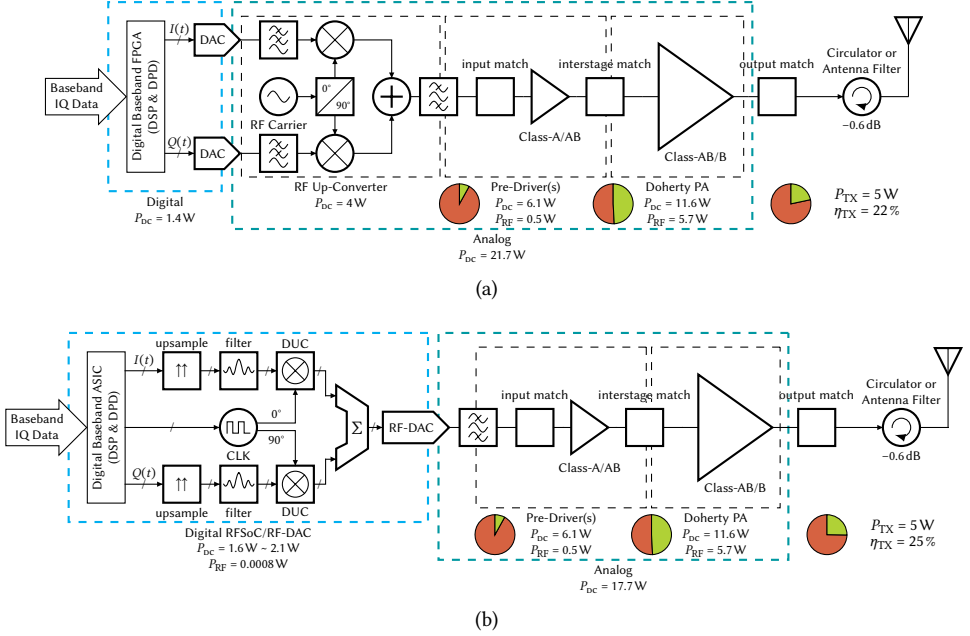
Figure 1.4: Average power generation and power consumption breakdown of the different analog TX stages: (a) Traditional analog-intensive TX line-up; (b) Modern mMIMO TX line-up, with digital-intensive RF signal generation. The digital power consumption for one TX channel is estimated using either a 1TX RF-DAC including interface [14] or a multi-T(R)X RF-DAC including interfacing and digital signal processing (DSP) [15, 16].

upsampling, digital up-conversion (DUC), filtering, and clock generation. As illustrated in Fig. 1.4b, it can offer significant power savings compared to their analog counterparts. However, the analog part currently still dominates the overall TX efficiency, spiking the interest in digital transmitters (DTXs) that benefit from the CMOS technology scaling by also digitizing the RF power generation.

One of the core assumptions of Dennard scaling is that an IC's supply voltage would be reduced by 30 % for every technology generation. While this is great for the power consumption of digital CMOS logic[3], it does not help RF power generation. The generation of RF power (P_{RF}) relies on a large enough voltage swing (for class-B an amplitude of V_{DD}, i.e., $2V_{DD}$ peak-to-peak) on a predefined load impedance (with resistive part R_L):

$$P_{RF} = \frac{V_{DD}^2}{2R_L}. \tag{1.1}$$

The typical supply voltages of digital-oriented CMOS technologies are simply too low to reach the power levels required for (analog) base stations, or require extremely low impedance levels that result in high output currents accompanying losses given the metal stack of CMOS technologies and makes it challenging to provide a (wideband) impedance match to 50 Ω. The market for high-power RF applications optimizes for vastly different parameters, such as minimizing drain loss and output capacitance while increasing

---

[3]While it lasted. The Dennard scaling law broke down in very small technology nodes due to increased leakage currents from the sheer number of transistors, also limiting the decrease of the supply voltage.

1

power density and gain. These needs caused IC technologies focused on high-power RF applications, such as LDMOS, GaAs, or GaN, to drift apart from digital CMOS.

Hence, digital transmitter (DTX) works published so far almost exclusively use advanced high-speed CMOS/SOI technologies with low breakdown voltages. As a result, all reported fully integrated DTX CMOS/SOI implementations are limited to at most 3 W of peak RF output power [17]. That is not enough to deliver enough average power for base station applications, even though the targeted TX output power is lowering due to the use of small cells and mMIMO. Furthermore, since CMOS/SOI technologies are optimized for digital applications and have limitations in their active devices and metal stack, they can only offer moderate drain efficiencies.

Hybrid DTX approaches, using combinations of CMOS/SOI with a high-power RF technology, are less frequently reported. These hybrid approaches often use an analog interface to drive the final output stage, yielding scaling limitations for higher frequencies or power levels. Alternatively, a dedicated high-breakdown CMOS driver can drive the high-power output stage digitally, which requires the output stage's input capacitance to be fully charged and discharged every RF cycle [18, 19]. As such, the digital operation of the RF output stage demands a constant input/drive power, which in deep power back-off can be even larger than the desired RF output power itself. As a result, the overall system efficiency will be low when handling signals with a high PAPR, as required in 5G.

1.4 Research Objective

Given the discussion above, there clearly is a performance gap between what digital CMOS can provide today and what is required for next-generation base stations. Even though transmitted RF power requirements tend to go down, CMOS technology is still unfit to provide either the RF output power or the efficiency levels needed. Benefiting from the increased functionality and power savings from the developments in digital CMOS while maintaining the power levels provided by high-power technologies—such as LDMOS—is taking the best of both worlds, giving rise to the following research objective:

> "How can digital-oriented low-power CMOS technology be combined with high-power RF technology such that energy-efficient operation of next-generation sub-7 GHz base stations can be achieved?"

1.5 Dissertation Outline

This dissertation is structured as follows.

Chapter 2 serves as background for the operation of DTXs. It does so by explaining common transmitter architectures, PA operating classes, PA efficiency enhancement techniques, and DTX topologies.

Chapter 3 presents the underlying technology and packaging challenges in combining digital CMOS and high-power technologies, which necessitate segmentation of the power device. It functions as a feasibility study for the concept of high-power DTXs and, as such, aims to answer the question: "What is necessary to make CMOS and segmented LDMOS/GaN physically and electrically compatible?"

Chapter 4 discusses the design aspects for high-speed digital drivers in CMOS. Namely, how to properly scale them for energy efficient operation, depending on the intrinsic parameters of the CMOS technology.

Chapter 5 explores the performance metrics of a DTX by proposing a mathematical definition of a DTX's transfer. It further explores how the segmented approach can be modeled and simulated to explain how a high-power DTX can be designed.

Chapter 6 proposes a power model of a DTX to explain its underlying power relations by first-order approximation. It is possible to perform hand calculations with the proposed power model, as well as to use its insights to optimize DTX designs.

Chapter 7 presents the designs and their measured results from several manufactured demonstrators that serve as the proof-of-concept for high-power DTXs. Using the experience from the design and measurement process it further discusses the lessons learned.

Chapter 8 uses these lessons as input to pioneer the next steps required. Doing so results in a high-density flip-chip assembly of CMOS and LDMOS devices to implement a high-resolution, high-power DTX.

Chapter 9 concludes the dissertation and projects the gained knowledge in an outlook for possible future high-power DTX implementations.
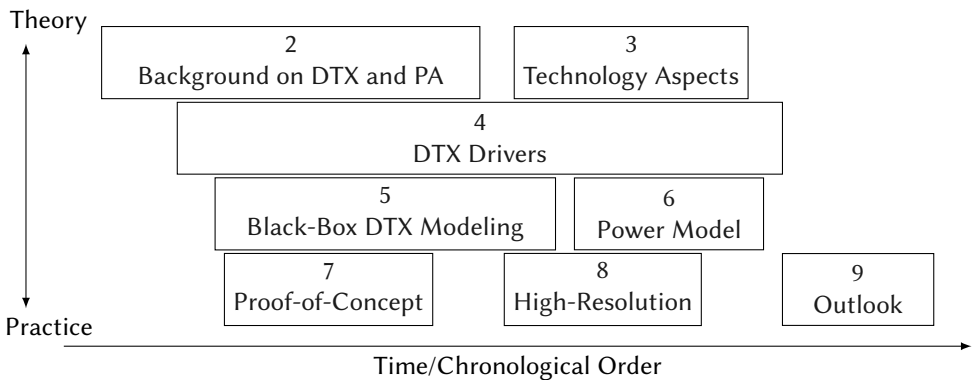
Figure 1.5: Chapter guide for this dissertation.

2

Background on DTXs and PAs

Chapter 1 showed that the demand for data capacity in mobile networks is increasing. This should not come at the expense of increased power consumption. However, in practice, this will be the case when capacity is increased by using smaller wireless cells and just reducing the base station's radiated power. This increased power consumption is due to the relatively large system overhead needed for the analog signal generation. Replacing the analog signal generation with more digital-intensive counterparts shows great promise in reducing power consumption. Therefore, pushing the boundary of the digital-to-analog conversion closer to the transmitter's output appears to be a good research direction. Ideally, in these digital transmitters (DTXs), only the output remains analog: the to-be-transmitted signal remains digital until the very end, implying that also the analog power amplifier (PA) is replaced.

To better understand the operating principles of a DTX, this chapter provides the background theory of digital-intensive transmitter operation. This background motivates later design choices and narrows down the design space for the implementation of high-power DTX configurations.

The structure of this chapter is as follows. First, a selection of transmitter architectures and amplifier classes relevant for digital-intensive operation are presented in Sections 2.1 and 2.2. Efficiency enhancement techniques can be applied to improve the efficiency of an amplifier, which are explained in Section 2.3. Lastly, Section 2.4 discusses three device topologies for implementing DTXs.

2.1 Transmitter Architectures

A conventional RF TX can be defined as a black box that modulates information from a (processed) baseband signal onto a high-frequency RF carrier as a passband signal. Next, it requires amplification to a specified power level before being transmitted by an antenna. The full conversion chain is depicted in Fig. 2.1. It needs to have a reasonably linear transfer from the baseband up to the antenna, such that its output has high spectral purity with minimal undesired spurious emissions. A transmitter can have several amplification or

---
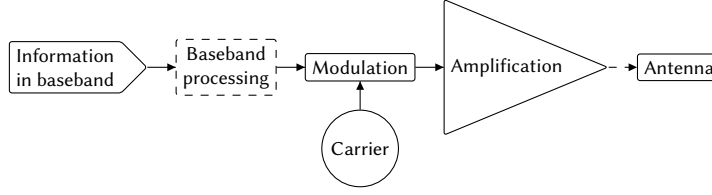
Figure 2.1: Transmitter block diagram.
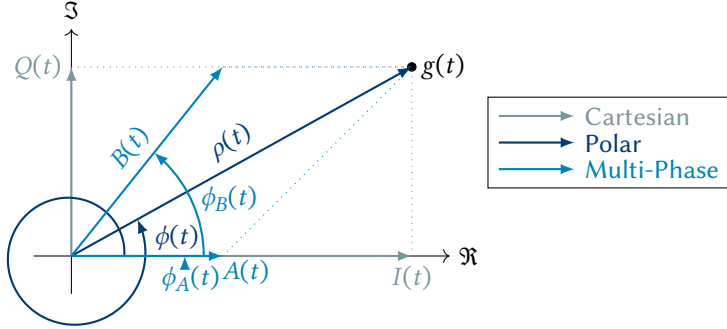
Figure 2.2: Different vector representations of a complex baseband envelope point.

modulation stages, such as within superheterodyne TXs. For simplicity, the TX architectures discussed next are limited to direct conversion architectures (i.e., a single modulation stage) as these are most relevant to digital-intensive TXs.

The baseband information can be encoded in the RF signal's amplitude, its phase/frequency, or both simultaneously. When both the amplitude and phase are modulated, this is called complex modulation. Many types of digital modulation make use of complex modulation. The full signal $s(t)$ can be described by $s(t) = \Re\{g(t)e^{j\omega_c t}\}$. Here, the complex baseband envelope is given by $g(t)$ that modulates the phasor $e^{j\omega_c t}$ at the carrier frequency $f_c = \omega_c/2\pi$. The complex baseband envelope can be described by a vector using Cartesian coordinates or by polar notation, as illustrated in Fig. 2.2. These closely relate to the Cartesian and polar transmitter architectures, which will be described first. Finally, a multi-phase transmitter is described, which finds a middle ground in terms of energy efficiency between the polar and Cartesian transmitter architectures.

2.1.1 Cartesian Transmitters

The complex modulated signal $s(t)$ can be reformulated using only real numbers by Euler's formula, resulting in

$$s(t) = \Re\{g(t)\}\cos(\omega_c t) - \Im\{g(t)\}\sin(\omega_c t). \tag{2.1}$$

Here, the complex baseband envelope $g(t)$ can be expressed by a time-varying in-phase component I and a quadrature component Q, resulting in

$$g(t) = I(t) + jQ(t) \tag{2.2}$$

$$s(t) = I(t)\cos(\omega_c t) - Q(t)\sin(\omega_c t). \tag{2.3}$$
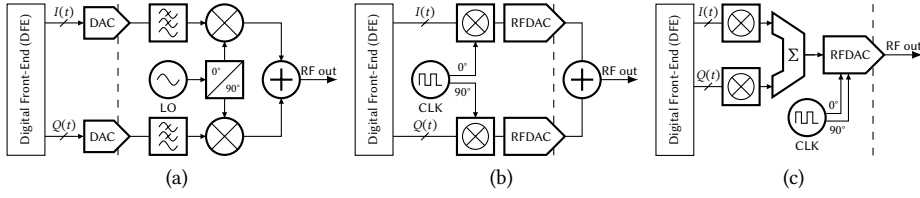
Figure 2.3: Cartesian transmitter architectures, where the functional blocks from (a) to (c) are increasingly digitized.

(a) Cartesian

(b) Polar
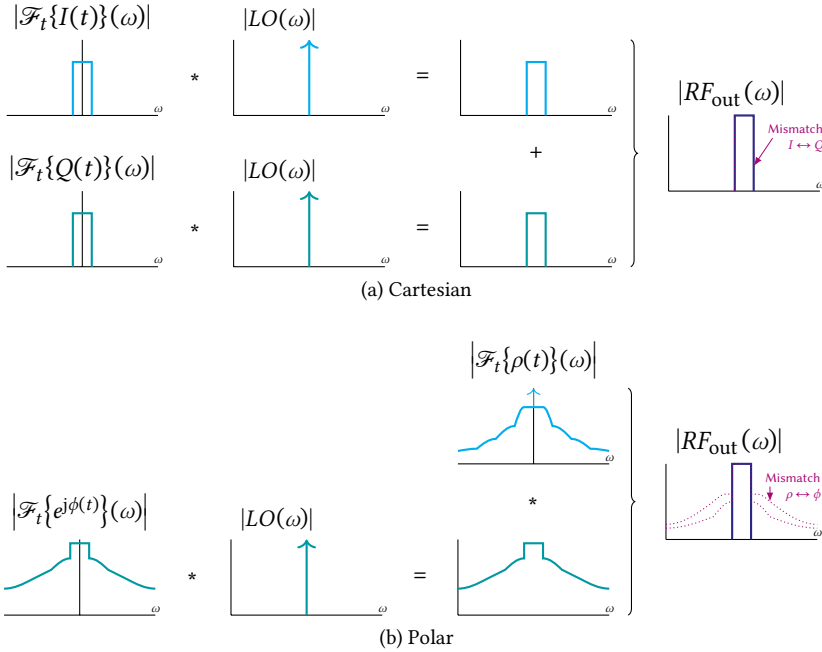
Figure 2.4: Spectra of the intermediate signals for different transmitter architectures.

This notation closely relates to the operating principles of a Cartesian transmitter, as shown in Fig. 2.3. There are two signal paths present, the I-path and the Q-path with (ideally) a 90-degree phase offset to the I-path. The I and Q signals do not mathematically interfere with each other as they are orthogonal and all operations are linear. This means that a band-limited channel also has band-limited I and Q signals, as illustrated in Fig. 2.4a. Even in the non-ideal case where the Q-path has a non-90-degree phase offset, the output spectrum is unaffected due to the linear operations involved.

When implementing two amplifier branches that have a linear efficiency relationship with their amplitude, their overall efficiency becomes dependent on both I and Q. The resulting normalized efficiency of a Cartesian transmitter can be captured in

$$\hat{\eta} = \frac{|g|^2}{|I| + |Q|} = \frac{I^2 + Q^2}{|I| + |Q|}. \tag{2.4}$$

2

(a) Cartesian   (b) Polar   (c) 8-Phase multi-phase
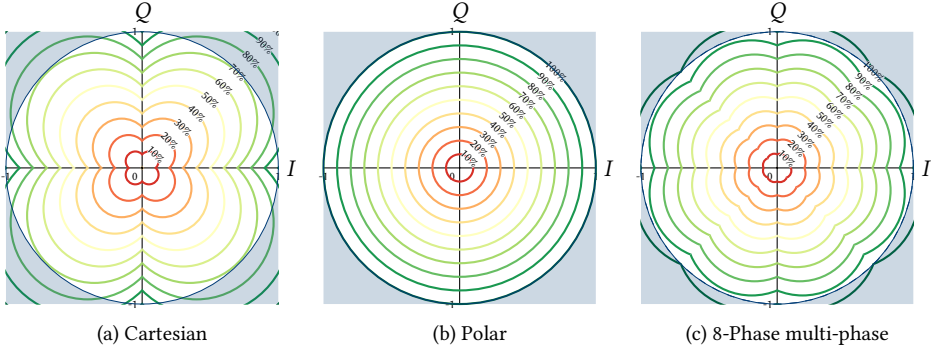
Figure 2.5: Normalized efficiency contours vs. complex modulation points for different transmitter architectures.

(a)   (b)   (c)
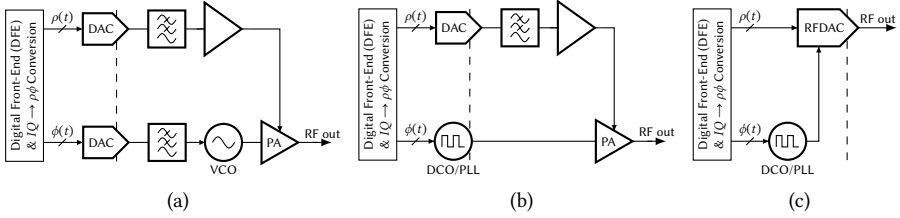
Figure 2.6: Polar transmitter architectures, where the functional blocks from (a) to (c) are increasingly digitized.

This results in the efficiency contours shown in Fig. 2.5a [21]. This relation can be explained by the fact that the output magnitude (the physical voltage amplitude) depends on the vector summation of I and Q, while the required power scales with their individual magnitudes. The physical voltage amplitude at the output of the (nonlinear) I and Q devices, combined with other possible nonidealities of the I and Q paths, result in a co-dependent transfer needing 2-dimensional digital pre-distortion (DPD) to correct it [22, 23].

2.1.2 Polar Transmitters

When using polar notation to express $s(t)$, we get

$$s(t) = \left| g(t) \right| \cos \left(\omega_c t + \angle g(t) \right). \tag{2.5}$$

Now, the complex baseband signal is represented by a magnitude and phase, which can be calculated from

$$\rho(t) = \left| g(t) \right| = \sqrt{I^2(t) + Q^2(t)} \tag{2.6}$$

$$\phi(t) = \angle g(t) = \text{atan2}\left(Q(t), I(t) \right) + 2n\pi. \tag{2.7}$$

The conversion from the orthogonal Cartesian base vectors to polar coordinates is clearly nonlinear, resulting in bandwidth expansion for the polar operation, as illustrated in Fig. 2.4b. This requires the use of components with a larger bandwidth to handle the ρ and ϕ paths of a polar transmitter, which are shown in Fig. 2.6. Only when the ρ and ϕ paths
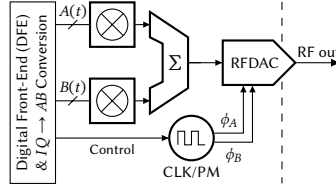
Figure 2.7: Example of a digital-intensive multi-phase transmitter architecture.

are combined is the resulting spectrum band-limited. If combining of the ρ and ϕ paths at TX stage is less accurate, for example, due to a timing difference, the resulting spectrum will suffer from a strong bandwidth expansion, which becomes worse when this mismatch increases.

On the other hand, the efficiency of polar transmitters only depends on their output magnitude,

$$\hat{\eta} = \frac{|g|^2}{\rho} = \rho, \tag{2.8}$$

yielding a straightforward efficiency contour, as shown in Fig. 2.5b. This provides a significantly higher phase-averaged efficiency than a Cartesian transmitter is capable of. The physical voltage amplitude at its output can be nonlinearly proportional to ρ and can also influence the actual output phase. These are the dominant nonlinearities in a polar transmitter, which as such only require two times a 1-dimensional correction for its DPD: one for AM–AM correction and one for AM–PM correction.

2.1.3 Multi-Phase Transmitters

Finding a middle ground between the benefits and drawbacks of Cartesian and polar transmitters is clearly something to strive for. Utilizing the possibilities provided by digital control logic, it becomes possible to map the complex baseband envelope $g(t)$ onto an arbitrary number of non-orthogonal base vectors. A logical choice for the number of base vectors would be a power of 2 due to its compatibility with binary logic. For example, using 8 of these vectors having 45-degree relative phase offsets. The real and imaginary parts of $g(t)$ can then be mapped onto these 8 vectors, where it is most straightforward to have only two of them active simultaneously, dividing the 8 vectors into two groups of 4, the A and B activations, where A and B have a relative phase of $\pm 45°$, depending on $g(t)$. Their magnitudes can be given by [24]

$$A(t) = \left\| \left| \Re\{g(t)\} \right| - \left| \Im\{g(t)\} \right| \right\| = \| |I(t)| - |Q(t)| \| \tag{2.9}$$

$$B(t) = \sqrt{2} \min \left[\left| \Re\{g(t)\} \right|, \left| \Im\{g(t)\} \right| \right] = \sqrt{2} \min \left[|I(t)|, |Q(t)| \right], \tag{2.10}$$

and their phases by

$$\phi_A(t) = \begin{cases} 0, & |I| \geq |Q| \wedge |I| \geq 0 \\ \pi/2, & |I| < |Q| \wedge |Q| \geq 0 \\ \pi, & |I| \geq |Q| \wedge |I| < 0 \\ -\pi/2, & |I| < |Q| \wedge |Q| < 0 \end{cases} \tag{2.11}$$

$$\phi_B(t) = \begin{cases} \pi/4, & |I| \geq 0 \wedge |Q| \geq 0 \\ 3\pi/4, & |I| < 0 \wedge |Q| \geq 0 \\ -3\pi/4, & |I| < 0 \wedge |Q| < 0 \\ -\pi/4, & |I| \geq 0 \wedge |Q| < 0. \end{cases} \tag{2.12}$$

The resulting structure is very similar to the digital-intensive Cartesian shown in Fig. 2.3c, with the addition of 45°, 135°, 180°, 225°, 270°, and 315° clocks. As apparent from Eqs. (2.11) and (2.12), the phases for A and B need dynamic selection, which can be implemented as a phase mapper [25], which is effectively a very low-resolution version of a polar phase modulator. While still prone to bandwidth expansions similar to polar transmitters, the defining equations for multi-phase operation are easier to calculate. Also, the timing requirements for the multi-phase amplitude and phase information are quantized, and thus it is possible to achieve perfect alignment by digital retiming: as long as the correct amplitude information remains aligned within the correct phase activation, the resulting fundamental channel is perfectly band-limited [26].

The related normalized efficiency contours are again dependent on the output magnitude, being the vector summation of A and B, over the individual magnitudes

$$\hat{\eta} = \frac{|g|^2}{A + B} = \frac{A^2 + \sqrt{2}AB + B^2}{A + B}, \tag{2.13}$$

which results in the normalized efficiency contours of Fig. 2.5c. These contours lie closer to the polar contours, providing a phase-averaged efficiency somewhere between the Cartesian and polar transmitters. Adding more phases to the multi-phase architecture moves the efficiency contours closer to that of the polar transmitter, with in the limit becoming identical to the polar efficiency. Adding more phases, however, increases the multi-phase architecture's complexity due to more discrete clocks being required, more strict timing constraints, as well as more complicated calculations for the magnitudes of A and B.

2.2 Amplifier Classes

Active devices (embedded in amplifiers) are required to bring the modulated signals to appropriate power levels for transmission. Power amplifiers (PAs) aim to convert their DC input power entirely to an amplified version of the power spectrum as presented at its input. However, it is very difficult to achieve both a perfect power conversion (i.e., efficiency) and a perfect (amplified) replication of the input signal (linearity). See Section A.2 for the definitions of efficiency and linearity used.

Efficiency is achieved by minimizing the simultaneous overlap between current through and voltage over the active device, typically a transistor. Namely, if both current and

(a)

(b)
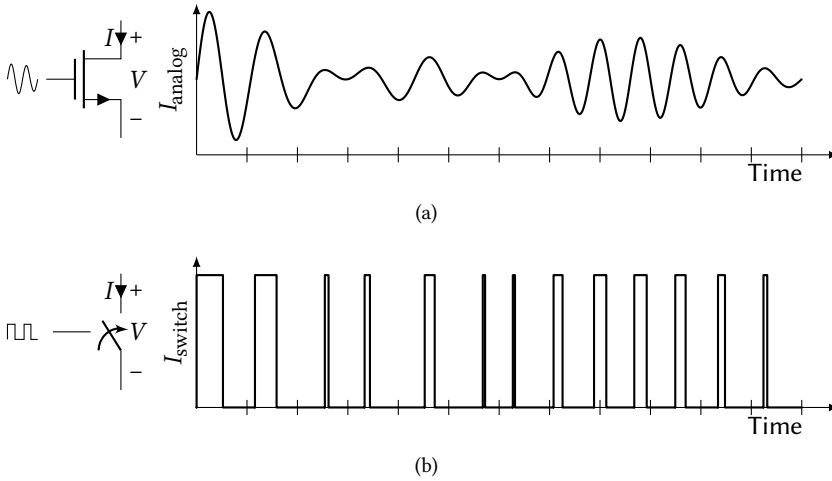
Figure 2.8: Illustrating the difference between "analog" and "digital" waveforms.

voltage are present at the same time, their product equals the power dissipated as heat in the transistor (Joule's law):

$$P_{\text{diss}} = I \cdot V. \tag{2.14}$$

At the heart of power amplifier design lies the art of designing the circuit networks around the active device such that the resulting output waveform is shaped for efficiency and the output spectrum meets the criteria for a certain signal bandwidth.

The operation of active devices can be categorized into several operating classes. The most well-known power amplifier classes are based on the transistor operating as a transconductance: classes A, AB, B, and C. Here, the input is a sinusoidal voltage around some DC bias, converted to a controlled output current. The output voltage is assumed to be sinusoidal, giving an unavoidable overlap between output current and voltage. The choice of DC bias sets this overlap and causes a trade-off between output power, efficiency, and linearity [27–31]. These transconductance classes are discussed first. When a rectangular voltage pulse with a small duty-cycle at the input is used instead of the sinusoidal voltage, a more favorable power–efficiency trade-off can be made [20, 30]. A rectangular voltage pulse is more closely compatible with digital circuits that benefit from technology scaling. The resulting output current pulses need to be scaled to control the output magnitude. Hence, these operation classes are called digital current scaling classes, which are discussed second.

Lastly, the active device could ideally be operated as a switch. The benefit of an ideal switch is that, the switch is either closed and the voltage across it is zero, or the switch is open and the current through it is zero. This means that the power dissipation of a switch is zero, and thus 100 % efficiency is theoretically possible. Using the active device as a switch is identical to how digital circuits operate, meaning it benefits maximally from technology scaling.

2
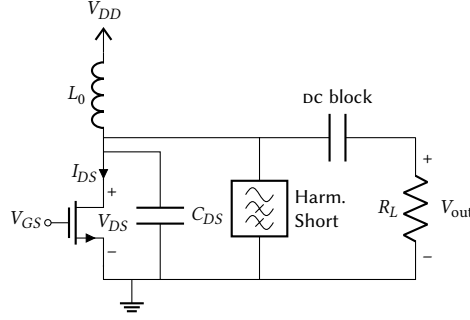
Figure 2.9: Basic circuit topology of transconductance amplifiers.

(a) Ideal device curve　　　　(b) Resulting device waveforms
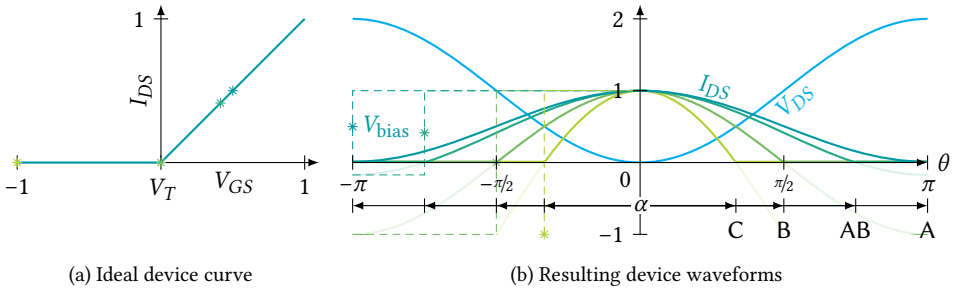
Figure 2.10: Idealized transconductance class operation, showing the device current for a range of conduction angles α with their respective bias voltages and currents.

2.2.1 Analog Transconductance Classes

The literature has extensively described the analog transconductance classes [27–31]. Here, a simple and normalized version for only peak output power is described, using an idealized MOSFET with a "linear" V_{GS}–I_{DS} curve, as shown in Fig. 2.10a. Also, all relevant voltages and currents are normalized, namely the transistor threshold voltage $V_T = 0\,\text{V}$, the maximum drain current $I_{DS,\text{max}} = 1\,\text{A}$, and the drain supply voltage is $V_{DD} = 1\,\text{V}$. The normalized 'time' that the transistor is conducting current is called the conduction angle α, which serves as a parameter that unifies the operation of all analog transconductance classes A, AB, B, and C. The resulting equations are provided in Section A.2.2.

For class-A operation, where $\alpha = 2\pi$, the MOSFET never turns off, and the drain current is a perfect sinusoid, as shown in Fig. 2.10b. For class-B operation ($\alpha = \pi$), the MOSFET is biased at the threshold voltage, is turned ON and OFF for half the time, and the drain current waveform is a perfect half-sine. Class-AB is the region between class-A and class-B operation ($\pi < \alpha < 2\pi$), and class-C is for all conduction angles smaller than in class-B, which requires an input bias below the threshold voltage. The resulting Fourier decomposition for all harmonics is plotted in Fig. 2.11. This shows that short circuit conditions are required for all harmonics to ensure a sinusoidal output voltage, with an exception for class-A, where no harmonic currents are generated, or for class-B, where there are only even harmonic currents present, which can be shorted practically using a $\lambda/4$ transmission line. Since perfect harmonic shorts are assumed, no power is generated or lost at these harmonics.
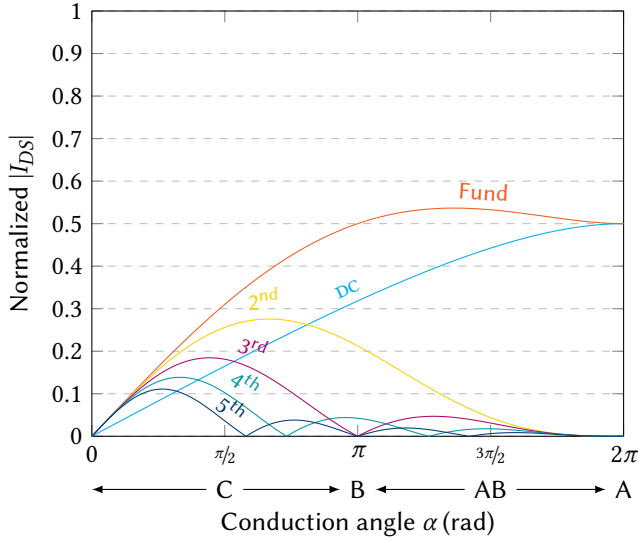
Figure 2.11: Harmonic currents of analog transconductance classes.

The output amplitude can simply be controlled by scaling the input voltage amplitude that drives the active device. It is important to note that the required device size increases with decreasing conduction angle to deliver the same output power, as is the case in class-C operation. The input voltage swing of V_{GS} (namely, $1 - V_{\mathrm{bias}}$) also increases asymptotically with decreasing conduction angle, lowering the amplifier power gain and requiring more input power.

With the matching network as given in Fig. 2.9, the output capacitance is resonated out only at the design frequency f_0. When significantly deviating from this frequency, the apparent load for the transistor gets a nonzero reactive part, degrading its performance. The higher the quality factor of the matching network, given by $Q = \dfrac{\omega_0 L_0}{R_L} = \dfrac{1}{\omega_0 C_{DS} R_L}$, the lower the bandwidth of the amplifier, giving a preference for a smaller C_{DS} for a given current capability. A way to extend the bandwidth of such an amplifier is to control the reactive impedance for the fundamental and second harmonic, yielding the (continuous) class-J-B-J*, which is also an analog transconductance class but is considered outside the scope of this dissertation. Furthermore, class-F operation is technically also a transconductance class, which can provide higher efficiency by controlling more harmonic conditions at the drain output to shape the drain voltage into a square waveform. Namely, by making the drain voltage more square, the overlap between the drain voltage and drain current gets reduced, lowering the power dissipation (see Section 2.2.4).

2.2.2 Digital Current Scaling Classes

When switching the transistor only ON and OFF, the resulting drain current waveform becomes rectangular. This switching can be achieved by using a very large sinusoidal V_{GS} drive to ensure quickly reaching $I_{DS,\mathrm{max}}$, or using a digital inverter to also provide a rectangular V_{GS} pulse to switch the transistor ON and OFF. Using a digital inverter as a
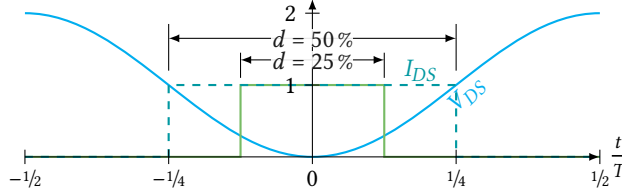
Figure 2.12: Ideal digital current scaling waveforms for different duty-cycles d.
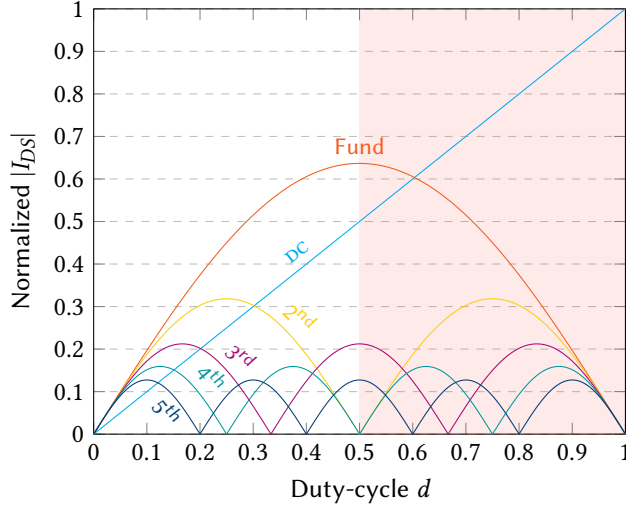
Figure 2.13: Harmonic currents of the family of digital current scaling classes.

driver is a far more interesting mode of operation in the context of digital-intensive TXs. The fraction of time the transistor is turned ON is called the duty-cycle d. The resulting ideal waveforms are illustrated in Fig. 2.12, and the drain current can simply be expressed by

$$I_{DS}(t,d) = \begin{cases} 1, & |t/T - n| \le d/2 \\ 0, & \text{elsewhere,} \end{cases} \quad (2.15)$$

which makes Fourier decomposition very straightforward, which for DC ($k = 0$) gives

$$I_{DS}(d)[0] = 2\int_0^{\frac{d}{2}} 1 \, dt = d, \quad (2.16)$$

and for all harmonics k gives

$$I_{DS}(d)[k] = 4\int_0^{\frac{d}{2}} 1 \cdot \cos(2\pi kt)) \, dt = \frac{2\sin(\pi kd)}{\pi k}. \quad (2.17)$$

The resulting harmonic current magnitudes are plotted in Fig. 2.13. Similar to the analog transconductance classes, all harmonic currents need to be shorted. For digital class-B
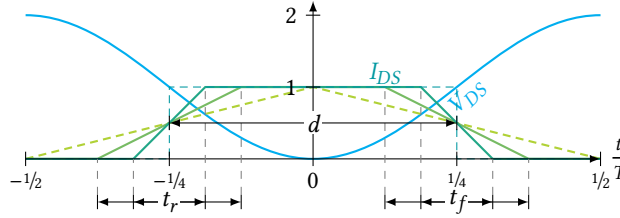
Figure 2.14: Digital current scaling waveforms with nonzero rise and fall times.

operation ($d = 50\%$), the transistor is turned ON and OFF for half of the time, making the drain current waveform a perfect square wave. Hence, only odd harmonic currents are generated, unlike analog class-B, which only has even harmonic currents.

Similar to the analog case, the drain efficiency and normalized optimum load can be calculated

$$\eta_D = \frac{I_{DS}(d)[1]}{2I_{DS}(d)[0]} = \frac{\sin(\pi d)}{\pi d} \tag{2.18}$$

$$R_{L,\text{opt}} = \frac{1}{I_{DS}(d)[1]} = \frac{\pi}{2\sin(\pi d)}. \tag{2.19}$$

The overall output current amplitude of a digital transmitter can be controlled by scaling it digitally through the activated width, as discussed in Section 2.4.3. This is why these operating classes are named 'digital current scaling.' It is again important to note that the required device size increases with decreasing conduction angle to deliver the same output power, such as for digital class-C operation. Anything beyond a duty-cycle of 50% makes no sense, as the fundamental current starts decreasing while the DC current keeps increasing.

Up to this point, ideal rectangular drain current pulses were assumed. In reality, these current pulses will have a nonzero rise time t_r and fall time t_f, as illustrated in Fig. 2.14. By defining the duty-cycle at the midpoint, the DC component remains identical. The harmonic content degrades to

$$I_{DS}\big(d, t_{rf}\big)[k] = \frac{2\sin(\pi k d)}{\pi k} \frac{\sin\big(\pi k t_{rf}\big)}{\pi k t_{rf}} \tag{2.20}$$

when assuming $t_r = t_f = t_{rf}$ and normalized to the RF period T for simplicity.

2.2.3 Comparing Analog Transconductance and Digital Current Scaling Classes

The analog transconductance and digital current scaling classes both use the transistor as a current source but with different current waveforms. Since they are so similar, they can be straightforwardly compared. First, a comparison at peak power operation is made, after which the operation at power back-off (reduced output magnitude) is compared.

Comparison at Peak Power

At peak output power, two metrics are of key interest. These are the fundamental output power and the drain efficiency. Analog class-B operation ($\alpha = \pi$) theoretically provides

2
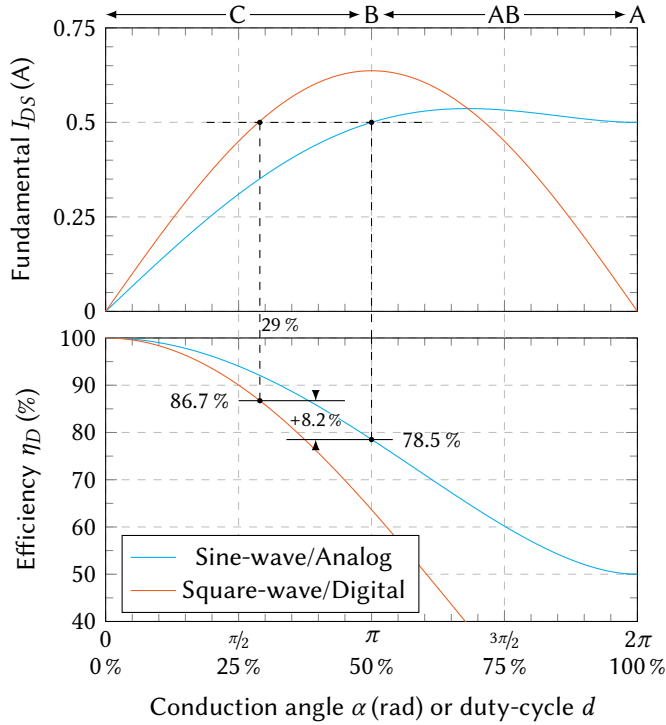
Figure 2.15: Analog and digital class-AB/C theoretical performance compared: Normalized output power and drain efficiency vs. conduction angle/duty-cycle [20].

78.5 % drain efficiency, whereas digital class-B operation ($d = 50$ %) has a relatively large overlap between $V_{DS}(t)$ and $I_{DS}(t)$, leading to an efficiency of only 63.6 % [20]. However, like in the analog case, reducing the conduction angle improves the achievable digital efficiency. This comes at the cost of a reduced RF output power capability. Figure 2.15 compares the output power and drain efficiency for the analog- and digital-driven output stage vs. conduction angle/duty-cycle. The square-wave current for the digital case yields a $4/\pi$ higher fundamental output power for the same $I_{DS,\text{max}}$, at a conduction angle of π rad. This provides a degree of freedom in trading off RF output power capability in favor of efficiency by reducing its duty-cycle/conduction angle. In doing so, we find that the digital drive has a significantly better efficiency–output power trade-off. Namely, at 29 % duty-cycle, digital current scaling has a drain efficiency of 86.7 % while providing the same output power as a traditional analog class-B-operated device. Further decreasing the RF duty-cycle to 25 %, which is a convenient choice for DTX implementations, the theoretical drain efficiency increases to 90 %, while its normalized output power reduces only from 0.50 to 0.45. In comparison, to achieve a 90 % peak efficiency using analog class-C operation, the related normalized output power reduces from 0.50 to 0.38. In Fig. 2.16, both cases' efficiency–output power trade-offs are visualized. Note that the digital "class-C-like" operation provides higher drain efficiency for a comparable, or even higher, output power than its analog counterpart.
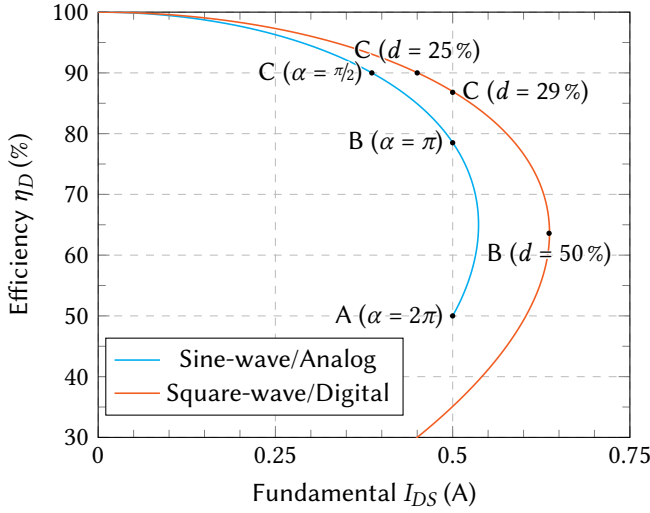
Figure 2.16: Theoretical drain efficiency for the analog and digital class-AB/C vs. fundamental output current [20, 30].

The previous section gave the theory of operation with nonzero rise and fall times. The resulting output power and drain efficiency can be visualized similarly (by plotting efficiency vs. fundamental current), now while varying both duty-cycle d and rise/fall times t_{rf}. This results in Fig. 2.17, where lines of constant duty-cycle are provided in gray, and lines of constant t_{rf} in varying shades of green to red. For reference, the analog transconductance classes are shown again in cyan. In the limit, at $d = t_{rf} \leq 50\%$, the output current becomes triangular, indicated by the dashed line in Fig. 2.14. It becomes clear that the rise and fall times should be sufficiently short to keep the advantage over the analog operation, e.g., better than $t_{rf} < 35\%$ when $d = 50\%$ or $t_{rf} < 15\%$ when $d = 25\%$. All duty-cycles beyond 50 % should be avoided since the fundamental current starts decreasing while DC current keeps increasing, but these are given here for completeness.

An observation was made by [30], who indicated that the rectangular current waveforms "have useful potential" but also recognized its very challenging drive(r) conditions. The digital drive concept (current scaling) is addressed in Section 2.4.3, and the drivers' implementation is discussed in more detail in Chapter 4.

Comparison at Power Back-Off

The rise and fall times of the digital current waveform include an important nonideality for digital operation. It is important to also recognize the nonidealities for analog operation. So far, an idealized MOSFET with a "linear" V_{GS}–I_{DS} curve (Fig. 2.10a) was considered. However, a device with such a sharp turn ON/OFF knee does not exist. A slightly more realistic curve is shown in Fig. 2.18, where even a class-B bias has a DC bias current through the amplifier, also called the quiescent current I_q. In Fig. 2.18, the bias current shown is $I_q = 0.1$ A and is hardly a concern at full power operation as the normalized input sinusoid has a signal swing from −1 V to 1 V. At no output power, the quiescent current is much more visible. This effect is illustrated in Fig. 2.19. In Figs. 2.19a and 2.19b, the drain voltage
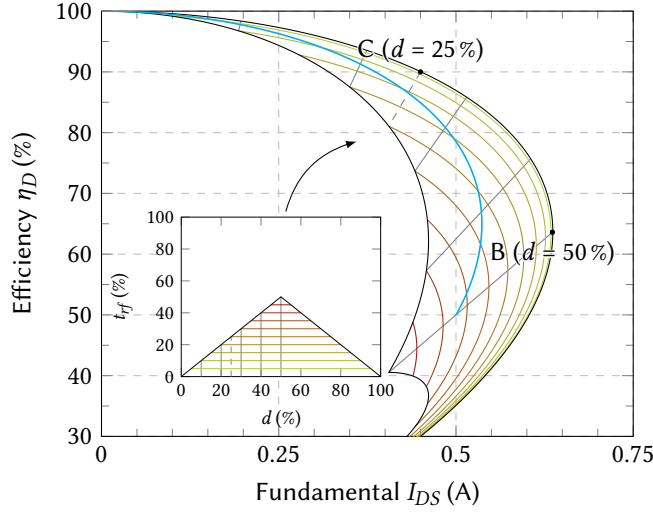
Figure 2.17: Drain efficiency for digital class-AB/C with varying rise/fall times vs. fundamental output current. The curve for analog is shown in cyan for reference.
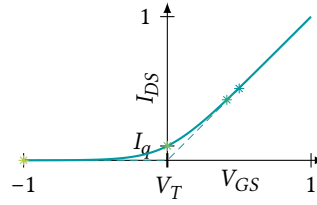
Figure 2.18: A more realistic device curve with a quiescent current I_q at $V_T = 0\,\text{V}$.

waveform is shown for full power operation and a power back-off point, respectively. The drain currents that cause these are then shown in Figs. 2.19c and 2.19d, with an analog class-B waveform with $I_q = 0.1\,\text{A}$ shown in red and a digital 8-phase multi-phase activation with a 12.5 % duty-cycle shown in green. The waveforms have been scaled such that their fundamental harmonic currents are identical, but their DC components are very different. Multiplying their drain voltages and currents and observing the area under these curves reveal the power dissipated in the transistor, as shown in Figs. 2.19e and 2.19f. This shows that the power dissipation caused by the analog I_q in power back-off can be quite significant.

The quiescent power consumption of analog amplifiers strongly depends on the operating class, see Fig. 2.20. Class-A has, by design, a constant DC power consumption regardless of output power (Fig. 2.20a). In contrast, ideal analog class-B has a linear DC power consumption and a quadratic RF output power, giving a linear drain efficiency relationship. However, with the transistor curve of Fig. 2.18 causing an I_q, the DC power consumption does not scale down to zero anymore for zero output power. This is clearly visible in Fig. 2.20c, where the solid lines indicate the non-ideal transistor curve, and dashed lines provide the idealized case.

To avoid the quiescent power consumption, the bias voltage can be lowered to operate
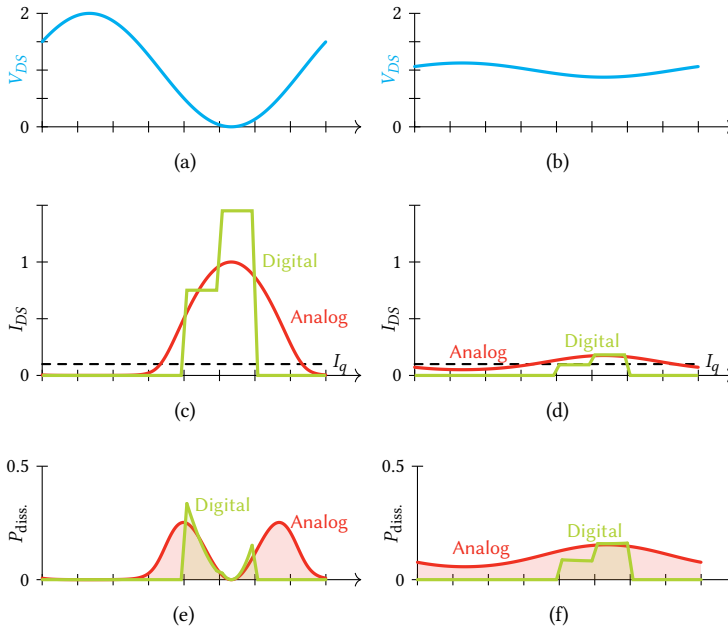
Figure 2.19: Comparing an example analog and digital current waveform, and their power dissipation for peak power condition and in power back-off.

the amplifier in class-C. However, as visible in Figs. 2.20d and 2.20e, the amplifier only starts to amplify once the input voltage reaches a large enough swing. Below this threshold, the amplifier has a very low gain. Overall, the gain of analog class-C amplifiers is not constant: they first experience gain expansion before going into gain compression. This could be a wanted behavior for Doherty amplifiers (Section 2.3.2). Still, the non-ideal device curve causes the DC power consumption to initially increase more rapidly than the RF output power, effectively resulting in overall drain efficiency loss with other Doherty branches.

In contrast, digital class-C operated DTXs do not experience gain expansion [20] (also see the measurments of Section 7.6). This is due to the distinct output magnitude control methods in place, causing the digital class-C to have an inherently linear transfer up to the amplifier compression point. Also, any quiescent bias currents can be avoided entirely in DTX, depending on the transistor's threshold voltage, as will be discussed in Section 3.1.1.

2.2.4 Harmonic Tuning and Switching Classes

Higher drain efficiencies can be achieved when harmonics are allowed in drain voltages, next to harmonic drain currents. For example, the switch-mode power amplifier (SMPA) classes aim to use the transistor as a switch: either the device is OFF or in the triode region. The ON-resistance R_{ON} is the limiting factor in reaching the 100 % theoretical efficiency that these switching classes can provide. The ON-resistance of several parallel devices can be employed to control the output power, hence also giving them the name of R_{ON}-scaling classes. Of equal importance is the OFF-resistance R_{OFF}, as a value that is too small will result in leakage currents. For DTX implementations, especially the classes where a transistor is
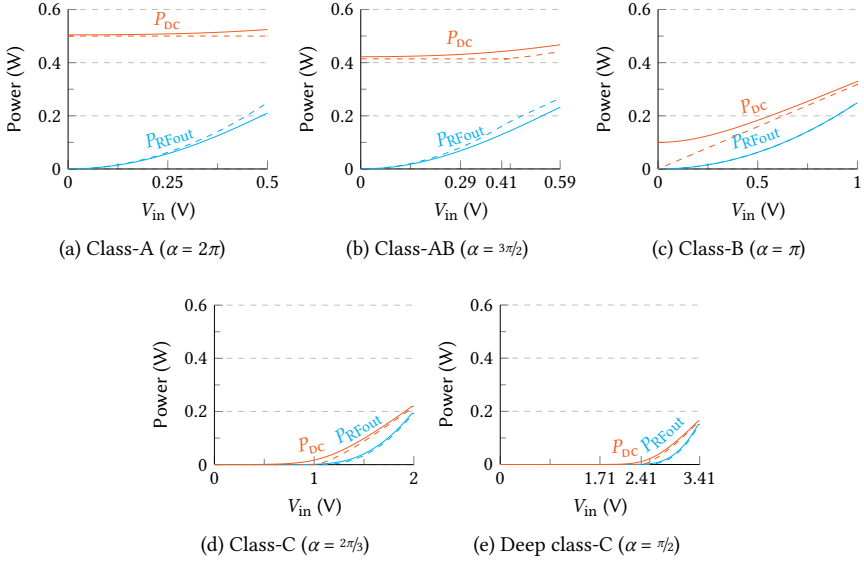
Figure 2.20: The impact in terms of power of an analog quiescent current I_q for different operating classes vs. input magnitude. The dashed line indicates idealized performance without I_q, while the solid line shows the performance using $I_q = 0.1\,\text{A}$.

operated as a switch are of interest due to their natural compatibility with digital logic.

Class-D and D^{-1}

A tuned class-D amplifier is a switching class amplifier that uses two complementary devices connected to a tuned output (Fig. 2.21a). That way, the output voltage and current still only contain the fundamental. The devices act as a 2-path, providing the necessary even order shorts to allow a half-sine current through either device, as shown in Fig. 2.21c. Further, all odd harmonics are open, allowing a square V_{DS} waveform. The theoretical drain efficiency is 100 % since there is no overlap between I_{DS} and V_{DS}. For RF purposes, tuned class-D is rarely used since the required complementary ᴘ-type device typically has inferior performance compared to an ɴ-type device: roughly twice the device width is required for the same R_{ON}, resulting in a 3× higher output capacitance than only using an ɴ-type device would have required. Class-D amplifiers are more widely used in non-tuned operations, such as low-pass filtered audio applications using oversampling techniques or as digital square-wave drivers with a capacitive load (without resonant filtering).

Inverse class-D, or class-D^{-1}, is used more often in DTXs as it can be implemented by two ɴ-type devices in a push-pull configuration (Fig. 2.21b) [32]. As typical for inverted classes, the harmonic shorts and opens are reversed. This means that for class-D^{-1}, there are even harmonic opens and odd harmonic shorts, providing square wave currents and half-sine voltages at the drain nodes, as shown in Fig. 2.21d.

Class-E

In a class-E switching amplifier, a single (ɴ-type) device is used, where the transistor output capacitance C_{DS} is explicitly used as the class-E shunt capacitance C_E. Class-E operation is
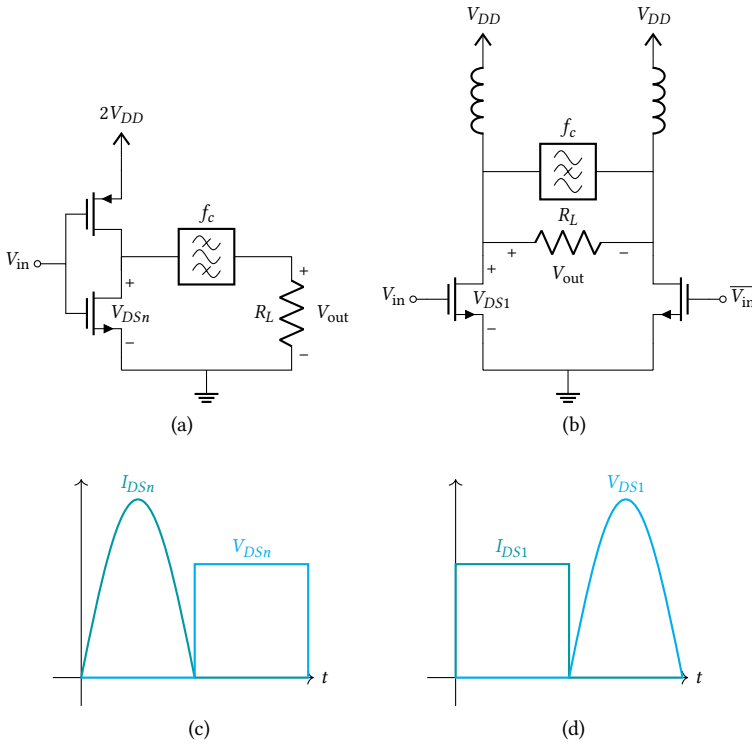
Figure 2.21: The circuit topology and resulting device waveforms for class-D and D^{-1} operation.

extensively analyzed in the literature, where the zero-voltage switching (ZVS) criterion is the critical requirement in reaching the theorized 100% drain efficiency [33]. A set of class-E normalization equations K can be defined depending on the RF duty cycle d and L_{DC}, where L_{DC} can range from finite to infinite[1] [34, 35]

$$q = \frac{1}{\omega_0 \sqrt{L_{\mathrm{DC}} C_E}} \tag{2.21}$$

$$K_C = \omega_0 C_E R_L \tag{2.22}$$

$$K_X = \frac{X}{R_L} \tag{2.23}$$

$$K_P = \frac{P_{\mathrm{RFout}} R_L}{V_{DD}^2}. \tag{2.24}$$

A technology-dependent upper frequency for class-E is given by

$$f_{E,\mathrm{max}} = \frac{I_{DS,\mathrm{max}}}{\alpha_E C_E V_{DD}} \tag{2.25}$$

---

[1]Acar *et al.* [34, 35] also defines a K_L to determine L_{DC}, but its calculation is straightforward from q and K_C.

2
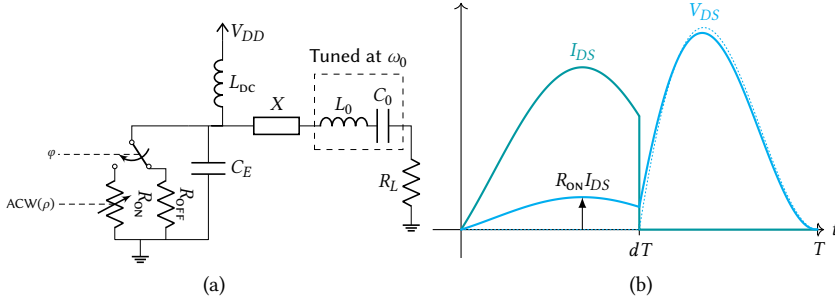
Figure 2.22: The circuit topology and resulting device waveforms for class-E.

where α_E is a constant ranging between 56 and 31, depending on the sub-class of class-E operation determined by the value q. Beyond this frequency, the ZVS criterion cannot be met anymore, and the active device operates (at least partially) in saturation rather than purely in its triode region. Conceptually, this can be understood by $I_{DS,\text{max}}$ being too small to discharge C_E fast enough. This also happens if $C_{DS} > C_E$, and shifts its operating class to the continuum between transconductance and switch-mode operation, also called class-BE operation or class-CE when $d < 50\%$ [36].

Class-F, F^{-1}, and E/F

Class-F amplifiers operate very similar to class-D amplifiers, but rather than using a 2-path approach to provide the even order shorts, these are explicitly implemented in the harmonic tuning. Only one (N-type) device is required, which is operated in the saturation region, making class-F also a transconductance-controlled class. The odd harmonics are open, thus again providing a square wave drain voltage waveform and a half-sine drain current waveform (Fig. 2.21c). If all the harmonics (infinite) are properly controlled, a class-F amplifier has a 100% peak drain efficiency, identical to class-D. For the even harmonics this is (theoretically) feasible by using a $\lambda/4$ DC feed line, but a lot tougher for the odd harmonics. Instead, only up to a few odd harmonics are controlled. For example, class-F$_3$ only has the proper third harmonic ($\eta_D = 88.4\%$), and class-F$_5$ has a proper fifth harmonic added ($\eta_D = 92.0\%$)[2].

Inverse class-F or class-F^{-1} has even harmonic opens and odd harmonic shorts instead. This gives a square wave drain current and a half-sine drain voltage (Fig. 2.21d). The device can be controlled by a square wave drive voltage, making class-F^{-1} of interest for DTX. Unfortunately, realizing a wideband class-F^{-1} matching network is near impossible when considering realistic devices with unneglectable output capacitance.

It is also possible to combine some of the harmonic conditions of class-F^{-1} with class-E switch-mode operation [37]. For example, second harmonic resonant class-E can be characterized as second harmonic open (as in class-F^{-1}) with the fundamental termination of class-E, and the other harmonic terminations that are capacitive in class-E. This combination of class-E and the second harmonic of class-F^{-1} can be denoted as class-E/F$_2$, which can be implemented using a push-pull structure [38]. Similarly, class-E/F$_3$ can be characterized

---

[2]Note that class-F$_1$ is equal to class-B, with $\eta_D = 78.5\%$.

by an added third harmonic short, which lowers the peak drain voltage present in normal class-E.

2.3 Amplifier Efficiency Enhancement Techniques

The discussion on amplifier classes mainly focused on a PA's operation at peak output power. A linear PA has an output power that is quadratic with its (linear) input quantity, as shown in Fig. 2.23. The required DC power is ideally proportional to the input quantity when no enhancement techniques are used, i.e., the drain efficiency is linearly proportional to the input. Possible enhancement techniques include either modifying the PA's supply voltage or the apparent load to the PA. These concepts are illustrated using the device load line in Fig. 2.24.

2.3.1 Supply Modulation/Switching

At lower output powers, an active device used in a PA has a reduced signal swing at its drain. The difference between the supply voltage bias and the output voltage amplitude can result in a higher loss in the active device (e.g., as in Fig. 2.19). These losses can be avoided by dynamically reducing the supply voltage bias, and higher efficiencies in power back-off can be achieved. The simplest version is to switch the supply between two or more values dependent on envelope amplitude. The PA uses a lower supply voltage when operating at low envelope amplitudes. As the envelope amplitude increases, the PA dynamically switches to the appropriate supply voltage. This is also called class-G operation.

Next is envelope tracking (ET) or class-H operation. With envelope tracking, the supply voltage continuously follows the envelope amplitude. This is typically combined with a linear amplifier, so the envelope tracker should provide enough voltage headroom not to affect the linearity.

Taking it one step further is envelope elimination and restoration (EER), where the supply voltage is entirely responsible for providing amplitude modulation. By its very nature, this type is a polar transmitter (Fig. 2.6), which means special care should be taken to align the AM and PM paths to avoid spectral impurities. As a benefit, a nonlinear amplifier with high drain efficiency can be used, such as class-E, which is otherwise unable to perform amplitude modulation.

Any of the supply modulation techniques operate using the amplitude ρ, which is subject to bandwidth expansion. This means the supply modulator should have a greater
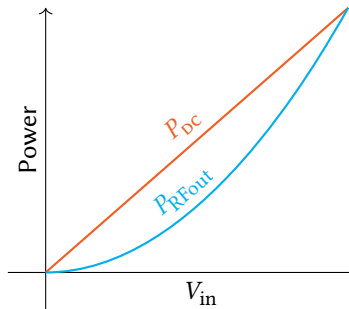
Figure 2.23: The power relations with input voltage of an ideal amplifier without efficiency enhancement.

(a) Supply modulation   (b) Load modulation
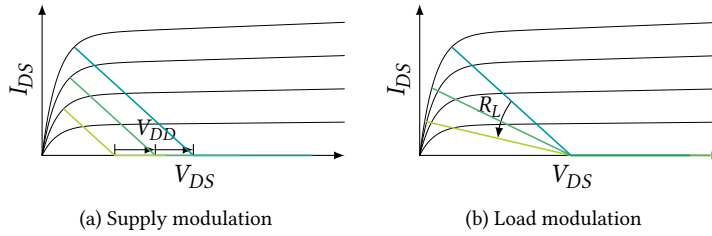
Figure 2.24: The change in device load lines caused by efficiency enhancement.

bandwidth than the signal's. Combine that with the same bandwidth requirements for the supply path(s) and device parasitics changing with voltage, and it becomes clear that it is hard to make very wideband amplifiers using these techniques.

2.3.2 Load Modulation

The apparent load to the device can be modified to avoid a small signal swing at the drain of a PA at lower output powers. Increasing the apparent resistive load can maintain the full signal swing even though the output current is reduced. The physical load remains the same, but the load modulation is achieved by the interaction of two or more PA branches. Several architectures achieve load modulation, which are discussed next.

Outphasing

The outphasing concept is based on two amplifier paths with an opposing signal phase offset, where this offset controls the output magnitude while the input drive levels remain constant. A possible implementation is shown in Fig. 2.25. The output signals ideally cancel perfectly at a 180-degree relative offset, resulting in no output power. The outphasing concept originates from linear amplification with nonlinear components (LINC), enabling nonlinear amplifiers (e.g., class-E) to also modulate amplitude. The two amplifier branches interact with each other, resulting in a load modulation ranging from $R_L/2$ to ∞ for each branch. The resulting normalized efficiency is shown in Fig. 2.26a as a dashed black line, and the loading of the top amplifier branch is shown in Fig. 2.26b.

By adding a compensation reactance X_c parallel to each branch, two high-efficiency points can be placed at a compensation angle ϕ_c, which are plotted in Fig. 2.26 with its resulting apparent loads [39, 40]. The result is an enhanced drain efficiency profile for
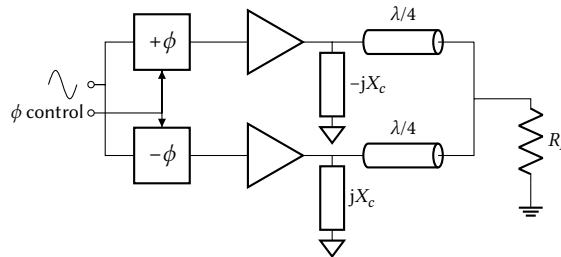
Figure 2.25: A typical outphasing circuit topology.
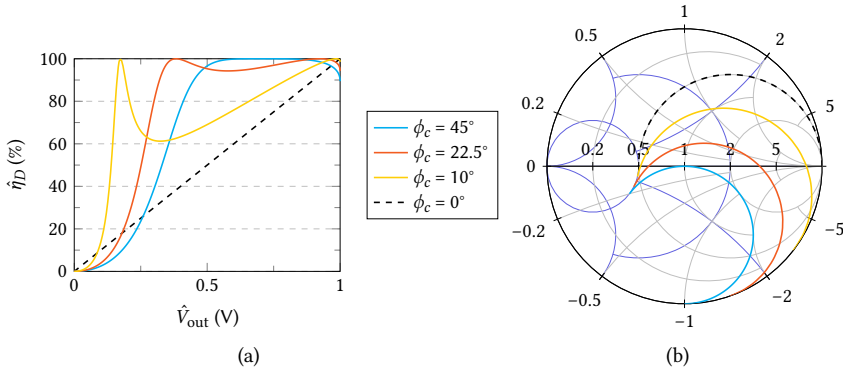
Figure 2.26: The efficiency and load conditions for a selection of outphasing compensation angles ϕ_c. The apparent load is only shown for the amplifier branch with capacitive compensation, the other branch has a loading curve mirrored over the resistive axis.
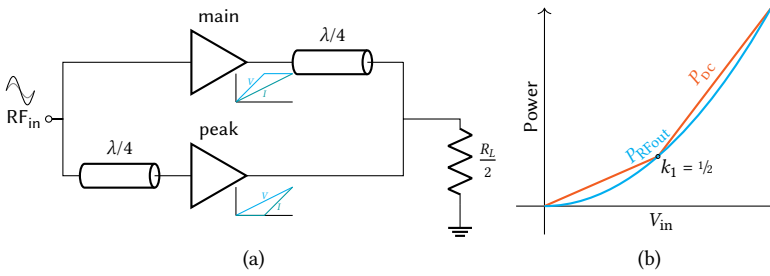
Figure 2.27: A typical symmetrical 2-way Doherty circuit topology, with its ideal power relations.

moderate output powers. However, the normalized drain efficiency is slightly lower due to reactive (mismatched) loading for peak power and very low output powers.

There are several other drawbacks. The fact that each branch has to be driven at constant amplitude results in a relatively large driver power consumption at low output powers, impacting system efficiency. The two amplifier branches must be very well matched to achieve very low output powers, and precise phase control is required. In general, outphasing has a very large load modulation ratio, which, combined with the reactive loading, makes the outphasing concept narrow-band.

Doherty

A Doherty amplifier uses one or more additional peaking amplifier branches to provide the higher output powers, whereas the main amplifier takes care of the low-power region. A typical topology is shown in Fig. 2.27 with the normalized power relations for a symmetrical Doherty. The peaking amplifier is turned OFF up to the first power back-off point k_1. The main amplifier reaches maximum voltage swing at this point. The peaking amplifier then turns ON, increasing the apparent load to the main branch. The main amplifier is connected through a $\lambda/4$ impedance inverter, thus lowering its apparent load. For a symmetric

2
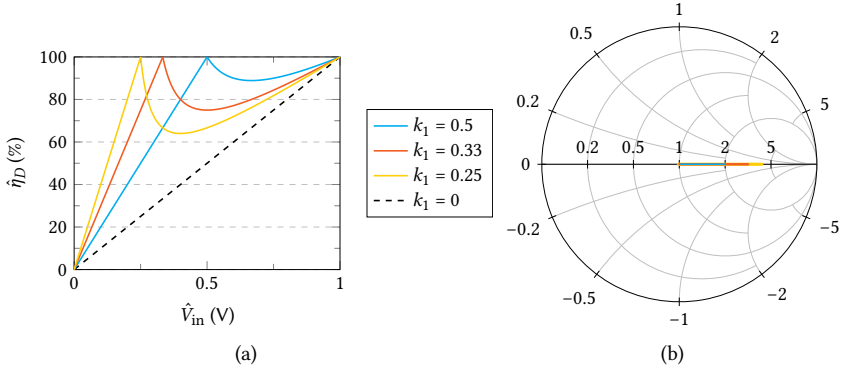
Figure 2.28: The efficiency of a Doherty amplifier and the load conditions for the main amplifier for a varying power back-off point k.

Doherty—with equal PA sizing—this load modulation ratio is 2: from $2R_L$ to R_L. The peaking amplifier then has a load modulation from ∞ to R_L.

Asymmetric 2-way Doherty PAs change the high-efficiency power back-off point by having two differently sized PAs. The resulting normalized efficiency curve is shown in Fig. 2.28, together with the loading conditions for the main amplifier. It shows that with increasing asymmetry, the first power back-off point k_1 shifts to lower amplitudes but increases the load modulation ratio with respect to its $R_{L,\mathrm{opt}}$ to $1/k_1$.

More peaking branches can be added to form an N-way Doherty that has high-efficiency points for $k_1, \ldots, k_{N-1}$, for which holds in general that each next branch reaches maximum V_{DS} at each consecutive k_n and

$$I_{DSm} = \rho I_{DSm,\mathrm{max}} \tag{2.26}$$

$$I_{DSp,n} = \begin{cases} \dfrac{\rho - k_{N-n}}{1 - k_{N-n}}, & \text{for } k_{N-n} \le \rho \le 1 \\ 0, & \text{elsewhere}, \end{cases} \tag{2.27}$$

providing a generalized piecewise normalized efficiency curve of

$$\hat{\eta} = \frac{\rho^2}{(k_n + k_{n+1})\rho - k_n k_{n+1}} \quad \text{for } k_n \le \rho \le k_{n+1}. \tag{2.28}$$

The Doherty concept provides a relatively moderate load modulation ratio, with no reactive loading at the design frequency, making it potentially wideband. Literature provides some design variations on the topology and circuit implementation to reduce losses and enhance the bandwidth [25, 41–44]. The main challenge in implementation is ensuring the input signal splitting results in the wanted output current from Eq. (2.27). One option is to bias the main amplifier in class-AB and the peaking amplifier(s) in class-C. The peaking amplifier, however, requires a larger gain and, as such, benefits from adaptive biasing circuits to improve performance. This becomes more challenging as more peaking branches are added. A Doherty amplifier is otherwise quite well-behaved, and its straightforward
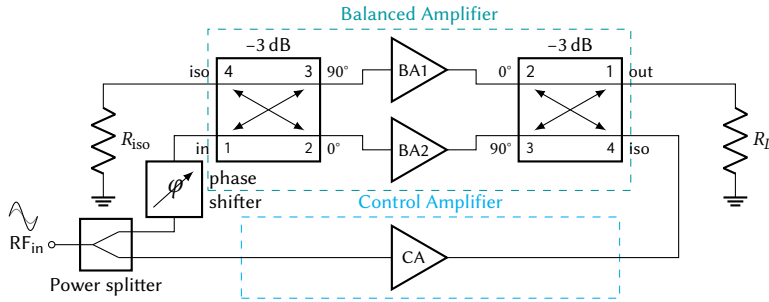
Figure 2.29: Circuit topology for a pseudo-Doherty load-modulating balanced amplifier.

implementation makes it the de facto efficiency enhancement technique that is employed in the industry.

Pseudo-Doherty using Balanced Amplifiers

Another way to achieve the Doherty efficiency curve is by using a load-modulating balanced amplifier (LMBA). The output combining network uses a 90-degree hybrid coupler, as shown in Fig. 2.29. The control amplifier takes care of the low-power regions and is fully turned on at the first power back-off point. As it is connected to the isolating port of the hybrid, it does not experience any load modulation by the balanced amplifiers (BA) connected to the 0-degree and 90-degree ports. Rather, its presence modulates the load of the balanced amplifiers; hence it is called a pseudo-Doherty technique: the PD-LMBA [45, 46]. Since it uses a 90-degree hybrid coupler instead of a $\lambda/4$ impedance inverter, it can be made very wideband. Aside from the very low to no load modulation ratio imposed on the control amplifier, the balanced amplifiers also have a limited load modulation of $1/k_1$. Additionally, the relative phase of the control amplifier also sets the phase of the load modulation of the balanced amplifiers, making it possible to compensate for reactive load variation over frequency. This potentially makes the PD-LMBA very wideband, only limited by the bandwidth of the individual PAs and the hybrid coupler.

A drawback is that the PD-LMBA requires even more control of the different amplifier drive conditions. Namely, the control amplifier should saturate at the power back-off level and not increase beyond that. Also, the hybrid coupler may be quite large when targeting large bandwidths with low insertion loss, while integrated coupler solutions are more prone to losses. The signal path of the control amplifier can be especially lossy, as it has to cross the coupler twice while being reflected by the balanced amplifiers. More efficiency points can be added by cascading a PD-LMBA functioning as a control amplifier in another PD-LMBA. However, this increases the complexity of controlling all different PA branches while being even more prone to losses.

2.4 DTX Topologies

Multiple transistor-level topologies are possible to implement the digital-to-RF-power conversion of a DTX, regardless of transmitter architecture or output operating class. There are three commonly used device topologies, which are discussed next.
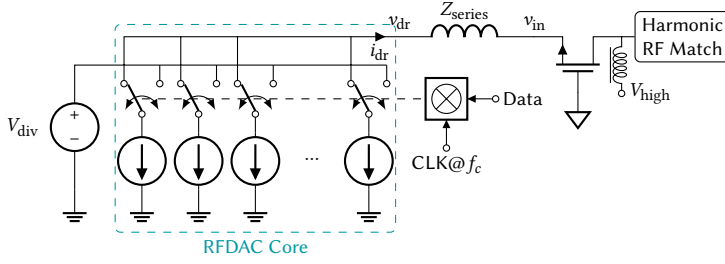
Figure 2.30: A current steering RFDAC controlling a common gate stage.

2.4.1 Common Gate

This topology is characterized by an external device with its gate (or base) connected to a signal ground. The common gate (CG) stage provides voltage gain to generate higher RF power, with a current controlling it at its source (or emitter). This current is typically generated by a current-steering (RF)DAC for its linearity [47, 48], as shown in Fig. 2.30. The bias voltage V_{div} makes sure that the diverted current also keeps the tail current sources saturated.

The combination can be extremely linear if implemented well. The CG stage should remain in saturation to benefit linearity. For that to happen, there should be a bias current present, which can be provided by adding (small) static current sources parallel to the current steering core at the cost of (small) added power consumption.

The small signal driver output voltage v_{dr} of the RFDAC core follows from the output current i_{dr} and the parasitic interconnect impedance Z_{series}

$$v_{\text{dr}} \approx i_{\text{dr}} \left(Z_{\text{series}} + \frac{1}{g_m} \right). \tag{2.29}$$

The driving voltage swing of v_{dr} at the output of the RFDAC core should be limited to be within the breakdown limits of the driver. The CG input voltage then follows from

$$v_{\text{in}} \approx \frac{\frac{1}{g_m}}{Z_{\text{series}} + \frac{1}{g_m}} \cdot v_{\text{dr}}, \tag{2.30}$$

where the v_{in} amplitude is preferably as high as possible [49]. The interconnect impedance is typically inductive, which increases with frequency. The output power is dependent on the voltage gain of the CG stage and the peak current the driver can provide. This yields scaling limitations towards higher frequencies and power levels; for example, generating powers above 10 W becomes very challenging.

2.4.2 Switched Capacitor

A switched-capacitor power amplifier (SCPA) is characterized by an array of unit cells containing an AC-coupled complementary driver [50]. Depending on the applied digital amplitude code word (ACW), a portion of these unit cells actively switches between its supply V_{low} and ground at the carrier frequency (f_c), whereas the remaining portion remains idle. This is conceptually illustrated in Fig. 2.32a. The ratio of active cells n to the total N
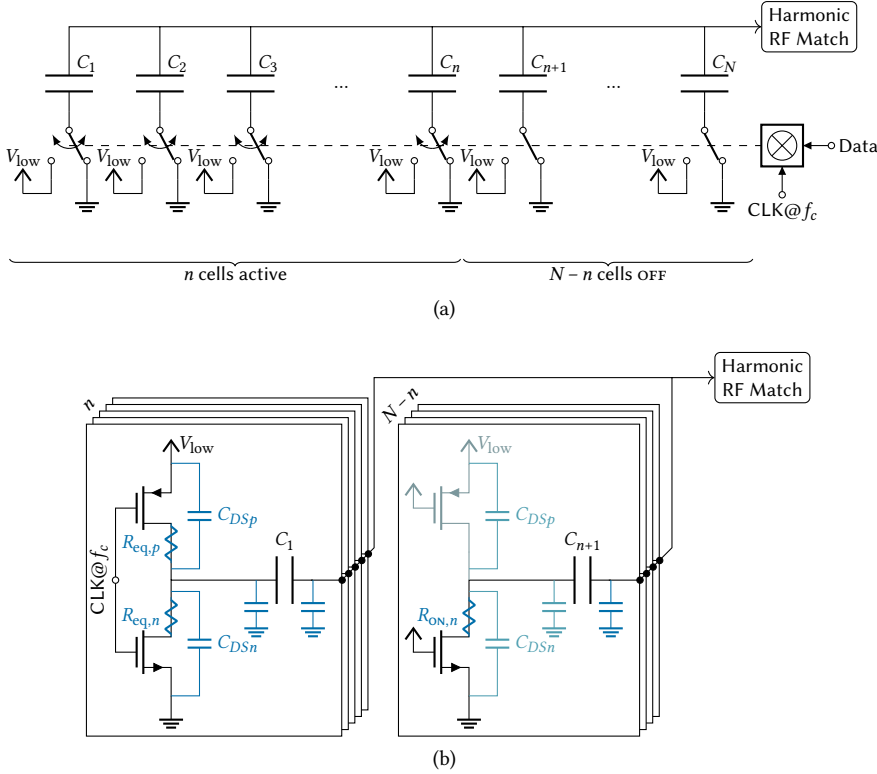
Figure 2.31: A switched capacitor topology and its parasitics.

determines the RF output amplitude. Generally speaking, a SCPA's efficiency is higher than that of a sole current steering RFDAC. The matching of on-chip capacitors is typically good, hence its linearity will be limited by any nonlinearities introduced by the complementary driver. The output power will still be limited due to low-breakdown digital-oriented CMOS technologies.

A possible circuit-level implementation is shown in Fig. 2.31b, together with highlighted parasitics. The unit cell implementation is effectively a CMOS inverter as complementary driver with a series 'switched' capacitor, with many of these unit cells in parallel. The series capacitors are typically identical and prevent a DC path from V_{low} to ground. The main source of nonlinearity is the CMOS inverter, which will typically exhibit a different effective switch resistance R_{eq} compared to its triode region ON-resistance R_{ON}, as well as changing behavior with overall output magnitude, though these can be minimized by proper design [51, 52]. Not all power supplied by V_{low} makes it to the RF output, with the main losses being introduced by the shunt device and parasitic capacitances, as well as the switch series resistances. In the OFF cells, the NMOS ON-resistance degrades the switched capacitor's effective quality factor (Q-factor). Typically, the total capacitance apparent to the output is resonated out by the matching network, for example using a resonant class-D operation. At low output powers or in a Doherty power back-off operation for example, the switched capacitors in OFF mode appear shunt to the output, increasing losses with

(a) Hybrid digital/analog   (b) Low-power digital

(c) High-power digital
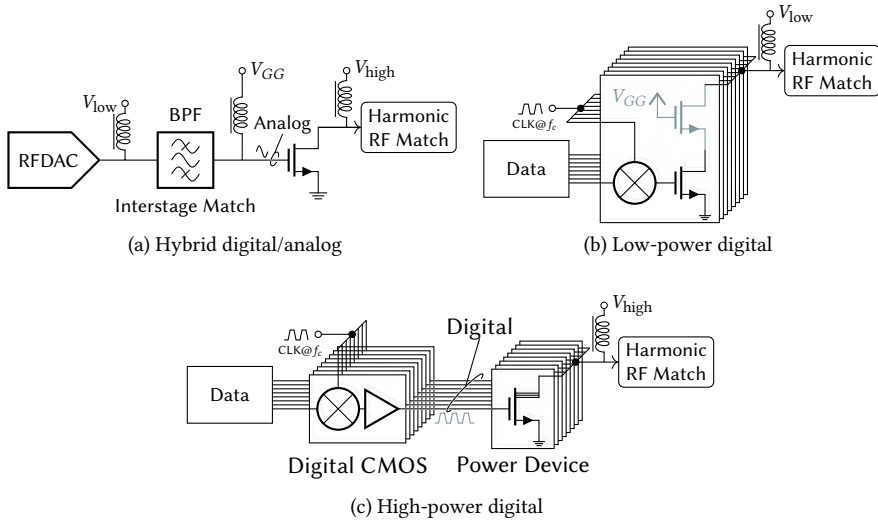
Figure 2.32: Examples of common source topologies.

worse Q-factors.

2.4.3 Common Source

It can be argued that the common gate topology (as shown in Fig. 2.30) is, in fact, a hybrid solution: an RFDAC with an external analog PA. This external PA can also be implemented using a common source (CS) topology, as shown in Fig. 2.32a. As a CS stage is controlled by its gate voltage, the power scaling issue can be avoided, albeit at the cost of requiring an inter-stage matching network. This matching network collapses the PA design back to a classical RF design, where a trade-off needs to be made between efficiency, linearity, bandwidth, and stability.

The CS topology can be used directly as the RF-generating component, as shown in Fig. 2.32b. The number of activated unit cells controls the output power. It does so by modulating the activated effective width of the device while the remaining width stays OFF. Compared to the switched capacitor approach, the efficiency can be higher, as only N-type devices with superior performance are required, combined with high-efficiency switch-mode operating classes. An optional cascode stage can be added to the CS stage, also drawn in Fig. 2.32b. This cascode is commonly used in DTX designs to increase the stack's breakdown voltage and, thus, output power. Still, even with the use of a cascode, V_{low} can only be in the order of 2 V, which might be considered high for a digital CMOS process, but is low in the context of RF power amplifiers.

An alternative is to use a dedicated high-breakdown technology to implement only the CS stage, which is then controlled digitally using a high-speed digital-oriented CMOS technology. This concept is shown in Fig. 2.32c. Its bandwidth, efficiency, and linearity are then limited only by the output stage and its matching network. The digital inter-face between the digital controller and the output stage has no matching requirements, theoretically removing any bandwidth constraints. The digital controller provides a low

effective impedance for the gate of the CS power stage, eliminating stability from the design equation. The output stage's supply voltage V_{high} considered can be in the order of 20 V to 50 V, which can enable true high-power DTX operation. This is why this topology is selected as the prime candidate for implementing a high-power DTX for this dissertation. However promising, this topology has never been implemented before due to challenges in practical implementation. Hence, these practical challenges are addressed in subsequent chapters to move towards a feasible implementation.

3

High-Power DTX Implementation Considerations

Chapter 2 introduced the operating principles of power amplifiers, both for analog and digital operation. However, generating the required multi-watt output power in fully digital implementations is challenging. Entirely integrated solutions are close to impossible due to the low breakdown voltages of the CMOS devices in advanced high-speed technologies. Therefore, a combination of high-speed digital CMOS and high-power RF technology is required. This chapter highlights the considerations unique to such a 'high-power DTX' combination.

To investigate the feasibility of a high-power DTX, we first explore the electrical constraints for high-power DTX in Section 3.1. Here, the electrical compatibility of CMOS with high-power RF technologies is addressed, and the requirements for DTX operation are formulated. Next, Section 3.2 addresses the practical obstacles in implementing such a topology, giving guidelines on how to arrange such a high-power DTX physically.

3.1 Electrical Considerations

In a power-DTX line-up, the (segmented) RF output stage uses binarily quantized driving signal(s) that puts the output stage (segments) in either 'ON' or 'OFF' mode. Achieving high efficiency from an RF output stage requires an N-type field-effect transistor (FET) in LDMOS or a N-type high-electron-mobility transistor (HEMT) in GaN technology. The driver of this output stage needs to provide very fast charging and discharging of the input capacitance of the RF output stage to guarantee digital and efficient operation. A class-D driver using complementary devices is considered the most logical candidate to perform this action. Unfortunately, in LDMOS or GaN technologies, no P-type devices are available

---
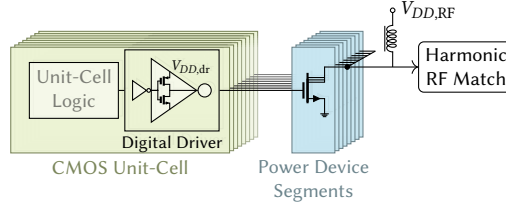
Figure 3.1: High-power DTX schematic using a common source configuration.

3

with adequate performance [54, 55]. In contrast, CMOS offers both high-performance (low-voltage) N-type and P-type devices, but, up to date, no high-voltage (e.g., $V_{DD,RF}$ > 20 V) RF-power devices, so monolithic integration is not an option.

As a result, we move to heterogeneous integration: a digital driver IC implemented in high-speed CMOS, interconnected to an RF power IC. This concept was introduced in Section 2.4.3, and the resulting diagram is shown in Fig. 3.1 with the relevant voltage domains: $V_{DD,dr}$ as supply for the digital CMOS driver, and $V_{DD,RF}$ being the drain bias of the power die. These two technologies have to be made electrically compatible with each other. The following sections discuss the requirements from the perspective of the power device.

3.1.1 Threshold Voltage Requirements

RF power FET technologies are typically optimized for their dominant high-volume application, namely base station Doherty operation. These are primarily analog Doherty PAs (see Section 2.3.2), with an output stage gain ranging from 13 dB to 22 dB and supply voltages ranging from 28 V to 50 V, based on gain, stability, and bandwidth considerations. In these applications, emphasis is placed on efficiency enhancement in power back-off operation, in combination with smooth AM–AM and AM–PM behaviors. Practical implementations aim to accomplish this by selecting the appropriate gate bias voltages to put the main device in class-AB and the peak device(s) in class-C operation. This design strategy has resulted in LDMOS and GaN technologies with high threshold voltages (V_T), e.g., a V_T above 2 V for LDMOS and below −2 V for GaN, while the V_{GS} voltage swing in these analog applications is in the order of 3 V_{pp} to 5 V_{pp} at the maximum drive level. This is illustrated for a few selected technologies in Fig. 3.2. In contrast, in digital-oriented high-speed CMOS technologies the supply voltage $V_{DD,dr}$ is limited to 1 V–2 V. In the targeted high-power DTX, the limited CMOS driver voltage should be sufficient to switch the LDMOS or GaN output stage (segments) between the 'on' and 'off' states. For GaN, this is somewhat more complicated since large negative voltages are required to turn the device completely off. Consequently, dedicated high-voltage CMOS devices and a related design technique have been developed for this purpose [56]. Establishing complete output stage switching with standard LDMOS or GaN devices is not practical when using commercially available high-speed CMOS technologies. Therefore, the V_T of RF power technologies for DTX applications needs to be reduced.

LDMOS technologies can offer some flexibility for lowering the threshold voltage. The downshifting of the LDMOS V_T can be done by selecting different doping concentrations or using thinner gate oxides, for example the low-V_T LDMOS plotted in Fig. 3.2 (see
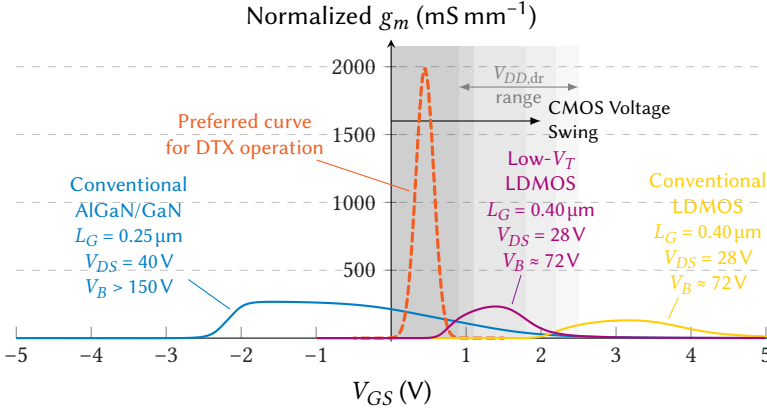
Figure 3.2: Example V_{GS}–g_m curves for some RF power technologies, with the shaded area for possible $V_{DD,\mathrm{dr}}$ ranges. Preferably the $g_m = 0$ for $V_{GS} = 0\,$V and $V_{GS} = V_{DD,\mathrm{dr}}$ and peaks in between to ensure complete 'OFF' and 'ON' switching.

Section 8.4.1 for its development). This is a delicate process since various other performance parameters like ruggedness[1] need to stay satisfied. To relax the V_T-shift requirements, thick-oxide CMOS devices can be considered to implement the drivers for the LDMOS output stage segments and their tapered buffer chains. Alternatively, a stacked ('cascode') driver design can be considered. The design considerations for these driver chains are discussed in Chapter 4. V_T engineering of GaN for DTX, as suggested in Fig. 3.2. still needs to start, while its negative V_T of depletion GaN devices imposes additional challenges in the driver design.

3.1.2 Gate Segmentation

High-efficiency SMPA operation, like class-E, is often entirely focused on the efficiency at peak power. However, when dealing with modulated signals with high PAPRs, control of the output power in an energy-efficient manner from a driver's perspective is needed. This can be achieved by segmenting the output stage [38, 58]. In low-power DTX implementations, this is done by putting many (scaled) unit cells in parallel, together with their embedded activation logic. This approach can be followed for both polar- [59] and Cartesian-oriented [60] DTX architectures. However, since we focus here on high-power DTX operation using separate dies for the CMOS controller and power stage(s), we cannot use (single chip) unit cells that include the RF-power-stage segment. To minimize layout and implementation parasitics, as well as to facilitate high output power and efficiency, segmentation of the gate width (W_G) of the power output stage itself is used. In such a gate-segmented power device (Fig. 3.1), all drain fingers are directly connected in parallel, and all sources are connected to the ground through highly-doped substrate plugs (LDMOS) or ground vias (GaN). Output power control is obtained by activating more gate segments, controlled by an ACW, scaling the effective R_{ON} (when operated in triode mode) or the effective g_m (when operated in saturation mode). R_{ON} control offers the highest theoretical efficiency (e.g., in class-E) but yields a nonlinear ACW–RF-output transfer. The latter can be handled

---

[1]Ruggedness is the ability to withstand a stress condition without degradation or failure [57].

by adopting a dedicated (nonlinear) segmentation technique [38] or by using DPD [61, 62]. Current mode or g_m-scaling can provide linear DTX operation up to compression [20], but this approach relies on a transconductance class of operation, yielding constraints in efficiency while being potentially more sensitive to variations of the driving voltage of the gate segments (see also the discussion on the measurement results in Section 7.5.3).

3.1.3 ESD Protection

When two dies are to be joined together, they must survive the die handling prior to bonding. Electrostatic discharge (ESD) is the biggest threat to the internal structures on a chip, especially for the (very thin) gate oxides. These gate oxides may already experience oxide breakdown in the order of one or a few volts, while an ESD event can be hundreds or thousands of volts. The drains of the on-chip devices are typically a bit more robust as they first enter non-destructive avalanche breakdown until thermal failure occurs; CMOS devices with a longer gate length can withstand up to 5 V or more.

Nonetheless, these on-chip devices need to be protected against potential ESD events. Several ESD protection strategies exist, the most common one is to place diodes from any chip input/output (IO) to the supply rails such that these diodes are in reverse bias when in normal operation. During an ESD event, these diodes divert the resultant ESD current while adding enough capacitance to the IO such that the resulting voltage at the ESD protected node never exceeds the allowed maximum. The level of protection is expressed in terms of the maximum static voltage that can be handled without damage, depending on the used ESD model. For the human body model (HBM), typical values are in the order of 1 kV to 8 kV [63]. For the charged device model (CDM), this is in the order of 125 V to 500 V [64]. The required protection level depends on the application[2]. For example, a packaged product needs a higher protection level than a sample which is only handled in ESD-protected areas.

The drains of RF power devices such as LDMOS have breakdown voltages (BV_{DSS}) beyond 70 V, or beyond 150 V in the case of GaN. Combined with the fact that the typical capacitance found at these devices' drains is relatively large means that ESD protection measures at these nodes may not be required. The Schottky gate barrier, as present in GaN HEMT, can be capable of dissipating ESD-induced currents. However, the gate oxides present in LDMOS, albeit thicker than CMOS gate oxide, may require protection depending on the capacitance present at the gate node (i.e., the total connected gate width).

The total connected gate width shrinks when gate segmentation is applied to the power device. This means that ESD protection is required for the outputs of the CMOS controller, as well as for the inputs of the RF power die. The capacitance related to the ESD protections adds to the total parasitic capacitance attributable to the interconnect. To avoid a disproportional reduction of the system efficiency, the total added switching power (including the upsized segment drivers) due to the ESD protections as a rule-of-thumb should not exceed 15 % of the targeted peak output power. As such, each gate segment should not be too small, i.e., each segment should be capable of providing at least 7× more output power than it costs to drive its ESD protection. The protection level required can be high in a configuration where the two dies are wire bonded together (see next section), as the dies and bond wires will be exposed to 'the outside world.' The protection requirement,

---

[2]Typical industry requirements are 1 kV HBM and 250 V CDM [65]. There used to be another model category, the machine model (MM) [66]. However, its use is discontinued as a technology qualification requirement in favor of the CDM [65, 67].

however, may be lowered for the flip-chip case since all connections between the two dies become unexposed internal nodes as soon as the two dies are joined together.

3.2 Physical Considerations

From the discussion of the electrical constraints above, it became clear that heterogeneous integration of a digital CMOS controller with an RF power die is required. This means that two dies must be physically connected, which is not straightforward to implement and yields practical constraints. In the following sections, these will be addressed.

3.2.1 Minimum Interconnect Pitch

RF power devices in analog applications typically have several gate fingers connected in parallel to a single gate bar, as shown in Fig. 3.3a. For a DTX, this single gate connection is cut into several individual gates when gate-segmenting the power device. Having as many gate segments as possible is preferable from the output signal perspective; namely, having more (digitally) activated segments offers finer control of the output amplitude and, thus, a better resolution of the power DTX. The digitally-controlled gate-segmented variant is shown in Fig. 3.3b.

The simplest interconnect method is using bond wires: one wire for each gate segment. Common wire bonding methods are ball–wedge bonding and wedge–wedge bonding. In ball–wedge bonding, a ball is first formed using a spark with the wire in a heated capillary. This ball is pressed onto a bond pad, and the wire is pulled towards the next receiving surface. This surface can be a pad on a printed circuit board (PCB) or another bond pad on a chip. This tail is then wedge bonded by thermosonicly attaching the wire on the receiving surface. This thermosonic wedge is placed using a combination of ultrasonic vibration, heat, and pressure to attach the wire. The resulting ball–wedge structure is shown in Fig. 3.4a. Wedge–wedge bonding uses the thermosonic wedge bond on both sides of the wire. Gold (Au) wires are often used in ball–wedge bonding and aluminum (Al) wires in wedge–wedge bonding, although either material or other materials such as copper are also possible.

The bond pad size dominates the minimum bond pitch, which mainly depends on the targeted bond wire diameter. The ball in a ball–wedge bond is roughly two to three times the size of the wire diameter, so the receiving pad on the chip should have a passivation opening of at least twice the wire diameter. Examples of common wire diameters are 18 μm, 25 μm, 38 μm, or 50 μm. The bond pad can be square to receive a ball bond. The bond pad for a wedge bond should be longer in the bonding direction, as well as have enough clear space to accommodate the tail of a wedge bond, roughly four times the wire's diameter or more. To minimize the pitch, thinner wires should be used. However, the bond wire interconnect also has inductive and resistive series parasitics, which tend to worsen for thinner wires. Conversely, very large bond pads have capacitive shunt parasitics to the IC substrate. This means there is a trade-off in wire diameter, pad size, and interconnect pitch.

When targeting a denser interconnect structure, staggered bond pads can be considered instead of in-line ones. Staggered bonds are shown in Fig. 3.4b. By placing two (or even three) rows of bond pads, the minimum bond pitch depends only on the bond wire diameter and the capillary size. Using such a small pitch causes the bond wires to inductively couple; the closer the wires are, the stronger this coupling. To minimize this coupling, close-by ground bond wires can be considered to minimize the overlap between current loops. These wires then function as ground return paths, as indicated in Fig. 3.3b.
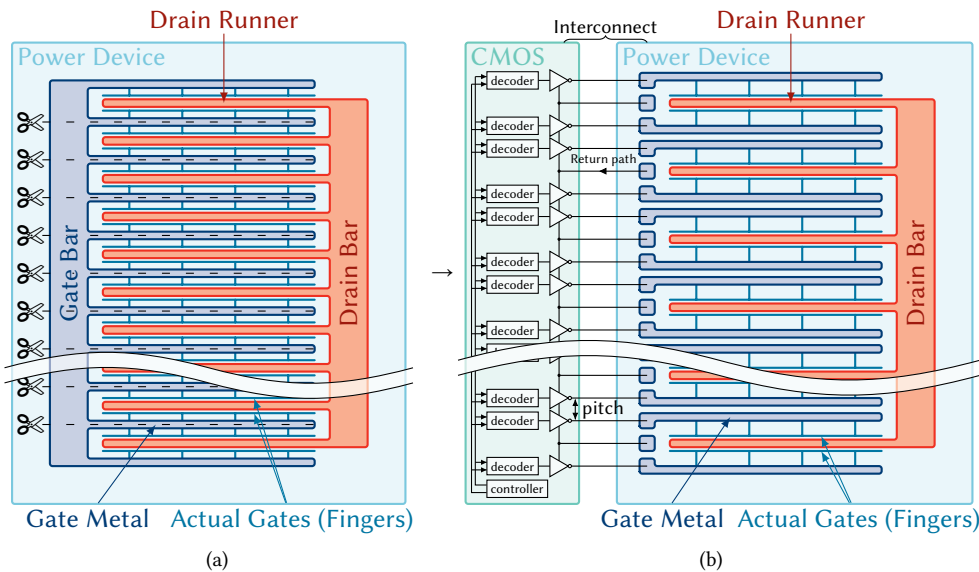
Figure 3.3: Segmenting a device layout commonly used for analog applications into a layout suitable for DTX operation.
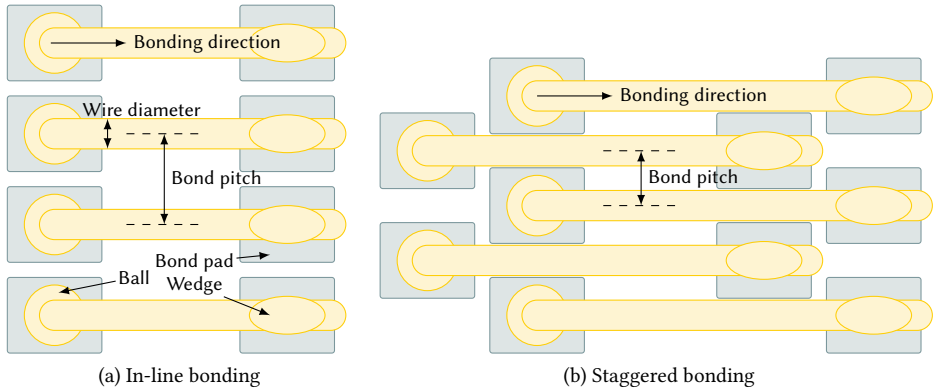
Figure 3.4: Examples of ball–wedge bonded structures.
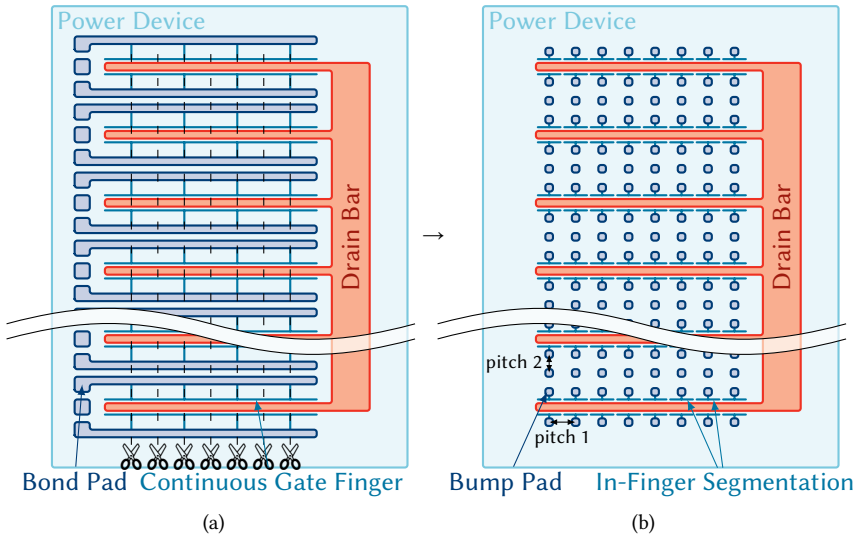
Figure 3.5: In-finger segmentation of a power device layout suitable for flip-chip bonded DTXs.

Further increasing the segmentation of the power device while minimizing interconnect parasitics requires more advanced packaging techniques. Using flip-chip bonding enables a full 2D grid of interconnections called bumps, whereas bond wires can only be placed at the chip edges. This further segmentation is illustrated in Fig. 3.5, where the gates are now also segmented within a finger. A continuous row of active areas can be achieved by placing the gate poly in-line, with a single connection for the drain. Each individual gate segment has its own bump pad connected to the digital CMOS controller. Typical foundry flip-chip processes support tin-silver (SnAg) bumps with pitches from roughly 170 μm down to 140 μm for more advanced processes, or even down to 80 μm using copper (micro-)pillars. Even more advanced, non-standard, flip-chip processes using an external bumping house can go down to a flip-chip pitch of 25 μm using SnAg micro-bumps [68]. Inductive parasitics can be minimized by providing ground bumps between the gate bumps. This can be even more effective compared to the bond-wire approach, as many more ground bumps can be placed between the digital controller and the gate-segmented power output stage segments. The inductance in the CMOS controller's power supply can also be minimized by alternating the ground bumps with supply bumps, lowering the supply-induced variations in the output voltage on the driver's control outputs (see Fig. 8.4 for an example).

3.2.2 Thermal Requirements

Attention should be directed towards power dissipated as heat, as we target a high-power DTX. Self-heating effects can degrade the performance of the RF power output stage, even though we aim to achieve a highly efficient operation. Hence, the thermal resistance of the power die to a heat sink should be limited. This is a standard consideration in RF PA packaging. Typical RF PA packaging involves thinning the power die (order of 50 μm) and attaching this thinned die to a metal flange that acts as a thermal interface as well as

the RF ground plane. This way, thermal resistances as low as $0.3\,\mathrm{K\,W^{-1}}$ from the junction to the case can be achieved [69]. A thin die has a lower source-to-ground inductance as added benefit. Copper (Cu) is a prime candidate for the flange material for its excellent electrical and thermal conductivities, if it weren't grossly mismatched with silicon (LDMOS substrate) or silicon carbide (typical GaN substrate) in its thermal expansion coefficient. The thermal expansion of copper is 6.5× the expansion of silicon (see Table 3.1), causing product reliability issues due to repeated heating and cooling, or assembly-related issues when heat is required (e.g., soldering or die attach). Molybdenum (Mo) or tungsten (W) can be added to the flange material to better match the thermomechanical properties of the flange to silicon. This can be done by making a Cu/W or Cu/Mo mixture, or by layering these metals together, depending on wanted properties. In the extreme case a flange entirely made of tungsten could be used, but this is rarely done as tungsten is hard to work.

Table 3.1: Electrical and thermal properties of relevant materials (at 293 K, where applicable) [70].

Material	Electrical Conductivity ρ (MS m^{-1})	Thermal Conductivity κ (W m^{-1} K^{-1})	Linear Thermal Expansion Coefficient α_L (10^{-6} K^{-1})
Silicon		149	2.56
Silicon Carbide		~340	~2.7
Copper	59.6	401	16.64
Molybdenum	18.7	138	5.10
Tungsten	18.9	173	4.42

3.3 Conclusion

In the sections above, we have explained the need for heterogeneous integration for a fully digital TX implementation capable of reaching a multi-watt RF output power. For its integration, we require a gate-segmented RF power die in a common source configuration, with its gates individually interconnected to a CMOS controller die. When using (staggered) bond wire interconnections, coarse segmentation is enabled, whereas flip-chip bonding can support a much finer segmentation. Finer segmentation is preferred to enable high DTX resolution, i.e., a finer output amplitude control.

Typical high-speed CMOS drivers are limited in their supply voltages and, therefore, cannot switch RF-power devices completely 'ON' or 'OFF.' Consequently, the threshold voltage (V_T) of RF power technologies for DTX applications needs to be reduced. ESD protection structures must be applied for each segment in the described multi-die scenario, likely on both sides of the interconnect. However, these ESD protections contribute additional parasitic capacitance, increasing the energy consumption of the gate-segment drivers, so this added capacitance must be minimized.

4

DTX Drivers

Chapter 3 showed the feasibility of a high-power DTX based on a hybrid technology featuring a high-speed digital CMOS driver/controller combined with a high-power RF LDMOS/GaN monolithic microwave IC (MMIC). In this configuration, the digital CMOS technology offers design flexibility and reconfigurability while each new CMOS node increases the switching performance, allowing the computational power of a DTX to be lowered. When using such a hybrid DTX solution (two separate dies), the wanted TX signal is constructed on the power MMIC. The controlling signals provided by the CMOS driver should have sufficient voltage swing to switch the segments on the power MMIC fully. Each output segment poses a capacitive load that needs to be charged and discharged at a speed as high (or higher) than the TX operating frequency. Digital drivers are used to perform this task. While typically not very different from a simple inverter, finding the best trade-off in speed and the power consumption of these drivers is very important to the overall DTX system efficiency. Therefore, the focus of this chapter is on these deceptively simple-looking devices.

In Section 4.1, the driver requirements are discussed. A simplified equivalent model is introduced to quantify the driver's impact on the overall system performance. From the theory in Chapter 2, we know that the driver's speed (in terms of rise and fall times) will affect the drain efficiency. Consequently, the driver's speed–power trade-off selection will have an optimum value for reaching the best system efficiency: the provided equivalent model gives directions on how to dimension the drivers. Next, in Section 4.2, these drivers with their actual implementation are considered in detail. Depending on the CMOS and power MMIC technologies available, the required voltage swing to drive the output stage might not be achievable with a 'standard' inverter driver design. Therefore, also other driver topologies are considered to achieve higher voltage swings at sufficient switching speed at the cost of increased driver complexity.
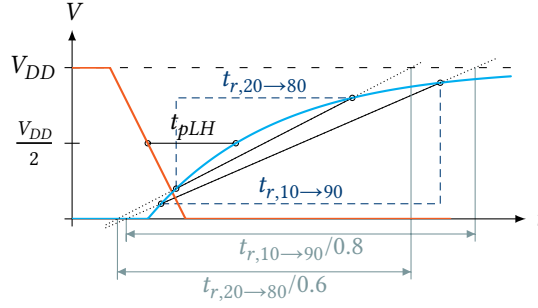
Figure 4.1: Propagation delay and rise time definitions illustrated using a linear (inverting) *RC* element.

4.1 Digital Driver Requirements

The purpose of a digital driver is deceptively simple: it must charge and discharge the gate capacitance of an output stage segment. A driver's load capacitance results in a switched capacitance loss of

$$P_{\mathrm{dr}} = f_0 C_L V_{DD,\mathrm{dr}}^2 \tag{4.1}$$

where f_0 is the switching frequency, C_L the load capacitance seen by the driver, and $V_{DD,\mathrm{dr}}$ the driver's power supply voltage. Equation (4.1) represents the absolute minimum power needed to fully switch the input capacitance of an RF-output stage segment in a controlled manner, assuming an ideal driver with no internal capacitances associated with it.

The driver's supply voltage ($V_{DD,\mathrm{dr}}$) must be chosen to satisfy the voltage swing requirements for the output stage as discussed in Section 3.1.1: it should sufficiently activate/deactivate (switch 'ON'/'OFF') the power RF device segments while not violating the reliability rules of the CMOS driver in a given technology. From Eq. (2.20), we know that a current-scaling DTX having equal normalized rise and fall times t_{rf}/T has the TX output stage's drain efficiency degraded by a factor $\mathrm{sinc}\left(t_{rf}/T\right)$. Consequently, the driver must provide sufficiently short rise/fall times not to significantly degrade the drain efficiency (e.g., not more than 10 % of the drain efficiency).

4.1.1 The Linearized CMOS Model

Any CMOS logic gate will have an equivalent switching resistance (R_{eq}) and input and output capacitances (C_i and C_o, respectively). Both the capacitances and equivalent resistance are highly nonlinear, which necessitates linearizing them to equivalent values before performing any analysis with them. Fortunately, methods of doing so are readily available in the literature [71]. From [71], all these 'linearized' quantities can be derived from variations of the propagation delay t_p.

The propagation delay for any logical circuit is defined by the time it takes for the output to change from its initial state to half its supply voltage ($V_{DD}/2$), referenced from when the input reached half the supply voltage. This is illustrated for an inverting logic function in Fig. 4.1. To be precise, it is the propagation delay for the output to go from logic 'low' to 'high': $t_{pL \to H}$ or simply t_{pLH}. Likewise, rise (t_r) and fall (t_f) times can be defined, of which the rise times are also illustrated in Fig. 4.1. Two common variants of rise time are used, namely the rise time from 20 % to 80 % of the targeted signal swing and 10 % to

90 %. Similarly, the propagation delay from 'high' to 'low' t_{pHL} can be defined. In a linear circuit, with an ideal R and C, these values can be found by

$$t_{pLH} = t_{pHL} = \ln(2)RC \tag{4.2}$$

$$t_{r,20\rightarrow80} = t_{f,80\rightarrow20} = \ln(4)RC \tag{4.3}$$

$$t_{r,10\rightarrow90} = t_{f,90\rightarrow10} = \ln(9)RC. \tag{4.4}$$

However, direct use of their value definitions yields inconsistent results when considering practical CMOS logic circuits having nonlinear resistances and capacitances. The propagation delays and rise and fall times are leading in the end, as they are the values that can be simulated or measured. Moreover, the linearized CMOS parameters are defined in terms of the propagation delay, which is also the industry standard for characterizing the delay of combinational logic. Consequently, only the propagation delay is used to derive the equivalent driver resistance $R_{dr,eq}$ (henceforth simplified to R_{dr}) and capacitance C_{dr}, which can then be related again to the rise and fall times and, thus, the drain efficiency.

4.1.2 CMOS Driver Model

Figure 4.2a shows the circuit and equivalent model of a CMOS inverter. For an inverter to function as a driver, it should be sized such that it can charge and discharge a capacitive load sufficiently fast. Whether this driver is inverting or not (being an inverter or a buffer) does not matter since another inverter can be added in front to get the desired logic behavior. A driver can be in one of the four following states.

- **Off**
  In the OFF-state, the output node (V_o) is connected to the ground through the ON-resistance of the driver's NMOS device.
- **Charging**
  The logic state of the input node (V_i) has just been changed, causing the driver's PMOS device to turn on. Its equivalent resistance ($R_{dr,p}$) charges the capacitive load (C_L) to the supply voltage ($V_{DD,dr}$).
- **On**
  In the ON-state, V_o is connected to the supply through the ON-resistance of the PMOS device.
- **Discharging**
  The logic state of V_i has just been changed, causing the driver's NMOS device to turn on. Its equivalent resistance ($R_{dr,n}$) discharges C_L to ground.

The 'drive strength' is determined by its equivalent resistance while switching. Typically, a driver is sized such that the equivalent resistances are equal, i.e., $R_{dr,n} = R_{dr,p} = R_{dr}$. This results in symmetrical charging and discharging behavior, i.e., $t_{pLH} = t_{pHL} = t_p$. Increasing the gate width of the NMOS and PMOS devices decreases R_{dr}. The lower R_{dr}, the 'stronger' the driver.

However, the driver itself also has an input ($C_{dr,i}$) and output capacitance ($C_{dr,o}$) that are proportional to its size. When (dis)charging the capacitive load, the driver also needs to (dis)charge its own $C_{dr,o}$ (its self-loading). This leads to the intrinsic propagation delay t_{p0} as a technology-dependent constant which is independent from transistor width, namely:

$$t_{p0} = \ln(2)R_{dr}C_{dr,o}. \tag{4.5}$$
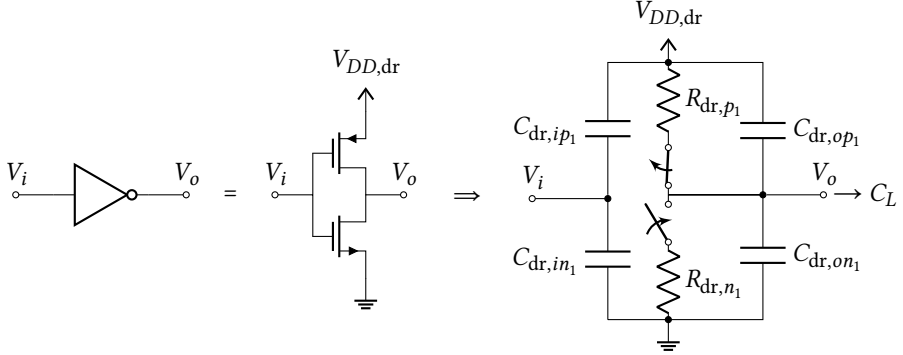
(a) Schematic and equivalent linear model of the final CMOS inverter ('driver')
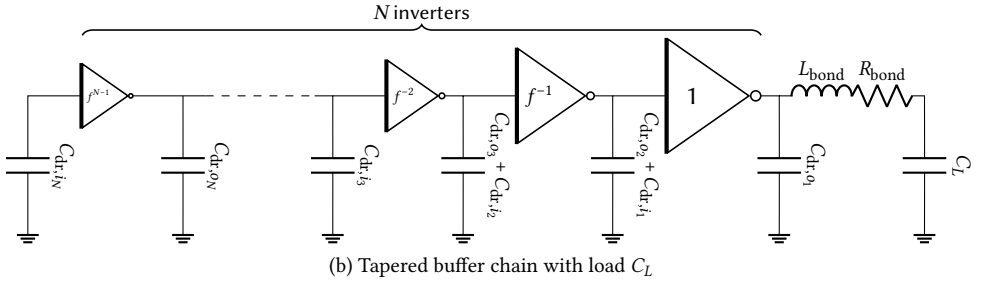
(b) Tapered buffer chain with load C_L

Figure 4.2: The driver chain for a single output stage segment.

For its definition, a stepped (zero rise or fall time) input signal is assumed [71]. When considering more inverters connected in a (buffer) chain, an inverter needs to drive both its own $C_{dr,o}$ as well as the $C_{dr,i}$ of the next inverter

$$t_{p,\text{int}} = \ln(2)R_{dr}\left(C_{dr,o} + C_{dr,i}\right), \tag{4.6}$$

using again the stepped input assumption. In view of this, the proportionality factor γ can be defined as a technology-dependent constant

$$\gamma = \frac{C_{dr,o}}{C_{dr,i}}, \tag{4.7}$$

which is close to 1 for most CMOS processes [71]. However, in practice, the input signal is not an ideal step in such a chain. When considering a non-stepped input, the propagation delay of an inverter becomes longer and depends on the propagation delay of the preceding inverter, this can be modeled by including a factor ς[1], yielding

$$t_{p,n} = t_{p0} + \varsigma t_{p,n-1}. \tag{4.8}$$

This behavioral model is strongly simplified. For example, the gate-to-source capacitances (C_{GS}) contribute to $C_{dr,i}$, but the effect of the gate-to-drain capacitances (C_{GD}) to the

---

[1]Rabaey *et al.* [71] defines this constant with the symbol η. Since this may be confusing with efficiency η, we here use the symbol ς instead.
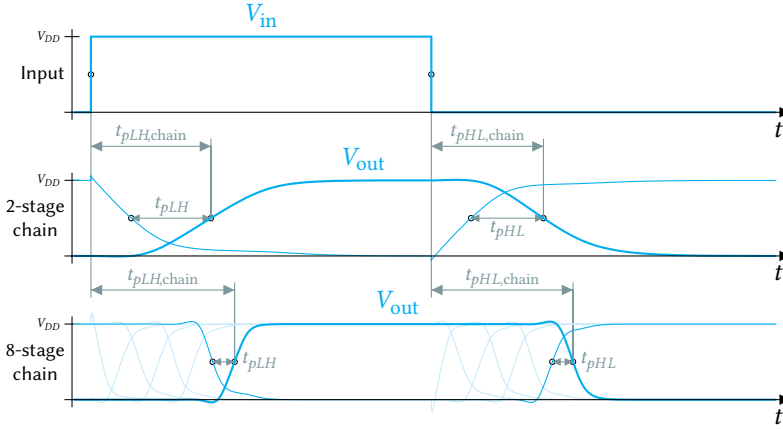
Figure 4.3: Illustration of different possible definitions of 'driver speed'. The waveforms resulting from the 2-stage chain would be the fastest one in traditional digital design definitions, while the 8-stage chain is faster in context of DTXs.

propagation times are less obvious. Namely, C_{GD} translates to an effective contribution to both $C_{dr,i}$ and $C_{dr,o}$. Also, the short-circuit current (I_{sc}) is modeled only indirectly through $C_{dr,o}$. Consequently, these linearized model parameters will vary with the operating conditions at which they are evaluated. The leakage current (or steady-state power) so far has not been accounted for but can be included as an OFF-resistance for the PMOS and NMOS devices if deemed necessary. Lastly, these model parameters and their scaling dependency on the transistor width are based on the intrinsic device. When including layout parasitics, these parameters change significantly since they strongly depend on the driver layout strategy used.

4.1.3 Trade-Off Between Power Consumption and Driver Speed

The driver's load capacitance consists of the input capacitance of the RF-output-stage segments, which is relatively large in the power-DTX concept. A chain of inverters of increasing size, referred to as a "tapered buffer chain," transfers the digital drive signal from the core logic to this large load capacitance. In digital design, it is common to size a tapered buffer chain for minimum total propagation delay or minimized power-delay-product [71]. In a DTX application, however, we are also interested in the achievable rise and fall times and their impact on drain efficiency. These rise and fall times depend mostly on the propagation delay *per stage* rather than having a low delay for the entire chain.

This is illustrated using simulated waveforms of two different chains as an example in Fig. 4.3. The chain using 2 inverter stages has a faster propagation delay for the entire chain, while the chain with 8 inverter stages has a faster propagation delay per stage but suffers from larger power consumption.

The power consumption of a driver chain can be analytically analyzed as a function of the propagation delay per stage using the linearized CMOS driver model. In Fig. 4.2 such driver chain is shown for a single output stage segment. The series interconnect parasitics between the CMOS controller chip and power die can be neglected since the operating frequency is (significantly) lower than the resonance frequency of L_{bond} and C_L,

and $R_{\text{bond}} \ll R_{\text{dr}}$. Next, each consecutive inverter is assumed a factor f smaller (seen from the load towards the input of the chain), and the overall effective fan-out is then defined as $F = C_L/C_{\text{dr},i_N}$. Consequently, the total number of stages becomes $N(f) = \left\lceil \log_f F \right\rceil + 1$. This implies a technology-dependent maximum driver speed as $\lim_{f \downarrow 1} N(f) \to \infty$. The total capacitance that needs to be switched for driving a single segment, including the tapered buffer chain using a fan-out f, is given by (see Section A.5.1 for the full derivation)

$$C_{\text{seg}} = C_L + \frac{C_L}{f}(1+\gamma)\frac{1-f^{-N}}{1-f^{-1}}. \tag{4.9}$$

This yields a driver-related power dissipation in the CMOS controller per segment of $P_{\text{seg}} = f_0 C_{\text{seg}} V_{DD,\text{dr}}^2$. Since P_{seg} is proportional to the total capacitance in the overall segment line-up, rather than only the input capacitance of a (unary) RF output stage segment and its related ESDs (C_L), we can define the capacitance multiplication factor

$$M = \frac{C_{\text{seg}}}{C_L}. \tag{4.10}$$

This factor M can be regarded as the reciprocal of the 'driver efficiency,' which is important in calculating the achievable DTX system efficiency for a given technology. Namely, the full driver-related power dissipation of the CMOS controller is

$$P_{DD,\text{dr}} = f_0 \cdot N_{\text{act}} M C_L \cdot V_{DD,\text{dr}}^2, \tag{4.11}$$

where N_{act} is the number of active segments. The impact of the driver power on the overall DTX system efficiency is discussed in more detail in Chapter 6. The factor M and the effective fan-out f are derived analytically in Section A.5.1, yielding

$$M \approx \frac{f + \gamma}{f - 1} \tag{4.12}$$

$$f = \gamma \frac{\ln(2) R_{\text{dr}} C_L}{t_{p0}}. \tag{4.13}$$

Using these equations, the DTX and its system efficiency can be quickly evaluated for a given technology, with the intended driver's speed in terms of its equivalent resistance R_{dr} as a free design parameter. Alternatively the desired propagation delay can be used directly, see Eq. (A.43). Consequently, the required number of stages in the tapered buffer chain follows as a result and not as a predefined constraint. This allows, for example, to only consider the equivalent resistance of the final driver stage with its output stage segment in a DTX simulation setup and make use of the M value (for a given CMOS controller technology) to include the power consumption of the forgoing tapered buffer chain. This approach is discussed in detail in Section 5.2 (see also Fig. 5.10), when introducing the DTX simulation models. For more accuracy the full device models of the final driver stage can be used. In that case, however, the C_{dr,o_1} of the final driver is already included and M needs to be replaced by M', which can be calculated as

$$M' = M\frac{C_L}{C_{\text{dr},o_1} + C_L} = M\frac{f}{f + \gamma} \approx \frac{f}{f - 1}. \tag{4.14}$$

4.2 Technology Considerations for Digital Drivers

A CMOS foundry can offer more gate oxide thicknesses in the same process node. The thinnest oxide is typically designated as "core oxide" or simply "thin oxide," providing the best device speed for logical circuits at low power consumption. A thick oxide device option has a higher gate dielectric breakdown voltage and can handle larger drain-to-source voltages. This oxide is often used for chip IOs that demand a larger voltage swing, hence referred to as "IO oxide" or "thick oxide." The use of thick oxide typically reduces the control on the channel and larger gate lengths are required to handle the increased drain–source voltage, which results in a lower speed for these devices.

4.2.1 Inverter-Based Drivers

The most straightforward method of implementing a segment driver is directly designing the whole tapered buffer chain using thick-oxide inverters. This chain should be connected to the core logic by a level shifter, shifting from a $V_{DD,\text{core}}$ to a $V_{DD,\text{dr}}$ swing. Such a level shifter should be DC coupled, so that a segment that should be in 'OFF-state' indeed remains 'OFF' no matter how long that may be. The design of such a chain directly follows from the method described in Section 4.1, but the question remains on how to attain the parameters for the linearized CMOS driver model and how to relate them to the (RF) rise and fall times, which we discuss next.

From Eqs. (4.5)–(4.7) we know that

$$\frac{t_{p,\text{int}}}{t_{p0}} = \frac{\ln(2)R_{\text{dr}}\left(C_{\text{dr},o} + C_{\text{dr},i}\right)}{\ln(2)R_{\text{dr}}C_{\text{dr},o}} = \frac{C_{\text{dr},o} + C_{\text{dr},i}}{C_{\text{dr},o}} = 1 + \frac{1}{\gamma}, \tag{4.15}$$

which means that γ can be derived from simulating t_{p0} and $t_{p,\text{int}}$

$$\gamma = \left(\frac{t_{p,\text{int}}}{t_{p0}} - 1\right)^{-1}. \tag{4.16}$$

From the definition of ς (Eq. (4.8)), this value can be determined by simulating with any nonzero input rise and fall time. This can be applied to a chain of equally sized inverters ($f = 1$) with constant propagation delays throughout the chain. Consequently, we define this propagation delay to be t_{p1} such that

$$\varsigma = \frac{t_{p1} - t_{p,\text{int}}}{t_{p1}}. \tag{4.17}$$

By its definition, t_{p1} is the fastest driver chain that can be implemented in a given technology, i.e., it acts like a figure of merit, namely $f_{\text{dr,max}} = 1/t_{p1}$. Similar to the f_T or f_{MAX} definitions for an RF technology, a digital driver's design-t_p should not be chosen close to its technology's t_{p1}.

A simulation example is given in Fig. 4.4. Here, a schematic-level inverter is implemented using the TSMC 40 nm RF device models in core oxide. These RF device models provide accurate simulation results at high frequencies for its provided model layout. As such layout parasitics are already included in the model, providing a more realistic estimate of what can be achieved in a physical implementation. The N:P device sizing targets equal $H \rightarrow L$ and $L \rightarrow H$ propagation delays for the t_{p1} case. For all cases, the average propagation delay
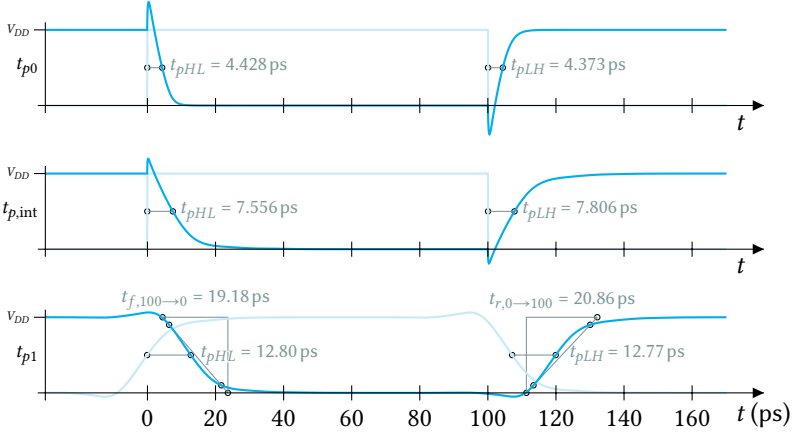
Figure 4.4: Example simulation result of an inverter using TSMC 40 nm devices with RF models in core oxide. These three propagation delays are then used to determine technology parameters t_{p0}, γ, ς and $r_{rf/p1,0-100}$.

is taken, giving here $\gamma = \left(\frac{7.681}{4.400} - 1 \right)^{-1} = 1.341$ and $\varsigma = \frac{(17.79 - 7.681)}{17.79} = 0.399$. This is repeated for an inverter of a standard cell library and a couple of different device types in the TSMC 40 nm technology, of which the results are provided in Table 4.1. We also repeated this for the GlobalFoundries 22FDSOI technology, whose results are reported in Table 4.2.

From Table 4.1 it is clear that layout significantly impacts device performance. Furthermore, the thicker oxide devices operating at higher supply voltages show slower performance than the core oxide devices.

One should know that higher-order effects may not be negligible when aiming for an accurate model. For example, from Eq. (4.4) and Fig. 4.1, the rise and fall time $t_{rf,0-100}$ (which is linearized to 0–100 %) should be linearly proportional to t_p by a factor of $5/4 \log_2 9 \approx 3.96$ when using the $t_{r,10 \rightarrow 90}$ definition of the rise and, similarly, fall times. However, a CMOS inverter is not an ideal RC combination: after saturation (in which R_{dr} is defined) it enters the triode region (R_{ON} region) causing the ratio $t_{rf,0-100}/t_{p1}$ to be smaller. This ratio must be included as an additional empirical parameter $r_{rf/p1,0-100}$, also included in Tables 4.1 and 4.2 (again linearized to 0–100 %). The resulting M-factor vs. t_{rf} is shown in Fig. 4.5 for three devices from Table 4.1 using post-layout model parameters. Although the rise and fall times dominate for the DTX efficiency, careful attention needs to be given to the propagation delays, as their differences can cause timing mismatches, or in the case of differences between the $L \rightarrow H$ and $H \rightarrow L$ propagation (unwanted) duty-cycle changes.

The tapered buffer chain's total propagation delay $t_{p,chain}$ should not be made arbitrarily large in favor of having a small per-stage delay, with fast rise and fall times. Obviously, having long chains is not beneficial for the overall power consumption of the chain, but it also makes it more sensitive to power supply variations. Namely, the variation in chain delay with respect to its nominal delay is inversely dependent on the supply voltage variation to the nominal supply voltage as

$$\frac{\Delta t_p}{t_{p,\text{nom}}} = \left(1 + \frac{\Delta V_{DD,\text{dr}}}{V_{DD,\text{dr,nom}}} \right)^{-\alpha} - 1, \tag{4.18}$$
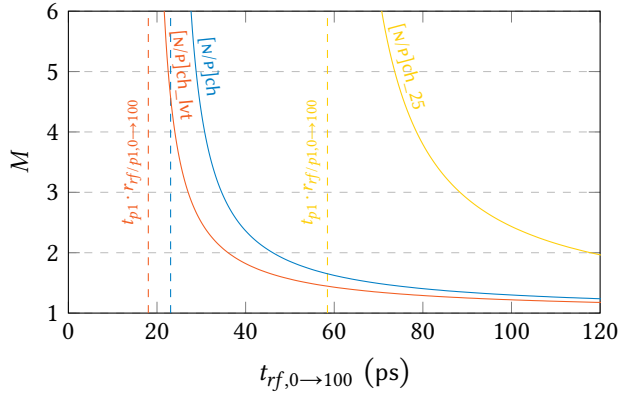
Figure 4.5: Resulting M-factor vs. rise/fall time (linearized to 0 % to 100 % based on 10 % to 90 %) using the model parameters from Table 4.1.

where α depends on the $V_{DD,\mathrm{dr}}$ relative to V_T, channel-length modulation, and the velocity saturation voltage, but typically is around 2 [71]. As a rule-of-thumb for small variations, any 1 % decrease in supply then causes a 2 % increase in delay (see also related measurements in Section 7.5.2). This clearly provides extra motivation to also minimize $t_{p,\mathrm{chain}}$.

4

Table 4.1: Simulated device parameters for TSMC 40 nm Bulk (TT25 corner).

	CKND1BWP ([n/p]ch)		CKND8BWPLVT ([n/p]ch_lvt)		[n/p]mos_rf	[n/p]ch_25		[n/p]mos_rf_25	[n/p]ch_25od33	[n/p]mos_rf_25od33
V_T	R	R	L	L	R	R	R	R	R	R
N:P (µm)	.270:.410	.270:.410	2.16:3.28	2.16:3.28	2.00:4.28	0.80:1.54	0.80:1.54	2.00:4.08	2.00:4.15	2.00:5.40
Parasitics	Intrinsic	Post-Layout	Intrinsic	Post-Layout	RF-Model	Intrinsic	Post-Layout	RF-Model	Intrinsic	RF-Model
V_{DD} (V)	1.1	1.1	1.1	1.1	1.1	2.5	2.5	2.5	3.3	3.3
t_{p0} (ps)	2.330	5.466	1.871	3.805	4.400	6.95	8.68	9.947	8.727	18.20
$t_{p,\mathrm{int}}$ (ps)	3.786	8.653	3.227	6.416	7.682	15.92	21.64	22.99	33.66	54.90
t_{p1} (ps)	6.391	15.16	5.107	10.29	12.79	21.26	29.31	30.11	47.57	76.70
$1/t_{p1}$ (GHz)	156.5	65.96	195.8	97.18	78.19	47.04	34.12	33.21	21.02	13.04
γ	1.601	1.715	1.380	1.458	1.341	0.774	0.669	0.763	0.350	0.496
ς	0.408	0.429	0.368	0.376	0.399	0.251	0.262	0.236	0.292	0.284
$r_{rf/p1,0-100}$	1.561	1.522	1.792	1.753	1.566	2.056	2.008	1.955	1.983	1.782

Table 4.2: Simulated device parameters for GF 22 nm FDSOI (TT25 corner).

	(slvt[n/p]fet)	HDB116SLT20_INV_S_32 (slvt[n/p]fet)		slvt[n/p]fet_rf	[n/p]fet	[n/p]fet_rf
V_T	SL	SL	SL	SL	R	R
N:P (µm)	0.170:0.227	2.16:3.28	2.16:3.28	2.40:3.20	0.170:0.270	2.40:3.56
Parasitics	Pre-Layout	Pre-Layout	Post-Layout	RF-Model	Pre-Layout	RF-Model
V_{DD} (V)	0.9	0.9	0.9	0.9	0.9	0.9
t_{p0} (ps)	1.511	1.440	1.459	1.298	1.703	1.466
$t_{p,\mathrm{int}}$ (ps)	2.451	2.337	2.360	2.203	2.760	2.519
t_{p1} (ps)	3.361	3.165	3.254	3.055	3.856	3.594
$1/t_{p1}$ (GHz)	297.5	316.0	307.3	327.3	259.3	278.2
γ	1.607	1.606	1.619	1.435	1.609	1.393
ς	0.271	0.262	0.275	0.279	0.284	0.299
$r_{rf/p1,0-100}$	1.703	1.704	1.955	1.800	1.612	1.709

Figure 4.6: The circuit of a stacked driver, which is then used as building block in a house-of-cards driver structure.

4.2.2 Stacked Device Drivers

A slightly more involved design can provide larger voltage swings while still using core oxide devices. The use of a cascode device to increase the voltage swing is quite common in analog design. Something similar is possible in digital design; by stacking two N and P devices together, a driver can be made that can reach twice the nominal $V_{DD,\text{core}}$, see Fig. 4.6a. Whether the driver is 'ON' or 'OFF', any V_{DS}, V_{GS}, or V_{GD} will never be larger than $1 \times V_{DD}$, given that the node between the stacked N or P devices settles to V_{DD} when the pull-down or pull-up networks (respectively) are inactive. For this purpose, optional devices can be added to these middle nodes to help them settle to the correct voltage when going into inactive mode. This helps both the reliability of this stacked driver, as well as to improve its speed. This variant is known as a "house-of-cards" driver [72]. It can be seen as an inverter driven by two other inverters at its supply and ground nodes, with a virtual supply around it. This can be repeated with more inverters, as shown in Fig. 4.6b for a stack of three. This concept can (in principle) increase its stacking (beyond three) until the bulk or well junctions reach their breakdown limits. The 'outside' devices are responsible for the main charge and discharge paths; the other devices can be made smaller.

A drawback is that the house-of-cards circuit of Fig. 4.6a requires two inputs, one 'low' input with a swing from V_{SS} to V_{DD}, and a 'high' input with a swing from V_{DD} to $2V_{DD}$. Like the thick-oxide inverter-based drivers, this requires a DC-coupled level shifter, but this time, it does not scale the signal's voltage swing; rather, it shifts the DC voltage of the pulse. This challenge increases further when a taller stack is built, although capacitive coupling techniques exist that may eliminate the middle node(s) [73].

Similar modeling as previously done for the (non-stacked) inverter-based drivers can be repeated here for the stacked drivers. The stacked driver also requires a tapered buffer chain at each of its inputs, necessitating 'splitting' the technology modeling parameters for the stacked driver and the chain, namely $t_{p0,s}$, γ_s, and ς_s for the stack, similarly $t_{p0,c}$, γ_c, and ς_c for the chain. However, as they now have two (or more) inputs, it is not possible to determine $t_{p,\text{int}}$ or t_{p1} to derive the γ_s and ς_s parameters from. Instead, the capacitance
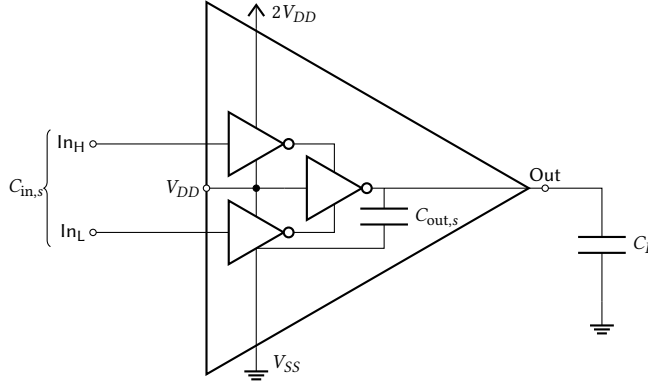
Figure 4.7: Model of the stacked driver for analytically determining the power–speed-trade-off.

found at all inputs can be summed together to present the stack's $C_{\text{in},s}$, as is illustrated in Fig. 4.7 for a stack of two. Capacitances $C_{\text{in},s}$ and $C_{\text{out},s}$ (and thus γ_s) can be found by integrating the current flow into (and out of) these nodes. The total $C_{\text{in},s}$ can then be modeled as the load of a single tapered buffer chain, even though it requires two in reality. Based on the simplifying assumption after Eq. (A.49), the tapered buffer chain is assumed to have an infinite number of stages. In that case, mathematically speaking, it does not matter whether two infinite chains each drive half of the capacitive load or one chain drives it in full. Parameters $t_{p0,s}$ and ς_s can be estimated by varying the rise and fall times at the inputs of the stack. However, when aiming to accurately estimate the stack's $C_{\text{out},s}$ based on the wanted $t_{p,s}$, it turns out that only using $t_{p0,s}$ yields significant modeling error. For example, one would expect that when $C_L = C_{\text{out},s}$ that this results in $t_p = 2t_{p0}$. This is, unfortunately, not true due to the nonlinear operation of the stacked devices. However, a new parameter $t_{p\text{eq},s}$ (typically smaller than $2t_{p0}$) can be introduced as the simulated propagation delay for when $C_L = C_{\text{out},s}$, which gives much better predictions. The values resulting from simulation for two technologies are provided in Table 4.3. Then, as a function of a wanted $t_{p,s}$

$$C_{\text{out},s}\left(t_{p,s}\right) = \frac{t_{p\text{eq},s} - t_{p0,s}}{t_{p,s} - t_{p0,s}} C_L. \tag{4.19}$$

In an inverter-based driver, the input capacitance $C_{\text{dr},i}$ has to be smaller than C_L,

Table 4.3: Simulated device parameters for house-of-cards drivers.

	40nm 2HoC		22nm 2HoC	
V_T	L	L	SL	SL
N:P (µm)	14.4:102.4	14.4:102.4	10.00:11.33	12.00:12.96
Parasitics	Intrinsic	Post-Layout	Intrinsic	RF-Model
$(n\times)\ V_{DD}$ (V)	2×1.1	2×1.1	2×0.9	2×0.9
$t_{p0,s}$ (ps)	8.560	16.10	4.154	4.051
$t_{p\text{eq},s}$ (ps)	18.01	31.58	7.025	6.725
γ_s	1.229	1.006	0.738	0.691
ς_s	0.191	0.150	0.117	0.109
$r_{rf/p\text{eq},0-100}$	1.684	1.801	1.376	1.337

otherwise the fan-out factor f would be smaller than 1, yielding an expanding chain instead of tapering down. The stacked driver does not have this constraint as long as its $C_{\text{in},s}$ can be driven by a preceding tapered chain. The propagation delay of the stack under the influence of its preceding chain is then given by

$$t_{p,\text{tot}} = \varsigma_s t_{p,c} + t_{p,s}. \tag{4.20}$$

The $t_{p,\text{tot}}$ can then be related again to the rise and fall time t_{rf} by an empirical ratio. There are multiple combinations of $t_{p,c}$ and $t_{p,s}$ possible for any given $t_{p,\text{tot}}$ (or t_{rf}), but only one combination yields minimum power consumption. In Section A.5.2 the optimum value for $t_{p,c}$ is analytically derived for a given wanted $t_{p,\text{tot}}$, as well as the total power consumption. These resulting values can be used directly to estimate DTX performance quickly also when using a stacked driver.

4.3 Conclusion

In this chapter, the requirements for a CMOS driver chain are given. Namely, it should provide sufficiently short rise/fall times to not significantly degrade the drain efficiency (e.g., not more than 10 %). However, making the driver chain faster also increases the power dissipated in the driver chain. To quantify this increase for a given CMOS technology, the capacitance multiplication factor M is introduced. The theoretical minimum power required to switch a load capacitance C_L is then multiplied by M to find the overall power of the chain.

Also, a driver model based on the linearized CMOS parameters is provided to calculate the M-factor based on the propagation delays in a given technology. An extra empirical factor $r_{rf/p1,0-100}$ is introduced to relate the propagation delays to the rise and fall times, which are relevant to the DTX drain efficiency. The driver chain should also be considered for its propagation delay. The propagation delays per stage are assumed constant throughout the chain and equal for propagating from high to low and vice versa to avoid timing glitches or (unwanted) duty-cycle changes.

If we assume that the drain efficiency should not be degraded by more than 10 %, we can find that $t_{rf}/T < 25$ %. Using the technology parameters from Table 4.1, we can determine the uppermost frequency a driver chain can operate at (the frequency where the fanout factor drops to unity, $f = 1$), which is 13.9 GHz for a core-oxide device driver chain in 40 nm CMOS operating at $V_{DD,\text{core}} = 1.1$ V, or 4.3 GHz for a thick-oxide device driver chain at $V_{DD,\text{dr}} = 2.5$ V. However, $f = 1$ results in a chain of infinite length, i.e., $M \to \infty$. Therefore, when we additionally define that M should be less than 4, we find the more realistic maximum frequencies of 8.1 GHz and 3.2 GHz, respectively. Note that a value $M = 4$ may still be high from the perspective of the overall DTX system efficiency, since it then tends to be dominated by its driver energy consumption. Therefore, more system-level context is required to determine the driver chain impact on the overall DTX efficiency, which we provide in Chapter 6 by introducing an overall DTX power model.

Furthermore, we can also observe that for more advanced technologies, such as 22 nm FDSOI at 0.9 V, using the same assumptions, the maximum driver chain frequency increases to 29 GHz, underlining the importance of having high-speed low-voltage drivers and the subsequent (custom) low-V_T segmented RF power devices to enable energy-efficient power-DTX applications (also see the conclusions of Chapter 3).

5

DTX Modeling

Previous chapters have discussed the technology aspects of implementing a high-power DTX using a combination of high-speed digital CMOS and high-power RF technology. To the best of the author's knowledge, no clear definition of the transfer of a DTX is available in the literature. This feeling is amplified by questions such as: "What is the gain of your DTX?" Or: "What is the PAE of your system?" This chapter aims to describe the input(s) and output(s) of a DTX system to define the DTX transfer clearly.

Further, mixed-signal systems, such as DTXs, are infamous for their (very) long simulation times to evaluate their system performance due to their high circuit complexity, high transient frequencies, and many actively switching components. Through a better understanding of the transfer of a DTX and its inputs and outputs, it is possible to define an enormously simplified DTX simulation model that allows better use of the frequency-domain techniques of circuit simulators, significantly speeding up the simulations and related decision-making.

In Section 5.1, the operation of a DTX is discussed. First, the numerical digital input of a DTX system is clarified. Using this insight, the transfer of a DTX can be unambiguously defined. This leads to the definition of the normalized digital transfer, complemented with an example of its use(fulness). Section 5.2 introduces a simplified simulation model for (gate-segmented) DTXs. First, a conventional simulation set-up using discrete components is discussed. Next, a new simulation model for DTXs using a current-scaling technique is proposed. This simplified model allows more extensive simulation studies, expanding the designer's capabilities to design a DTX efficiently.

5.1 DTX Black-Box Operation

A DTX has one or more numerical input(s) that control the RF output. This is in contrast to an analog PA, where the input and output quantities are typically considered in terms of power or as waves in terms of voltages or currents in a defined impedance. A transfer function (h) can be used to define the relationship between a linear system's input (x) and output (y)

$$y(t) = (x * h)(t) \triangleq \int_{-\infty}^{\infty} x(t - \tau) \cdot h(\tau) \ \mathrm{d}\tau. \tag{5.1}$$
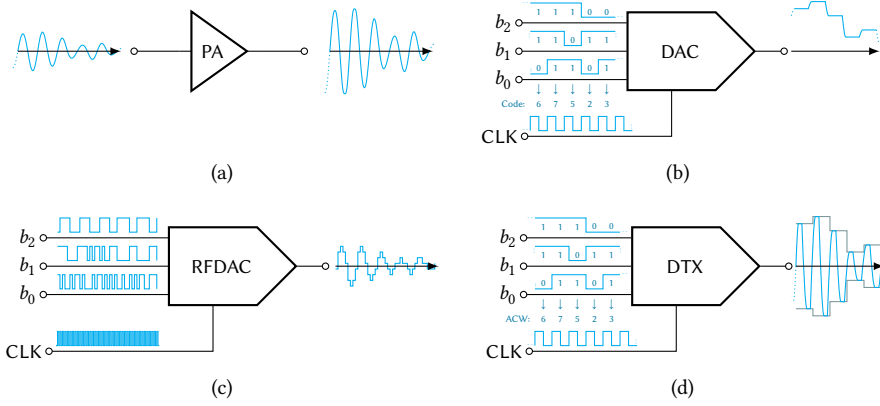
Figure 5.1: Conceptual comparison of different systems' inputs and outputs: (a) power amplification only; (b) digital-to-analog conversion only; (c) high-frequency digital-to-analog conversion; (d) both digital-to-analog conversion as well as modulating operation, with a large output magnitude.

5

Alternatively, it can be expressed in the (Fourier) frequency domain by the ratio of the output quantity divided by the input quantity

$$H(\omega) = \frac{Y(\omega)}{X(\omega)}. \tag{5.2}$$

These definitions can be used for analog PAs and digital TXs, but some differences exist. First, this difference is conceptually highlighted using a DTX's basic binary input representations, which also implicitly explains why a digital *TX* is not a digital PA. This insight becomes evident when considering the mathematical description of the DTX concept. Namely, it is possible to define the transfer of a DTX and use it similarly as we describe analog PA metrics in terms of AM–AM, AM–PM, or DAC metrics such as integral nonlinearity (INL) and differential nonlinearity (DNL).

5.1.1 Bits-in RF-out

In contrast to an analog PA, a DTX is a mixed-signal system (compare Fig. 5.1a and Fig. 5.1d). Therefore, it is less practical to express its transfer in terms of *S*-parameters and their related power gain definitions, especially since a DTX's input is a dimensionless numerical value at baseband, as in (RF)DACs. As such, traditional DAC linearity metrics such as INL and DNL could be considered. However, these (non)linearity measures are static, while a DTX has a modulated RF output centered around a carrier (ω_0) and optionally other harmonic or spectral replicas. In that sense, a DTX provides both the modulation and amplification functionality of a TX chain (see also Section 2.1). The different concepts are shown in Fig. 5.1 in a very simplified manner: in the PA of Fig. 5.1a both the input and output are at RF, while the inputs of the DAC and RFDAC in Figs. 5.1b and 5.1c are in digital baseband. The RFDAC differs from the DAC by a much higher sampling rate, as such able to construct an RF waveform without using a mixer, but with limited output magnitude. The DTX in Fig. 5.1d has an input in digital baseband, provides mixing with the clock, and outputs an RF waveform with a large magnitude.

(a) $101_2 = 5$ (b) $0011111_1 = 5$

(c) $0011111_1 101_2 = 5_8 5_8 = 45$

(d) $101_2 0011111_1 = 5_8 5_8 = 45$

(e) $0111_1 00111_1 = 3_5 3_6 = 21$
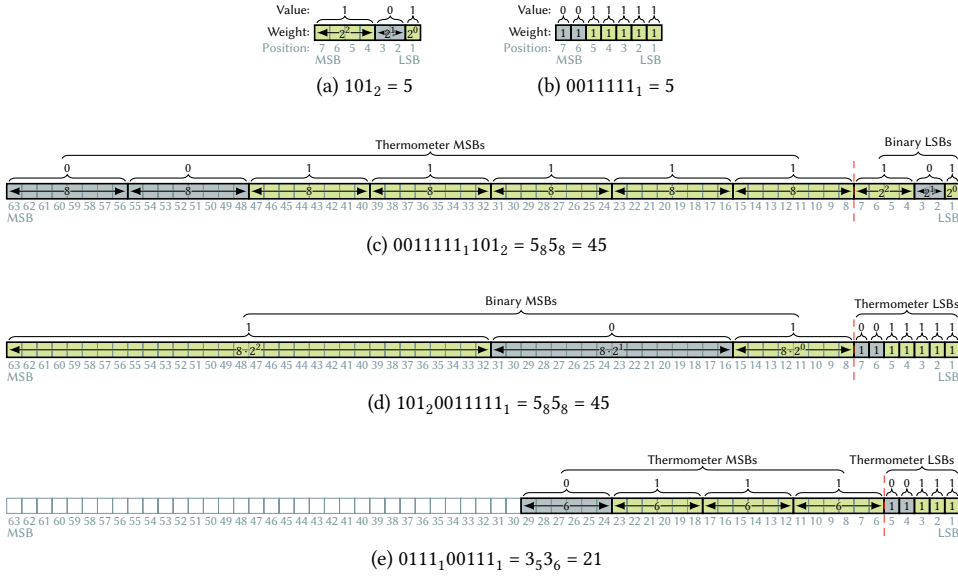
Figure 5.2: Digital number representations.

Bits-in Notation

To describe the system-level behavior/transfer of a DTX, it is often more convenient to use the numerical value that the ACW represents than the ACW's bits themselves. Several digital notations can be used in a (RF)DAC/DTX, like binary-coded, thermometer-coded, or a combination of both. These notations often closely relate to their practical implementation. Namely, a DTX consists of many parallel unit cells, which can have an assigned weighting similar to these binary coding notations.

In a straightforward unsigned binary-coded notation, each signal is binary weighted as

$$\text{ACW} = \{b_{N-1}, \ldots, b_1, b_0\}_2 = \sum_{n=0}^{N-1} 2^n b_n. \tag{5.3}$$

This means that N bits represent 2^N possible integer values in the range $\left[0, 2^N - 1\right]$. In the 3-bit digital systems shown in Figs. 5.1b and 5.1d, an example code is $\text{ACW} = \{b_2, b_1, b_0\}_2 = \{1, 0, 1\}_2 = 2^2 \cdot 1 + 2^1 \cdot 0 + 2^0 \cdot 1 = 5$. This value is graphically shown in Fig. 5.2a. Here, b_2 represents the largest magnitude, making it the most significant bit (MSB), whereas b_0 represents the smallest magnitude, making it the least significant bit (LSB). Alternatively, it can be notated using radix notation, where the base of the number is indicated by a subscript, namely $\text{ACW} = 101_2 = 5_{10}$ or 5_{dec}.

In a thermometer coding, each bit is unary weighted, as

$$\text{ACW} = \{b_{N-1}, \ldots, b_1, b_0\}_1 = \sum_{n=0}^{N-1} b_n. \tag{5.4}$$
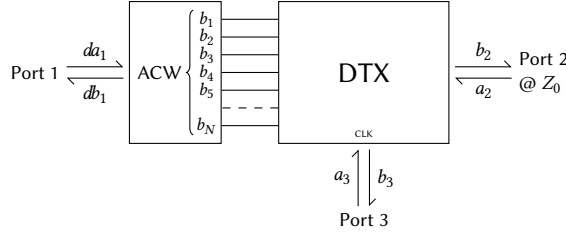
Figure 5.3: The (3-port) black-box representation of a DTX. It is a mixed-signal component, where port 1 is the numerical (baseband) input, port 2 is the RF output port following classical RF definitions as used in S-parameters, and port 3 serves as the modulation carrier, with a frequency (f_c) and phase (θ_c) reference.

In this case, N signals represent $N + 1$ possible integer values in the range $[0, N]$. Using 5 again as an example in a 7-bit thermometer-coded system, its representation would be $\{b_6, b_5, b_4, b_3, b_2, b_1, b_0\}_1 = \{0, 0, 1, 1, 1, 1, 1\}_1$, or 0011111_1, as graphically shown in Fig. 5.2b.

Combinations of binary and unary weighting are also possible, or multiple 'levels' of unary weighting can be used. For example, the MSBs can be thermometer coded, and the LSBs binary coded. For this latter case, when there are N thermometer-coded MSBs and M binary-coded LSBs, we can write

$$\text{ACW} = \{b_{M+N-1}, \ldots, b_{M+1}, b_M\}_1 \{b_{M-1}, \ldots, b_1, b_0\}_2 = \sum_{n=0}^{N-1} 2^M b_{n+M} + \sum_{n=0}^{M-1} 2^n b_n. \qquad (5.5)$$

This means that $N + M$ signals represent $(N + 1)2^M$ possible integer values. As an example, with $N = 7$ and $M = 3$, the ACW $= 0011111_1 101_2$ (mixed-radix notation) represents $5_8 5_8 = 5 \cdot 8 + 5 = 45$ (Fig. 5.2c). In a multi-level thermometer coding, one level has a higher significance than the next level. For example, a two-level thermometer code with N bits in the first layer and M in the second layer

$$\text{ACW} = \sum_{n=0}^{N-1} M b_{n+M} + \sum_{n=0}^{M-1} b_n. \qquad (5.6)$$

Here $N + M$ signals represent $(N + 1)(M + 1)$ possible integer values, for example, with $N = 4$ and $M = 5$, the ACW $= 0111_1 00111_1$ represents $3_5 3_6 = 3 \cdot 6 + 3 = 21$.

Regardless of notation, a resolution (of a DTX) is always reported as a binary-weighted number of bits N_b, i.e., a 10-bit system has 2^{10} unique numerical input values, so from 0 to 1023.

5.1.2 Introducing D-Parameters for DTX

With the numerical value of the digital input defined, also the virtual digital input (power) wave can be defined, which allows the specification of a mixed-signal transfer in terms of D-parameters [74]. The D-parameters can be considered as the mixed-signal equivalent of the analog-oriented S-parameters. In [74], D-parameters were introduced to describe a digital RX. Here, we rework them to allow defining a DTX transfer.

The black-box representation of a DTX is given in Fig. 5.3. First, the digital input da_1 at port 1 should be normalized to provide a quantity that is dimensionally compatible with

the output. This yields the normalized ACW

$$da_1 = \frac{\text{ACW}}{2^{N_b} - 1} \sqrt{2 P_{\text{norm}}}, \tag{5.7}$$

where the first term is the numerical value of the ACW scaled to be in the interval $[0, 1]$ by dividing it with its maximum range $2^{N_b} - 1$, where N_b is the resolution in the number of bits of the ACW. The next term is a normalization power P_{norm} to ensure da_1 has the dimension $\sqrt{W} = \frac{V}{\sqrt{\Omega}} = A \sqrt{\Omega}$. Useful values for P_{norm} can be 1 mW or 1 W, so the resulting transfer can be expressed in dBm or dBW. The incident wave a_2 and outgoing (or reflected) wave b_2 at port 2—the DTX output—follow the classical RF definitions as used in S-parameters. The digital 'reflection' component db_1 can be ignored, since we're describing a transmitter. We start with the linear definition of D-parameters (not yet considering upconversion)

$$db_1\big|_{a_3 = A e^{j\theta_c}} = D_{11} da_1\big|_{a_3 = A e^{j\theta_c}} + D_{12} a_2 \tag{5.8}$$

$$b_2 = D_{21} da_1\big|_{a_3 = A e^{j\theta_c}} + D_{22} a_2 \tag{5.9}$$

which is quite similar to how the transfers of an analog PA in S-parameters would be described, but now also explicitly includes the clock as a reference signal with its amplitude and phase for which the resulting D-parameters are valid. D_{22} relates to two analog waves and is, thus, identical to S_{22}.

Next, the frequency conversion should be explicitly included in the modulating action based on the clock at port 3 (Fig. 5.3). Namely, the input at port 3 (a_3) serves as the modulation carrier, with a frequency (f_c) and phase (θ_c) reference. A square wave clock is typically used when considering a DTX, in which the fundamental frequency is assigned to harmonic index 1. The output b_2 is controlled for its envelope by the ACW at port 1 and for its upconversion by the (reference) signal at port 3. Consequently, the output signal at the fundamental frequency can be generally described using the describing function $F_{2[1]}$

$$b_{2[1]} = F_{2[1]}\big(da_{1[0]}, da_{1[1]}, \ldots, da_{1[k]}, a_{2[1]}, a_{2[2]}, \ldots, a_{2[k]}, a_{3[1]}, a_{3[2]}, \ldots, a_{3[k]}\big). \tag{5.10}$$

Here, the subscripts $[k]$ denote the kth harmonic of the clock's f_c. Some simplifying assumptions can be made to specify this $F_{2[1]}$ for the ideal case. First, the DTX is assumed to be perfectly matched at the output, hence $a_{2[k]} = 0$ for all k. Next, da_1 is considered as a bias (DC or baseband) value, hence $da_{1[k]} = 0$ for all $k \neq 0$. Finally, the clock at port 3 is only considered to be a phase reference, i.e., its amplitude is supposed to have no influence, while $F_{2[1]}$ is time invariant. Defining the phase/delay of the reference signal at port 3 at its fundamental frequency as

$$P = e^{j \cdot \text{Arg}(a_{3[1]})} = \frac{a_{3[1]}}{\big|a_{3[1]}\big|} \tag{5.11}$$

and shifting Eq. (5.10) by P^{-1} in time, and removing all zero-valued variables gives

$$b_{2[1]} P^{-1} = F_{2[1]}\big(da_{1[0]}, a_{3[1]} P^{-1}, a_{3[2]} P^{-2}, \ldots, a_{3[k]} P^{-k}\big). \tag{5.12}$$

Since none of the clock's harmonics are supposed to contribute to the fundamental output, further simplification yields

$$b_{2[1]} = F_{2[1]}\big(da_{1[0]}, 1\big) P. \tag{5.13}$$
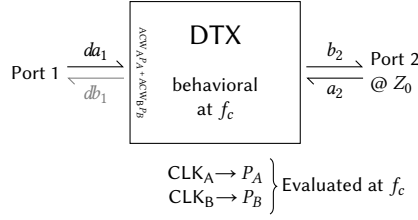
Figure 5.4: Simplified 2-port representation of a DTX at the fundamental frequency (f_c). The phase reference(s) at f_c are now included by making the baseband input da_1 complex-valued.

Consequently, the only remaining variable is $da_{1[0]}$, and in the ideal case $b_{2[1]}$ is linearly proportional with it, such that

$$b_{2[1]} = D_{2[1];1[0]} da_{1[0]} P \Big|_{a_2=0,\ a_3=P} \tag{5.14}$$

where $D_{2[1];1[0]}$ is a complex-valued constant that describes the conversion from the digital baseband signal applied at port 1 to the fundamental RF signal emanated from port 2, with the phase reference P determined by port 3. In the nonideal case, however, the value of $D_{2[1];1[0]}$ will be a function of $da_{1[0]}$ and the clock will have harmonics. For readability purposes, the further derivations that consider the above dependencies are provided in Section A.3 by expanding the number of ports and drawing inspiration from the large signal S-parameters (also called X-parameters).

5.1.3 Normalized Digital Forward Transfer

Now that a mathematical description is established, it is time to translate it to metrics useful for design. For example, the gain in analog PA design depends on the input/output power. This can be represented by mathematical formalisms as $\left|X^{(S)}_{2[1],1[1]}\left(\left|a_{1[1]}\right|\right)\right|^2$. But not many people are doing that in practice. In fact, most people connect a device under test (DUT) to a vector network analyzer (VNA), perform a frequency or power sweep, plot the measured S-parameters over frequency, and use the output power versus the input power to define the compression point or plot the gain over input or output power as an AM–AM curve, often accompanied by an AM–PM curve. However, using the previous formalism, these AM–AM/PM curves *are* the magnitude and phase of the $X^{(S)}_{2[1],1[1]}$ parameter. Consequently, something similar can be done for DTX, resulting in some sort of "gain" definition of a DTX.

Definition of the Normalized Digital Forward Transfer

The main simplification is to use the digital input as the numerical value of the complex envelope they try to represent. Therefore, the wanted transfer from the baseband to the RF carrier is of interest. As such, it is no longer necessary to consider the examples of Cartesian or multi-phase DTXs as 5 port systems. In fact, in this simplified approach the digital input can be represented as a complex number, or as its equivalent two-dimensional vector, corresponding to the wanted RF phasor at the DTX's output port. This is illustrated in Fig. 5.4. The complex number's argument of the digital input is set by the phase of its
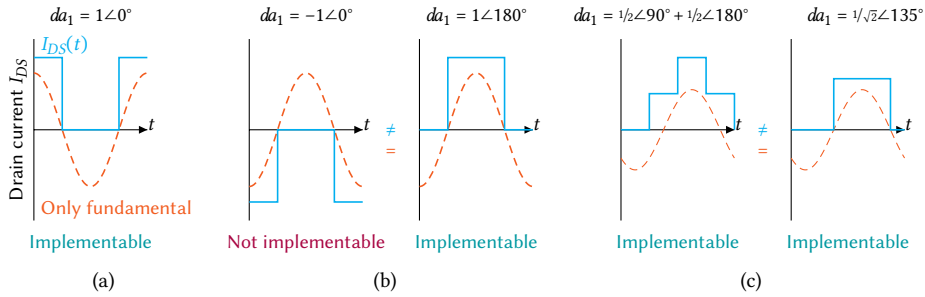
Figure 5.5: Illustrating the sign/phase inequality when considering all harmonics of a square wave drain current. The fundamental is shown in dashed orange, where the full waveform is represented in cyan. (a) shows the 0° reference, where (b) shows the "negative" and 180° shifted cases. In (c), the difference between summing two vectors (e.g., signed-Cartesian operation) and phase shifting (e.g., polar operation) are highlighted, even though the resulting fundamental contents are again the same.

reference clock, such that a phase reference can be defined for each clock

$$P_A = e^{j\theta_A} \text{ and } P_B = e^{j\theta_B}. \tag{5.15}$$

The digital input wave then becomes

$$da_{1[0]} = \frac{\text{ACW}_A \cdot P_A + \text{ACW}_B \cdot P_B}{2^{N_b} - 1} \sqrt{2P_{\text{norm}}}. \tag{5.16}$$

This validates the vector diagram for DTX, as was used in Fig. 2.2, and the use of complex numbers explaining the transmitter architectures for DTX purposes in general, as was done in Section 2.1.

It is vital to remember that, in this representation, there is a difference in meaning between $-1\angle 0°$ and $1\angle 180°$, since they are only *numerically* identical when considered at the fundamental after mixing (Fig. 5.5b). In fact, the ACW cannot be negative in a practical high-efficiency, high-power switch bank. Note that this is conceptually not different than in the case of energy-efficient analog operation, such as class-B (e.g., Fig. 2.19). The only way to represent a 'negative' envelope value is by shifting the clock waveform's phase by 180°. The same holds for complex envelope values, such as 90° which also needs to be implemented by clock phase shifts in the DTX switch banks, as is illustrated in Fig. 5.5c.

Nonetheless, the complex-valued simplification for the digital baseband input ($da_{1[0]}$ from Eq. (5.16)) allows us to collapse a DTX into a behavioral two-port which is more convenient in practical measurements. In such a case, the outgoing wave can then be plainly described by

$$b_{2[1]} = D^{(S)}_{2[1]1[0]}\left(da_{1[0]}\right) da_{1[0]}\Big|_{a_2=0}. \tag{5.17}$$

The transfer from input to output is then simply given by

$$D_{21} = \frac{b_2}{da_1}, \tag{5.18}$$

where, for further simplicity, all harmonic subscripts are dropped. Unless otherwise specified, D_{21} relates the digital baseband (representing a complex envelope) to the RF

output: D_{21} forms the normalized digital forward transfer, similar to how S_{21} is the forward transmission for an analog PA.

Note that there is no numerical output for a DTX; thus, there is no digital 'reflection' component db_1, hence

$$db_1 = 0. \tag{5.19}$$

This makes a DTX a true unilateral transmitter as $D_{12} = 0$, which agrees well with its physical nature. Also $D_{11} = 0$, making a DTX unconditionally stable as long as $|D_{22}| < 1$, which should be the case always. This is to say that the numerical input value cannot be changed by any other means than to change the underlying amplitude code words themselves. In theory, the signal integrity of the electrical lines can be disturbed. However, proper digital design with sufficient timing margins and voltage headroom should be able to avoid that from happening.

Use of the Normalized Digital Forward Transfer
Using the normalized digital forward transfer D_{21} in a DTX is functionally identical to using S_{21} or the ratio power gain G_P for an analog PA. To illustrate this, a simulation result of a simplified DTX (based on the DTX design discussed in more detail in Chapter 8) with a realistic device model is presented in Fig. 5.6. In this simulation, the 12-bit DTX uses 9 bits in a first-layer thermometer coding and 3 bits in a second layer, making the maximum possible ACW = $111111111_2 111_2 = 511_{10} 7_8 = 4095$. First, in Fig. 5.6a, the outgoing wave b_2 is plotted for all possible ACWs, yielding a large set of complex numbers with both a magnitude and a phase.

The related powers, $P_{DD,\mathrm{RF}}$ (DC power), P_{RFout} (RF output power), and P_{dr} (switched capacitor loss of the RF output stage) are plotted in Fig. 5.6b. The RF output power (P_{RFout}) shows an overall quadratic relationship up to the compression point, and the RF output stage's DC power ($P_{DD,\mathrm{RF}}$) has a linear relationship with the applied ACW (see also Section 2.2.3). The maximum $P_{\mathrm{RFout}} = 19.0\,\mathrm{W}$. The power required to switch the input of the RF output stage (P_{dr}, as defined in Eq. (4.1)) has a linear relationship over the full range of applied ACWs (as expected from Section 4.1.3 and Eq. (4.11)).

The digital input wave is, in this case, given by $da_1 = \frac{\mathrm{ACW}}{4095} \sqrt{2P_{\mathrm{norm}}}$. Using ACW = 4095 as an example, the outgoing wave $b_2 = 6.17\angle 59.7° \sqrt{\mathrm{W}}$. That makes $10\log|D_{21}|^2 = 10\log\left(\frac{6.17}{\sqrt{2\cdot 10^{-3}}}\right)^2 = 42.8\,\mathrm{dBm}$, by choosing $P_{\mathrm{norm}} = 10^{-3}$ the resulting transfer is given in dBm. This indeed corresponds to the 19 W peak RF output power found before. The magnitude in dBm of the normalized digital forward transfer D_{21} is shown in Fig. 5.6c for all possible ACWs. This graph is then an 'AM–AM' curve for a DTX; more accurately, it is an ACW–AM/ACW curve. This curve visualizes that compression behavior sets in at around ACW = 2048. For low values of ACW, the transfer $D_{21} \approx 44.7\,\mathrm{dBm}$, indicating that the DTX is in compression by $44.7 - 42.8 = 1.9\,\mathrm{dB}$ at maximum ACW drive. Finally, the phase of the normalized digital forward transfer is shown in Fig. 5.6d.

The normalized digital forward transfer is functionally identical to the analog AM–AM (or gain) and AM–PM transfers. Just as with the analog transfers, D_{21} can be plotted in a variety of ways. For example, the x-axis for the digital input wave da_1 can be adjusted, as is done in Fig. 5.7a. By setting the maximum value of da_1 as 0 dB, the ACW–AM/ACW curve is now represented versus input back-off relative in dB to the input's full scale (dBFS). Using this scaling, the slight mismatch of the LSBs is better visualized. Especially when

(a) Outgoing DTX power wave vs. ACW

(b) DTX powers vs. ACW

(c) ACW–AM/ACW Curve

(d) ACW–PM Curve
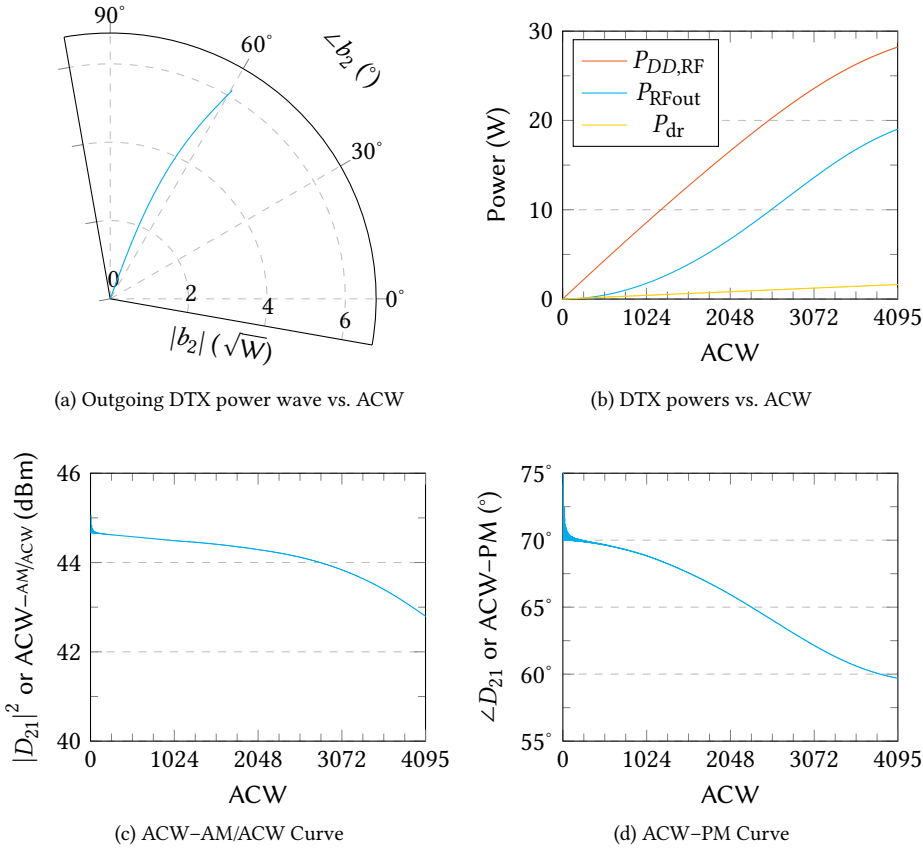
Figure 5.6: Visualizations of the normalized digital forward transfer and its inputs, using a simplified DTX simulation model with realistic device models, based on the high-resolution DTX introduced in Chapter 8.

compared to the plotting vs. ACW as done previously in Fig. 5.6, where the tiny magnitude entirely hides it. By choosing the normalization power $P_{\mathrm{norm}} = 29.2\,\mathrm{W}$ the peak D_{21} value now corresponds to 0 dB, so the DTX's compression can also be visualized well (Fig. 5.7b).

It should be stressed that the normalized digital forward transfer is not a measure of gain, as a DTX's input is numerical. Namely, for analog PAs, the input is power, such that the power gain G_P is actually the transfer of an analog system. For a DTX, dividing the RF output power by the switching losses (P_{dr}) of the input capacitance of the switch bank gate segments (which include, in this particular example, the ESD diodes and interconnect parasitics), is still possible. The related results are shown in Fig. 5.8 using a linear and logarithmic scales. Since the LSBs consume slightly more power than the MSBs (due to the capacitive overhead), these 'gain' curves zigzag. This highlights that such a 'gain' curve does not provide any linearity information about the DTX. Another gain definition used in analog PAs is the slope gain G_{sP} (Eq. (A.9)). To find an equivalent, we must consider the discrete nature of a DTX's input. It is possible to define something similar to the DNL or INL used in DACs. However, in such a definition, it should be considered that b_2 is a
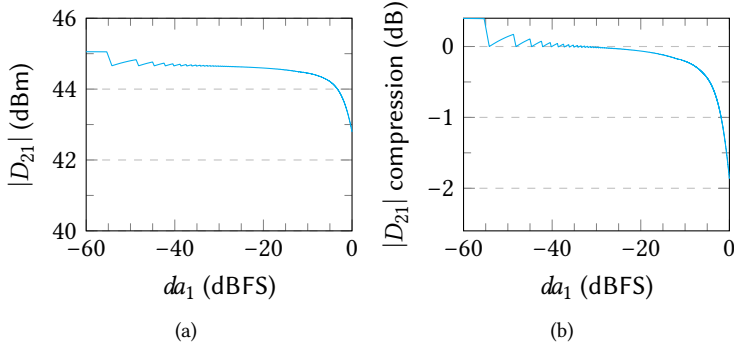
Figure 5.7: Potentially useful alternative magnitude visualizations of the normalized digital forward transfer of Fig. 5.6c: small mismatches are better visualized by plotting the input relative to full scale (dBFS). By choosing a different normalization power, in (b) the DTX's compression is well visualized.

5
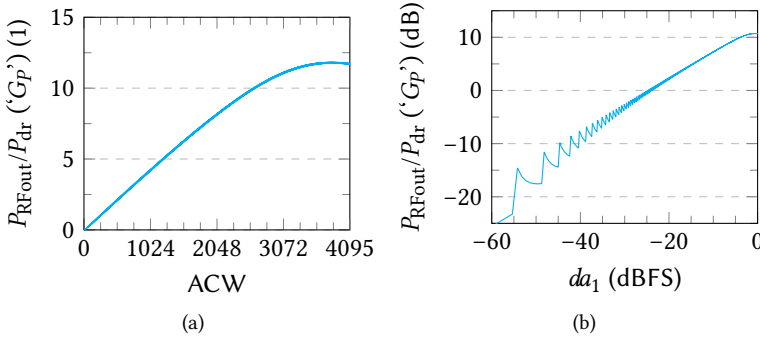
Figure 5.8: Example of RF output power over input switching loss in a DTX, which is the closest equivalent to analog power gain. It is not the transfer of a DTX, nor does it provide any linearity information.

complex number. For example, the difference in magnitude is different than the magnitude of the difference. Depending on the definition used, the 'DNL equivalent' may not sum to zero, and statements about monotonicity from its values may not be possible. As such, the 'DNL' and 'INL' equivalents for a DTX should not actually be called the DNL or INL to avoid possible confusion.

5.2 Simulating a Gate-Segmented DTX

Simulating a DTX with many segments or unit cell elements requires special attention to do so efficiently. For example, simulating with many discrete unit cell elements might yield the most accurate result but is also severely limited by the circuit simulator's capabilities, resulting in large simulation times. This approach is explained first.

Most circuit simulation tools for RF use a frequency-domain approach, which is challenging in the DTX case due to the many independent segments and, thus, high circuit complexity. A well-chosen DTX model can strongly reduce the circuit complexity for
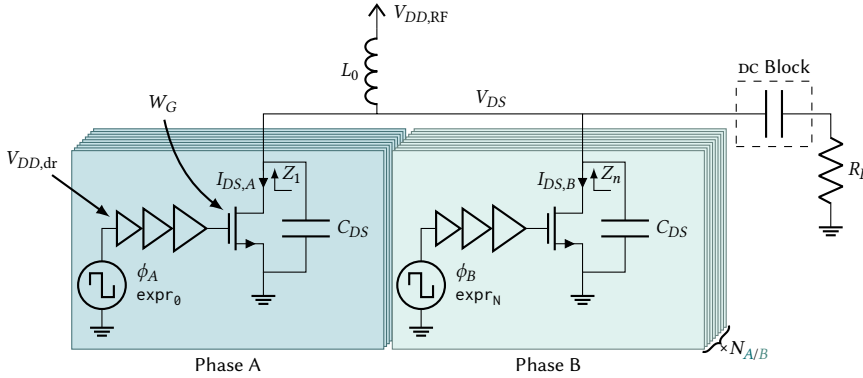
Figure 5.9: Schematic of a DTX simulation model using discrete segmented devices.

simulation purposes, which is discussed thereafter. It is important to reiterate that a DTX is inherently a large-signal mixing structure (its input is at baseband, while its output is at RF) with many frequency components present. The simplified model can use a system-level description of a DTX based on the harmonic superposition principle discussed in the previous section and Section A.3.

5.2.1 Discrete Simulation Model

One possibility is to use a simulation set-up that includes all individual unit cells with the segmented power devices. This is conceptually shown in Fig. 5.9. It closely relates to how a DTX will be implemented but has a drawback: it requires many nodes and instances. Namely, each unit cell requires its individual segment and a (voltage) source driving that segment. For example, implementing a simulation set-up of an 8-bit DTX using thermometer coding requires 256 individual unit cells and each of the included sources requires a unique expression, making it time consuming to implement. This becomes even more troublesome when more than one activation phase is present.

In general, it is not practical to simulate a modulated signal with so many individual unit cells in the frequency domain. Namely, each individual source expression will have a poorly defined harmonic 'steady state,' especially with complex modulated signals. The same holds when a digital controller or encoder is included. Hence, a transient simulation is the most feasible simulation type for such a DTX model when aiming to study its handling of modulated signals. Alternatively, a static input signal can be simulated using harmonic balance, e.g., to determine the DTX's steady-state transfer using a point-by-point ACW sweep.

Additionally, not all transient simulators can handle transmission lines or components defined by (EM simulated) S-parameter models equally well. Distributed inductive components are a known cause of introducing timestep and trapezoidal integration issues.

5.2.2 Current Scaling Simulation Model

The simplified DTX simulation model can use a single instance of the power device model in the simulation set-up. This is achieved by scaling the current into the drain (mainly I_{DS}) by the (normalized) ACW, effectively modulating the activated width of the device. This
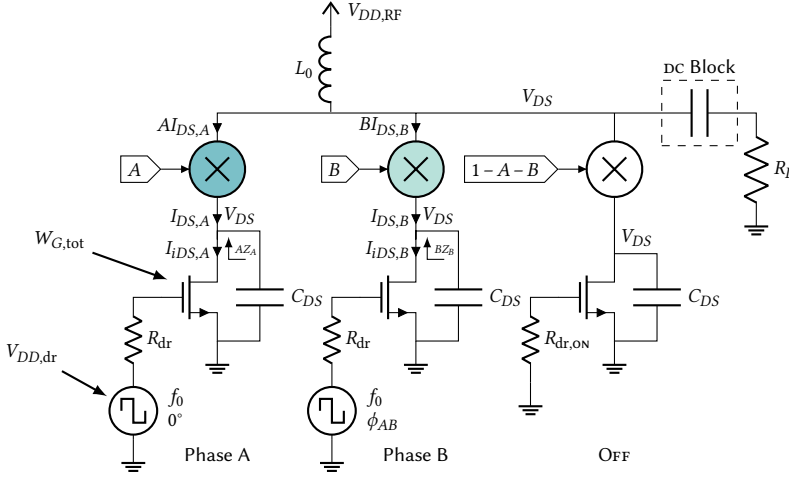
Figure 5.10: Schematic of a simplified DTX simulation model using explicit current scaling and the simplified driver model using its equivalent switch resistance R_{dr}.

directly corresponds to the actual operation of a DTX, regardless of operating class. When using a normalized ACW (range $[0,1]$) the full (unsegmented) $W_{G,\mathrm{tot}}$ should be used for the device model, whereas the ACW should not be normalized when using a single segment's W_G. To accurately reflect the entire operation of a DTX, the inactive (OFF) segments should be added. For simple simulations using ideal current sources this is not necessary. However, the total connected width should remain constant when using a transistor model with (nonlinear) capacitances. This situation is schematically shown in Fig. 5.10. All V_{DS} nodes are linked together to ensure the correct operation of the device, i.e., only the currents are scaled. It should be noted that, even though this simulation approach is using current scaling, its application is not limited to current-scaling DTXs. For example, switch-mode operation with R_{ON} modulation can also be used. A schematic-level implementation of the current scaler is provided in Section B.2.1.

In the schematic of Fig. 5.10, the CMOS drivers are replaced with a (frequency domain) pulse voltage source and the driver's equivalent switch resistance R_{dr} or its ON-resistance $R_{\mathrm{dr,ON}}$ (see Section 4.1.2). In Fig. 5.10, the full gate width $W_{G,\mathrm{tot}}$ is assumed for each device, such that the scaling factor is in the normalized range $[0,1]$. Doing so means that the related driver resistances should be scaled accordingly. The power required for a complete driver chain is then estimated by multiplying the switched capacitance loss (the power dissipated in R_{dr}) with the factor M, which is set by the driver's equivalent resistance R_{dr} (using Eqs. (4.12) and (4.13)).

It may be difficult to find the switched capacitance loss in a frequency-domain simulation. A switch with a resistance R_{dr} toggling between ground and $V_{DD,\mathrm{dr}}$ (e.g., the SPDT_Dynamic component described in Section B.2.3) might alleviate this issue; the switched capacitance loss is then simply the DC power delivered by the $V_{DD,\mathrm{dr}}$-source. Alternatively, the actual driver schematic can be used to drive a single segment for more accuracy. Long driver chains may impact frequency domain convergence. However, it is possible to shorten such a chain to only the final driving stage in simulation while keeping the driver power

consumption accurate. This method is also described in Section 4.1.3. No matter which approach is used, the driver's supply current should be scaled by the activation, similar to the drain current of the power device, to reflect the correct power consumption.

The peak voltage of the pulse voltage source should be the targeted CMOS supply voltage. This peak voltage should remain constant, and the amplitude modulation is set by the value of A. Phase modulating this source is possible, provided the peak voltage remains correct. This is very difficult to define in the frequency domain. Instead, it is much easier to add a discrete phase modulation step to the simulation set-up by duplicating the active current scaling model and providing a pulse voltage source with a different phase. This can then be repeated for all possible phases in the DTX upconversion architecture (Section 2.1).

The example in Fig. 5.10 uses two phases with corresponding activations. Note that the numerical baseband inputs A and B here directly relate to the definition of $da_{1[0]}$ in Eq. (5.16), here for the situation of Fig. 5.10. Choosing $\phi_{AB} = 90°$ corresponds to the first quadrant of a signed Cartesian DTX, for example. This makes phase reference $P_B = $ j, giving $da_{1[0]} = (A + \mathrm{j}B) \sqrt{2P_{\mathrm{norm}}}$. In other words, $A = I|_{I>0}$ and $B = Q|_{Q>0}$. Of course, the segments that should remain OFF now are weighted by $1 - A - B$, such that the total width again remains the same. The output impedance seen by the intrinsic device plane can then be visualized per activation phase, for example, for phase A by

$$Z_A = \frac{V_{DS}}{AI_{iDS,A}}, \tag{5.20}$$

where $I_{iDS,A}$ is the internal controlled current source node for the power device controlled for phase A by the A activation.

Using this simulation model greatly simplifies setting up system-level simulations for a DTX, which is suitable for, e.g., designing output matching networks. A complete frequency domain solution of the network can be found using harmonic balance simulations when all nodes in the circuit can reach a harmonic steady state. This is the case when the A and B signals have well-defined harmonic contents, which is the case for static values and simple periodic signals, such as two-tone simulations. For arbitrary A and B signals, an envelope simulation can be used, where the A and B signals are defined in baseband for discrete time-steps, while the RF fundamental and its harmonics are solved in the frequency domain. Additionally, quantization noise can be included in this simulation by explicitly quantizing the A and B signals; otherwise, the simulation accuracy is limited by the floating-point accuracy, the SPICE tolerance settings, and the number of harmonics considered (order). These kinds of simulations are simply not possible with the discrete simulation model. Table 5.1 summarizes the simulation models and their capabilities.

5.2.3 Example of System Level Simulations: DTX Two-Tone Operation

To see the benefits of using the current scaling simulation model, we describe an example of a two-tone simulation of a DTX system. In this scenario, the baseband input signal can be described as

$$da_{1[0]} = A\cos\left(\frac{\omega_{\mathrm{TT}}t}{2}\right) + 0\mathrm{j} \tag{5.21}$$

which is shown graphically in Fig. 5.11. Here $\omega_{\mathrm{TT}} = 2\pi f_{\mathrm{TT}}$ provides the two-tone spacing and A its amplitude. As was emphasized in Section 5.1.3, this only represents the wanted

Table 5.1: Comparison between the discrete and current scaling models.

	Discrete	Current Scaling
Elements	many	few
Harmonic steady state	✖	✔
Simulation time	high	low
Accuracy	high	medium
Phase shift handling	good	poor*
Possible simulation Modes		
DC	✔	✔
AC/SP	✔	✔
Transient	✔	✔
HB	≡	✔
Env	✖	✔

*) Can be mitigated by adding more elements, adding one for each possible phase.
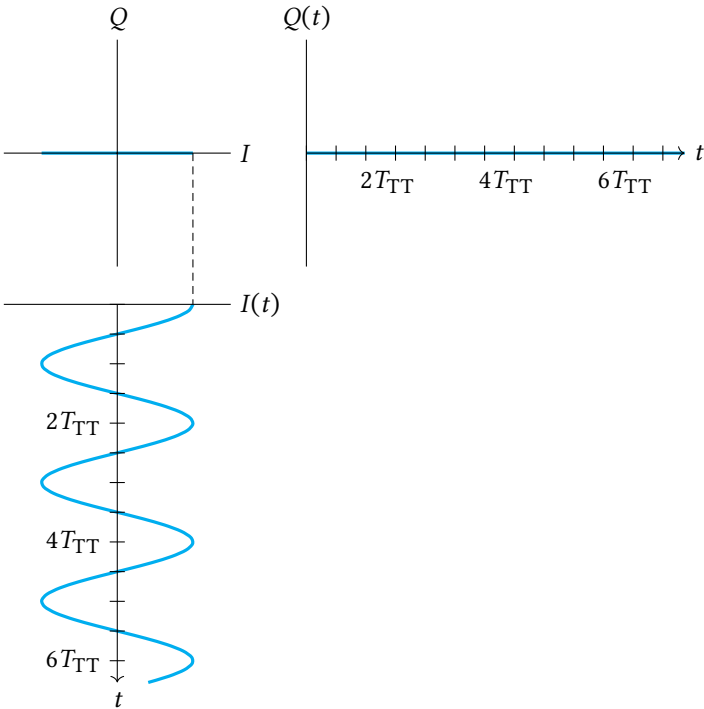
Figure 5.11: Baseband I and Q representation of the digital input signal da_1 for a two-tone simulation.
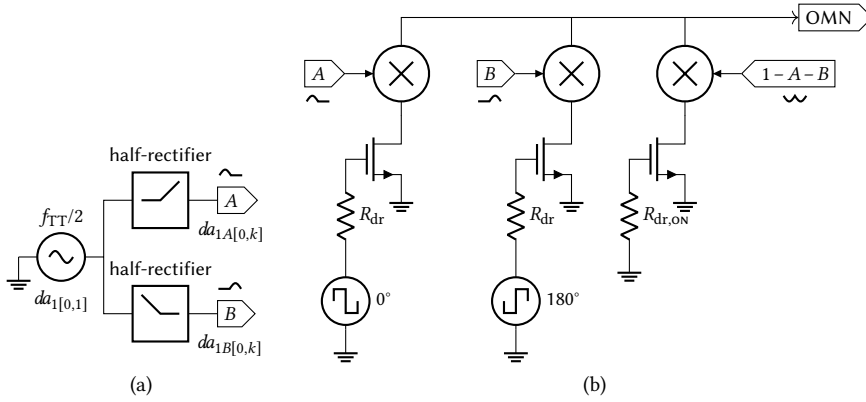
Figure 5.12: Schematics for simulating a DTX with a two-tone input, suitable for harmonic balance simulation. The output matching network is provided in Fig. 5.14.
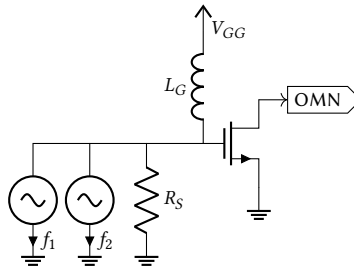
Figure 5.13: Schematics for simulating an analog PA with a two-tone input, suitable for harmonic balance simulation. The output matching network is provided in Fig. 5.14.

phasor at RF, the actual underlying amplitude code words (ACWs) cannot have a negative value. In fact, the negative values of $da_{1[0]}$ are implemented by a 180-degree phase shift. As such, all positive values of $da_{1[0]}$ are mapped onto ACW_A with phase reference $P_A = 1$ and all negative values onto ACW_B with phase reference $P_A = -1$. This is conceptually shown in Fig. 5.12b. A normal sine voltage source is used to generate the inputs A and B, followed by two half-rectifiers (e.g., LinearActivation from Section B.2.5) shown in Fig. 5.12a. The normal sine is then the representation of the fundamental tone at the baseband, so the harmonic index $[0, 1]$. The remaining OFF signal can be acquired simply by adding and subtracting voltages. The resulting time domain waveforms are shown in Fig. 5.15a. Alternatively two half-sine voltage sources could be used to generate A and B, provided these are available in the used simulation software. This clearly yields (unavoidable) bandwidth expansion of the baseband signal, of which the resulting harmonic spectrum is shown in Fig. 5.15b.

A DTX with a real power device model and a realistic matching network as provided in Fig. 5.14, including transmission-line-based harmonic shorts, is simulated. Using a harmonic balance set-up with 2070 frequencies and 65536 time-samples (equivalent to
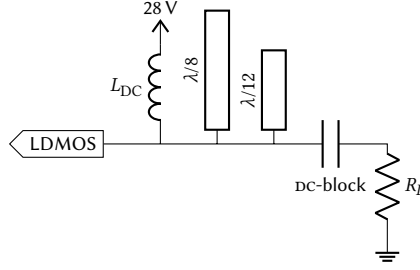
Figure 5.14: The output matching network (OMN) used in the two-tone simulations, both for the analog PA and the DTX.

100 ns), this simulation is finished in 28 seconds on a laptop[1]. In comparison, a 20 ns transient simulation of a schematic-level DTX on a server[2] may take up to 40 minutes. This simulation yields the intrinsic drain current harmonic spectrum that is $AI_{DS,A} + BI_{DS,B}$, as shown in Fig. 5.15c. A copy of $da_{1A} + da_{1B} = |da_1|$ appears directly around DC as mixing terms with harmonics $[0, 2k]$, which are shown in cyan. The resulting fundamental current is then shown in orange. This current is injected into the resonant output-matching network (Fig. 5.14), which results in the outgoing wave b_2 into R_L (Fig. 5.15d). It shows significant IM$_3$ products, as no DPD is applied here, and this DTX is driven 3 dB into compression.

For comparison, an analog class-AB PA is also simulated for the same nominal output power, output matching network with harmonic shorts, and driven into the same compression level. The simulation structure is shown in Fig. 5.13 and the identical output matching network of Fig. 5.14, where the intrinsic drain current harmonic spectrum is shown in Fig. 5.16a. Here, the harmonic index $[1, 0]$ refers to the first tone with $f_1 = f_c - \frac{f_{TT}}{2}$ and $[0, 1]$ refers to the second tone with $f_2 = f_c + \frac{f_{TT}}{2}$. As such, it is not possible to divide the spectrum into harmonic origins with multiples of f_c, but rather into odd and even orders of mixing products.

This highlights a couple of things. First, in the DTX two-tone scenario the baseband current yielded a harmonic expansion of the absolute value of the digital input wave. In the analog class-AB case, the exact same thing happens as even-order intermodulation products. This can be intuitively explained by the PA's gate bias being around the threshold voltage, and as such, it only conducts the positive part of the time domain waveform. The only difference is that for a DTX, this explicitly happens at the input signal splitting in baseband, whereas, for the analog case, this happens at RF and is subjected to the used power device's $V_{GS}-I_{DS}$ curves. Moreover, the third-order intermodulation products (IM$_3$) for the analog case are given by harmonic indices $[2, -1]$ (namely $2f_1 - f_2 = 2\left(f_c - \frac{f_{TT}}{2}\right) - \left(f_c + \frac{f_{TT}}{2}\right) = f_c - \frac{3f_{TT}}{2}$) and $[-1, 2]$ (vice versa). For a DTX, however, since the base band input is explicitly mixed with the upconverting clock the IM$_3$ products are found at $[1, -3]$ (namely $f_c - \frac{3f_{TT}}{2}$)

---

[1]Running Keysight ADS 2021U2, Windows 10 Pro, with a 10th generation Intel Core i7-10750H (2.6 GHz to 5.0 GHz, 6 cores, 12 threads) CPU, and 32 GiB RAM.
[2]Running Cadence Virtuoso ICADVM20.1, CentOS 7.9, with 2 Intel Xeon Gold 6136 (3.0 GHz to 3.7 GHz, total 24 cores and 48 threads) CPUs, and 377 GiB RAM.
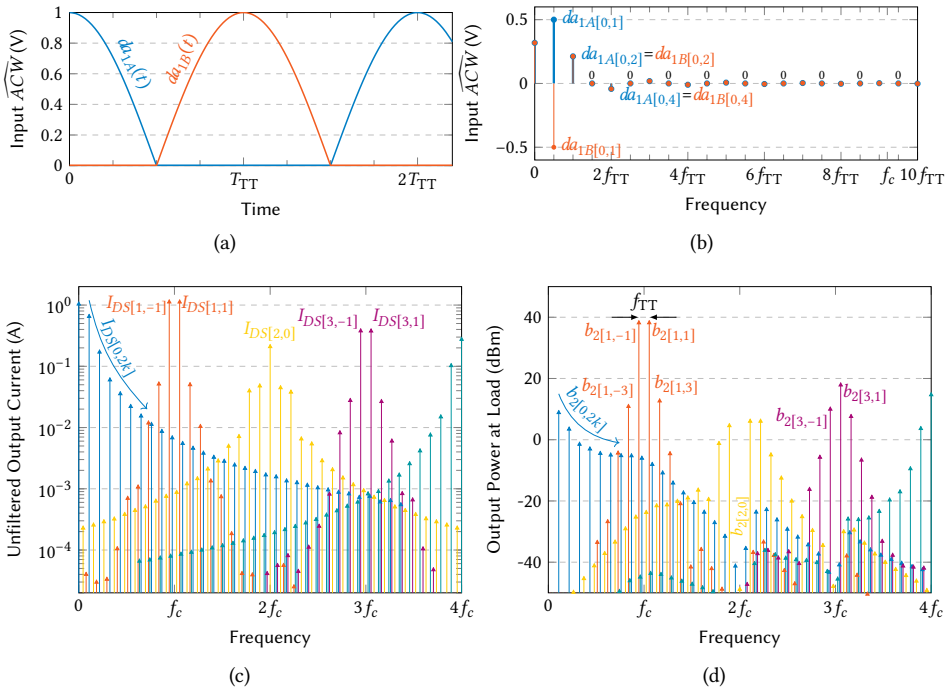
Figure 5.15: Digital inputs and output spectra of a digital system using a large-signal two-tone excitation for realistic device models (without DPD), showing the frequency relations in a DTX.
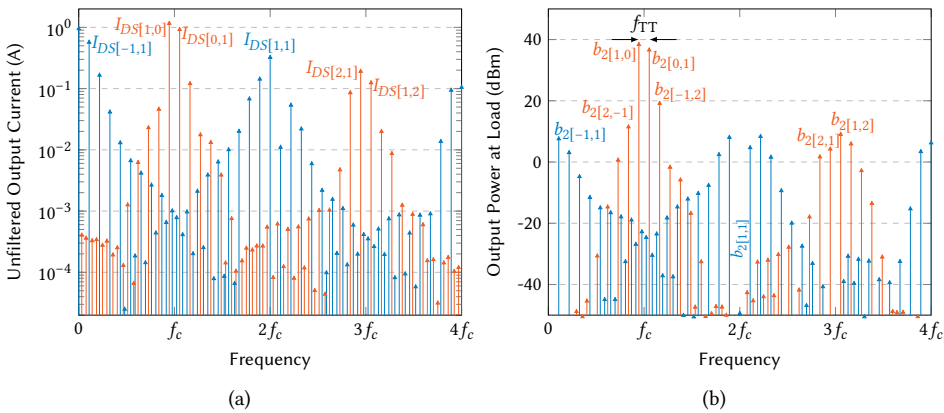
Figure 5.16: Output spectra of an analog system using a large-signal two-tone excitation for a realistic device model (without DPD). Note the very similar output spectrum to the digital case, but with increased asymmetry of the two tones and their IM_3 products.

and $[1,3]$ (namely $f_c + \frac{3f_{TT}}{2}$) instead, which in practice yields the same output frequencies, but depend on different mixing products (the indices). Even though this is a single (quite specific) example, it does highlight the operation of a DTX. Furthermore, secondary IM_3 mixing yields IM_3 asymmetry for the analog case due to the reactive nature of the input biasing network, which is replaced by a (very) low source impedance R_{dr} from the drivers in the DTX case.

5.3 Conclusions

The transfer of a DTX is discussed in this chapter. Namely, a (baseband) digital input wave da_1 is defined, which is mixed to a modulated (analog) output wave b_2 under the influence of one (in case of polar upconversion) or more (in case of signed Cartesian or multi-phase upconversion) reference clocks. Evaluating the input da_1 at baseband with its reference phase at the fundamental results in a complex value (Eq. (5.16)) representing the wanted complex modulation envelope. The normalized digital forward transfer D_{21} is then simply defined by dividing the output wave b_2 (envelope value at the fundamental frequency) by the complex da_1 (at baseband). This connects very well to the testing situation in a practical measurement set-up, where b_2 can be measured using a vector signal analyzer (VSA), while the DTX's numerical input is a given. More frequencies and harmonics are present in a DTX system, which, using the harmonic superposition principle, can be further generalized, which is done in Section A.3.

The superposition of harmonics helps us understand the full operation of a DTX, which is then used to introduce a simplified DTX simulation model. In this simulation model, the drain current of the output power stage is scaled by the baseband input while being driven by an RF clock with the corresponding input's reference phase. By doing so, this set-up can be evaluated using frequency domain techniques. Aside from providing over 400× speed-up, it also better helps a designer understand the operation of a DTX by visualizing the different contributions to the output by harmonic index. However, using this simplified current-scaling DTX simulation set-up, the effects of any digital logic, such as encoders and decoders, or any timing mismatches/glitches cannot be simulated. These circuits have to be evaluated separately to ensure correct operation. Assuming their correct operation, their output should be identical to the input signal applied to the simplified current-scaling set-up. Then, the current-scaling set-up can be used to effectively design harmonic output matching networks, even when including 3D electromagnetic (EM)-simulated S-parameter models, for such a high-power DTX.

6

Estimating the DTX Output Power and Efficiency

As a key objective, this dissertation is focused on combining digital-oriented low-power CMOS with high-power RF technologies. The big question is how to realize the power-efficient operation of these DTXs.

To do so, it is essential to understand its underlying powers and their relations, as illustrated in Fig. 6.1, and described Section 6.1. The related drain and system efficiencies of a DTX can be calculated based on these powers, similar to traditional analog PAs. Using these insights and the theory of current-scaling DTXs (Section 2.2.2), it is possible to develop a simplified power model for a DTX, allowing a quick estimation of its efficiency performances. This is done first in Section 6.2 for single line-up DTXs and extended later for efficiency enhancement techniques like Doherty, which use in each branch a DTX line-up (Section 6.3). Two DTX system examples are given in Section 6.4 to showcase the proposed DTX power model, demonstrating the performance potential of DTXs.
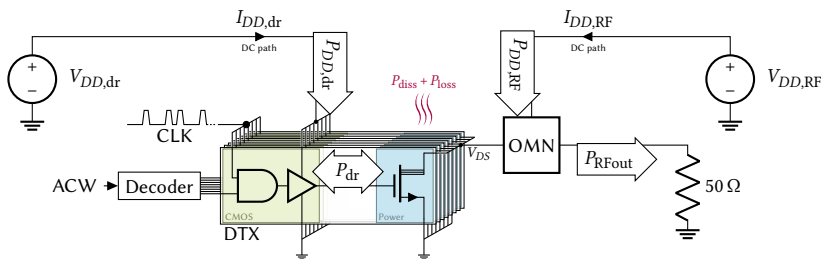
Figure 6.1: The power flows in a DTX.

6.1 DTX Power and Efficiency Definitions

There are four power relations of interest in a DTX, as shown in Fig. 6.1. First and foremost, P_{RFout} is the DTX output power at RF, which is typically delivered to a load of 50 Ω. This is defined as usual

$$P_{\text{RFout}} = \Re\left\{ \frac{V_{\text{RFout}}\overline{I_{\text{RFout}}}}{2} \right\} = \frac{\Re\{V_{\text{RFout}}^2\}}{2 \cdot 50}.$$ (6.1)

Only the wanted power around the fundamental is included in P_{RFout}, since all harmonic currents are assumed to be perfectly shorted, so $P_{\text{RFout}} = P_{\text{RFout}}$[1]. Next is the power consumed by the power stage, which is the power delivered by its DC-bias source, straightforwardly calculated as

$$P_{DD,\text{RF}} = V_{DD,\text{RF}}I_{DD,\text{RF}}$$ (6.2)

as $P_{DD,\text{RF}}$ is a DC-only power.

The digital driver, as introduced in Chapter 4, delivers the input signal to the power stage. It needs to charge and discharge the input capacitance C_{GG} of the output segment between the driver's supply voltage $V_{DD,\text{dr}}$ and ground. This driver is implemented on the CMOS die and can have a higher supply voltage than the nominal CMOS supply; as such, we differentiate between the driver's supply voltage $V_{DD,\text{dr}}$ and the voltage $V_{DD,\text{core}}$ which provides the DC power for the digital core devices. The power needed to drive the power stage follows from the number of activated segments N_{act} multiplied by the capacitive switching loss of the input capacitance (from Eq. (4.1))

$$P_{\text{dr}} = N_{\text{act}} \cdot f_0 C_{GG}V_{DD,\text{dr}}^2.$$ (6.3)

From this equation, it is clear that there's a linear relation with operating frequency and input capacitance while it scales quadratically with the driving voltage. More interestingly, it scales linearly with the number of activated segments, saving power at lower activation levels. When no segments are activated, no power is consumed at all.

The power required to switch the input capacitance is not the only power consumed by the driver. This driver also must drive the parasitic interconnect capacitance and the driver has its own input and output capacitance, which might require a preceding tapered buffer chain. The total DC power consumed by the driver(s) follows from Eq. (4.11) and is then given by

$$P_{DD,\text{dr}} = V_{DD,\text{dr}}I_{DD,\text{dr}} = N_{\text{act}} \cdot f_0 M C_L V_{DD,\text{dr}}^2 + P_{\text{cont}}.$$ (6.4)

The power consumed by the driver and tapered buffer chain with their capacitive load C_L is proportional to MC_L, in which the factor M (as defined in Section 4.1.3, Eq. (4.10)) indicates the total capacitive overhead with respect to C_L. The "continuous" power dissipation (P_{cont}) can be attributed to, for example, level shifter(s) and static leakage currents and can be considered extremely small compared to the dynamic power dissipation.

The voltage-to-power transfer of the segments of the output power device makes the classical efficiency definitions based on the "operating power gain" (G_P) useless. However, when considering the power consumption of the tapered buffer chain, the (system) efficiencies of interest can still be defined similar to those of analog PAs (see Section A.2.1), starting with the drain efficiency of the power output stage

$$\eta_D = \frac{P_{\text{RFout}}}{P_{DD,\text{RF}}}.$$ (6.5)

The difference between $P_{DD,\text{RF}}$ and P_{RFout} is dissipated as heat in the intrinsic device

$$P_{\text{diss}} = P_{DD,\text{RF}} - P_{\text{RFout}}. \tag{6.6}$$

Further, the total efficiency of the power output stage

$$\eta_T = \frac{P_{\text{RFout}}}{P_{\text{dr}} + P_{DD,\text{RF}}}. \tag{6.7}$$

Note that P_{dr} should not be confused with the (almost zero) input power P_{in} in a conventional impedance-matched RF situation where the input capacitance of the power device is resonated out, although they serve a similar function in the expression. In that sense, the most closely related measure to 'power gain' would be $\frac{P_{\text{RFout}}}{P_{\text{dr}}}$, although it is important to stress again that this metric does not convey any linearity information of a DTX (see Section 5.1.3), nor does it capture a fair comparison to analog PAs, as their efficiency metrics typically do not include any (pre-)driver related power(s). Lastly, the system efficiency is defined as

$$\eta_S = \frac{P_{\text{RFout}}}{P_{DD,\text{core}} + P_{DD,\text{dr}} + P_{DD,\text{RF}}} \tag{6.8}$$

in which $P_{DD,\text{core}}$ is the power consumed at the CMOS controller's core supply voltage. These powers include the drivers and their tapered buffer chains, which also include the input capacitive switching loss of the power stages' gate segments (P_{dr}), as well as possible static (leakage) powers and any continuous power consumed by, e.g., continuously running RF clocks.

6.2 DTX Power Model in a Single Line-Up

The provided theory of operation and accompanying equations allow a simplified power model for a DTX. It aims to get a quick first-order approximation of what can be achieved and intuitively describe how device parameters influence DTX energy performance. Such a model uses several simplifications, blocking it from being a replacement of actual circuit simulation. However, it can provide a reasonably accurate estimation when provided with the correct input parameters. The simplifications applied are:

- The output class is based on digital-current scaling with all harmonics shorted (see also Section 2.2.2).
- No output stage compression is considered.
- Uniform unary segmentation is assumed.
- Only (shunt) capacitive interconnect parasitics are considered; any series parasitics are considered negligible.
- Phase uniform modulation, such as OFDM, is assumed.

Only the model's governing equations for power and (system) efficiency are described here; practical examples and their interpretation are provided in the next section.

A reference (ideal) peak fundamental output power per unit gate width W_G can be derived from analog (polar) class-B operation (see Figs. 2.15–2.17 and Eq. (A.18))

$$\hat{P}_{\text{classB}} = \Re\left\{ \frac{V_{DS}\overline{I_{DS}}}{2} \right\} = \frac{1}{2}V_{DS,\text{max}}\left(\frac{1}{2} \cdot \hat{I}_{DS,\text{max}} \right). \tag{6.9}$$

The value of $\hat{I}_{DS}$ and $\hat{P}_{classB}$ for a given technology are given per unit 'width' (typically mm). The current $\hat{I}_{DS}$ should be provided using a V_{GS} swing of $V_{DD,dr}$ considered in large signal operation, but, when $V_{DD,dr}$ and the power technology are chosen properly, it should have little small-signal dependency on the actual drive voltage (i.e., $\hat{I}_{DS} = \hat{I}_{DS,max}$, see Fig. 3.2). Ideally $V_{DS,max} = V_{DD,RF}$ at full swing, but typically a 'correction factor' is needed to include intrinsic device losses and the V_{DS}-knee voltage. The dominant intrinsic device loss mechanism depends on the power technology used. For LDMOS, for example, the dominant loss mechanism can be modeled by its ON-resistance (e.g., $V_{DS,max} = V_{DD,RF} - R_{ON}I_{DS,max}$) and parasitic R_{DS}, the series resistance of the (nonlinear) shunt output capacitance C_{DS} [75], providing an operating-frequency dependent loss.

A target output power should be defined to calculate the required total gate width $W_{G,tot}$ of the output stage, along with selecting an upconversion architecture and RF duty cycle (operating class). The used DTX architecture leads to a bank current utilization factor U_{bank}, as derived and defined[1] in [21]. This U_{bank} depends on the phase angles between possible output activation phasors ϕ_{AB} and the bank implementation as

$$U_{bank}(\phi_{AB}) = \begin{cases} 1 & \text{single bank (polar),} \\ 1/2 & \text{separate banks,} \\ \cos\left(\frac{\phi_{AB}}{2}\right) & \text{bank sharing,} \\ 1 & \text{bank interleaving.} \end{cases} \tag{6.10}$$

There is only one bank needed for polar operation (since it is driven with a continuous changing phase), thus $U_{bank} = 1$ serving as the reference case. The required total gate width $W_{G,tot}$ depends on the fundamental drain current, which in turn depends on the bank current utilization, the used RF duty cycle (Eq. (2.17)) and the rise/fall times with respect to the RF period (Eq. (2.20)). It can be calculated as

$$W_{G,tot} = \frac{P_{RFout,target}}{\hat{P}_{classB}} \cdot \frac{1}{U_{bank}} \cdot \frac{\pi}{4\sin(\pi d)} \cdot \frac{1}{\text{sinc}(t_{rf}f_0)}. \tag{6.11}$$

Here $P_{RFout,target}$ is the targeted fundamental peak output power in watts. Note that when using bank interleaving, the RF duty cycle is limited to $\frac{\phi_{AB}}{2\pi}$.

With the size of the switch bank now fixed, other quantities related to the output stage can be determined. These quantities are shown in Fig. 6.2. This includes the total capacitive load C_L to be driven. This load C_L here consists of two parts. The first part is related to connecting the CMOS controller to the segmented power device, and includes any interconnect and ESD protection diodes. This part is represented in Fig. 6.2 as C_{intcon}. The second part is the capacitance intrinsic to the power device, earlier (Eq. (6.3)) referred to as C_{GG}. Here

$$C_{GG} = C_{GS} + C_{GD}\left(1 + \frac{2V_{DS}}{V_{DD,dr}}\right). \tag{6.12}$$

---

[1]Mul *et al.* [21] actually defines an upconversion current utilization factor F_{up}, which describes the ratio between the vector summation and summation of their absolute values (Eqs. (2.4), (2.8) and (2.13)), in the context of a bank sharing implementation, which also depends on the modulation phase θ. Its minimum value (for $\theta = \phi_{AB}/2$) is $F_{up,min}(\phi_{AB}) = \cos(\phi_{AB}/2)$, and specifies the required relative $W_{G,tot}$. Here it is renamed to U_{bank} for clarity in the context of the different bank implementations.
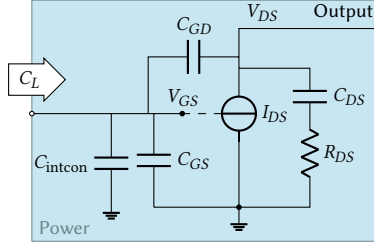
Figure 6.2: Relevant capacitances for the power output stage.

The feedback capacitance C_{GD} appears larger at the input due to the change in drain–source voltage V_{DS} of the power devices, which is in the worst case twice the fundamental magnitude (i.e., $2V_{DS,\text{max}}$). Miller's approximation is used here to represent this, which is a good approximation at the operating frequency: only a little phase shift from input to output is expected, since the output capacitance is resonated out, and the input has no impedance matching applied. All these capacitances may be nonlinear, but their effective value can be found in simulation by integrating their input currents when switching from ground to $V_{DD,\text{dr}}$ (or the other way around, using the conservation of charge). When normalizing them again to a capacitance per unit gate width ($\hat{C}_{GG}$ and $\hat{C}_{\text{intcon}}$), the total capacitive load is given by

$$C_{L,\text{tot}} = W_{G,\text{tot}} \left(\hat{C}_{\text{intcon}} + \hat{C}_{GG} \right). \tag{6.13}$$

Especially $\hat{C}_{\text{intcon}}$ can be difficult to estimate as it is an empirical value and highly dependent on the DTX implementation. Still, it is important to consider, as it can contribute $30\,\%$ to $120\,\%$ additional capacitance compared to C_{GG}. The output parasitics can be estimated similarly by

$$C_{DD} = C_{DS} + C_{GD} \left(1 + \frac{V_{DD,\text{dr}}}{2V_{DS}} \right) \approx C_{DS}. \tag{6.14}$$

The approximation $C_{DD} \approx C_{DS}$ can be made when the feedback capacitance is small and the input-to-output voltage ratio is small, which both are typically the case for RF power technologies. Based on the total gate width

$$C_{DS} = W_{G,\text{tot}} \hat{C}_{DS} \tag{6.15}$$

and

$$R_{DS} = \frac{\hat{R}_{DS}}{W_{G,\text{tot}}}. \tag{6.16}$$

Assuming full signal swing ($V_{DS,\text{max}}$), the output stage's losses at peak output power can be estimated as

$$P_{\text{loss}} = \frac{V_{DS,\text{max}}^2}{2} \cdot \frac{R_{DS}}{R_{DS}^2 + \frac{1}{\omega_0^2 C_{DS}^2}}. \tag{6.17}$$

The P_{loss} is proportional to the output power, since it is in parallel to the output, and both the loss and the output power depend on the square of V_{DS}. From Eqs. (2.19) and (2.20) we

know the optimum load, which including the bank current utilization factor gives

$$R_{L,\text{opt}} = \frac{V_{DS,\text{max}}}{W_{G,\text{tot}}\hat{I}_{DS,\text{max}}U_{\text{bank}}} \cdot \frac{\pi}{2\sin(\pi d)} \cdot \frac{1}{\text{sinc}\left(t_{rf}f_0\right)}. \tag{6.18}$$

Since we also know the output capacitance and operating frequency, the maximum fractional bandwidth (FBW) of the DTX's output matching network can be estimated from the quality factor and the maximum tolerable reflection coefficient $|\Gamma|_{\text{max}}$ [76]. The Q factor is

$$Q = \omega_0 R_{L,\text{opt}} C_{DS} = V_{DS,\text{max}}\frac{\omega_0 \hat{C}_{DS}}{\hat{I}_{DS,\text{max}}U_{\text{bank}}} \cdot \frac{\pi}{2\sin(\pi d)} \cdot \frac{1}{\text{sinc}\left(t_{rf}f_0\right)} \tag{6.19}$$

which is independent of the used $W_{G,\text{tot}}$. The fractional bandwidth then is

$$\text{FBW} = \frac{\Delta\omega}{\omega_0} = \frac{1}{Q\sqrt{\frac{1}{|\Gamma|^2_{\text{max}}} - 1}} \tag{6.20}$$

Typically the -3 dB bandwidth is used where $|\Gamma|_{\text{max}} = 1/\sqrt{2}$, such that $\text{FBW} = Q^{-1}$, although in PA and TX design -1 dB is also commonly used which relates to $|\Gamma|_{\text{max}} \approx 0.45$ and $\text{FBW} \approx 0.51Q^{-1}$.

The expected DTX drain efficiency is a linear function of amplitude (ρ), assuming no output stage compression, and is described by

$$\eta_D(\rho) = \frac{P_{\text{RFout}}(\rho)}{P_{DD,\text{RF}}\left(\rho,\phi_{AB}\right)} = \frac{\rho^2\left(P_{\text{RFout,target}} - P_{\text{loss}}\right)}{\rho P_{DD,\text{RF}}\left(\phi_{AB}\right)} \propto \frac{\rho^2}{\rho} = \rho, \tag{6.21}$$

where $P_{DD,\text{RF}}$ depends on the RF duty cycle (see Eq. (2.18)) and the chosen upconversion architecture. Namely, it depends on the phase-averaged current utilization factor[2] from [21]

$$F_{\text{up,avg}}\left(\phi_{AB}\right) = \frac{\phi_{AB}}{2}\cot\left(\frac{\phi_{AB}}{2}\right). \tag{6.22}$$

This then gives

$$P_{DD,\text{RF}}\left(\rho,\phi_{AB}\right) = \rho V_{DD,\text{RF}}I_{DS,\text{max}}U_{\text{bank}} \cdot \frac{d}{F_{\text{up,avg}}\left(\phi_{AB}\right)}, \tag{6.23}$$

or by substituting Eq. (6.23) into Eq. (6.21), as $I_{DS,\text{max}}U_{\text{bank}} \propto P_{\text{RFout,target}}/V_{DS,\text{max}}$

$$\eta_D\left(\rho,\phi_{AB}\right) = \rho\left(1 - \frac{P_{\text{loss}}}{P_{\text{RFout,target}}}\right)\frac{V_{DS,\text{max}}}{V_{DD,\text{RF}}}\text{sinc}(d)\,\text{sinc}\left(t_{rf}f_0\right)F_{\text{up,avg}}\left(\phi_{AB}\right). \tag{6.24}$$

The peak input power to the power stage follows from substituting Eq. (6.13) into Eq. (6.3)

$$P_{\text{dr,tot}} = f_0 C_{L,\text{tot}}V_{DD,\text{dr}}^2. \tag{6.25}$$

---

[2]Note that in the case of polar modulation $F_{\text{up,avg}} = 1$.

Table 6.1: Approximate values for the upconversion current utilization factors for the various bank implementations, from [21].

	ϕ_{AB}	U_{bank}	$F_{\text{up,avg}}$	$F_{\text{dr,avg}}$
Polar	0	1	1	1
Separate Banks	$\pi/2$	1/2	0.785	0.785
	$\pi/4$	1/2	0.948	0.948
	$\pi/8$	1/2	0.987	0.987
Bank Sharing	$\pi/2$	0.707	0.785	0.785
	$\pi/4$	0.924	0.948	0.948
	$\pi/8$	0.981	0.987	0.987
Bank Interleaving	$\pi/2$	1	0.785	1.111
	$\pi/4$	1	0.948	1.281
	$\pi/8$	1	0.987	1.320

This is the power when the full $W_{G,\text{tot}}$ would be switched. The actual required phase-averaged driver power is again dependent on the phase-averaged current utilization factor, similar to $P_{DD,\text{RF}}$. However, the bank interleaving case is special. The drivers' power consumption in bank interleaving mode depends on the maximum value of either activation vector, instead of their sum. This is hard to express in a universal expression, so analytical values are provided next instead. The phase-average driver power utilization factor is then given by

$$F_{\text{dr,avg}}\left(\phi_{AB}\right) = \begin{cases} F_{\text{dr,avg,interleaving}}(\phi_{AB}) & \text{bank interleaving,} \\ F_{\text{up,avg}}(\phi_{AB}) & \text{other implementations,} \end{cases} \tag{6.26}$$

where for signed-Cartesian and 8- and 16-phase multi-phase operation

$$F_{\text{dr,avg,interleaving}}(\phi_{AB}) = \begin{cases} \frac{\pi\sqrt{2}}{4} & \phi_{AB} = \frac{\pi}{2}, \\ \frac{\pi}{8}\left(\cos\left(\frac{\pi}{8}\right) + \sin\left(\frac{\pi}{8}\right) - 1\right)^{-1} & \phi_{AB} = \frac{\pi}{4}, \\ \frac{\pi}{16}\left(\sin\left(\frac{\pi}{16}\right) - 2\left(1 + \sqrt{2}\right)\sin^2\left(\frac{\pi}{32}\right)\right)^{-1} & \phi_{AB} = \frac{\pi}{8}. \end{cases} \tag{6.27}$$

Table 6.1 provides an overview of its approximate values. The phase-averaged input drive power as a function of amplitude is then

$$P_{\text{dr,tot}}\left(\rho, \phi_{AB}\right) = \rho \cdot f_0 C_{L,\text{tot}} U_{\text{bank}} V_{DD,\text{dr}}^2 \cdot \frac{1}{F_{\text{dr,avg}}\left(\phi_{AB}\right)}. \tag{6.28}$$

From Eq. (6.4) then follows the total DC power consumed by the drivers to be

$$P_{DD,\text{dr}} = M\left(t_{rf}\right) P_{\text{dr,tot}}\left(\rho, \phi_{AB}\right) \tag{6.29}$$

where M as a function of t_{rf} follows from Eqs. (A.44) and (4.12), which increases with faster drivers (lower t_{rf}) for a given CMOS driver's technology parameters.

The total efficiency is provided by

$$\eta_T(\rho) = \frac{P_{\text{RFout}}(\rho)}{P_{\text{dr,tot}}\left(\rho, \phi_{AB}\right) + P_{DD,\text{RF}}\left(\rho, \phi_{AB}\right)} \propto \rho. \tag{6.30}$$

All DC powers are linearly proportional to ρ, which means that for $\rho = 0$ no DC power is consumed or dissipated at all. The drain and total efficiencies are also linear with ρ, meaning that both the output stage and the drivers follow a class-B-like efficiency curve. This shows the benefit of using the gate segmentation in the high-power DTX concept over traditional analog TXs, as their drivers typically follow a class-A/AB efficiency curve. When considering the overall DTX system efficiency, also the core power consumption ($P_{DD,\text{core}}$) in the low/core CMOS voltage domain $V_{DD,\text{core}}$ needs to be included, involving: continuously running RF clock lines, memories, interfaces, or computational blocks (e.g., DUC, digital signal processing (DSP), DPD), yielding the system efficiency

$$\eta_S(\rho) = \frac{P_{\text{RFout}}(\rho)}{P_{DD,\text{core}} + P_{DD,\text{dr}}\left(\rho, \phi_{AB}\right) + P_{DD,\text{RF}}\left(\rho, \phi_{AB}\right)}. \tag{6.31}$$

Some DTX upconversion architectures provide a higher spectral purity than others or need more DSP power to achieve so, which is covered in [26] in detail. In general, a DTX's power consumption decreases with technology scaling.

When considering modulated signals such as OFDM, we can introduce a Rayleigh probability density function (PDF) as an aid in calculating the average efficiency [77, 78]. Namely, the Rayleigh PDF is a function of the amplitude ρ depending on the signal's variance σ, namely

$$f_{\text{Ray}}(\rho, \sigma) = \frac{\rho}{\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} \tag{6.32}$$

with

$$\sigma \approx \frac{1}{\sqrt{2\text{PAPR}}} = \frac{10^{-\frac{\text{PAPR}_{\text{dB}}}{20}}}{\sqrt{2}}. \tag{6.33}$$

The approximation made here lies in the fact that, in a DTX, $\rho \in [0, 1]$, while the PDF is defined for $\rho \in [0, \infty)$. For large enough PAPR, this error is small; for example, for PAPR = 8 dB, the deviation is 0.05 dB. The average system efficiency can be given by (numerically) integrating the product of the PDF with the powers over ρ, resulting in

$$\eta_{S,\text{avg}} = \frac{\int_0^1 f_{\text{Ray}}(\rho, \sigma) P_{\text{RFout}}(\rho) \, d\rho}{\int_0^1 f_{\text{Ray}}(\rho, \sigma)\left(P_{DD,\text{core}} + P_{DD,\text{dr}}\left(\rho, \phi_{AB}\right) + P_{DD,\text{RF}}\left(\rho, \phi_{AB}\right)\right) d\rho}. \tag{6.34}$$

6.3 DTX Power Model Using Efficiency Enhancement

Multiple DTX line-ups are required when considering efficiency enhancement. Using Doherty load modulation as an example, one DTX line-up is used for each branch. Most of the considerations above for a single line-up can be repeated here for each DTX branch. However, an adaptation needs to be made for including the Doherty drive profile and its related load modulation. Namely, all equations involving the amplitude ρ ($\propto I_{DS}$) should be reconsidered for the input and output of each individual branch, which result in a current I_{DSm} for the main DTX branch, and $I_{DSp,1}, \dots, I_{DSp,n}$ for n peaking DTX branches. The only exception is the loss Equation (6.17), which should be considered for the voltage swing at the output of each branch.

Figure 6.3 shows three examples of possible driving profiles for Doherty configurations in terms of normalized drain voltage, normalized drain current, and normalized efficiency,
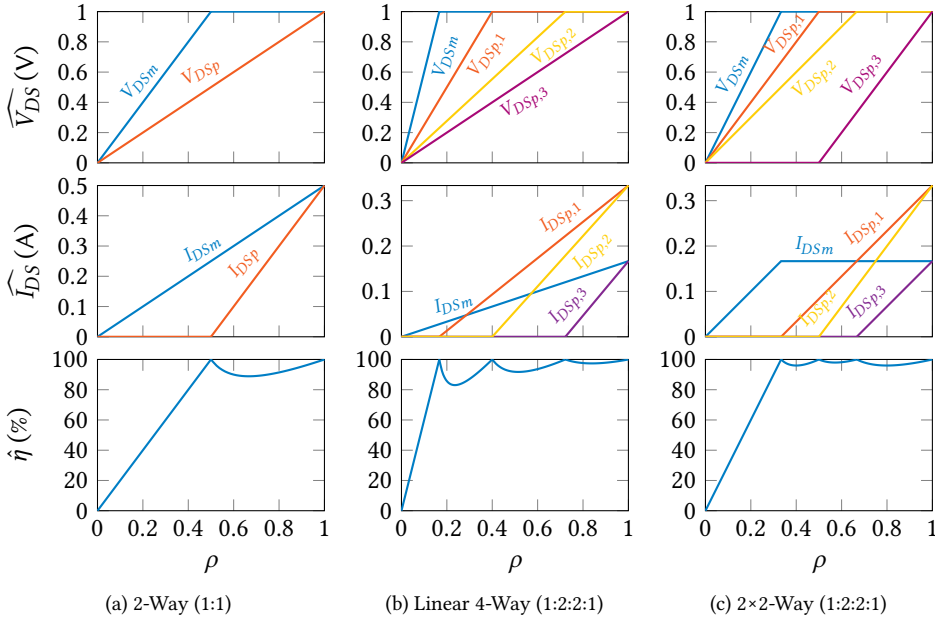
Figure 6.3: Three examples of Doherty driving profiles.

from top to bottom, respectively. The simplest is the symmetrical 2-way Doherty, which means that both branches should be sized for identical output power. For DTX, applying the driving profile is trivial, as it is digitally set by the input ACW. However, this is more difficult for analog implementations since the peaking branch should have twice the transconductance compared to the main branch. This complexity increases for a 4-way Doherty (see Fig. 6.3b), requiring four distinct driving profiles for two different device sizes. Analog implementations either require complicated biasing schemes and input power splitters, or power hungry multiple input upconversion strategies using individual (pre-) drivers. Implementing such a structure using DTXs is comparatively easy: two differently sized DTX branches should be designed, while the driving profile comes basically for free. This gained freedom can support more exotic driving profiles and Doherty configurations, such as the 2×2-way Doherty (Fig. 6.3c) [44]. This configuration, due to its main branch's early current saturation, provides a constant load impedance for the main branch from the first back-off efficiency point onwards, allowing this configuration to be extremely wideband.

Note that by defining $\sum_n \widehat{I_{DS,n}}(\rho = 1) = 1$, most equations introduced in the previous section do not need modification. All quantities are then provided in terms of the total current of all branches summed together. Using the symmetrical 2-way Doherty as an example, setting a $P_{\text{RFout,target}}$ will result in both branches delivering $P_{\text{RFout,target}}/2$ each. Also, each branch will have a $W_{G,\text{tot}}/2$, when identical settings in terms of upconversion architecture, duty cycle, or rise/fall times are assumed. However, note that the main and peak DTX branches are allowed to have different settings.

With the above in mind, we need to make modifications to Eqs. (6.21), (6.23) and (6.29).

First, the total RF output power is given by

$$P_{\text{RFout,tot}}(\rho) = \sum_n \widehat{I_{DS,n}}(\rho)\widehat{V_{DS,n}}(\rho)P_{\text{RFout,target}} - \widehat{V_{DS,n}^2}(\rho)P_{\text{loss},n}\widehat{I_{DS,n}}(1). \qquad (6.35)$$

For the other two equations, simply ρ has to be replaced by the driving profile's $\widehat{I_{DS,n}}(\rho)$ and summed for all branches such that

$$P_{DD,\text{RF,tot}}\left(\rho, \phi_{AB}\right) = \sum_n P_{DD,\text{RF},n}\left(\widehat{I_{DS,n}}(\rho), \phi_{AB}\right) \qquad (6.36)$$

and

$$P_{DD,\text{dr,tot}}\left(\rho, \phi_{AB}\right) = \sum_n P_{DD,\text{dr},n}\left(\widehat{I_{DS,n}}(\rho), \phi_{AB}\right). \qquad (6.37)$$

The total (average) efficiencies can then be found using these updated total power equations.

6.4 Example Calculations with the DTX Power Model

Though the introduced power model may seem complicated, judging from its explanation spanning six pages and 26 equations, all equations are either related to technology constants or simple multiplications and divisions that can be performed by hand (or using a hand calculator). The average efficiency of Eq. (6.34) may be the only exception as it requires integration, but even this can still be performed using numerical integration in standard spreadsheet software by summing over an array of varying amplitudes. Of course, the provided model is strongly simplified and cannot compete with the accuracy of SPICE or EM simulations. However, it can provide qualitative and quantitative insight into how certain parameters relate to others.

The best way to showcase this model's strengths is to provide a few examples. This coincidentally showcases the performance potential of DTXs. The first example starts simple, using a polar single-ended DTX line-up. The second example is provided for a symmetrical 2-way Doherty DTX and compared for its efficiency to a typical analog TX.

6.4.1 Calculation Example for a Single-Ended DTX

To provide a first example, we assume an idealized LDMOS technology (e.g., Fig. 7.2 for a V_T of 0.2 V) for the RF power output stage driven by a 40 nm CMOS controller. Here, we use the post-layout parameters of the N:P 0.8:1.54 μm thick oxide devices at 2.5 V from Table 4.1 for the CMOS driver model. These parameters are repeated here in Table 6.2, together with the assumed LDMOS parameters. Further, we target an output power of 20 W using polar upconversion so that a single bank suffices. The interconnect parasitics between the CMOS and LDMOS die (pads and ESD protection) are assumed to be 0.6 pF per mm width of the LDMOS' output stage. This is also provided in Table 6.2, together with the resulting upconversion current utilization factors, which are all 1 for polar operation.

With the values provided in Table 6.2, we can start the calculations using the power model. We evaluate the model for an operating frequency of 2 GHz, an RF duty cycle of 25 %, and at maximum power. The driver's speed is left as a free parameter so that we can evaluate its impact on system performance. We know from this CMOS technology's parameters that the fastest possible driver chain can reach an output t_{rf} of 58.8 ps, which at 2 GHz corresponds to 11.7 % of an RF cycle. With the assumed duty cycle of 25 %, this

Table 6.2: Assumed technology and model parameters for the single ended DTX example, using a modified 400 nm LDMOS technology (see Fig. 7.2 for $V_T = 0.2$ V), driven by thick oxide 40 nm CMOS (from Table 4.1).

LDMOS Parameters

$V_{DD,\mathrm{RF}}$	28	V
$\hat{R}_{\mathrm{ON}}$	9	Ω mm
$\hat{I}_{DS,\max}$	0.22	A mm^{-1}
$@V_{GS}$	2.5	V
$\hat{C}_{GS}$	1.04	pF mm^{-1}
$\hat{C}_{GD}$	7	fF mm^{-1}
$\hat{C}_{DS}$	0.296	pF mm^{-1}
$\hat{R}_{DS}$	2.3	Ω mm
$V_{DS,\max}$	26	V
$\hat{P}_{\mathrm{classB}}$	1.43	W mm^{-1}
$\hat{C}_{GG}$	1.193	pF mm^{-1}
$\hat{C}_{DD}$	0.303	pF mm^{-1}

Specifications

f_0	2	GHz
$P_{\mathrm{RFout,target}}$	20	W
Upconversion	Polar ($\phi_{AB} = 0$) Single bank	
U_{bank}	1	A A^{-1}
$F_{\mathrm{up,avg}}$	1	A A^{-1}
$F_{\mathrm{up,avg,dr}}$	1	A A^{-1}

CMOS Parameters

$V_{DD,\mathrm{dr}}$	2.5	V
t_{p0}	8.68	ps
γ	0.669	
ς	0.262	
$r_{rf/p1,0-100}$	2.008	

Interconnect Parameters

$\hat{C}_{\mathrm{intcon}}$	0.6	pF mm^{-1}

Table 6.3: DTX power model results and intermediate calculation results for a single line-up DTX with the provided values of Table 6.2.

Variable Model Inputs

d	25	%
$t_{rf}f_0$	17.1	%
ρ	1	

LDMOS Model Parameters

$W_{G,\mathrm{tot}}$	16.308	mm
$C_{L,\mathrm{tot}}$	29.234	pF
C_{DS}	4.827	pF
R_{DS}	0.141	Ω
P_{loss}	0.175	W
$R_{L,\mathrm{opt}}$	16.9	Ω
Q	1.025	
$\mathrm{BW}_{-1\,\mathrm{dB}}$	0.993	GHz

CMOS Model Parameters

t_{rf}	85.52	ps
t_p	42.59	ps
R_{dr}	1.127	Ω
f	1.753	
M	3.215	

Estimated phase-averaged DTX performance at peak power

P_{RFout}	19.82	W
$P_{DD,\mathrm{RF}}$	25.11	W
η_D	78.9	%
$P_{\mathrm{dr,tot}}$	0.365	W
η_T	77.8	%
$P_{DD,\mathrm{dr}}$	1.175	W
η_S	75.4	%

Estimated DTX performance with OFDM signals

PAPR	6	dB
σ	0.3544	
$1 - F_{\mathrm{Ray}}(1,\sigma)$	1.87	%
$P_{\mathrm{RFout,avg}}$	4.61	W
$P_{DD,\mathrm{RF,avg}}$	10.85	W
$P_{DD,\mathrm{dr,avg}}$	0.507	W
$\eta_{D,\mathrm{avg}}$	42.5	%
$\eta_{S,\mathrm{avg}}$	40.6	%

6

Figure 6.4: Maximizing the DTX system efficiency by adjusting the driver's strength, and the resulting powers vs. amplitude ρ.

would theoretically provide an idealized drain efficiency of $\mathrm{sinc}(0.25)\,\mathrm{sinc}(0.117) = 88\,\%$. Evaluating this in our power model including LDMOS R_{ON} and R_{DS} related losses, we find a drain efficiency of 81 %. However, we find an infinitesimal value when we evaluate the related system efficiency, since infinite power is required to provide a driver of such speed in this technology. Sweeping the driver from the maximum speed ($t_{rf} = 58.8\,\mathrm{ps}$) to the allowed minimum ($t_{rf} = d/f_0 = 125\,\mathrm{ps}$, to meet the specified duty cycle requirement) results in the efficiencies as shown in Fig. 6.4a. Here, we can observe that an optimum system efficiency of 75.4 % can be reached for a $t_{rf} = 85.5\,\mathrm{ps}$. All other results using the power model equations are provided in Table 6.3. In Fig. 6.4b the resulting powers are graphed vs. amplitude, for which we can evaluate the average efficiencies. Assuming a complex modulated signal (with a uniform phase distribution, such as OFDM) with a PAPR of 6 dB, we can find its distribution to have a variance of 0.3544 (Eq. (6.33)). The remaining tail of the distribution that exceeds the range of ρ can be found by evaluating the complementary cumulative distribution function (CCDF). This tells us that 1.87 % of the signal would be hard-clipped to $\rho = 1$, which, for simplicity, we simply ignore. Its probability is quite small, and it would give an (ever so slightly) too optimistic value for the efficiency if we had this remainder to contribute beyond the highest efficiency points of the DTX. This means that evaluating $P_{\mathrm{RFout,avg}}$ gives 4.61 W, even though we would expect 4.98 W provided an 6 dB PAPR signal in this configuration. This then results in an $\eta_{D,\mathrm{avg}}/\eta_{S,\mathrm{avg}}$ of 42.5 %/40.6 %, suggesting it would already beat the typical commercial analog Doherty transmitter line-up as described in Fig. 1.4b, even though this is only a single line-up (!), although idealized. For example, no additional matching losses are assumed. However, the estimated peak drain efficiency is not unrealistic. It's close to the results reported for on-wafer load-pull measurements of this LDMOS technology at these frequencies.

With this model in place, we can now reverse the equations to reiterate the question of the maximum possible frequency of operation of a DTX. This was briefly handled in the conclusion of Chapter 4, which suggested that this type of driver could maximally operate at 4.3 GHz, assuming $t_{rf}f_0$ does not exceed 25 %. Assuming $d = t_{rf}f_0 = 25\,\%$, we can sweep the operating frequency and observe the resultant system efficiency, which is plotted in
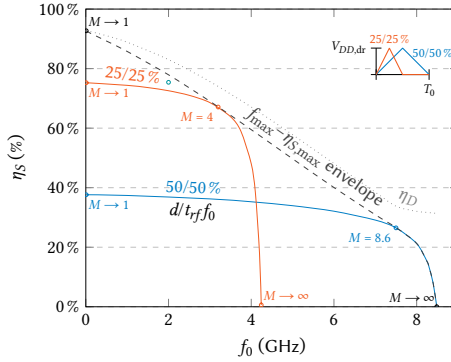
Figure 6.5: Maximum system efficiency vs. DTX operating frequency for the provided technology parameters (thick oxide 40 nm CMOS + 400 nm LDMOS in a polar operation, from Table 6.2) when varying duty cycle and rise and fall times.

Fig. 6.5. We can indeed observe that the very maximum frequency is 4.2 GHz, according to our power model. A more realistic value, assuming the driver chain may not multiply the switched input capacitance loss (M) by more than 4×, was set at 3.2 GHz, which is annotated by an orange circle in Fig. 6.5.

However, assuming a duty cycle and relative rise/fall time of 25 % is not the limit in achieving high frequency performance. We could also consider the (limit) situation of using 50 % for both. In doing so, we find that the maximum frequency of such a DTX (with the provided technology parameters) can 'operate' at most at a frequency of 8.5 GHz. We can find an 'envelope' of maximum peak system efficiencies versus possible operating frequencies, which is shown dashed in Fig. 6.5. The result of Table 6.3 is indicated as a green circle, indeed falling within this envelope. Going towards very low frequencies, we can observe that the model predicts a maximum system efficiency of 93 %, which also shows the limitations of the model: this efficiency is dominated by the LDMOS R_{ON} (i.e., $V_{DS,\mathrm{max}}/V_{DD,\mathrm{RF}}$), while the 'optimum' d and $t_{rf}f_0$ tend to zero. However, we could consider increasing R_L beyond $R_{L,\mathrm{opt}}$ to force the DTX into compression, increasing the efficiency, as well as to consider other operating classes (such as class-D, D^{-1}, E, or F) that can provide higher efficiencies, especially at lower frequencies.

6.4.2 Two-Way Doherty DTX

As a second example of the DTX power model, we will assume the use of the modified LDMOS technology of Chapter 8, which, compared to the V_T–shifted RF LDMOS of Chapter 7, has a thinner gate oxide. Its I_{DS} and g_m behaviors vs. V_{GS} are shown in Fig. 8.12.

Furthermore, to control the LDMOS segments, we assume stacked drivers with a 2.2 V supply voltage implemented in 40 nm TSMC CMOS technology. Table 6.4 gives the technology input parameters used in this example. A DTX with 40 W peak RF output power operating at 1.8 GHz is targeted. A symmetrical 2-way Doherty DTX configuration is selected to improve efficiency. Its branches use bank-sharing in combination with 8-phase multi-phase upconversion. To be realistic, we include a 0.3 dB matching loss for the DTX output power combiner. A duty cycle of 25 % is chosen to allow a straightforward generation of the required clock phases. To create these clocks, we assume $P_{DD,\mathrm{core}}$ = 273 mW.

Table 6.4: Assumed technology and model parameters for the 2-way Doherty DTX example, using a custom thin oxide 400 nm LDMOS technology (see Fig. 8.12), driven by stacked core oxide low-V_T 40 nm CMOS (from Table 4.3).

Specifications		LDMOS Parameters			CMOS Parameters		
	Main & Peak		Main & Peak			Main & Peak	
f_0	1.80 GHz	$V_{DD,RF}$	28	V	$V_{DD,dr}$	2.2	V
$P_{RFout,target}$	39.7 W	$\hat{R}_{ON}$	12	Ω mm	$t_{p0,s}$	16.10	ps
Matching Loss	0.30 dB	$\hat{I}_{DS,max}$	0.23	A mm^{-1}	$t_{peq,s}$	31.58	ps
	8-phase multi-phase	@V_{GS}	2.2	V	γ_s	1.006	
Upconversion	($\phi_{AB} = \pi/4$)	$\hat{C}_{GS}$	1.90	pF mm^{-1}	ς_s	0.150	
	Bank Sharing	$\hat{C}_{GD}$	15.5	fF mm^{-1}	$r_{rf/peq,0-100}$	1.801	
$P_{DD,core}$	0.273 W	$\hat{C}_{DS}$	0.326	pF mm^{-1}	$t_{p0,c}$	3.805	ps
		$\hat{R}_{DS}$	6.0	Ω mm	γ_c	1.458	
U_{bank}	0.924 A A^{-1}	$V_{DS,max}$	25.2	V	ς_c	0.376	
$F_{up,avg}$	0.948 A A^{-1}	$\hat{P}_{classB}$	1.447	W mm^{-1}	Interconnect Parameters		
$F_{up,avg,dr}$	0.948 A A^{-1}	$\hat{C}_{GG}$	2.270	pF mm^{-1}	$\hat{C}_{intcon}$	0.77	pF mm^{-1}
		$\hat{C}_{DD}$	0.342	pF mm^{-1}			

Table 6.5: DTX power model results for the provided values (sym. 2-way Doherty, see Table 6.4) for the main and peaking DTX branches separately, and the model averages for both branches together.

Variable Model Inputs		
	Normalized	Real
d	25 %	25 %
$t_{rf}f_0$	10.5 %	10.5 %
ρ	0.5	1

LDMOS Model Parameters		
	Normalized	Real
$W_{G,tot}$	33.59 mm	16.80 mm
$C_{L,tot}$	102.12 pF	51.06 pF
C_{DS}	10.951 pF	5.475 pF
R_{DS}	0.179 Ω	0.357 Ω
P_{loss}	0.867 W	0.434 W
$R_{L,opt}$	7.975 Ω	15.951 Ω
Q	0.988	0.988
BW$_{-1dB}$	0.927 GHz	0.927 GHz

CMOS Model Parameters		
	Per DTX	Total
t_{rf}	58.33 ps	
$t_{p,tot}$	32.39 ps	
$t_{p,c}^{(min\,P)}$	22.56 ps	
$t_{p,s}$	29.00 ps	
R_{dr}	0.323 Ω	0.161 Ω
f_c	3.937	
M_s	1.199	
M_c	1.837	
$M_{P,tot}$	2.747	

Estimated phase-averaged DTX performance at peak power		
	Per DTX	Total
P_{RFout}	19.42 W	36.24 W
$P_{DD,RF}$	26.35 W	52.70 W
η_D	73.68 %	68.76 %
$P_{dr,tot}$	0.433 W	0.867 W
η_T	72.49 %	67.65 %
$P_{DD,dr}$	1.191 W	2.381 W
η_S	70.50 %	65.47 %

Estimated DTX performance with OFDM signals	
PAPR	8 dB
σ	0.2815
$1 - F_{Ray}(1, \sigma)$	0.18 %
$P_{RFout,avg}$	5.74 W
$P_{DD,RF,avg}$	10.91 W
$P_{DD,dr,avg}$	0.493 W
$P_{DD,core,avg}$	0.273 W
$\eta_{D,avg}$	52.6 %
$\eta_{S,avg}$	49.2 %

6

Figure 6.6: The resulting powers vs. amplitude ρ (using the model calculations from Table 6.5) and the maximum DTX system efficiency vs. operating frequency, for provided technology parameters (stacked core oxide lvt 40 nm CMOS + thin oxide 400 nm LDMOS in an 8-phase multi-phase operation, from Table 6.4).
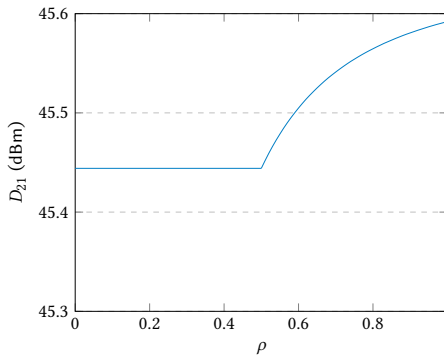
Figure 6.7: Transfer of the 2-way Doherty DTX using the power model with drain losses. In analog Doherty PAs, this effect is also visible as gain expansion due to the device losses at their drains.

Again, we can evaluate the power model equations for each branch (see Table 6.5). The resulting powers vs. amplitude are given in Fig. 6.6a. Finding the maximum possible operating frequency is also repeated for this Doherty DTX (Fig. 6.6b, where also a green circle is placed for the result of Table 6.5). Compared to the previous calculation example, the maximum operating frequency has improved to 12.6 GHz since this Doherty DTX uses a more advanced driver topology. Setting the duty cycle and the driver's relative rise and fall times to 25 %/25 %, we can now find a maximum frequency of 7.8 GHz, and for 50 %/50 % a maximum of 11.8 GHz. A maximum of 12.6 GHz can be found using $t_{rf}f_0$ = 42 %, which provides 15 % more fundamental output current compared to $t_{rf}f_0$ = 50 %, as such decreasing the required $W_{G,\text{tot}}$. The fact that the highest possible operating frequency can here be found for a faster driver shows that, in this Doherty DTX scenario, the output losses P_{loss} due to the R_{DS} of the power devices are limiting compared to the CMOS driver, although it may be an artifact of using a too simplistic loss model for our power model (Eq. (6.17)) for
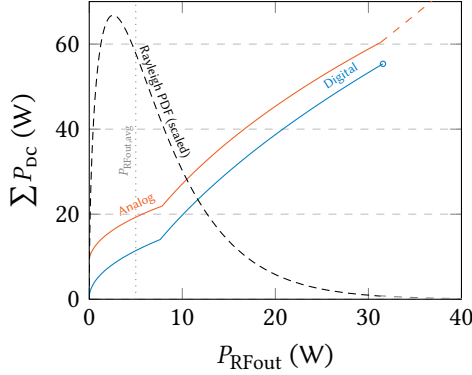
Figure 6.8: Comparing the full DC power consumptions of an analog transmitter (case from Fig. 1.4b) to the 2-way Doherty DTX example with assumed matching and circulator loss.

such a high frequency.

Further, when evaluating the average efficiencies, additional attention should be paid to the DTXs' transfers when integrating over the signal amplitude. Since now both the main and peak DTX branches introduce losses, the transfer becomes nonlinear, as seen in Fig. 6.7. Note that this effect is also present in any (analog) Doherty implementation. Although the effect here is moderate, with increasing losses or frequencies this effect may no longer be negligible. Hence, the PDF should not be taken with respect to the input amplitude, but rather to the output amplitude. This could be regarded as look-up table (LUT)-based calibration, or 'static DPD,' which is now included in the model average of Table 6.5. We can now find the average system efficiency to be 49.2 %, which includes full upconversion and phase modulation.

6.4.3 Comparison to an Analog TX Line-Up

We can use the Doherty DTX calculation above in a comparison to the example state-of-the-art analog TX discussed in Chapter 1 (Fig. 1.4b). To make this a more honest comparison, we must include 0.6 dB of additional loss in the DTX case to include the effects of the circulator. We compare their performance by plotting the total DC input to the system versus the RF output power, as done in Fig. 6.8. It shows that the reduced drive power due to gate segmentation pays off, especially at lower output power levels. Evaluating for system efficiency provides a value of 25.8 % for the analog TX, assuming a peak η_D of 68 %, while it is allowed to be driven into compression over the remaining 0.18 % of the CCDF of the Rayleigh distribution. For the 2-way Doherty DTX, we can find $\eta_{S,\mathrm{avg}}$ = 42.8 %, which now includes the circulator loss. The continuous power consumption of the DTX is much lower since the power-hungry analog pre-driver(s) are replaced by digital drivers that scale with output magnitude, and (drain) quiescent currents are eliminated.

6.5 Conclusion

With the proposed DTX power model, we can perform simple (hand) calculations to estimate the expected system efficiencies for a given CMOS driver and RF power technology. In a calculation example, a 2-way Doherty DTX was compared to a modern mMIMO TX line-up

using also a Doherty PA. It shows that the scaling of the input DTX drive power with the RF output magnitude and eliminating any (drain) quiescent currents can give better performance in terms of power consumption.

Understanding the power relations in a DTX allows for optimizing designs, as well as quantifying what is possible with a given technology in a DTX application. The CMOS controller clearly benefits from a lower intrinsic propagation delay t_{p0}, which is the technology constant that relates driver output capacitance to its equivalent switch resistance. This is typically the key focus of any improvement in CMOS technology (i.e., Moore's law), and as such, it is an extra motivation for going digital with TXs in general. From an RF power technology perspective, we prefer to have the ratio $I_{DS}/C_{GG}V_{DD,\mathrm{dr}}^2$ as high as possible (preferably with $I_{DS} = I_{DS,\mathrm{max}}$ evaluated at $V_{GS} = V_{DD,\mathrm{dr}}$). Addressing these technology aspects will boost the DTX system performance, both in operating frequency and efficiency.

6

7

The Proof-of-Concept for High-Power DTXs

Previous chapters have elaborated on the topology for high-power DTXs, how to model them for simulation, and how to interpret their simulation results. That discussion was still conceptual, a 'prerequisite,' so to speak. The next challenge is to actually fabricate DTX demonstrators as a proof-of-concept, showing that high RF output power with a digital transmitter is indeed feasible.

This chapter discusses the design of these demonstrators. Three different demonstrators are fabricated, each focusing on a different performance aspect, hence having unique output impedance matching networks or power combiner(s). However, they use very similar LDMOS or GaN output ICs and make use of the same CMOS DTX controller IC at their core. Since designing large, high-performance CMOS ICs is costly and time-consuming, the DTX controller was designed to be as universal in its use as possible to support various high-power DTX implementations. Section 7.1 first provides the aimed functionality and requirements for the demonstrator. Next, the design of the LDMOS and GaN output stages is discussed in Sections 7.2 and 7.3, respectively. The design details of the CMOS controller are provided in Section 7.4. Together, they form the basis for the three high-power digital transmitter demonstrators.

Next, in Sections 7.5 through 7.7, the three demonstrators' design aspects and their measurement results are highlighted. The first design, discussed in Section 7.5, aims to show

---

Parts of this chapter are based on published works:

[79]: R.J. Bootsman, D.P.N. Mul *et al.*, "An 18.5 W Fully-Digital Transmitter with 60.4 % Peak System Efficiency," *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 1113–1116, doi: 10.1109/IMS30576.2020.9223942.

[20]: D.P.N. Mul, R.J. Bootsman *et al.*, "Efficiency and Linearity of Digital 'class-C Like' Transmitters," *2020 50th European Microwave Conference (EuMC)*, Utrecht, Netherlands, 2021, pp. 1–4, doi: 10.23919/EuMC48046.2021.9338122.

[17]: R.J. Bootsman, D.P.N. Mul *et al.*, "High-Power Digital Transmitters for Wireless Infrastructure Applications (A Feasibility Study)," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 5, pp. 2835–2850, May 2022, doi: 10.1109/TMTT.2022.3153000.

[80]: R.J. Bootsman, Y. Shen *et al.*, "A 39 W Fully Digital Wideband Inverted Doherty Transmitter," *2022 IEEE/MTT-S International Microwave Symposium (IMS)*, Denver, CO, USA, 2022, pp. 979–982, doi: 10.1109/IMS37962.2022.9865405.

high peak output power with high drain and system efficiency. It serves as the first proof-of-concept for high-power DTX, employing a class-BE output match at 2.1 GHz. Section 7.6 demonstrates the second design. This DTX design aims towards efficient operation while supporting wideband modulation. For that reason multi-phase operation with a digital class-B output match at 1 GHz is selected. Digital class-C operation can also be enabled, with this demonstrator, by reducing the duty cycle of the modulating RF clock. The final DTX design, using this chip set, aims for improved average efficiency over a large RF bandwidth. This design is discussed in Section 7.7, showing a DTX employing digital class-C operation with an inverted Doherty power combiner, having a center frequency of 2.0 GHz.

7.1 Aimed Functionality and Requirements

Since the realization of a power-DTX prototype is costly, our aimed demonstrator must support various DTX operation conditions for testing purposes. So, its architecture is intended to support polar, signed Cartesian, and 8-phase multi-phase operation, of which the latter two are implemented as separate banks for the *IQ* or *AB* vectors, and thus demand two separate baseband-to-RF upconverting TX line-ups, which is conceptually shown in Fig. 7.1. This dual TX line-up topology can also be used to support testing of two-way Doherty, out-phasing, or dual carrier (carrier aggregation) prototypes in the future. Consequently, the output stage segments are grouped in two independently controlled switch banks. Furthermore, the output stage(s) of the power-DTX must be flexible in their operating class and compatible with a rectangular (digital) drive signal. The drive signal is generated in a CMOS controller, which is implemented in TSMC 40 nm technology. To relax somewhat the V_T requirements for the custom low-V_T LDMOS (Section 3.1.1), thick-oxide CMOS devices are selected to implement 2.5 V drivers for the output stage segments and their tapered buffer chains (see Chapter 4). Logically, these devices are slower than core oxide devices, which we use in the control logic.

For the output stage, two technologies are available: a 0.4 μm silicon LDMOS process by Ampleon, and a 0.25 μm GaN on SiC process by Fraunhofer IAF. The nominal supply voltage $V_{DD,RF}$ of the LDMOS technology is 28 V and a breakdown voltage $BV_{DSS} > 72$ V, whereas the supply for the GaN technology is 40 V and $BV_{DSS} > 150$ V. More parameters relevant to the design are provided in Sections 7.2 and 7.3 for the two respective technologies.

The standard LDMOS die size for power devices is 4.9 mm, which we adopt as the
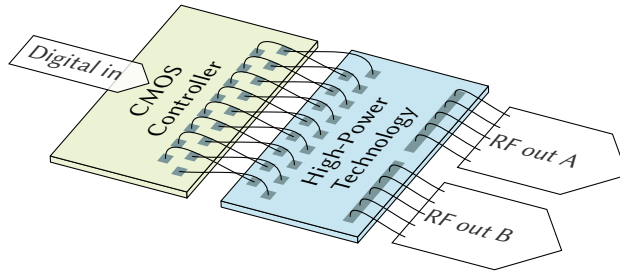
Figure 7.1: Conceptual diagram of the proposed RF high-power mixing-DAC configuration with a dual TX line-up topology using a CMOS controller and a gate-segmented high-power output stage.

maximum power die width in this work. The number of interconnections between the CMOS and power dies has been maximized to allow as many DTX segments as possible, to reach maximum DTX resolution and dynamic range. This is achieved by using the minimum available bond-wire pitch (see Section 3.2.1) of 80 μm, in combination with staggered bond pads and 25 μm diameter gold bond-wires with special ball–stitch-on-ball bonding to minimize bond-wire loop height (see also Section 7.2.2).

Wedge–wedge bonds with 50 μm diameter aluminum wires with a definable loop shape are used at the output of the power die to achieve low output losses and a predictable connection to the PCB output matching network. This requires a minimum bond-wire pitch for these aluminum wires at the output of 130 μm and a bond bar on the power die of at least 167 μm wide. The minimum trace width on the PCB is 100 μm for a 35 μm thick copper layer. Finally, to make our feasibility study of interest to future DTX based mMIMO base stations, the power-DTX prototype should operate up to at least 3 GHz transmit frequency while being capable of delivering at least 20 W of peak RF output power with high system efficiency.

7.2 LDMOS Implementation

LDMOS technologies can offer some flexibility for lowering the threshold voltage, as discussed in Section 3.1.1. Therefore, in this work, we use a 0.4 μm 28 V LDMOS technology, combined with a high-speed digital controller based on 40 nm CMOS technology. The downshifting of the LDMOS V_T can be done by selecting different doping concentrations and/or using thinner gate oxides. This is a delicate process since various other performance parameters like ruggedness need to stay satisfied. Furthermore, when shifting the V_T down, the LDMOS process technology is no longer calibrated, so there will be uncertainty in the actual realized V_T.

Figure 7.2 provides the drain current (I_{DS}) and transconductance (g_m) vs. gate voltage (V_{GS}) for the considered LDMOS technology. We observe that setting the V_T close to 0 V allows the device to switch entirely 'on' and 'off' using a 2.5 V driver signal. However, choosing a very low value for the V_T causes a drop in R_{OFF} (Fig. 7.3), which also affects the peak drain efficiency. This is even more prominent in power back-off operation when a part of the DTX segments will be deactivated. Consequently, the optimum LDMOS V_T for a power-DTX implementation ranges from 0.4 V–1.4 V. For now, we will assume $V_T = 0.8$ V. When the details of the power-DTX line-up have been determined, this V_T choice is reevaluated in Section 7.5.1.

When assuming a $V_T = 0.8$ V and a drive signal swing of 2.5 V, the modified LDMOS technology has an $I_{DS,max} \approx 0.17$ A mm^{-1} (Fig. 7.2). This current density is somewhat limited by the lower voltage swing from the CMOS driver. Furthermore, in class-E operation, the supply voltage needs to be lowered to 20 V to avoid breakdown, lowering the output power capability. Due to requirements on the effective R_{ON} and to allow for some design flexibility (e.g., use of Cartesian operation, requiring I and Q banks, or flexibility to use operating classes with lower output power capability, such as digital class-C, in later implementations), we assume for 20 W RF output power a doubling of the minimum bank size with respect to an analog class-B implementation at $V_{DD,RF} = 28$ V, giving a minimum LDMOS total gate width $W_{G,tot} \geq 33.6$ mm, which is split over the two banks (Fig. 7.4).
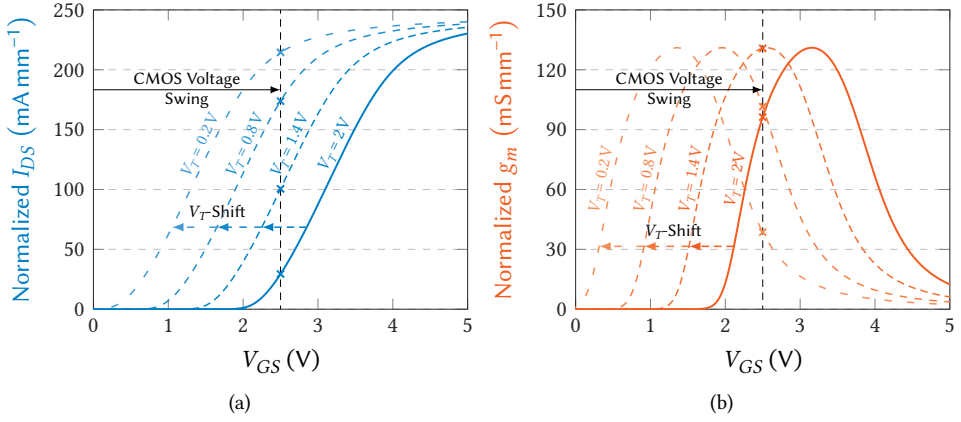
Figure 7.2: V_{GS}–I_{DS} and $-g_m$ curves for the LDMOS process when V_{DS} = 28 V, while varying the V_T.
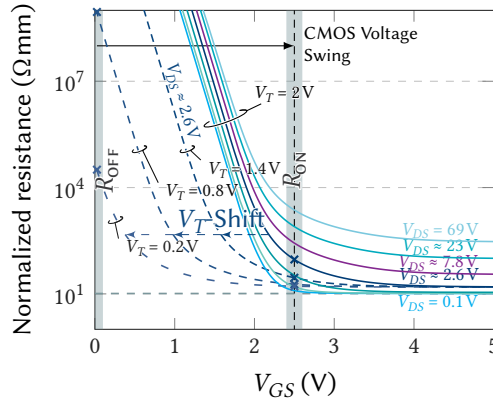
Figure 7.3: Modeled drain ON/OFF resistance shown versus V_{GS} for different values of V_{DS}.

7.2.1 Unary and Binary Weighted Segments

The maximum number of segments is limited by the number of interconnections between the CMOS controller and the LDMOS output stage. Naive division (i.e., neglecting things such as pad sizes and bank spacing) of the available die size of 4900 μm by the minimum bond-wire pitch of 80 μm yields 62 interconnections in total, so 31 for a single bank. Some of these interconnections are reserved for ground return paths (see also Sections 3.2.1 and 7.2.2), leaving 23 per bank available for LDMOS segments. Fully thermometer coding would yield a too small dynamic range for the DTX to sufficiently handle modulated signals. Therefore, a hybrid approach was selected, resulting in 15 unary-weighted MSB segments (4-bit resolution) and 7 binary-weighted LSB segments for each bank (see also Section 5.1.1). The size of the least significant binary bit is limited by the minimum W_G of the LDMOS device, being 2 × 5.1 μm. This least significant binary-weighted segment was doubled to make use of the remaining space, allowing some extra redundancy in testing,

(a)

(b)
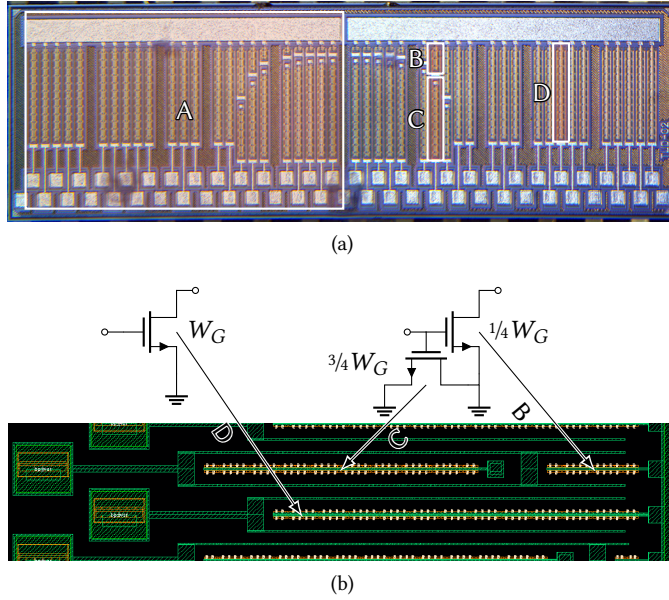
Figure 7.4: Layouts of the segmented LDMOS power die. (a) Micrograph of the die ($4.9 \times 1.6\,\mathrm{mm}^2$) having a total gate width ($W_{G,\mathrm{tot}}$) of 41.472 mm distributed over two switch banks (the surrounding box indicates the left switch bank A), each bank featuring 7 binary-weighted segments, (B) indicates such a binary segment with dummy device (C), and 15 unary segments representing 4-bit, denoted by D. (b) Layout detail showing the use of a dummy LDMOS device (C) for equalizing $C_{\mathrm{in}} = C_{\mathrm{ESD2}} + C_{GS} + C_{GG,\mathrm{dummy}}$. Here the 2$^{\mathrm{nd}}$ MSB binary segment is shown, and below a unary segment.

yielding an overall segment (bond) pitch of 83.5 µm. The above makes the summed gate width of all binary-weighted segments equal to that of a single unary segment. Therefore, $W_{G,\mathrm{tot}} = 2 \times 16\,W_G$, with W_G being the gate width of a unary segment. To comply with the output power requirement $W_G \geq 1.05\,\mathrm{mm}$, which for this LDMOS technology is spread over two gate fingers per segment.

Aside from the LDMOS gate capacitance, the loading capacitance for an LDMOS segment driver also consists of ESD protection diodes on both dies. To reduce the disproportionate ESD contribution and to have a W_G that is decently divisible by 2, the W_G was increased to 1.296 mm, yielding a unary segment loading capacitance of $C_L = C_{\mathrm{ESD1}} + C_{\mathrm{ESD2}} + C_{GS} \approx 0.81\,\mathrm{pF} + 0.50\,\mathrm{pF} + 1.10\,\mathrm{pF} = 2.41\,\mathrm{pF}$ and $W_{G,\mathrm{tot}} = 41.472\,\mathrm{mm}$. This gives a total LDMOS output capacitance of $C_{DS} \approx 12.3\,\mathrm{pF}$ when $V_{DS} = 28\,\mathrm{V}$. The use of binary-weighted segments raises the issue that C_{GS} reduces by a factor of 2 with each binary step. To avoid delay mismatches caused by different loading of the segment drivers, the lower C_L of these binary segments is compensated by adding dummy LDMOS devices (Fig. 7.4b), with both source and drain tied to ground. Following this strategy, the total driver-connected W_G of a binary segment is targeted to be equal to that of a unary segment. As the gate capacitance is nonlinear with drain voltage, the load capacitance will not match perfectly. Implementing an equivalent (more linear) metal–insulator–metal (MIM) capacitance was briefly considered; however, this would yield unknown matching with process variations and a larger mismatch overall than accepting the mismatch from the grounded drain LDMOS dummies.
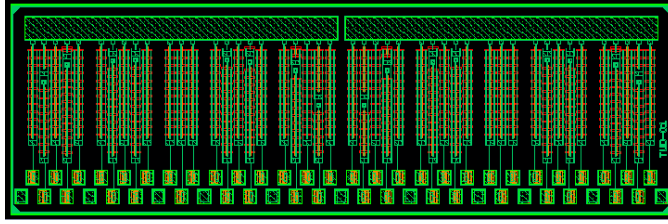
(a) TUD-01: distributed LSBs
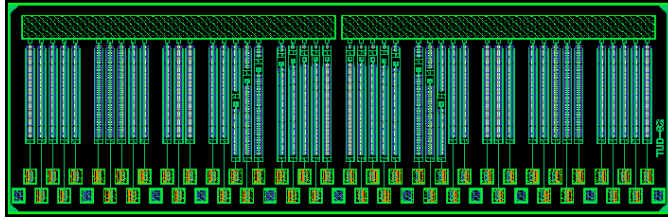
(b) TUD-02: inside LSBs
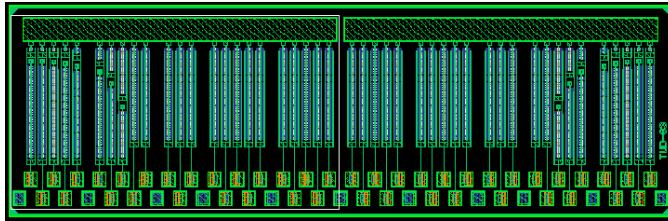
(c) TUD-03: outside LSBs

Figure 7.5: Variations on the LDMOS layout, where the position of the LSBs is varied with respect to the MSBs.

The physical location of the LSBs can also impact the performance of the DTX. First, the series inductance of gate metallization and the drain runner should be considered, but no device models capable of simulating these effects were available. Instead, the effect of series inductance in the drain path is assumed to be more important for DTX performance than the effects of additional series inductance to the gate connection. Especially since the dummy segments' capacitance make the interconnects more capacitive, lowering the lines' equivalent characteristic impedance. Hence, the active part of the LSBs is placed closest to the drain bar output, which also makes shorting the dummy drain and routing the active drain to the output easier.

The position of the LSBs along the drain bar can also matter, but during the design phase it is unknown which aspect is most important. To be effective, three different LDMOS layout variations have been taped out, which are shown in Fig. 7.5. The first variant (Fig. 7.5a) spreads the LSBs uniformly among the MSBs, aiming to minimize output matching network current redistribution effects. The next variants (Figs. 7.5b and 7.5c) are devised with the input bond wires in mind. These bond wires couple together, causing the middle bond wires to have a lower self-inductance than the outer ones. However, the middle bond wires might be more influenced by all surrounding wires than the outer ones. The second variant
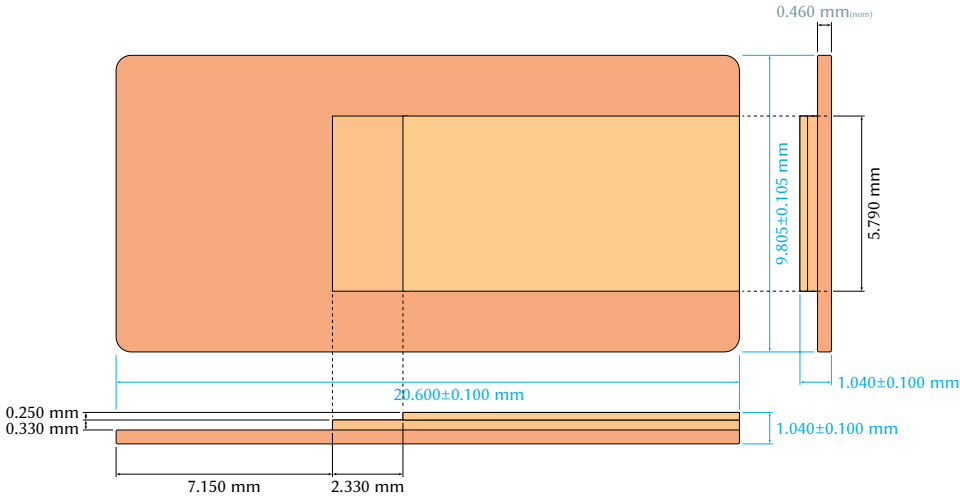
Figure 7.6: The flange as used in a SOT1275-1 package, modified to accommodate different die thicknesses.

aims to have minimal self and mutual inductances for the LSBs, whereas the last variant aims for the lowest interaction between all MSBs and the LSBs.

7.2.2 Assembly of the Demonstrator

The bond wires between the CMOS controller die and the LDMOS power die should be as short as possible to minimize parasitic series inductance. The different thicknesses of the two dies pose a challenge for that. Namely, the thickness of the CMOS controller IC is 300 µm, while the LDMOS die has been thinned to 50 µm to minimize source inductance. The LDMOS die typically has to be attached to a metal flange that serves both as a ground plane and a low thermal resistance (see Section 3.2.2). To accommodate the different die thicknesses, a standard copper flange used in a SOT1275-1 package is modified by milling a recess where the PCB and CMOS die are planned. The flange dimensions are given in Fig. 7.6. The highest plateau is where the LDMOS die will be attached and has to remain unscathed. Namely, the flange has a gold metal finish, allowing for a good thermal and electrical connection using either gold eutectic die attach or silver epoxy glue.

An artist's impression of the assembly (targeting demonstrator I, Section 7.5) is shown in Fig. 7.7. The CMOS controller die's surface is at the same height as the LDMOS power die's surface, as well as the metal surface of the first PCB (green). A cavity is made in this PCB to accommodate the flange, where sidewall metallization is applied to have a good ground available in the top PCB metal, which can then be connected to the CMOS controller. The output PCB (shown in white) is placed on top as a separate PCB to allow for different output-matching network designs while reusing the first PCB's design. To ensure a good electrical connection to the output PCB, the die-to-flange orientation has been rotated 90° with respect to how this flange is typically used. Namely, the resulting top plateau is longer, providing more area for soldering the top PCB to the flange. An integrated passive device (IPD) is placed in the DC path of the output matching network to provide an RF (and baseband) short for the integrated shunt inductor (inshin) wire bonds.
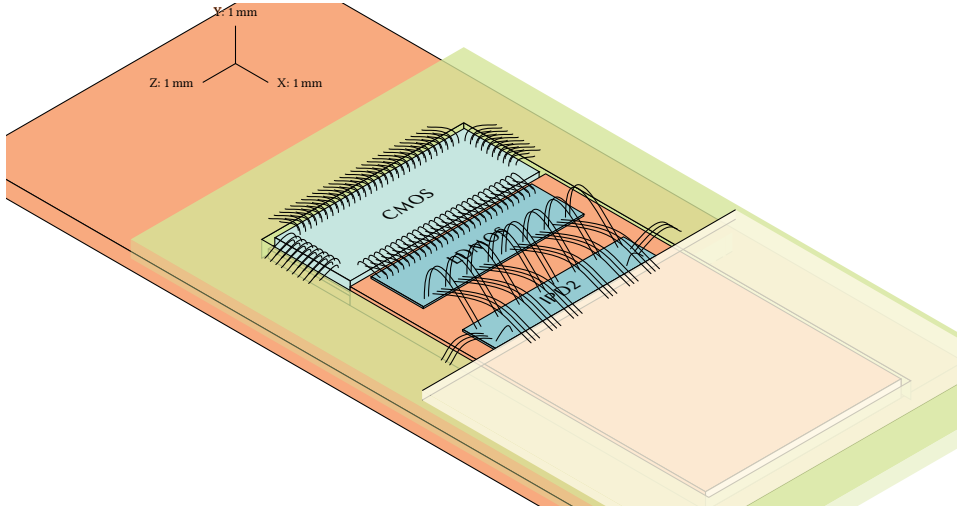
Figure 7.7: Artist's impression of the assembly.

A third PCB (not shown in Fig. 7.7) with a cavity for the entire flange assembly is added underneath for mechanical stability (e.g., for board connectors) and to bridge the distance to an aluminum heat sink.

Special attention is given to the bond wires connecting the CMOS controller and LDMOS power die. These have been EM simulated stand-alone to evaluate their self-inductance and mutual coupling. A 3D view of the simulation setup is shown in Fig. 7.8. The dies can be placed close together since the bonding surfaces are at the same height, which minimizes the bond wire distance. The loop height of the bond wires (combined with the short bond wire distance) can be minimized using special ball–stitch-on-ball bonding. Staggered bond pads are used to realize the minimum bonding pitch of 80 μm, which causes concern for mutual coupling between the bond wires. To minimize this coupling, the staggered bond wires are placed in alternating directions, minimizing the overlapping loop area.

Additionally, grounded pads were placed on the LDMOS power die to provide current return paths to the CMOS drivers through bond wires. These grounded pads can also be identified in Fig. 7.5 on the bottom row of pads without LDMOS segment attached to them. Fig. 7.8b shows the self-inductance values from the EM simulation. Only half the bond wire array could be simulated due to limitations in memory and computation power, but it still provides some insight into the distribution of self-inductance and coupling values. On average, the self-inductance of each bond wire is found to be 563 pH, where the wires on the outside show higher values than those on the inside. The coupling values for directly neighboring wires (opposite bonding directions) are 0.49, every other wire (same bonding directions, not including ground) is 0.45, and those with grounds in between them on average 0.40. The resulting *S*-parameter model is used to simulate quasi-static performance, which is discussed together with the measurements in Section 7.5.2.

(a)

(b)
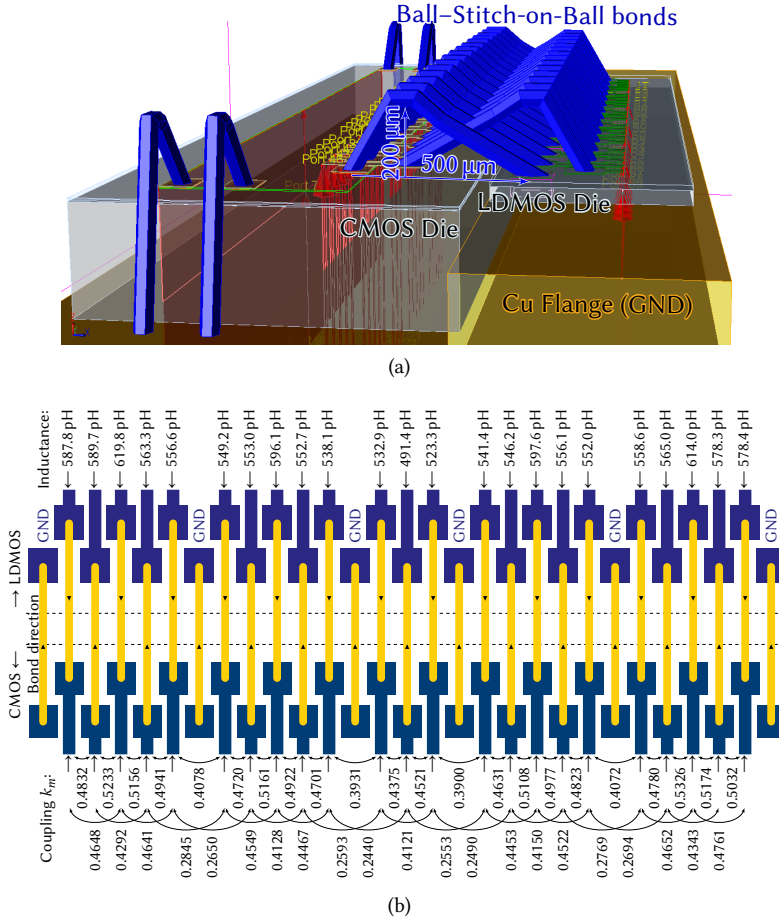
Figure 7.8: 3D view of the FEM simulation setup of the bond wires between the CMOS controller and LDMOS power die, and the found self inductance and coupling values.

7.3 GaN Implementation

The negative threshold voltage of the available GaN technology poses the main challenge for implementing a power DTX. The DC level of the drive signal has to be shifted down, somehow, since the CMOS controller targets a (positive) swing from 0 V to 2.5 V. For that purpose, inspiration is taken from passive oscilloscope probes. These use a passive *RC* structure similar to the schematic shown in Fig. 7.9a, such that the time constants of the two parallel *RC* combinations are identical. There is a constant voltage transfer over frequency (in both magnitude and phase) this way. A similar (but reversed) all-pass structure can apply a DC shift to the GaN gates, at the cost of some drive signal attenuation.

The resulting structure with (nominal) values is shown in Fig. 7.9b. The GaN gate capacitance forms the shunt capacitance and the bias resistor forms the shunt resistance. Implementing a 5-to-1 ratio in terms of capacitance was feasible, resulting in an input peak-to-peak voltage swing of 2.08 V (factor 0.833). The GaN gate is a nonlinear capacitance,

(a) An $1 : (N+1)$ attenuator   (b)
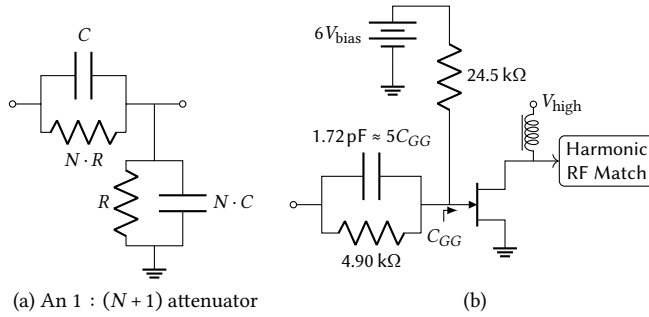
Figure 7.9: GaN input match using an RC all-pass.
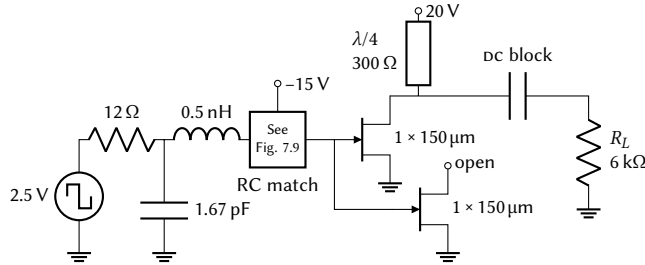
Figure 7.10: Transient GaN input simulation setup, scaling the RC input match such that no input DC drift occurs.

depending on the operating region and the voltage changes at the drain side (due to the feedback capacitance). The simulation setup shown in Fig. 7.10 mimics a simplified but typical use case to determine a suitable series capacitance. The series capacitor is implemented using a MIM capacitor made using a SiN dielectric layer. The resistors can be implemented using either a NiCr thin film resistor or the GaN epitaxial layer. The epitaxial resistor has a 10× higher sheet resistance than the thin film resistor but can handle less current and is nonlinear. Since we want as little wasted DC bias power as possible, these resistors should be large and conduct little current as a consequence. Hence, the epitaxial resistors are chosen. The nonlinear nature of these resistors is not significant for very little current. Also, the exact values of these resistors do not matter, as long as their ratio is correct.

The segmentation and assembly of the GaN output stage are very similar to the LDMOS case (Section 7.2.1). Namely, the GaN output stage is segmented into 15 unary-weighted MSBs per bank and 8 binary-weighted LSBs, including the dummy devices. These dummy GaN devices have floating drains. Only one variation could be fabricated, where the distributed LSB layout similar to LDMOS variation 01 was chosen. In terms of sizing, one unary-weighted MSB consists of 2×150 μm fingers, making the $W_{G,\text{tot}} = 9.6$ mm.

7.4 CMOS Controller Architecture

Next, we discuss the design of the CMOS controller. From this demonstrator's functionality requirements, we know it has to support polar upconversion, as well as a baseband DAC, a
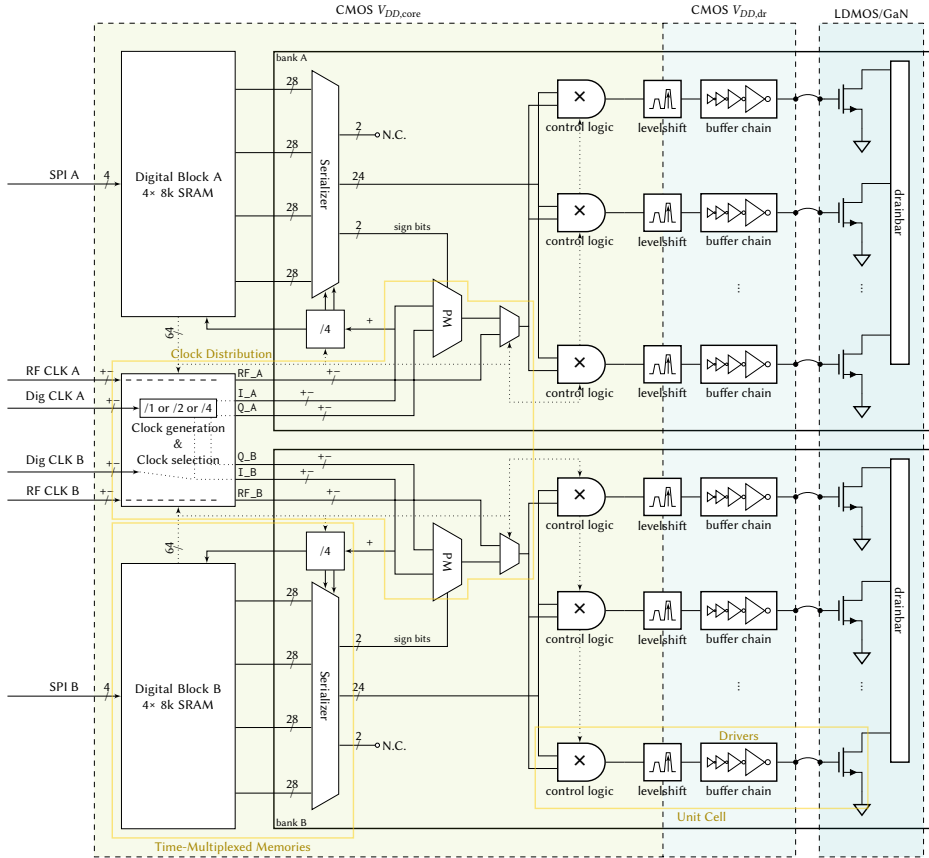
Figure 7.11: Block diagram of the CMOS controller, showing the dual line-up.

dual line-up operated in signed Cartesian and 8-phase multi-phase upconversion, or for two-way Doherty, out-phasing, push-pull, or dual carrier operation. This leads to certain design choices described per category in the next subsections.

7.4.1 Overview

The controller can be split into a couple of functional blocks: the unit cells, digital blocks with memories connected to a serializer, clock generation, and the supply distribution network. An overview of these blocks is shown in Fig. 7.11. The unit cells consist of controlling logic, a level shifter from $V_{DD,\text{core}}$ to $V_{DD,\text{dr}}$, the drivers, and an LDMOS segment.

The drivers are discussed first in Section 7.4.2 since these drivers profoundly impact the system's performance. Next, the rest of the unit cell is discussed in Section 7.4.3. This section focuses on the CMOS part, as the LDMOS segments have been discussed in the previous section. Section 7.4.4 focuses on the serializer connected to the SRAMs, enabling a sampling rate equal to the RF carrier frequency. The clock generation, division, and distribution are discussed in Section 7.4.5. The supply distribution network and capacitive decoupling are discussed in Section 7.4.6. The last remaining global features are discussed

next, as they do not require an entire section of their own.

This DTX controller has a dual line-up, meaning that it is split into two banks: bank A and bank B. These are functionally identical to each other but mirrored in layout. Each is provided with three pseudo-differential clock signals[1] and has its own digital block. The digital blocks are synthesized from an HDL description consisting of a serial peripheral interface (SPI), four parallel SRAMs, and a memory controller. This serial peripheral interface (SPI) has 49 registers of 8 bit. Eight of these registers (so 64 bit) are reserved for configuring the chip's functionality. Their addresses and purposes are provided in Table 7.1. Registers `0x19`–`0x1B` control the clock generation & selection block. Registers `0x1C`–`0x20` relate to the control logic in the unit cells or selecting clock phases for the digital sampling clocks. Since each bank has its own SPI, a total of 128 bit is available to configure the operation. The four SRAMs each have a depth of 8192 words and an IO width of 28 bit. Only 25 are needed in the design, but this width is not available from the SRAM compiler. Since it is available anyway, one additional bit is used for an additional `TRIG` output pin. This pin can be used for synchronization and testing purposes. Table 7.2 provides the use of these 26 connections. The SRAM size is selected such that their cycle time is fast enough to support a minimum operating frequency of 750 MHz, such that, after serializing four of them, a sampling frequency f_s of at least $3\,\text{GSa\,s}^{-1}$ is supported.

7

---

[1]Pseudo-differential refers to a clock signal pair consisting of a positive and negative (or inverted) polarity, originating from CMOS logic. True differential signaling uses a differential pair, actively rejecting any common mode. Here, these pseudo-differential clocks remain aligned using digital phase aligners (cross-coupled inverters) and by using symmetrical layouts.

Table 7.1: Available SPI register addresses and their purpose.

Address	Bit7 (MSB)	Bit6	Bit5	Bit4	Bit3	Bit2	Bit1	Bit0 (LSB)
0x00	←			Control Register				↑
0x01–0x02	←			SRAM write address				↑
0x03–0x04	←			SRAM→SPI read address				↑
0x05–0x14	←			SRAM 1–4 input<0:3>				↑
0x15–0x18	←			SRAM 1–4 write enable				↑
0x19	7 MemDuty<5>	6 memDuty<4>	5 MemDuty<3>	4 MemDuty<2>	3 MemDuty<1>	2 MemDuty<0>	1 E_Mem_Corr	0 S_diff_inv
0x1A	15 RFDuty<5>	14 RFDuty<4>	13 RFDuty<3>	12 RFDuty<2>	11 RFDuty<1>	10 RFDuty<0>	9 E_RF_Corr	8 ↑
0x1B	23	22	21 S_RF_Phase	20 S_RF_AB	19 S_Mem_AB	18 E_clkdiv	17 S_div2div4	16 S_clkphase
0x1C	31 E_Retime	30 E_CS	29 E_RTZ	28 S_RFCLK	27 S_DCLK<1>	26 S_DCLK<0>	25 S_digclk<1>	24 S_digclk<0>
0x1D	39	38	37	36	35	34	33 E_TRIG_Retime	32 E_TRIG_RTZ
0x1E	47 CS_06	46 CS_05	45 CS_04	44 CS_03	43 CS_02	42 CS_01	41 CS_00	40 CS_TRIG
0x1F	55 CS_14	54 CS_13	53 CS_12	52 CS_11	51 CS_10	50 CS_09	49 CS_08	48 CS_07
0x20	63 CS_22	62 CS_21	61 CS_20	60 CS_19	59 CS_18	58 CS_17	57 CS_16	56 CS_15
0x21–0x30	←			SRAM 1–4 → SPI output<0:3>				↑

Table 7.2: Connection numbering from SRAM to IO.

SRAM	27:26	25	24:2	1	0
MUX	27:26	25	24:2	1	0
Connected to	NC	Output tree	Output tree	PM +180°	PM +90°
OutChain	-	TRIG	<22:0>	-	-
Pad label	-	TRIG	AIO_ <1:23>	-	-
IO name	-	TRIG_A	RF_A_ <0:22>	-	-

7

The general floorplan is given in Fig. 7.12. The dimensions in the x-direction are set by the CMOS–LDMOS interconnect (see Section 7.2.1), the width of the LDMOS die, and some overhead needed for the CMOS IO ring. The drivers' outputs can be found on the top of the design as well as the ground return path pads. All digital inputs, clocks, and core supply voltages can be found at the bottom of the design. Having short wires for the high-frequency input clocks is more critical than for the low-frequency (< 16 MHz) SPI. Hence, these high-frequency input clocks are aligned with the traces on the PCB (typical advanced PCB processes allow a trace pitch of 200 μm). That leaves the sides available for the driver supply voltage connections. As a rule of thumb, one supply pad is suitable for 100 mA to 125 mA. As the drivers are expected to require up to 1.6 A, this necessitates 13 to 16 pads divided over the two sides. This sets the y dimension of the controller.

7
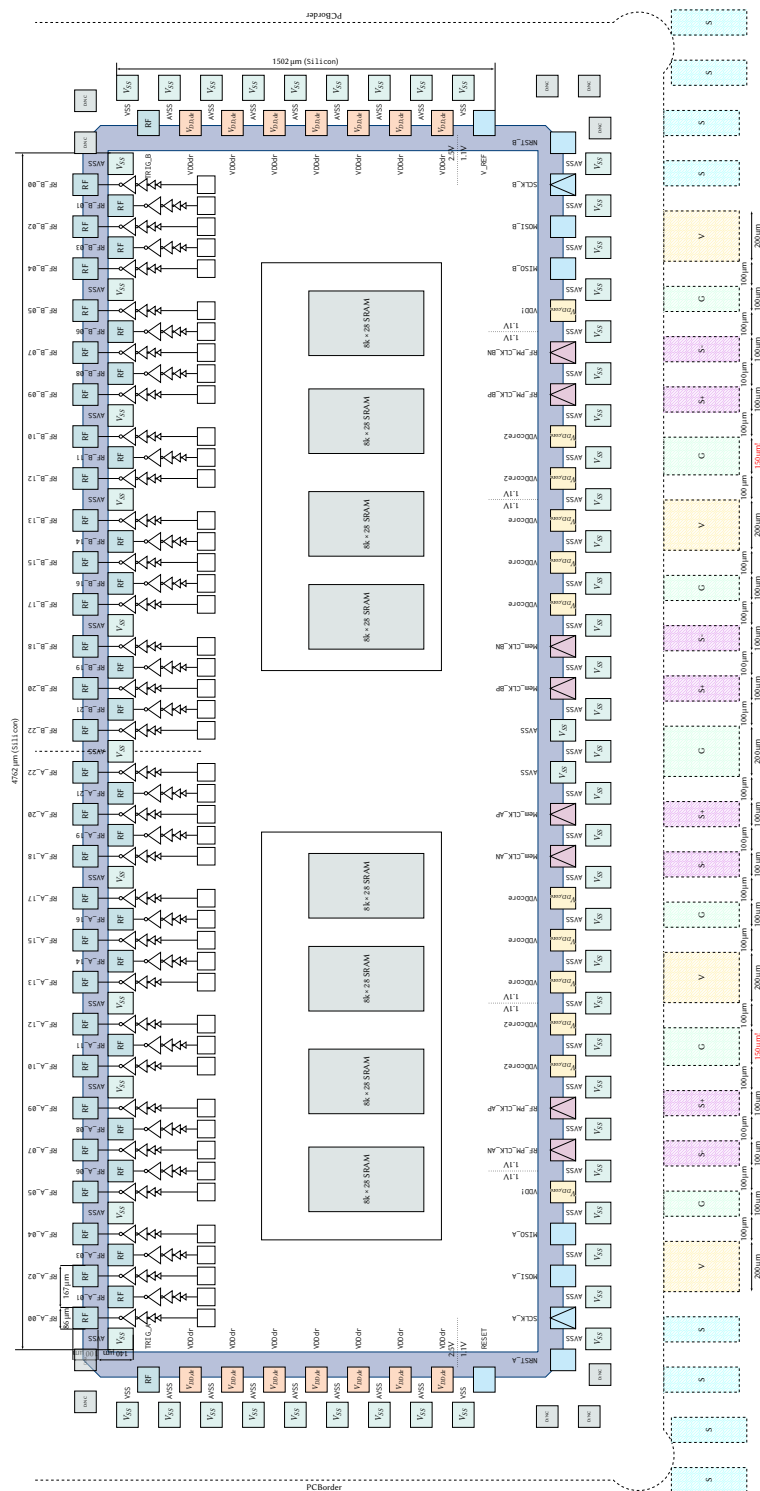
Figure 7.12: Physical input/output (IO) positioning of the CMOS controller layout. The controller's dimensions are fixed from the IO requirements.

7
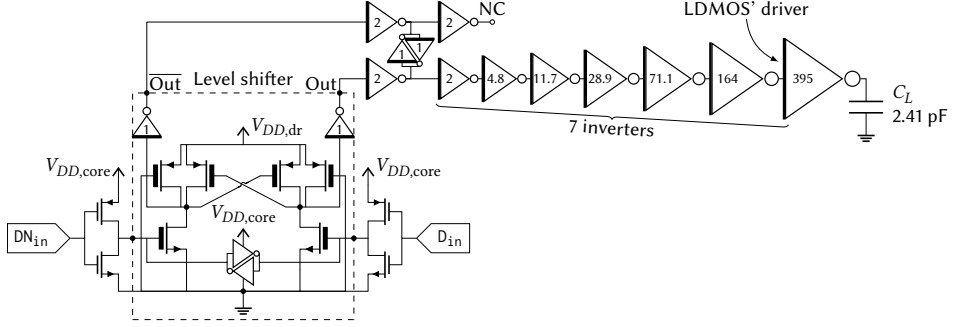
Figure 7.13: Level shifter and tapered buffer chain.

7.4.2 Drivers

From Section 7.2.1, we know that the LDMOS' driver (the final inverter of the tapered buffer chain) has to drive a segment load of $C_L = 2.41\,\mathrm{pF}$. The empirical constants extracted from the thick-oxide device models in the CMOS 40-nm technology, including layout parasitics, are found to be (see Table 4.1): $t_{p0} = 8.68\,\mathrm{ps}$, $\gamma = 0.669$ and $\varsigma = 0.262$. However, aiming to drive such a large load, additional care has to be taken to ensure low series resistances from the driver to the output and the positive and ground supply rails, and to meet current density-related metallization reliability rules. Also, the driver output must be routed from the lowest metal layers to the topmost metal layer of the IO pad. The IO pad capacitance is included in the load capacitance, but the additional metal to ensure the low series resistance increases this driver's output capacitance, impacting t_{p0}. This gives the following empirical parameters for the LDMOS' driver: $t_{p0} = 10.76\,\mathrm{ps}$, $\gamma = 1.092$ and $\varsigma = 0.264$.

When optimizing the DTX line-up for 3 GHz class-BE operation (Section 7.5.1) with realistic output losses, optimum system efficiency can be found for $R_{\mathrm{dr}} = 16.87\,\Omega$ (per segment). This requires $f = 2.88$ (Eq. (4.13)) and $N = 7.5$ for the number of driver stages, yielding a capacitance multiplication factor $M = 2.114$ (Eq. (A.49)). In the implementation a chain of 7 stages could be realized. The TRIG outputs are driven using the same drivers as used for the LDMOS gate segments.

7.4.3 Unit Cell

The LDMOS' driver and tapered buffer chain in the unit cell are preceded by a level shifter. This level shifter translates the $V_{DD,\mathrm{core}}$ voltage level used in the unit cell logic to the drivers' supply voltage $V_{DD,\mathrm{dr}}$. Its schematic is provided in Fig. 7.13. It needs to be DC-coupled to avoid voltage level drift when a segment is required to be off for an extended time. The heart of the level shifter is a cross-coupled thick-oxide PMOS latch, needing a differential input. The RF clock-tree features pseudo-differential clock lines, allowing these differential inputs to be generated.

In Fig. 7.14, the schematic of the unit cell logic is provided. In Fig. 7.14a, the clear/set logic is given, used for resetting the unit cell and for bypassing the data input D_{in} as a fallback option in case the SRAM logic does not work properly. Figure 7.14b provides the rest of the unit cell logic. First, the data input is retimed by the data clock and made differential. Since the Q and $\overline{\mathrm{Q}}$ outputs have different propagation delays, potentially impacting the RF duty cycle, this differential signal is retimed a second time. From there
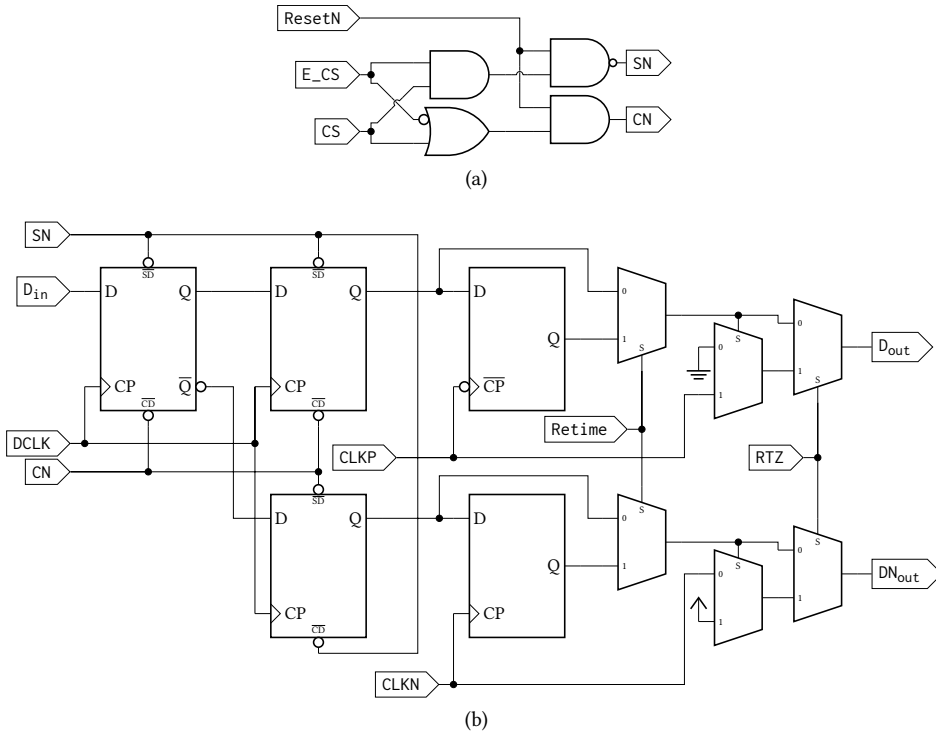
Figure 7.14: Simplified unit cell logic (buffering and delay equalization removed), connected to the differential input of the levelshifter (Fig. 7.13). (a) Unit cell reset control logic and data by-pass using a static clear/select signal (CS) with a separate enable signal (E_CS) (b) Symmetrically designed using MUX logic and a retimed differential data generator to the data clock (DCLK). Optionally, the data can also be retimed to the pseudo-differential RF clock (CLK) and the upconversion/mixing with the RF clock is controlled by the return-to-zero (RTZ) signal.

on, both signals should see identical propagation delays, so mux logic is used instead of conventional logic gates. As the DTX should support different upconversion modes (polar, signed Cartesian, etc.), another retiming step with the RF clock is optional, which is enabled by the Retime setting signal. This optional retime step can avoid glitches due to the (phase modulated) RF clock differing in timing from the ('static') data sampling clock by ensuring the 'effective' data cannot change during an activation period. Since the requirements also specify the option to operate as a baseband DAC, the upconversion must also be optional. This is signaled by the return-to-zero RTZ input.

The trigger outputs feature the same unit cell logic, but have TRIG_RTZ and TRIG_Retime inputs separate from the 'normal' unit cells (see Table 7.1). This ensures maximum flexibility in testing and synchronizing the DTX line-up. The SRAM bypass flag E_CS is universal for both the 'normal' and trigger cells.

7.4.4 Time-Multiplexed Memories

None of the available memories from the TSMC SRAM compiler can support sampling rates up to $3\,\mathrm{GSa\,s^{-1}}$. Hence, four SRAMs are placed in parallel, so they only have to run at $750\,\mathrm{MHz}$. These four signals then have to be time-interleaved, or serialized. A clock

frequency of 3 GHz is too fast for (multi-rate) HDL synthesis, but it is challenging to reach the timing requirements even for manual design.

In order to make it work, a 2-bit Gray counter (as shown in Fig. 7.15a) is used as a divide by 4, ensuring only one selection bit changes at a time. The bank's positive *I* clock is used as the input clock for this circuit (Fig. 7.11). The counter itself will have unacceptably skewed timings between the different output clock phases, necessitating a retime step with the original input clock. The serializer is then implemented as a multiplexer with the $f_s/4$ Gray counter clocks as selection bit inputs. All muxes are implemented as pass-gates, with minimal buffering to ensure fast enough propagation times. The output delay of the HDL synthesized block is prone to PVT variations and thus assumed to be unknown, which poses a potential timing hazard. To avoid issues in post production, either of the Gray counter's output clocks can be selected as the input clock of the digital block using the S_digclk<1:0> SPI signals (see Table 7.1 and Fig. 7.11).

7.4.5 Clock Generation, Division and Distribution

To support all different operating options, several clocking options have to be supported. Four input clocks can be used therefor. These clocks are provided from off-chip differentially since the DTX targets high-power operation, which could result in a varying on-chip ground potential, influencing the clocks' common modes. The schematic of the clock generation & selection block (Fig. 7.11) is given in Fig. 7.16.

First, polar operation needs to be supported, which requires a phase-modulated RF clock. In addition, both banks should be able to operate with individual clocks to support Doherty, outphasing, push-pull, or carrier aggregation upconversion architectures, which result in the RF_PM_A and RF_PM_B input clocks. These are fed into a differential pair for common-mode rejection, and the resistive terminations are placed in shunt. Since an off-chip balun has to be used, which potentially has an amplitude and phase imbalance, a perfect 50 % clock duty cycle cannot be guaranteed. Hence, a duty cycle correction loop can be used in the clock generation chain, which is enabled using the E_RF_Corr signals and is shown in Fig. 7.17 (courtesy of Mohsen Hashemi [81]). It is also possible to have both banks synchronized, or in opposite phase from each other, which can be controlled using the S_RF_AB and S_RF_Phase signals from the SPI registers (right side of Fig. 7.16).

Next, the on-chip memories should be supplied with a sampling clock that is never phase-modulated. Again, both banks should be able to have their own sampling clocks, resulting in the final two input clocks: Mem_CLK_A and Mem_CLK_B. It has to be possible to generate 90 or 45-degree offset phase clocks from these input clocks to also support signed Cartesian and 8-phase multi-phase operation, for which digital quadrature clock dividers are used. For signed Cartesian (90 degrees), this requires a divide by 2 operation. A divide by 4 operation is required for 8-phase multi-phase (45 degrees), which is implemented as two sequential divide by 2 operations. This means these clock inputs should support higher frequencies than the (phase-modulated) RF input clocks. The differential-pair-based input buffers have 3 GHz as their upper limit, depending on their bias current (set using V_{REF}, right bottom Fig. 7.12). Inverter-based clock input buffers are used as such. These support a much higher input frequency, up to 9 GHz, which means after a 4 times division, the maximum operating frequency remains 2.25 GHz. These inverter-based clock input buffers are DC-coupled, so they also do not have a lower frequency bound, whereas the differential-pair-based ones are AC-coupled. The resistive terminations are shared between
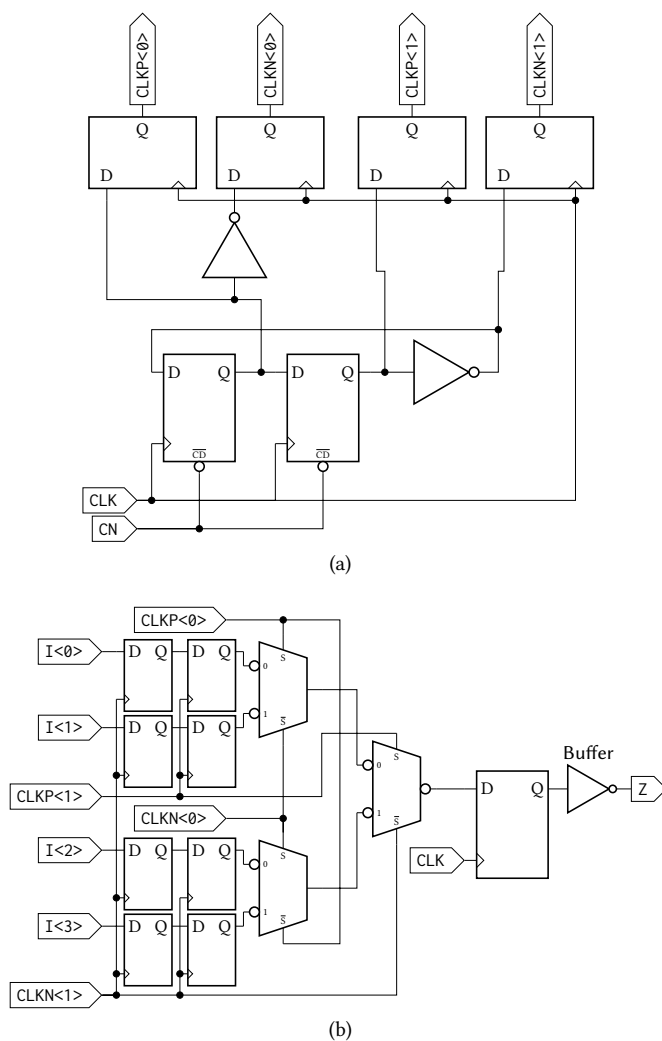
Figure 7.15: Schematics used to achieve 4:1 serializing or time-multiplexing operation of the memory data: (a) the 2-bit Gray counter with additional retiming to prevent skewing of the 4 resulting clock phases; (b) the serializer schematic, implemented using pass-gate muxes. The Gray counter drivers 28 of these 4-bit serializers, resulting a $3\,\mathrm{GSa\,s^{-1}}$ digital signal that is 28 bit wide.

Figure 7.16: The clock input routing and division schematic. All clocklines are executed in pseudo-differential fashion using the RF mux from Fig. 7.18, for simplicity only the positive polarities are drawn. All control signals labeled A_ originate from digital block A (see Fig. 7.11 and Table 7.1), similar for those labeled B_ and digital block B.

7

Figure 7.17: Duty cycle loop [81].

Figure 7.18: Schematic of the multiplexer used for the RF clocks. Additional pull-down transistors have been added to improve isolation between the two clocks.

the two input options, and their selection is made using the S_diff_inv signals from the SPI registers. Also, these inputs are provided with the duty cycle correction loops.

An 'RF' clock mux is designed for the many selection possibilities that need to be made, which schematic is shown in Fig. 7.18. Also these clocks are handled pseudo-differentially, on top of symmetrical layouting and equalizing low-to-high and high-to-low propagation delays using transmission gate muxes, to ensure signal integrity. Different clock frequencies may be present at the inputs, e.g., in the case of carrier aggregation operation or after clock division. These may not influence the timings of the muxing operation, which could happen due to forward capacitive coupling. Hence, an additional 'pre-mux' stage with pull-down transistors is added to each signal input to ensure the input for the final mux stage is quiet enough not to influence its timings. All muxes shown in Fig. 7.16 are these 'RF' muxes.

After all selection options, three (pseudo-)differential clocks are routed to each bank. As already mentioned in Section 7.4.4, the positive I clock is used to generate the digital clock at $f_s/4$. The data is eventually retimed to a data clock DCLK in the unit cells. This DCLK can be selected from any of the four digital clock phases (IP, IN, QP, or QN) provided to the bank and is selected using the S_DCLK signals from the SPI registers. These four clock phases are also fed into a phase mapper (indicated as PM in Fig. 7.11) controlled by two 'sign bits' provided by the SRAMs. The phase mapper is implemented using the RF mux of Fig. 7.18. Choosing between the phase-modulated RF clock or the phase-mapped digital clock is done using the S_RFCLK signal from the SPI registers.

Next, these pseudo-differential clocks should be routed over the chip in such a way that signal integrity is preserved: the clocks should not be influenced by external factors, such as capacitive or inductive coupling to the clock lines, as well as having the lines and their drivers fast enough to support clocks up to 3 GHz. This means the clock lines should be somehow shielded. The higher metal layers (M6 and up) are thicker and, thus, have lower resistance. However, these are mainly reserved for supply routing, yielding limited shielding opportunity, and have relatively poor pitch and spacing rules, which would yield very bulky clock lines. Hence, the lower metal layers are considered to implement a coax-like shield around the clock lines. Line parameters can be estimated using transmission line calculation methods. Two options are selected for more thorough EM simulations, shown in Fig. 7.19 as an example, namely, either using a double coax structure or using a twinax

(a) Dual Coaxial   (b) Twinaxial
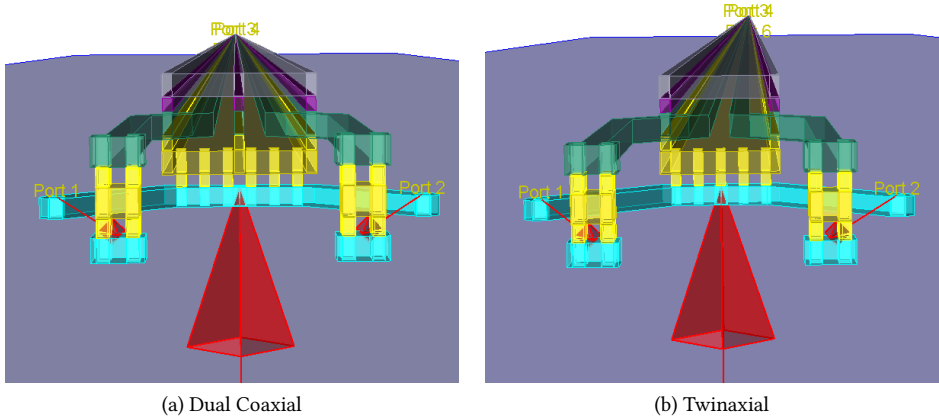
Figure 7.19: 3D EM views of two possible shielded RF clock line implementations using the lower metal layers of the TSMC 40 nm CMOS technology.

structure. From the EM simulation, the odd-mode capacitance of the lines was evaluated, indicating the twinax structure has 25 % lower odd-mode capacitance than the double coax structure while also needing less space. This twinax structure can be simulated relatively easily by assuming it to be an edge-coupled coplanar waveguide with top and bottom covers. For these lower metal layers, the resistive component is much more dominant in line propagation than its inductive component. A trade-off has to be made in line capacitance (which relates to the power consumption to drive it), line resistance (which relates to the line propagation speed), and spacing. First, using metal layers 1 and 5 for shielding while using metal layer 3 as a conductor yields a much lower capacitance than when layers 2 and 4 are used as shields (as is the case in Fig. 7.19). A reasonable trade-off can be found when using a 270 nm wide conductor with a space and clearance of 775 nm, which is also compatible with the height of standard cell clock buffers and inverters. This results in an odd-mode capacitance of 126 fF mm^{-1} and a 1.21 kΩ mm^{-1} series resistance. The series inductance of these lines is found to be 336 pH mm^{-1}, which at 3 GHz has an impedance of 6.3j Ω, showing that the series resistance is indeed dominant. Using the propagation delay of a distributed *RC*-line [71], this gives 0.38*RC* = 58 fs/mm^2. To satisfy the second signal integrity condition, having the driver and the lines fast enough to support clocks up to 3 GHz (with reasonable margin), the maximum line length is limited to 720 μm. A binary clock tree is then used to ensure identical timing to all unit cells, which is shown in Fig. 7.20. Ideally, a binary clock tree branches out to a power of 2 'leaves.' Here, the unit cells are not distributed uniformly due to the positioning of the ground return pads. The binary clock tree is routed to these spots to equalize line loading. We are still 4 leaves short, which is compensated for by a dummy buffer in an effort to equalize line loadings and delays. Post-layout simulation with 190 Monte Carlo samples for PVT variations (including supply routing) shows that the average propagation delay of the full clock tree is between 367 ps and 501 ps, with an expected value of 422 ± 20 ps. More important is the matching between the different tree branches, which in the worst process corner was 9.2 ps, and an expected
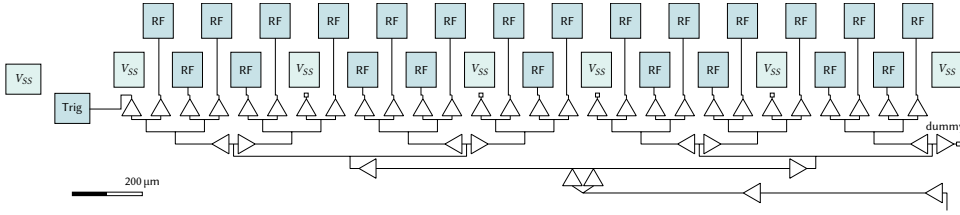
Figure 7.20: The binary clock tree of bank A, where the *x*-dimension is to scale. The RF clocks are routed using shielded twinaxial lines. The data, data clock, unit cell settings, and the $V_{DD,\text{core}}$ supply are routed along with the RF clock tree.

value of 4.8 ± 1.2 ps. The largest delay deviations could be found in the unit cells connected to the rightmost tree branch that uses the dummy buffers, because of unequal power supply loading of that branch, despite best efforts to equalize supply paths and *IR*-drop.

7.4.6 Supply Decoupling

The CMOS controller has 4 supply voltage domains and one shared ground domain. One high-voltage domain $V_{DD,\text{dr}}$, primarily used for the drivers' supply at a nominal voltage of 2.5 V. There are three core voltage domains: $V_{DD,\text{core}}$, $V_{DD,\text{core2}}$, and 'VDD!', all at a nominal voltage of 1.1 V. These domain separations are indicated on the IO ring in Fig. 7.12. The VDD! domain is used for the synthesized digital blocks with the SRAMs to be able to increase their supply voltage separately in case the memories run too slowly. The $V_{DD,\text{core2}}$ domain is only used for the phase-modulated RF clock generation. Lastly, the $V_{DD,\text{core}}$ domain is used for all other purposes, including the sampling clock generation, clock division and selections, clock tree, muxes, and unit cell logic. The bias currents for the differential pairs are generated from the $V_{DD,\text{dr}}$ using a separate V_{REF} input. Next, the important design details for the separate supply domains are highlighted.

The high-voltage driver supply may have the strictest specifications since it is the domain with the highest expected current and power consumption. The driver chains are designed to have an equivalent switch resistance of 16.87 Ω each, at a peak instantaneous current consumption of 97 mA, meaning that activating all 46 of them in parallel yields an equivalent drive resistance of $R_{\text{dr}} = 0.367\,\Omega$ towards the LDMOS and up to 4.6 A of peak instantaneous current, if only for a very brief period (in the order of tens of picoseconds). Any supply and ground rail resistances appear in series to R_{dr}, potentially degrading the drivers' performances (*IR* drop). Even more stringent is the change rate $dI/dt > 100\,\text{GA s}^{-1}$, even with a moderate series inductance of 1 nH, resulting in over 100 V dynamic voltage drop, which is unacceptable for a 2.5 V supply. To combat this, local capacitive decoupling is used. A total of 16.29 nF is implemented on-chip for this supply domain. This capacitance is distributed over the chip and between the unit cells, with a narrow (0.45 μm) single lower metal layer (M5, 200 mΩ $\square^{-1}$) connecting them in a grid to increase their series resistance. If no series resistance were present, these decoupling capacitors would result in a perfect resonator with the supply feed inductance. The grid is connected to higher metal layers irregularly to have variation in the present series resistances, resulting in different *RC* time constants, improving the decoupling effectiveness. Overall supply rail low series resistance is achieved by placing 20 parallel maximum width ultra thick metal (M7, 5 mΩ $\square^{-1}$) lines, 10 for $V_{DD,\text{dr}}$ and 10 for V_{SS}, alternating to decrease inductance. This amounts to a total of

215 μm wide copper, using up 315 μm of space along the unit cells to meet metal density requirements. Connecting from either side of the die, each travels a maximum of 2.36 mm, meeting in the middle, yielding a worst-case series resistance of 55 mΩ for either supply rail. But, since there are ground contacts interleaved within the unit cells, the ground resistance is expected to be lower than this.

The other supply domains have less stringent requirements but still do require decoupling. The synthesized digital block automatically includes decoupling fill cells between the logic gates, which resulted in 1.12 nF per digital block. Outside each block, an additional 1.23 nF was laid out, bringing the total on-chip decoupling capacitance for the VDD! domains to 4.72 nF. Another 4.41 nF is laid out for the $V_{DD,\text{core2}}$ domain, and 8.26 nF for the $V_{DD,\text{core}}$ domain since it has more connections. This brings the total decoupling capacitance available on-chip to 33.67 nF.

Since on top of the clock tree no ultra-thick metal is available anymore (all is used to route $V_{DD,\text{dr}}$ and V_{SS}), the $V_{DD,\text{core}}$ domain is routed along the binary clock tree using a medium-thickness metal layer (M6, 23 mΩ □$^{-1}$). This is no problem, since very little power consumption is expected for this domain. At the root of the clock tree (the two parallel line drivers in Fig. 7.20), these supplies are split into 4 separate supply lines, each serving a quarter of the tree's leaves. The fourth, serving also the dummy branch, has been decreased in width in an effort to equalize *IR* drop between the tree branches, as discussed briefly in Section 7.4.5.

7.5 High-Power DTX Demonstrator I: On-Resistance Modulation – Class-BE

This very first DTX prototype targets to show the feasibility of high RF output power while simultaneously having high drain and system efficiencies. Considering the ON/OFF activation of the RF output stage, a resonant SMPA operating class seems to be the most logical choice to achieve high energy efficiency. In view of this, class-E is favored over (inverse) class-F operation since it is challenging to realize harmonic open conditions for a high-power device with a large output capacitance (e.g., approximately 0.3 pF W^{-1} for the considered LDMOS technology). Furthermore, class-E stands out with a theoretical drain efficiency of 100 % while benefiting from a simple circuit topology (see Section 2.2.4). Practical implementations, however, will be restricted in their peak drain efficiency by limitations imposed by the used technology in relation to the targeted operating frequency and output impedance level in terms of f_T/f_0 and R_L/R_{ON}. Namely, these ratios need to be large for high drain efficiency at peak output power [82]. For that reason, we will explore the following high-power DTX single-ended polar class-E technology demonstrator, with its conceptual diagram given in Fig. 7.21.

7.5.1 Class-BE output match

Using the LDMOS dimensions provided in Section 7.2, next, we will consider the class-E output matching network. In a polar class-E transmitter, the segmented output stage can be modeled as a switch that toggles between the effective R_{ON} and R_{OFF}, in parallel with the output capacitance C_{DS}, which is the sum of all segments' output capacitances (Fig. 7.21). The ACW in this polar configuration provides the amplitude information ρ and defines the number of activated LDMOS segments. The clock phase φ controls the switching moment
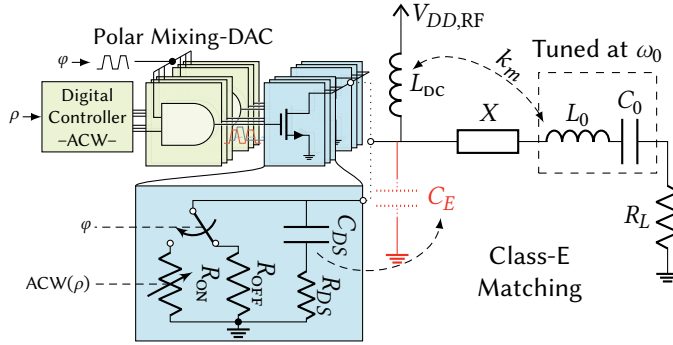
Figure 7.21: Digital polar class-E DTX configuration using output stage segmentation. The digitally controlled segments can be modeled as a single switch with an ACW-controlled R_{ON}.

Table 7.3: Class-(B)E design sets and driver sizes used for generating the data points of Fig. 7.22.

V_T	$V_{DD,\mathrm{RF}}$	q	K_C	K_X	R_{dr}	M
2.00	28.00	0.983	7.189	1.167	10.31	3.781
1.67	28.00	0.987	7.996	0.960	12.69	2.809
1.33	28.00	1.017	8.169	0.500	15.44	2.289
1.00	28.00	1.296	0.209	−2.280	14.43	2.441
0.80	28.00	1.368	0.155	−3.071	14.39	2.448
0.60	27.11	1.407	0.137	−3.265	14.67	2.401
0.40	25.75	1.448	0.122	−3.398	14.62	2.410
0.20	24.58	1.486	0.111	−3.426	14.46	2.436
0.00	23.50	1.525	0.102	−3.365	13.84	2.548

of these segments. The capacitance C_{DS} is used as the class-E shunt capacitance C_E. Its equivalent series resistance R_{DS} relates to the losses that occur when an RF signal is applied at the output of the LDMOS segments, a condition of particular interest when dealing with Doherty or high frequency operation. The resistance R_{OFF} models the static losses due to the output stage's bias. By applying the class-E theory from Section 2.2.4, we can find that the upper class-E frequency $f_{E,\max}$ (Eq. (2.25)) for the considered LDMOS technology is between 510 MHz and 920 MHz for $V_{DD,\mathrm{RF}} = 20\,\mathrm{V}$. Since our targeted operating frequency is 3 GHz for our power LDMOS DTX, class-BE operation has to be used for its design.

A simulation study is performed to explore the requirements on LDMOS V_T, where we optimize our power-DTX's drain and system efficiencies for class-BE operation at various V_T values of the LDMOS segments. For this purpose, the equivalent resistance of the LDMOS' driver is modeled by R_{dr} (see Section 4.1.2). Selecting the R_{dr} value affects the switching speed of the LDMOS segments, and as such, the LDMOS drain efficiency. Furthermore, the chosen R_{dr} value automatically sets the required size of this driver for given CMOS technology, as specified in Section 4.1.3 by Eqs. (4.12) and (4.13). Here we assume the number of stages N in the related tapered buffer chain to be continuous. Finally, the output matching network (Fig. 7.21) is assumed to be lossless.

Using the class-E theory (Section 2.2.4) and the DTX efficiency definitions (Section 6.1),
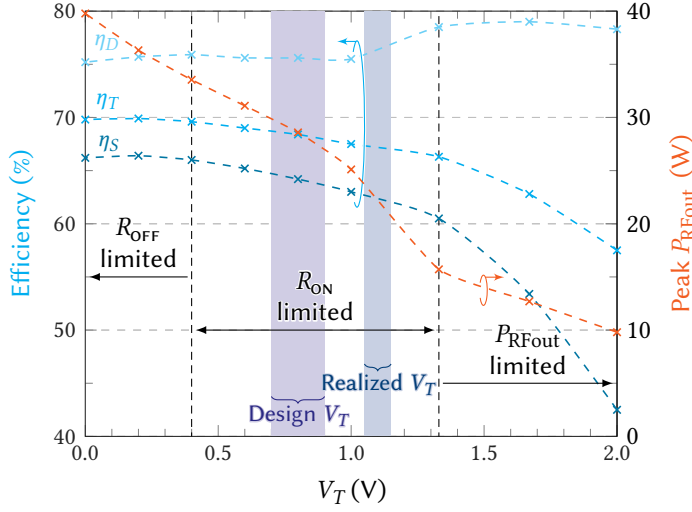
Figure 7.22: Theoretical full power DTX performances for the optimized class-BE design sets for varying values of the V_T of the applied LDMOS technology. The design parameters used to generate this graph are: L_{bond} = 0.6 nH, R_{bond} = 0.2 Ω, Q_{L_0} = 10, k_m = 0.2264, f_0 = 3 GHz, d = 50 % and $V_{DD,dr}$ = 2.5 V, the remaining parameters depend on V_T and are given in Table 7.3.

the normalized class-E design parameters (see Table 7.3) have been determined by numerical optimization for each V_T point in the graph of Fig. 7.22 to maximize the related efficiencies and output power. Interestingly, for low V_T, also the optimum R_L ($\propto K_C$) is low, which can be explained by the increased maximum current capability of the LDMOS. For this condition, also the optimum value for R_{dr} drops, implying that the driver provides faster rise and fall times to reach high drain efficiency. This low R_{dr} comes at the cost of an increased $P_{DD,dr}$, which can be afforded from a system efficiency perspective since P_{RFout} is also higher for low V_T. For high V_T, the RF output power is severely limited, and rise and fall times need to be sufficiently short to achieve any respectable output power at all. This also requires a low R_{dr}, further decreasing the achievable system efficiency.

It appears from Fig. 7.22 that a customized LDMOS technology, having a threshold voltage of V_T = 0.2 V, yields the optimum system efficiency. However, such a low V_T requires very high doping concentrations in the LDMOS fabrication process, impacting other device parameters like ruggedness. Also, the LDMOS OFF-resistance R_{OFF} (Fig. 7.3) then becomes so low that the efficiency in power back-off would be severely affected. Therefore, based on practical considerations, a V_T of 0.8 V is targeted for this demonstrator.

Since this V_T value deviates significantly from the standard LDMOS process flow, it comes in practice with some uncertainty. To handle this V_T uncertainty in the demonstrator, duty cycle (d) and supply voltage ($V_{DD,RF}$) adjustment can be used in the testing phase of the power-DTX. Hence, the post-production tuning possibilities of the DTX demonstrator are evaluated next. Lowering the duty cycle gives the output stage more time in the 'OFF'-state, resulting in a lower drain voltage at the switching moment, yielding lower switching losses. Assuming the V_T = 0.8 V and varying both d and $V_{DD,RF}$ results in the efficiency dependencies as shown in Fig. 7.23a. Note that the drain efficiency can be slightly
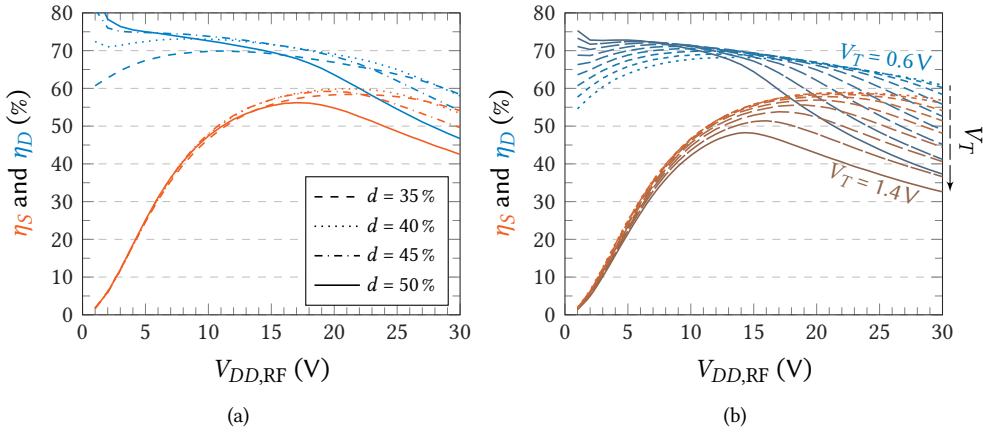
Figure 7.23: System and drain efficiencies versus $V_{DD,\mathrm{RF}}$ for: (a) post-production duty cycle adjustment assuming a realized V_T of 0.8 V; (b) potential realized V_T-shifts assuming a duty cycle of 50 %.

improved by decreasing d to 40 %. Decreasing d further degrades the efficiency due to the finite rise and fall times of the driver. The impact of V_T is evaluated in Fig. 7.23b for a d = 50 %, showing that, in general, the drain efficiency increases for lower values for $V_{DD,\mathrm{RF}}$. However, since the peak output power also drops, while $P_{DD,\mathrm{dr}}$ for peak output power remains constant, the overall system efficiency drops. Decreasing V_T results in more output power, yielding higher system efficiencies. For higher values of $V_{DD,\mathrm{RF}}$, aside from more switching-related losses, the nonlinear LDMOS C_{DS} changes to a lower value, resulting in a mismatch with the desired class-BE loading, also causing a decrease in drain efficiency.

Based on these observations, the DTX is re-optimized for the targeted V_T = 0.8 V with the output matching network now including realistic losses. This results in the optimum R_{dr} = 16.87 Ω as mentioned in Section 7.4.2. This requires f = 2.88 and N = 7.5 for the number of driver stages, yielding a capacitance multiplication factor M = 2.114 (M' = 1.532). The simulation result of the power consumption of the CMOS controller using these values is shown in Fig. 7.26 and will be discussed in more detail together with the measurement results in the next section.

With the design study above, the high-power DTX configuration is defined within the technology options available. The within this project targeted DTX peak output power (larger than 20 W) and peak system efficiency (close to or above 60 %) seem feasible.

7.5.2 Demonstrator Realization and Measurement Results

The TSMC 40 nm CMOS controller and custom 0.4 µm 28 V low-V_T LDMOS power die have been fabricated, targeting the optimum design values discussed previously. The DTX's detailed schematics and a photograph of its implementation are shown in Figs. 7.24 and 7.25. Here LDMOS layout 02 (inside LSBs) is used for the assembly. The single-ended class-BE output matching network, centered around f_0 = 2.1 GHz, was implemented using discrete capacitors, bond wires, and PCB transmission lines as inductances, and a quarter-wave transformer to match R_L to 50 Ω. DC-based verification of the realized LDMOS V_T-shift indicated a V_T of 1.10 ± 0.05 V. This is higher than intended, but the availability of on-chip
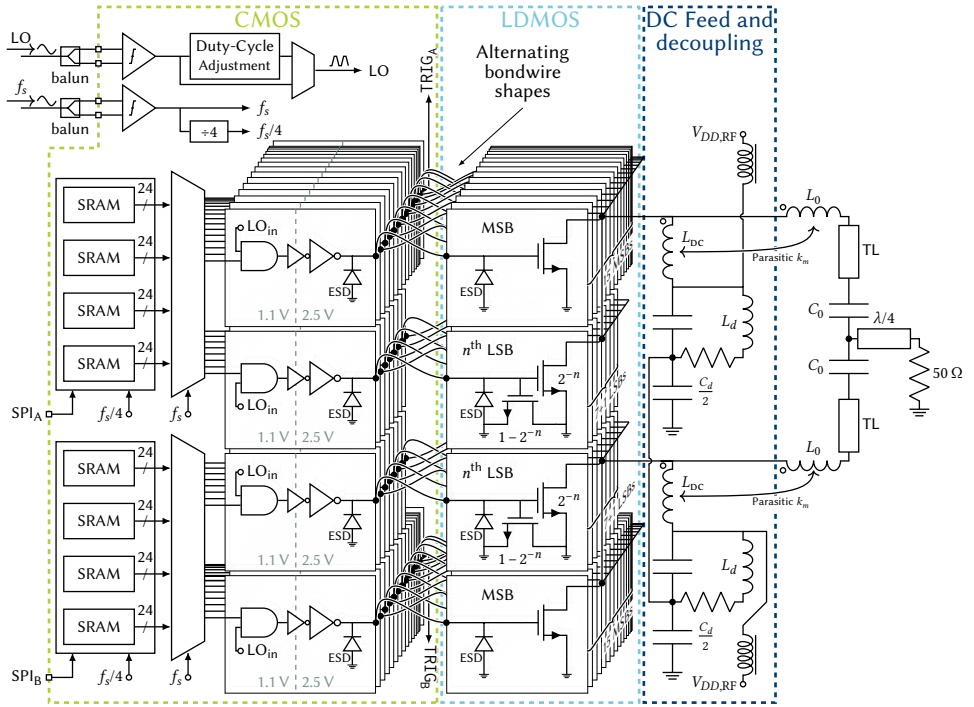
Figure 7.24: Realized schematic of the power-DTX configured for polar operation with the single-ended 2.1 GHz class-BE output matching. The prototype features a 40 nm CMOS controller and a V_T-shifted segmented LDMOS output stage, featuring two switch banks that take their ACW data from time-multiplexed on-chip memories. The power-DTX prototype also has extra trigger output pins to monitor its output independent of the LDMOS die.
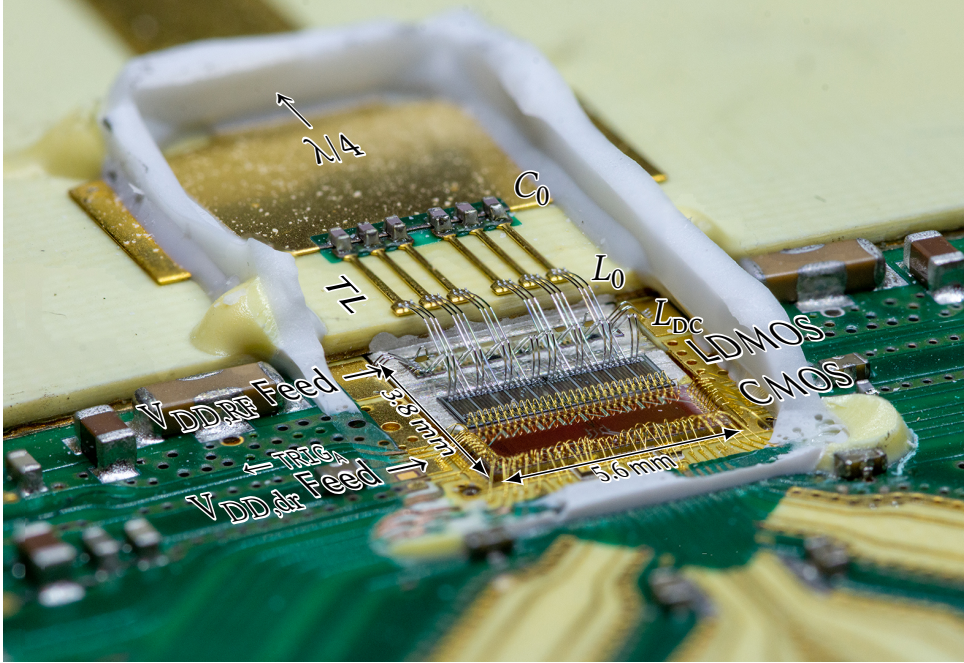
Figure 7.25: Photograph of the power-DTX with the ceramic cap removed.

duty-cycle control and flexibility in the LDMOS supply voltage can handle this deviation (see Fig. 7.23).

Dynamic Power Consumption of the Controller

First, the simulated 2.5 V power consumption of the CMOS controller (simulated $M' \cdot P_{DD,\mathrm{dr}}$) is compared to the measured data, shown in Fig. 7.26. Clearly visible are the fluctuations resulting from the hybrid unary and binary-weighted implementation of the segments with equalized input capacitances (see Section 7.2.1 and Fig. 7.4). That aside, the power dissipated does show the overall linear proportionality with ACW. We can compare the simulated $P_{DD,\mathrm{dr}}$ with the total controller power consumption needed to drive the LDMOS segments using an empirical fit for the capacitance multiplication factor $M' = 1.291$, which is 15.7 % lower than what was theoretically predicted in the previous section (Section 7.5.1). This can be attributed to the slightly more aggressive scaling in the implemented tapered buffer chain, resulting in a 7 stage chain, and the fact that the (nonlinear) device capacitances are not 100 % charged and discharged each cycle. Still, the accuracy of this estimation can be observed from its almost perfect correlation with the measured power dissipation of the 2.5 V supply domain. Furthermore, if the ESD capacitances could have been avoided, the required drive power would have been a factor of 2.2 lower, given the same driver speed (see Section 7.2.1). The above suggests that close to perfect scaling of the consumed DTX power with the RF output signal is in reach.
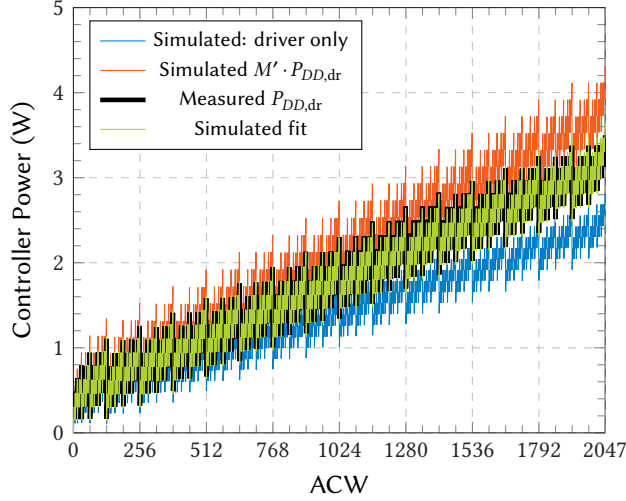
Figure 7.26: Simulated drive power including $C_{dr,o1}$, this power is multiplied by the theoretically calculated $M' = 1.532$ to provide the total power consumption of the driving drive chain, which is compared to its actual measured values.
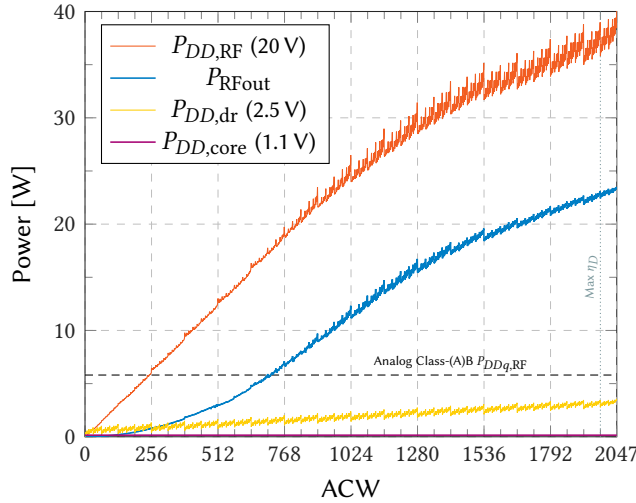
Figure 7.27: Measurement of the fully digital-TX line-up ($W_{G,tot} = 41.472$ mm, 11-bit) in pulsed CW operation at 2.1 GHz, using 15 % duty-cycling to lower thermal effects, and $V_{DD,RF} = 20$ V. Plotted are: the RF output power (P_{RFout}), the total DC power consumed by the segmented LDMOS devices ($P_{DD,RF}$), the DC power consumed by the driver (2.5 V domain), the continuous DC power of the low-voltage (1.1 V) CMOS circuitry, and (dashed) the static quiescent DC power for an LDMOS device with the same $W_{G,tot}$ when operated in an analog class-AB bias condition for linearity.
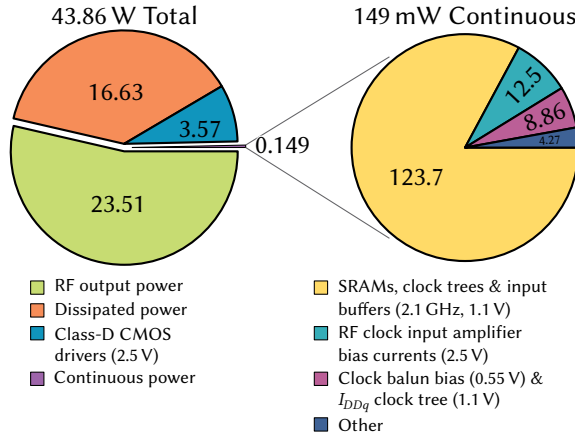
Figure 7.28: The dynamic and continuous DC power consumption breakdown in the implemented DTX at peak RF output power conditions (ACW = 2047).

Digital-to-RF Transfer of the Power DTX

The digital-to-RF transfer is confirmed by performing static power measurements versus ACW using a power meter. A 30 dB attenuator was included in the measurement setup to protect the instruments, whose losses have been de-embedded. Pulsed continuous wave (CW) measurements are performed using a 15 % time duty-cycle to reduce thermal effects. The optional RF duty-cycle adjustment is first bypassed, so the RF duty cycle in these measurements is close to 50 %. For ACW = 1920 peak drain and system efficiencies occur (see Fig. 7.27), being 62.6 % and 58.1 %, respectively.

The total power balance at maximum ACW is shown in Fig. 7.28, showing that the static power consumption basically only occurs in the low-voltage domain (1.1 V) and is a tiny fraction of all the powers involved. This is in strong contrast with an analog-oriented TX line-up biased for linear class-AB operation. The related quiescent current would, in combination with a typical LDMOS supply, result in static power consumption ($P_{DDq,\mathrm{RF}}$) of 5.8 W for an LDMOS device of comparable size. This level is indicated in Fig. 7.27 by the dashed line.

As indicated in the previous section (Section 7.5.1), the efficiency and output power can be optimized by changing the RF duty cycle and supply voltage. Using $V_{DD,\mathrm{RF}}$ = 20 V, the best peak system efficiency measured was 60.4 %, using an RF duty cycle (d) of 43 %, with a maximum output power of 18.5 W and drain efficiency of 66.7 %. Maximum output power was observed for $V_{DD,\mathrm{RF}}$ = 28 V and d = 50 %, being 29.8 W. Due to the nonlinear LDMOS C_{DS} and higher switching losses, the drain and system efficiencies under these conditions were 40.0 % and 38.1 %, respectively. This downwards trend was already expected from simulations regarding the post-production tuning (see Fig. 7.23).

Further CW efficiency measurements with varying $V_{DD,\mathrm{RF}}$ have not been done, as the demonstrator sample used for these measurements was damaged after finishing these peak power measurements. This was caused by lacking a proper shut-down procedure for the power-DTX in this set of measurements, as the DTX could go from full power to zero in a single sample when downloading new amplitude data to the SRAMs or when aborting a
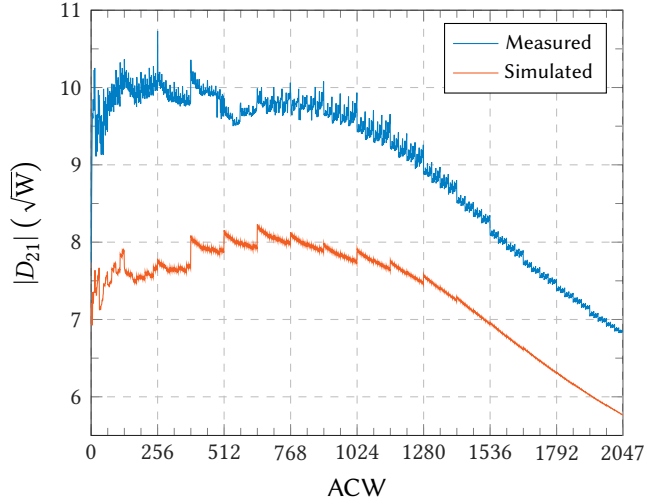
Figure 7.29: Simulated and measured static normalized digital transfers (D_{21}) compared. Both clearly show the effect of the hybrid unary and binary-weighted implementation of the segments, especially when switching to a unary weighted segment (every 128 ACWs).

measurement. Such a sudden change in the related DC current can cause large V_{DS} spikes, even over a moderate bias feed inductance. The remaining measurements in this paper were done before this breakdown, or performed on a different prototype sample. To avoid damaging these remaining samples, further CW measurements or measurements using $V_{DD,RF}$ = 28 V were avoided.

In Fig. 7.29, the simulated and measured static (CW, with stepped ACW, using a power meter) digital-to-RF transfer of the entire DTX system is given for $V_{DD,RF}$ = 20 V and d = 50 %, in terms of the normalized digital transfer D_{21} (as defined in Section 5.1.3). The simulation setup used includes all known parasitics and imperfections in the realized power-DTX demonstrator. The first parasitics to be included are the mutual inductances of the staggered bond-wire array between the CMOS controller and LDMOS output stage, and limited bond-wire quality factor, by using an S-parameter EM model for the bond-wires (see Fig. 7.8). The effective self-inductance ranges from 490 pH to 620 pH with a coupling factor up to 0.52, depending on bond-wire location within the array. This coupling is stronger between the first and third bond-wires than between the first and second bond-wires, due to their staggered orientation. The remaining parasitics to be included are lumped component parasitics, as well as the underestimated C_{GS} of the drain–source short-circuited dummy LDMOS devices included in the binary segments as compensation capacitances, yielding a 15 % higher dummy C_{GS} value than intended. This deviation was caused by the used compact model of the LDMOS device, which was optimized for analog transconductance operation, and, therefore, not for the V_{DS} = 0 V condition. It can be concluded from the simulations at low ACW, that the impact of the mutual inductances of the bond-wires between the CMOS and LDMOS dies is quite pronounced. Activating the thermometer segments at every 128 ACWs yields small jumps, except for the first two transitions. This is caused by aforementioned mutual coupling, which increases the effective RF drive voltage (V_{GS})
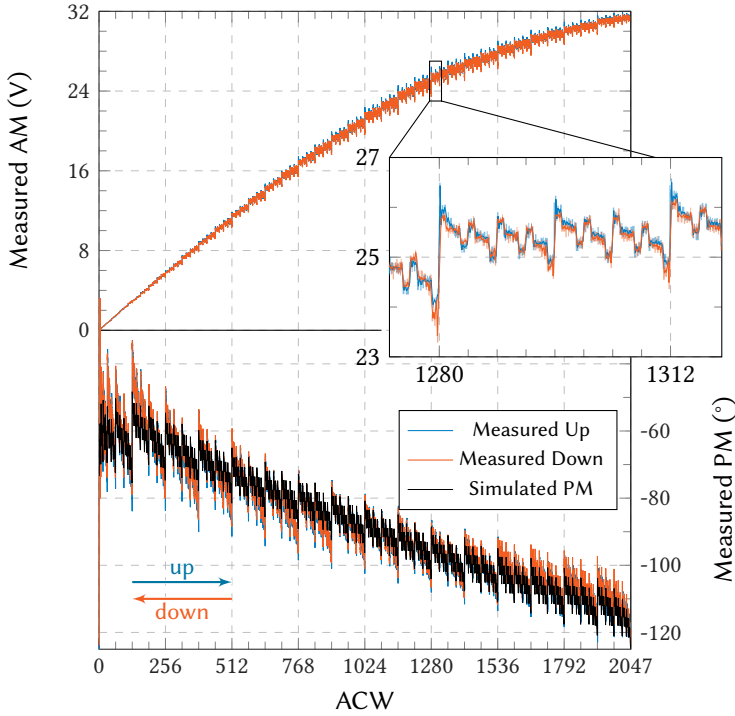
Figure 7.30: Measured dynamic response for a linear ACW triangle-shaped envelope signal, centered around 2.1 GHz with the f_s and VSA analysis bandwidth set both to 525 MHz, leading to a cycle time of 62.4 µs. The ramp down is flipped and placed over the ramp up to enable comparison. This shows some hysteresis in the form of overshoots when going up and undershoots when going down. For reference also the simulated static ACW–PM is plotted, adjusted for the delay of the full tapered buffer chain and resistive supply variation (see also Fig. 7.34.)

of an LDMOS segment when neighboring segments are activated, yielding an unwanted bit-to-bit interaction in the output power. In this simulation, the unary-weighted segment located the farthest away from the grouped binary-weighted segments (see Fig. 7.5b) is activated first. After ACW ≈ 1024, the DTX starts to enter compression, evident from the transfer decreasing in magnitude.

Our forgoing driver analysis assumed that the LDMOS input capacitance of the segments remains perfectly constant, independent of what happens at the drain terminal of the RF output stage. However, the C_{GS} is dependent on the operating region of the LDMOS device. In addition C_{GD} undergoes Miller multiplication and also appears at the input of the segment. Although in RF-oriented LDMOS technologies C_{GD} is very small, this yields some dependence on the output operating condition. These effects modulate the effective loading capacitance of the LDMOS' driver and its delay, appearing as (additional) ACW–PM distortion in the DTX line-up.

In Fig. 7.30, the measured (dynamic) ACW–AM/PM curves are shown for a triangular (linear) ACW ramp, measured using a spectrum analyzer in VSA mode. For this purpose, the baseband sampling rate f_s is reduced to 525 MHz, an integer factor 4 lower than the RF center frequency, to be within the baseband sampling rate. Every SRAM value directly
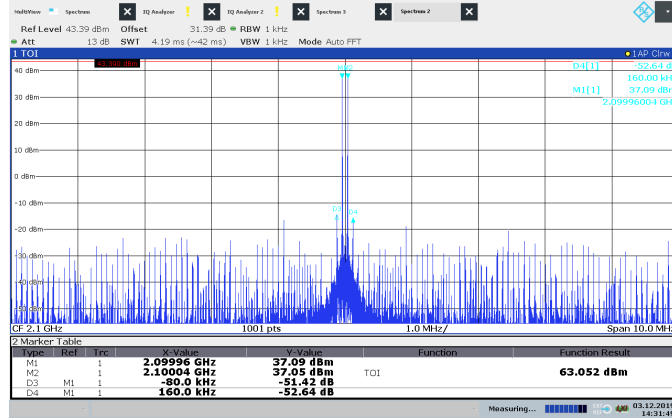
Figure 7.31: The measured spectrum of the 80 kHz two-tone signal, showing an $IM_3 \leq -51.4$ dBc after static LUT calibration, only using the unary segments with pulse density modulation.

corresponds to one demodulated *IQ* sample, and every ACW value is repeated eight times in the memory, giving a cycle time of 62.4 µs. For reference, also the simulated static ACW–PM curve is plotted. In this simulation only the final stage of the tapered buffer chain was included (see Section 4.1.3) to aid simulation speed and convergence. To compensate for the delay variation of the full chain due to the resistive supply variation (*IR*-drop, see Fig. 7.34), thus additional ACW–PM variation, the on-chip $V_{DD,\text{dr}}$ is used to adjust the simulated curve. The overall difference between simulated and measured ACW–PM can be attributed to the aforementioned use of a compact model of the LDMOS device, optimized for analog transconductance operation, not including the severe C_{DS} nonlinearity when operating close to $V_{DS} = 0$ V. The non-monotonic behavior of the binary-weighted segments can be clearly seen in this measurement. The triangular ramps up and down have been folded on top of each other to compare them. Doing so, some memory effects also become visible. Namely, the ramp up has some overshooting when switching to a higher ACW value, while in the ramp down undershooting happens when switching to a lower ACW value: at ACW = 1280, a unary-weighted segment is being switched ON/OFF, and at ACW = 1312 an 2^{-2}-weighted segment is switched ON/OFF. This hysteresis/memory effect is attributed to timing glitches, which are caused by delay differences due to C_L mismatch in the binary-weighted LDMOS devices, as well as to current redistribution effects related to the physical location difference of these segments.

Power-DTX operation with Modulated Signals
For measurements with modulated signals, the power-DTX was used in its polar operation mode, and a Keysight M8190A AWG was used to provide the phase-modulated RF clock centered around $f_c = 2.1$ GHz. The on-chip duty-cycle adjustment circuit was found to generate some spectral dirt caused by its analog control. Therefore it was bypassed and hence $d = 50$ %. Static LUT calibration has been applied to compensate for the ACW–AM and ACW–PM DTX nonlinearities. This approach was attempted first for the full range of possible ACWs; however, the ACW–PM variations introduced by the use of the binary-weighted segments could not be compensated such that the effective number of bits (ENOB)
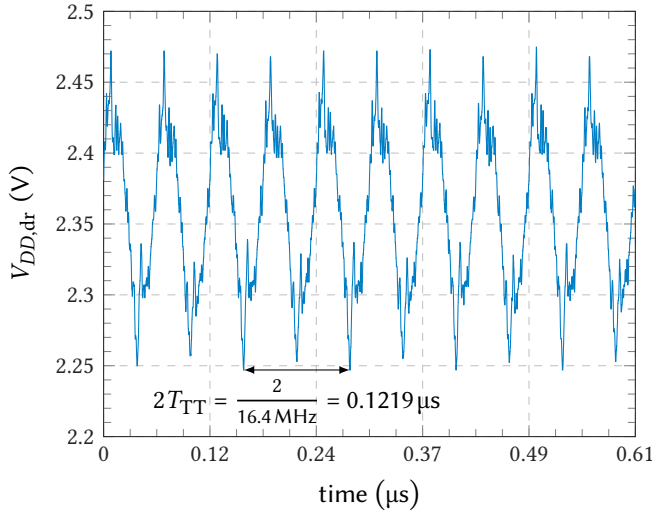
Figure 7.32: Measured $V_{DD,\text{dr}}$ through the TRIG pins using a 2.1 GHz two-tone test with 16.4 MHz tone spacing.

was increased. Clearly, the dynamic effects in our DTX prototype dominate here; henceforth, only the unary-weighted segments are used. To compensate for the reduced resolution and still improve the DTX's dynamic range and noise floor, pulse density modulation techniques using the remaining least significant bit were applied. Using this approach, a two-tone signal with a narrow tone-spacing of 80 kHz was generated and measured (Fig. 7.31), yielding an $IM_3 \leq -51.4$ dBc.

When moving to signals with video bandwidths larger than 10 MHz, the linearity degraded more than what theoretically can be expected based on the ENOB and sampling speed. Especially a two-tone with a tone spacing of 16.4 MHz proved to be problematic. Closer inspection indicated that a resonance in the DC feed of the CMOS controller caused this linearity degradation. This resonance hypothesis is confirmed, by measuring the related on-chip driver voltage $V_{DD,\text{dr}}$ through the TRIG test pin outputs with an oscilloscope. These TRIG pins were included in the DTX demonstrator to increase its testability (see Fig. 7.24). The on-chip driver supply should ideally be constant at 2.5 V. However, the measured on-chip $V_{DD,\text{dr}}$ in Fig. 7.32 shows around 240 mV$_{pp}$ variation with the two-tone envelope frequency (see also Section 5.2.3). It proved not possible with the current hardware demonstrator to correct this defect. Besides the quality of the driver supply decoupling, the supply variation also depends on the resistance in the DC supply path, which in our setup (cables, PCB, bond-wires, and on-chip interconnect) is estimated to be 94 mΩ in total. Due to the considerable overall propagation delay ($t_{p,\text{tot}}$) of the tapered buffer chain and its dependence on the supply voltage ($V_{DD,\text{dr}}$), any variation in the controller's supply voltage will cause an unwanted phase change in the DTX output, yielding unwanted phase modulation.

The resulting ACW–PM distortion is confirmed by the measured down-converted *IQ*-constellation diagram for this 16.4 MHz two-tone (Fig. 7.33), which strongly deviates from the ideal straight line. The variation in delay due to the $V_{DD,\text{dr}}$ variation was also measured using an oscilloscope by putting a square wave in the SRAM for the TRIG pins
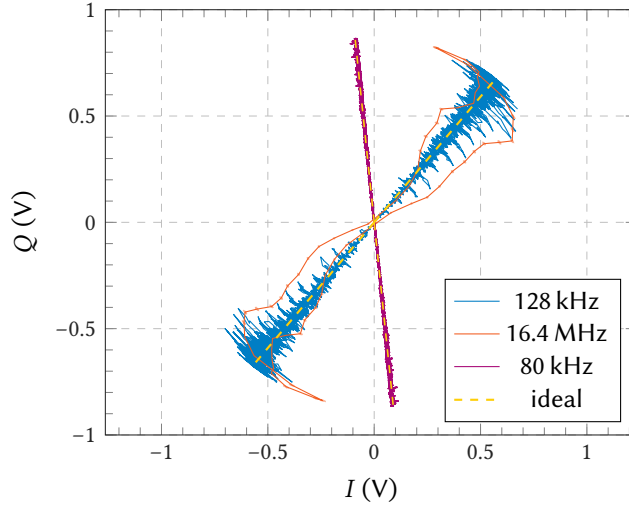
Figure 7.33: Measurements using a 2.1 GHz two-tone test with varying tone spacing: down-converted *IQ*-constellations, after static DPD, showing hysteresis/memory-effects for 16.4 MHz tone spacing. For reference, also measured constellations using 128 kHz tone spacing, and the 80 kHz tone spacing only using the unary segments with pulse density modulation are shown (with arbitrary phase reference).

7
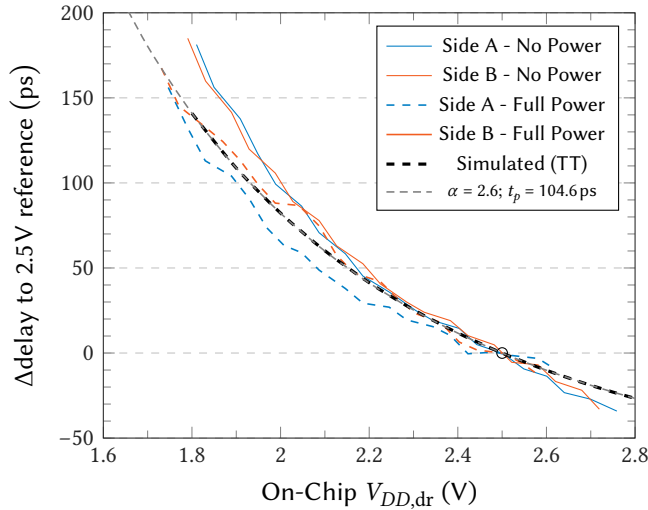
Figure 7.34: Change in propagation delay of the CMOS tapered buffer chain compared to the $V_{DD,\mathrm{dr}}$ = 2.5 V reference case versus the measured $V_{DD,\mathrm{dr}}$ which includes the *IR*-drop (averaged measurements, the ±1σ measurement variation is 14.4 ps). The simulated propagation delay is shown thick dashed and a fitted line according to theory in dashed gray.
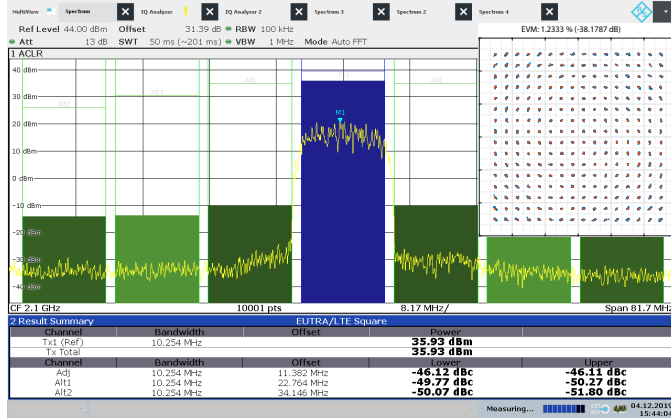
Figure 7.35: The measured spectrum and constellation of the 10 MHz 256-QAM signal, showing an ACLR = −46.1 dBc and EVM = 1.2 % after static DPD.

and comparing the delay between its rising edge and a reference marker of the AWG. The measured and simulated results for the buffer chain are shown in Fig. 7.34, showing a decreasing delay with increasing (on-chip) supply voltage. All delays are referenced to the nominal $V_{DD,\mathrm{dr}}$ = 2.5 V case, for which the simulated tapered buffer chain delay is $t_{p,\mathrm{tot}}$ = 605.8 ps. It can be seen that the measured delay differences follow the simulated values very well. It also follows the theory from Section 4.2.1 very well when α = 2.6 and the varying nominal $t_{p,\mathrm{nom}}$ = 104.6 ps. This confirms that a DC supply resonance around 16 MHz indeed causes the phase deviations in the RF output. Henceforth, we restrict ourselves to lower bandwidths for the following measurements.

In Fig. 7.35, the measured result for a 10.254 MHz 256-QAM signal around a carrier of f_c = 2100 MHz is shown. The baseband sampling rate has been multiplied by 5/8 to f_s = 1312.5 MHz, to comply with the 2^{15} baseband samples in the SRAM and to have the video bandwidth just below the supply resonance. The signal was filtered by a RRC with a roll-off factor of α = 0.22, and static (memoryless) LUT calibration was applied (again only using a LUT for the unary-weighted segments with pulse density modulation for the LSB). This yielded for a signal with 7.2 dB PAPR, a measured adjacent channel leakage ratio (ACLR) of −46.1 dBc, and an root-mean-square (RMS) error-vector magnitude (EVM) of 1.2 % (−38.2 dB, averaged over 4096 constellation points). The average output power was 3.9 W, with average drain and system efficiencies at 28.1 % and 25.6 %, respectively.

7.5.3 Key Take-Aways

This work aims to explore the feasibility of high-power fully-digital TXs that allow for high integration and low (static) standby power. Within the design constraints of today's LDMOS and CMOS technologies, our first high-power DTX demonstrator achieved decent performance in terms of output power and energy efficiency, which are the main objectives of this study. Although the dynamic range of the DTX, in terms of its implemented number of bits, could not be fully reached, reasonable linearity figures (IM$_3$, ACLR, and EVM) could be obtained for a complex modulated TX signal with not too large bandwidths. This bandwidth was mainly restricted by unfortunate resonance in the driver's DC supply bias.

Table 7.4: Performance summary and comparison with the state-of-the-art DTXs and digital PAs.

	This work			RFIC'16 [83]	PAWR'16 [84]	JSSC'17 [85]	IMS'17 [19]	IMS'11 [18]
Technology	40 nm CMOS + LDMOS			180 nm CMOS SOI + GaN	180 nm CMOS SOI	130 nm CMOS	GaN	65 nm CMOS + GaN
Architecture	Polar DTX			Broadband Polar DTX	Polar Doherty DTX	Multi-Phase SC-PA	Broadband EER DTX	Outphasing PA
f_c (GHz)	2.1			0.9	0.9	1.8	2.1	1.95
Supply (V)	1.1/2.5/20		1.1/2.5/28	1.7/4.1/15	1.7/4.1	3	17	1.2/5.0/26
P_{max} (W)	18.5[1]	23.5[2]	29.8[3]	2.8[4]	2.0	0.398	2.5	19
Peak Drain Efficiency (%)	66.7[1]	62.6[2]	40.0[3]	57.7	55.4	n/a	51.7[4,5]	68.5[4]
Peak System Efficiency (%)	60.4[1]	58.1[2]	38.1[3]	n/a	n/a	24.9	n/a	n/a
Bandwidth (MHz)	n/a	10	n/a	5.0	No dynamic signals used	10	4.0	3.84
Signal Type		256-QAM		16-QAM		64-QAM LTE	256-QAM	WCDMA
EVM (%)		1.2		n/a		3.5	1.1[4,6]	n/a
ACLR$_1$ (dBc)		−46.1		−36.6		−30.3	−44[4,6]	−47

[1]43 % LO duty cycle [2]Uncorrected LO duty cycle ($\approx 50\,\%$) [3]Maximum output power [4]Estimated from figure [5]EER stack efficiency [6]No DPD

However, other (lower power) DTX work [25] shows that large bandwidths ($> 300\,\mathrm{MHz}$) with good linearity are in reach using a fully-digital approach.

The measured performance of the presented DTX is summarized in Table 7.4 and compared to 'high-power' prior-art solutions. In [84] and [85] fully integrated digital solutions are demonstrated, Diddi *et al.* [83] shows a hybrid digital/analog solution, and McCune *et al.* [19] shows a digital-intensive EER solution. The work presented offers a 10× higher output power compared to these works, while also providing higher (system) efficiency and comparable or better modulated performance. The work presented in [18] can be regarded as a digital-intensive PA and is added for completeness. In contrast to our proposed DTX concept, this solution is not based on a segmented output stage. Therefore, the operation of its two amplifier branches is limited to a constant envelope, and thus requires constant drive power, limiting the achievable system efficiency [86].

Overall, valuable lessons for future power DTX implementations were learned from this study, which we summarize below.

- Use fully saturated operation for the output stage segments, i.e., low g_m. Doing so minimizes the dependency of the DTX bit-to-RF transfer on the actual drive voltage offered to the gate of these segments (V_{GS}). This requires a relatively low V_T for the power stage.
- Avoid mutual inductance in digital-to-RF-power die interconnections to stay away from bit-to-bit interaction.
- Reduce delay variation in the driver chain due to variation of the on-chip voltage ($V_{DD,\mathrm{dr}}$) by adopting the following strategies,
  - Take the CMOS controller supply decoupling extremely serious, even when it is supposed to be "just" a digital circuit;
  - Make the tapered buffer chain as short as possible (lower $t_{p,\mathrm{tot}}$, so lower absolute variation with $V_{DD,\mathrm{dr}}$).

- Use a finer (multilevel) thermometer coding, resulting in better linearity and ENOB (similar to DACs).
- Duty-cycle control should be implemented in a very robust way to avoid jitter; therefore, analog duty-cycle control loops can be best avoided.
- Based on its digital amplitude data, the DTX can go from full power to zero in a single sample, yielding a very rapid change in the supply current, which, due to the always present self-inductance in the DC bias network, can cause a large overvoltage condition on the RF output stage that can be easily damaged. So soft shut-down in a testing phase is essential.

In future implementations, when optimized LDMOS/GaN technology and high-density interconnect (such as high-density flip-chip) are available, the effective DTX resolution (ENOB) and related linearity and bandwidth can be drastically increased by using more and smaller segments. Also, the overall DTX system efficiency would benefit strongly from avoiding inter-die ESD protection, which results in (unnecessary) additional capacitive loading of the CMOS drivers. In such a situation, the overall TX system efficiency will closely approach the drain efficiency of the RF power technology used. Equally important is the almost perfect scaling of its total power consumption with the RF output voltage delivered. Namely, it is this property that will enable significant energy savings when applied in mMIMO base stations. Within the technology constraints for this demonstrator, also encouraging results were obtained for modulated signals, showing an ACLR of −46.1 dBc and an RMS EVM of 1.2 % for a 10 MHz 256-QAM signal, and an $IM_3 \leq -51.4$ dBc for an 80 kHz two-tone. Although these are promising results, more work is still needed. Namely, efficiency enhancement techniques like Doherty or (mixed-mode) outphasing need to be applied to improve average efficiency. Furthermore, to improve linearity, future DTX implementations should offer a higher number of effective bits. This can be achieved by improving the DC-supply decoupling of the digital controller to avoid any V_{GS}–I_{DS} variation, as well as minimizing the g_m in the 'ON' state by even better shaping the $I_{DS}(V_{GS})$ relation of LDMOS or GaN technologies used in the segmented output stage for the available driver voltage swing. Finally, more advanced interconnect schemes need to be applied between the CMOS controller and the segmented output stage, yielding lower interconnect parasitics while allowing more thermometer bits. When these issues are addressed, high-power DTX solutions are ready to take over conventional analog solutions and will offer higher system integration and higher functionality, at lower costs and lower energy consumption.

7.6 High-Power DTX Demonstrator II[2]: Introducing Current Scaling – Digital class-C

This second demonstrator aims to investigate suitable operation classes for segmented power output stages in DTX architectures that target higher RF bandwidths, yielding the theory on current-scaling DTX classes discussed in Section 2.2.2. Emphasis is placed on achieving the best overall performance regarding drain and system efficiency, output power, bandwidth, and ACW-to-output transfers in terms of ACW–AM.

---

[2]This section describes work done in coorporation with Dieuwert Mul, published in [20]: D.P.N. Mul, R.J. Bootsman *et al.*, "Efficiency and Linearity of Digital 'class-C Like' Transmitters," *2020 50th European Microwave Conference (EuMC)*, Utrecht, Netherlands, 2021, pp. 1–4, doi: 10.23919/EuMC48046.2021.9338122. Figures are cited with [20], where applicable.

The use of class-E has the benefit of a theoretical peak drain efficiency of 100%, but there are also some drawbacks. First, class-E is constrained for its operating frequency (Eq. (2.25)), as well as having a relatively large V_{DS} swing relative to its supply voltage bias, hence limiting the achievable output power at a given RF frequency for a practical power device. Secondly, its ACW–AM and ACW–PM transfers are not constant over frequency, due to its reactive loading. This sets an upper limit in handling signals with large modulation bandwidth without using more advanced DPD techniques [62]. Lastly, class-E demands some over-dimensioning of the output stage and overdrive conditions to make R_{ON} sufficiently low. This puts high demand on the drivers used for these power device segments in terms of their output voltage swing and drive current.

A different strategy for a DTX is to always keep the segmented output stage in current-mode operation. This can be achieved by using the saturation region of the segmented power output stage. The resulting drain current is close to a square wave, which is fed to the (harmonic) output matching network. In ideal current-mode DTX operation, the amplitude of this square-wave current is directly proportional to the applied ACW, indicating an intrinsically linear digital ACW-to-RF transfer. The square-wave current resulting from the DTX output, at first glance, suggests the use of inverse class-F operation. However, when considering realistic power devices with non-neglectable output capacitance, it is close to impossible to realize a wideband inverse class-F matching network that can handle the aimed bandwidths of 5G. In contrast, class-B operation/matching uses only short-circuited conditions for its baseband and higher harmonics, yielding relatively easy implementation. Furthermore, these shorts will limit the voltage stress of the output stage and provide excellent decoupling, as such enabling relatively large RF and modulation bandwidths, even for devices with a large output capacitance. These advantages come without any penalty in the operating frequency or linearity, explaining the widespread popularity of class-B operation in commercial applications.

7.6.1 Designing for Class-B Multi-Phase Operation

This design targets class-B operation for which the power output stage should see a resistive load at the fundamental frequency of 1 GHz, and all harmonics need to be shorted. Two power output stage technologies are targeted: a version using LDMOS and a version using GaN. Their matching networks are shown in Fig. 7.36 [87]. The device output capacitance C_{DS} is resonated out using the shunt inductor L_{DC} and then matched to $50\,\Omega$ using a $\lambda/4$ transformer. The even-order harmonic shorts are provided using a $\lambda/4$ shunt short-circuited stub. The expected RF bandwidth using an LDMOS output stage is lower than when using GaN, due to LDMOS having a larger output capacitance. An optional $\lambda/2$ shunt short-circuited stub is included to increase the bandwidth for the LDMOS variant. This stub has a capacitive reactance for frequencies lower than the center frequency and an inductive reactance for higher frequencies. Due to the placement after the $\lambda/4$ transformer, these reactances are inverted, i.e., they appear to the intrinsic device as an inductive reactance for frequencies lower than the carrier frequency and vice versa. This counteracts the off-center reactance of the C_{DS}–L_{DC} resonator, extending the bandwidth.

Complex modulation capability is evaluated next. The output amplitude can simply be set by the active gate width W_G, controlled by the ACW. Phase modulation is achieved using the two DTX line-ups, each provided by their own activation phase from the phase mapper. This causes interaction between the two drain currents, which shows as a reactive
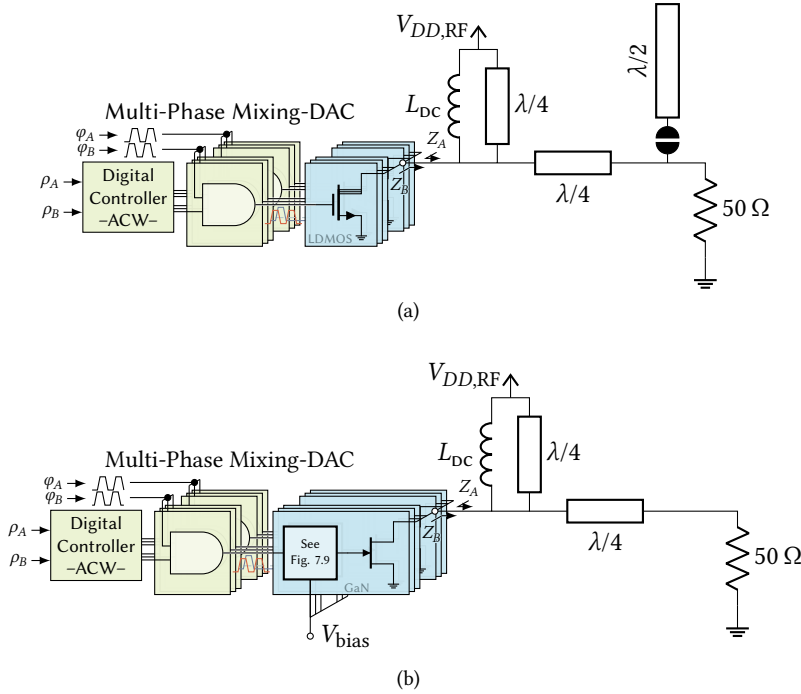
Figure 7.36: Class-B matching networks targeting digital current-mode operation at 1 GHz, implemented as a resonator for the fundamental with a parallel $\lambda/4$ short circuited stub as even order harmonic short and a series $\lambda/4$ transformer for matching the resistive load. In (a) the network for the LDMOS implementation is shown, which has an optional $\lambda/2$ line to extend the bandwidth. The GaN implementation in (b) has a lower device output capacitance, which does not need this.

load impedance to each power device's drain. The resulting load impedances to the intrinsic device (using Eq. (5.20)) are displayed in Fig. 7.37. Here, it shows that using signed-Cartesian operation (Fig. 7.37a, with a phasor offset $\phi_{IQ} = 90°$), the impedance seen by the I-branch is pulled to the inductive region by the Q-branch activation when increasing the DTX output phase from 0° to 90°, whereas the Q-branch observes the I-branch to impose a capacitive load. This phasor interaction causes a drop in drain efficiency (see Fig. 2.5), which is most severe when both branches are equally active (i.e., $\theta = 45°$). In a multi-phase system the phasor offset is reduced. The resulting impedances are shown for 8-phase multi-phase ($\phi_{AB} = 45°$) in Fig. 7.37b, when increasing the DTX output phase from 0° to 45°. The A- and B-branches (as indicated in Fig. 7.36) still see reactive loads under each other's influences but are strongly reduced compared to the signed-Cartesian operation, resulting in a drain efficiency closer to polar operation.

The individual branches are connected to a PCB using bond wires, where the shunt inductor is implemented as a short-circuited stub on the PCB. These bond wires are EM simulated (Fig. 7.38) for their influence on the output matching and the PCB transmission lines were adjusted accordingly.
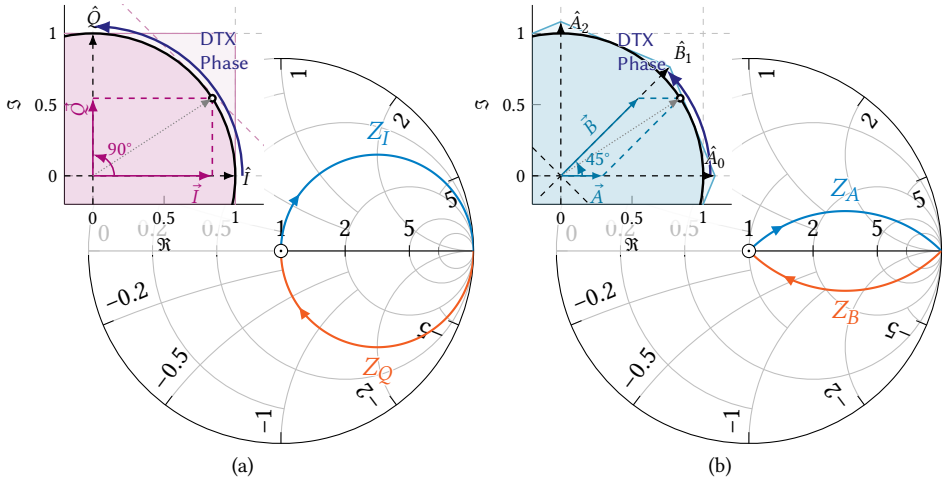
Figure 7.37: Impedances seen by the intrinsic devices in (a) signed-Cartesian system and (b) 8-phase multi-phase systems. The device interaction leads to reactive loading, thus lowering drain efficiency over modulation phase. The interaction can be lowered by decreasing the relative activation phasor angle from 90° to 45°, increasing the drain efficiency.
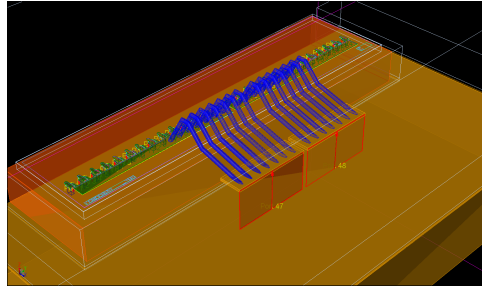
Figure 7.38: A 3D view of the bond wires from the GaN die to the PCB, for use in FEM simulation.

7.6.2 Duty-Cycle Reduction and Linearity

Even though the drain efficiency is now less impacted by the activation phasors, it is still limited to a maximum of 63.6 % due to the class-B duty cycle of 50 %. Fortunately, the CMOS controller features a duty-cycle-adjustment loop (Section 7.4.5). A reduced duty cycle provides a higher efficiency (see Section 2.2.2 and Eq. (2.18)). This effectively moves toward a 'digital class-C' operation, for which we can analyze its linearity over output amplitude and compare it to analog class-C operation.

For a fictitious ideal active device with a linear V_{GS}-to-I_{DS} relation for $V_{GS} > 0$, ($I_{DS} = 0$ for $V_{GS} < 0$, Fig. 2.10a), the analog class-B operation with a conduction angle of $\alpha = \pi$ results in a perfectly linear signal transfer [30]. Moving towards class-C yields gain expansion and, consequently, distortion of the output signal. This linearity degradation in the analog class-C operation, even in an ideal device, can be intuitively understood by considering that the input signal (V_{GS}) first needs to reach the threshold voltage before any activation happens.

(a) Analog class-C ($\alpha = \pi/2$)   (b) Digital "class-C like" ($d = 25\,\%$)
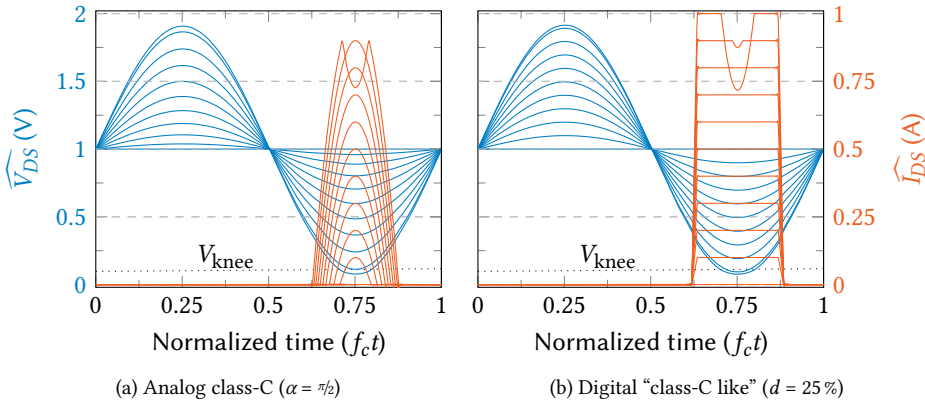
Figure 7.39: Time domain waveforms for varying input quantity, all normalized. The analog case shows an expanding conduction angle with increasing activation, while in digital operation the duty cycle remains constant [20].

Once the device threshold voltage is reached, not only the instantaneous amplitude of I_{DS} linearly changes with V_{GS}, but also its effective conduction angle, as shown in Fig. 7.39a. The latter issue yields an extra push in fundamental output power and shows up as (nonlinear) gain expansion. In contrast, the DTX effectively uses dynamic scaling of its power device width by activating the segments proportional to the applied ACW, while the segment drive is the same for all segments. As the drive voltage waveforms are independent of the ACW, this yields perfect linearity as shown in Fig. 7.40, even for a nonlinear (e.g., quadratic) V_{GS}-to-I_{DS} relation, under the condition that the active device remains in the current-mode region. It is worth mentioning that the analog class-B ($\alpha = \pi$) case, using a perfect square-law device (biased for $g_{m3} = 0$), is special and can still provide linear amplification [88–90]. For all operation modes, at larger signal excursions, compression starts to occur when the load line reaches the triode region of the output stage.

7.6.3 Measurement Results

The realized physical design is used to verify the forgoing theory. A photograph of the realized hardware is given in Fig. 7.41. For this purpose both TX line-ups have been synchronized (i.e., the same activation phase) such that they effectively act like one larger unified segmented output stage. The duty cycle of the activation pulse can be varied between 28 % and 51 %.

The LDMOS output stages' drains are biased at 28 V, the thick-oxide CMOS drivers and the digital controller use a 2.5 V and 1.1 V supply, respectively. Pulsed RF envelope operation with a 10 % envelope duty cycle is used to avoid excessive heating, while applying the described RF pulse duty-cycle control, at 930 MHz. A spectrum analyzer is used to measure the in-pulse RF output by doing a zero-span measurement using large radio and video bandwidths (RBW and VBW). The absolute read-out of the spectrum analyzer is referenced to a power meter. In addition, the average currents and voltages of the power supplies are used to obtain the drain and system efficiencies according to their definitions (Eqs. (6.5) and (6.8)), in which the latter includes the total power consumption of the digital
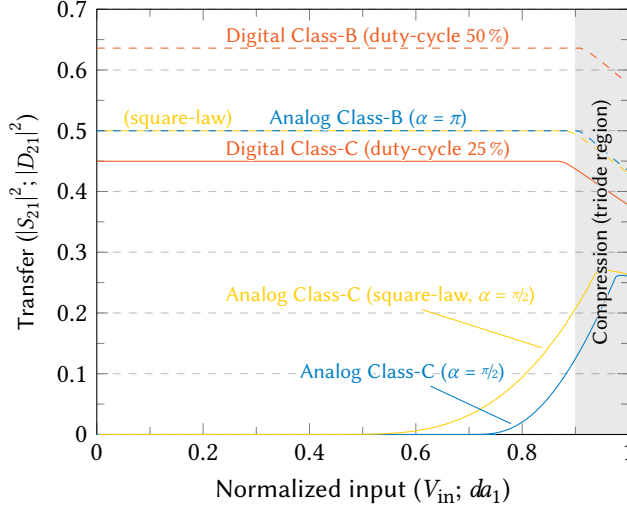
Figure 7.40: Transfers of analog and digital transconductance/current-scaling classes, using normalized input and output quantities ($I_{DS,\mathrm{max}}$ = 1 A, $R_L = R_{L,\mathrm{opt}}$, and P_{norm} = 0.5 W). Solid lines show class-C operation ($\pi/2$ conduction angle/25 % duty cycle), clearly indicating gain expansion in the analog cases.

controller and the CMOS drivers, corrected for the envelope pulse duty cycle used in the pulsed measurement. In Fig. 7.42a, the measured drain and system efficiencies vs. RF duty cycle are given, Fig. 7.42b provides the measured drain efficiency vs. output power, and Fig. 7.42c gives the measured DTX transfer, which is proportional to the fundamental drain current I_{DS}/ACW. As one can observe, a maximum drain and system efficiency for the LDMOS device of respectively 75.7 % and 72.9 % is achieved at 25.9 W output power, while the expected linear transfer can be clearly noted until triode region related compression of the output stage sets in. Note that the high system efficiency is a direct consequence from the digital nature of the proposed DTX approach, which consumes practically no static power, while its dynamic power consumption is dominating and proportional to the ACW (e.g., Eq. (6.29)).

Dynamic measurements using a narrowband two-tone are performed to also verify the linear transfer using modulated signals. The resulting ACW–AM/ACW and ACW–PM are shown in Fig. 7.43, together with the output spectrum. These show that the ACW–AM/ACW transfer is indeed linear, while the ACW–PM dominates due to the nonlinear LDMOS C_{DS}. Without any correction an $\mathrm{IM}_3 < -31.1$ dBc was measured, which improves to -48.6 dBc when static look-up table (LUT) calibration is used. When an 8.8 MHz 256-QAM signal at 1.125 GHz is applied, using only the linear region, measurements show an ACLR of -36.3 dBc without DPD, -43.7 dBc with static phase correction only, and -48.3 dBc with static LUT calibration, combined with an EVM of 3.0 %, 1.5 %, and 1.0 %, respectively. The measured output spectra are shown in Fig. 7.44. These measured ACLR levels again show that the ACW–PM nonlinearity is dominant.
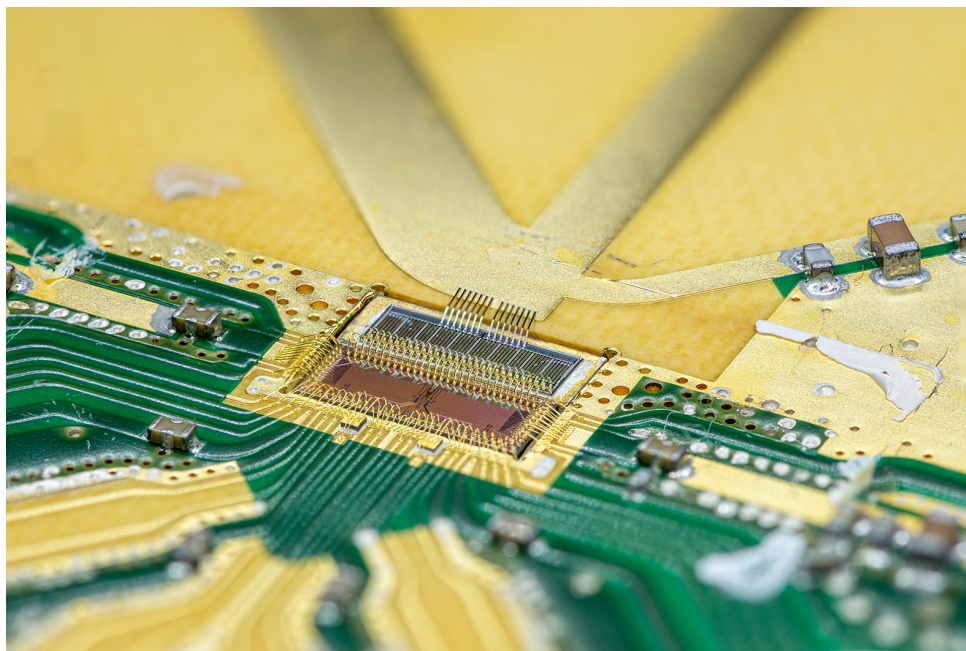
Figure 7.41: Photograph of the bits-in–RF-out, high-power DTX featuring the segmented LDMOS output stage with the class-B output match (Fig. 7.36a). To the left is the $\lambda/4$ stub for even order harmonic termination, in the middle the $\lambda/4$ transformer, and to the right the short-circuited inductive stub.

(a)

(b)

(c)
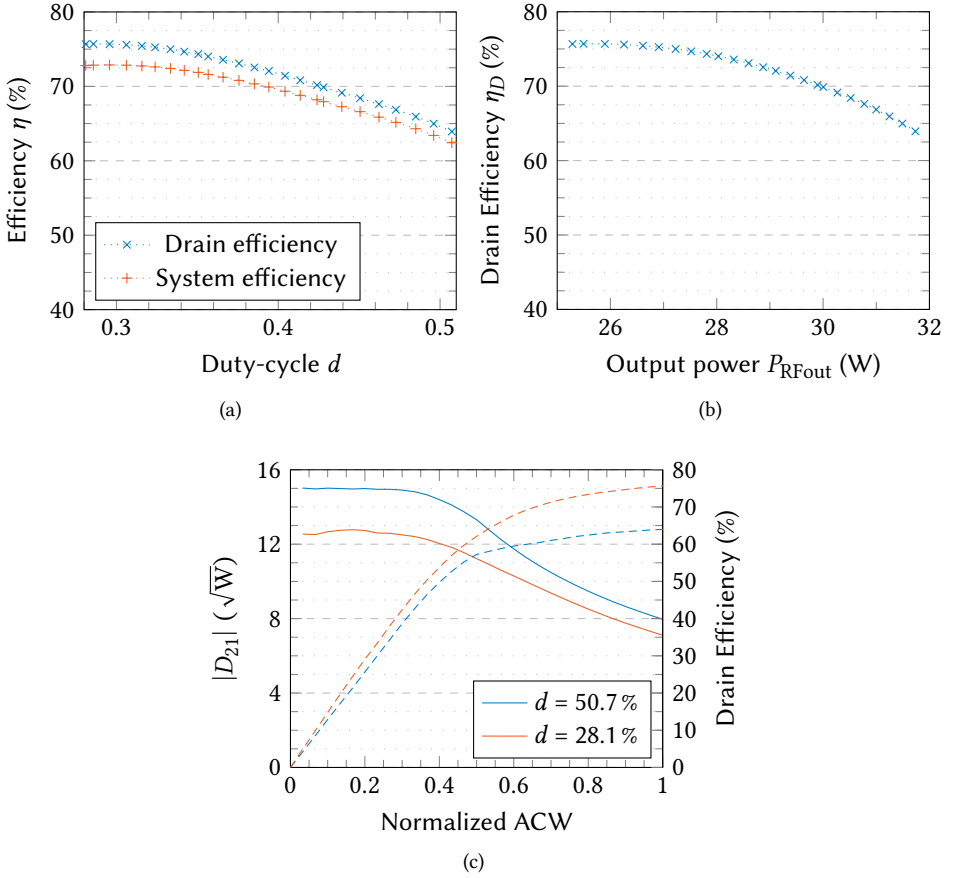
Figure 7.42: Pulsed envelope RF measurements (10 % envelope duty cycle) using digital "class-C like" operation at 930 MHz with segmented LDMOS power devices: (a) Drain and system efficiency vs. RF duty cycle; (b) Drain efficiency vs. output power for an RF duty cycle ranging from 30 % to 52 %; (c) Normalized digital forward transfer (dashed: drain efficiency) [20].

(a)

(b)

(c)
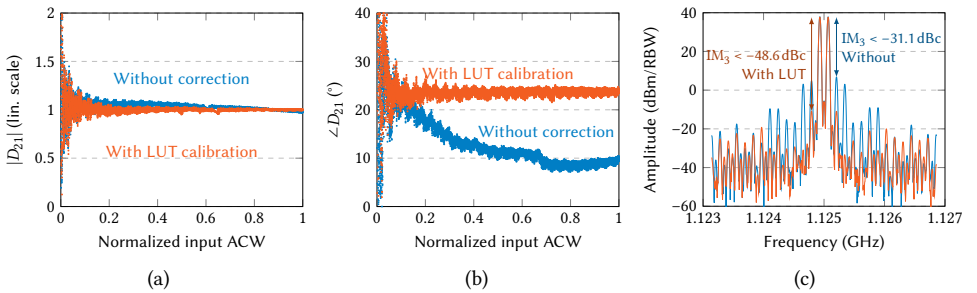
Figure 7.43: Dynamic ACW−AM/ACW and ACW−PM transfers, and the output spectrum of the digital class-C setup using narrowband two-tone signals. The blue curves show the transfer and spectrum without any correction, whereas for the orange curves (static) LUT calibration is used.
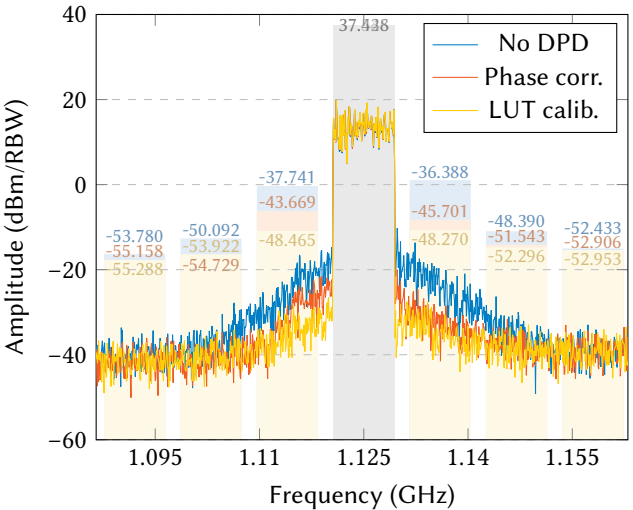
Figure 7.44: Output spectrum of the digital class-C setup using 8.8 MHz QAM signals, with annotated the measured channel power and ACLR levels [20].

7.7 High-Power DTX Demonstrator III: Wideband Digital Class-C Doherty

The previous two DTX demonstrators have focused on achieving high DTX RF output power and peak efficiency, offering promising benefits over traditional analog TXs: they do not require any input impedance matching or quiescent currents for their output stages. Furthermore, they completely eliminate stability issues and are flexible in their activation profile and output matching. Consequently, the DTX concept can offer very high system efficiencies, especially at PBO, while its RF BW is only limited by the design of the applied output matching network. However, neither demonstrator offers enhanced efficiency in PBO. With this demonstrator, we target the world's first fully-digital high-power Doherty transmitter using the existing CMOS and LDMOS hardware.

After some demonstrator I class-BE samples were damaged during measurement anyway, potentially destructive testing of the drivers could be performed without further penalty. Its power supply voltage $V_{DD,dr}$ was increased beyond its nominal (2.5 V) and maximum (2.75 V) values to find the breakdown limits. It was found that no short-term degradation occurred when scaling the $V_{DD,dr} > 3.5$ V. As such, we have decided that it should be safe enough to design this demonstrator for $V_{DD,dr} = 3.0$ V, which makes sure the LDMOS output stage operates closer to its $I_{DS,max}$. This way, the influence of the CMOS–LDMOS bond wires' mutual inductance on the LDMOS drain current is lowered, resulting in less bit-to-bit interaction and, thus, increasing the effective resolution of this demonstrator.

7.7.1 Design of the Harmonic Output Match

Digital class-C operation is selected as this demonstrator's operating class for the Doherty DTX (DDTX) output stages; its capability to handle a relatively large output capacitance, its linear operation, limited voltage swing, and excellent efficiency–output power relation motivated this choice. Lowering the duty cycle in digital class-C operation increases the efficiency at the cost of RF output power. For the Doherty DTX (DDTX) hardware used, the minimum achievable duty cycle is 40 % at the targeted RF operating frequency of 2.0 GHz by using the on-chip duty-cycle adjustment. This sets the maximum theoretical achievable efficiency for these conditions to 75.7 %, while providing 21 % higher output power compared to analog class-B, assuming an ideal device and lossless output matching network (see Section 2.2.2). Similar to the analog transconductance classes, digital class-C operation demands close to short-circuited conditions for its higher harmonics. For an activation duty cycle less than 40 %, the resulting internal LDMOS current waveform will be dominated for its higher harmonics by the 2nd and, to a lesser extent, the 3rd harmonic. However, realizing wideband short-circuited conditions for the 2nd harmonic without reducing the bandwidth of the fundamental matching is challenging. Therefore, in this design, the use of explicit 2nd harmonic shorts at the drains of the active devices has been omitted. To achieve wideband operation, shunt inductors have been used at the drains of the switch banks to resonate out their (7.9 pF) output capacitance around the center frequency of 2 GHz. The relative high C_{DS} of the switch bank capacitances tend to approximate AC short-circuited conditions for the higher harmonics.

Inverted Doherty Power Combiner
The inverted Doherty topology has been selected for its improved power and efficiency bandwidths in power back-off operation over the classical Doherty topology [42]. Since
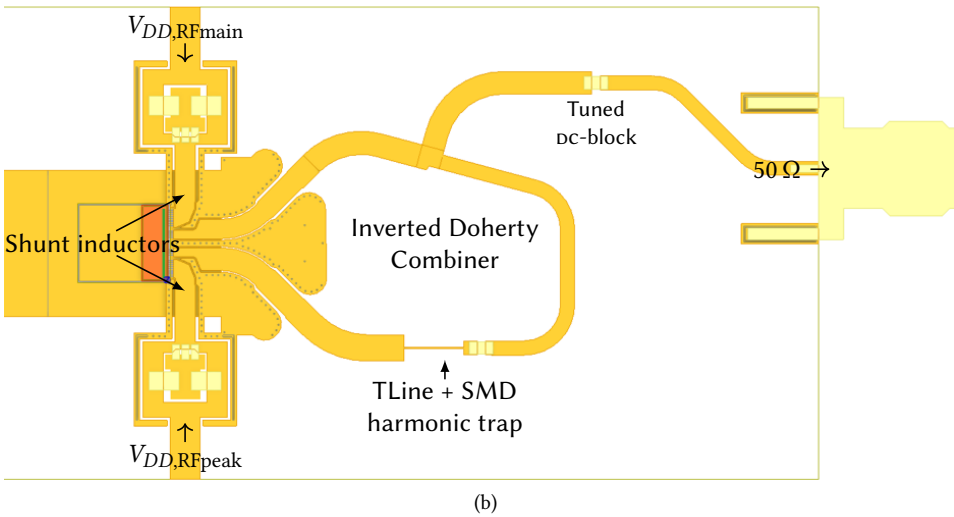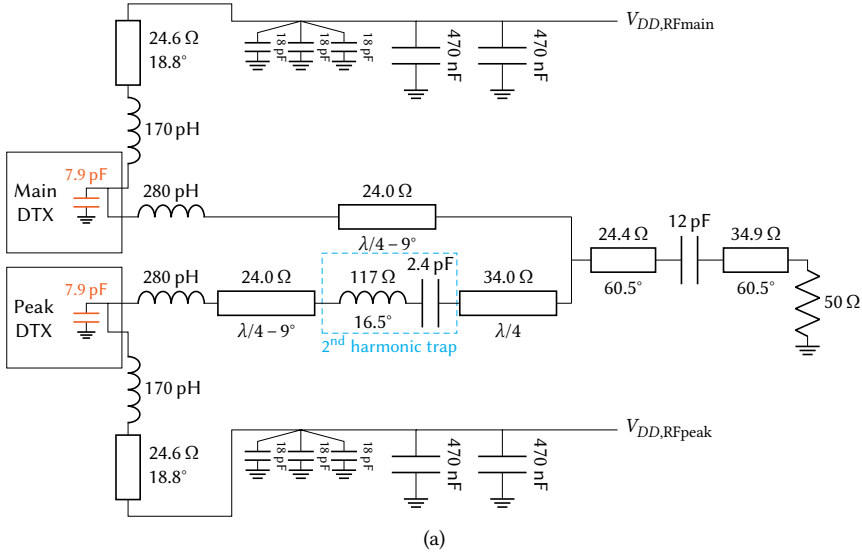
Figure 7.45: Schematic (a) and layout (b) of the proposed inverted Doherty power combiner featuring a low-Q 2nd harmonic trap in the peak path to guarantee smooth output power and efficiency vs. frequency. The second harmonic trap is implemented by a high characteristic impedance transmission line section and an SMD capacitor.

both switch banks have the same total $W_{G,\text{tot}}$, a symmetric inverted Doherty is designed. The DDTX switch banks have a nominal load impedance of $R_L = 17\,\Omega$ with a supply of $V_{DD,\text{RF}} = 28\,\text{V}$. The schematic and layout used for the inverted Doherty power combiner is shown in Fig. 7.45. The bond wire configuration was EM simulated, and its extracted effective values (see Section A.4.2) are shown in the schematic. The inductance of the bond wires to the power combiner is absorbed in the connecting $\lambda/4$-lines lines by slightly shortening the lines. The PCB also provides the bond-wire-connected stubs that are used to feed the DC bias and resonate out the C_{DS}. The number of bond wires for each connection (either the power combiner or the shunt inductor) is selected based on the fundamental current's magnitude. The shunt inductor resonates with the C_{DS} in a relatively low impedance environment, leading to a higher fundamental current magnitude, thus more bond wires are allocated for that connection to minimize loss. The mutual inductance between each connection is only $9\,\text{pH}$ ($k_m \approx -0.05$) by providing enough spacing between them, and can therefore be considered negligible. Between the two DDTX branches is more space, while also having a grounded via fence between the two branches. This avoids unwanted coupling between the two DDTX branches. The $\lambda/2$-line of the inverted Doherty is split into two $\lambda/4$-lines with different impedances to further enhance the RF bandwidth [42, 43]. Finally, the output of the inverted Doherty combiner is matched to $50\,\Omega$ using a two-section impedance transformer, including a DC blocking capacitor.

Second Harmonc Trapping

The main and peak DTXs are driven by signals that are delayed $90°$ from each other at the center frequency. The total electrical length between the main and peak DTXs is $3\lambda/4$ for the fundamental at the center frequency, fulfilling the requirement for the Doherty load modulation. In Figs. 7.46c and 7.46d, the fundamental impedance seen by the main switch bank is shown at full and back-off powers. However, since there is no explicit 2nd harmonic short applied, the C_{DS} susceptance of the peak DTX at the 2nd harmonic is transferred by the $3\lambda/4$ network and directly seen by the main DTX. For frequencies slightly higher than the center frequency of the design, the transformed C_{DS} susceptance of the peak DTX yields a large inductive susceptance, which is in parallel with the output capacitance of the main DTX. At this frequency, it results in an undesired parallel resonance in the 2nd harmonic termination of the main DTX (see Fig. 7.46c). This leads to a sharp increase in the 2nd harmonic of the main, causing a sharp dip in output power and efficiency (see Figs. 7.46a and 7.46b). To prevent this phenomenon and avoid degradation of the usable bandwidth of the DDTX, a 2nd harmonic trap was added by placing a series LC resonator in the peak path. This resonator is designed to have a low Q-factor ($Q \approx 1$) at the fundamental to not degrade the bandwidth of the DDTX. By placing it after the first $\lambda/4$-line, the undesired 2nd harmonic parallel resonance for the main path no longer occurs (see Fig. 7.46d), while it has the added benefit of also providing a short for the 3rd harmonic for the peak DTX.

7.7.2 Activation Pattern

As the bond wire structure and the matching network have been EM simulated, we can also use this information to select a suitable LDMOS variant. LDMOS variant TUD-03 was chosen (Fig. 7.5c) as it provides a slightly higher efficiency, while the inside MSBs suffer less from mutual coupling. The activation pattern then remains as a free variable.

This section focuses on the influence of the physical location of an LDMOS segment on
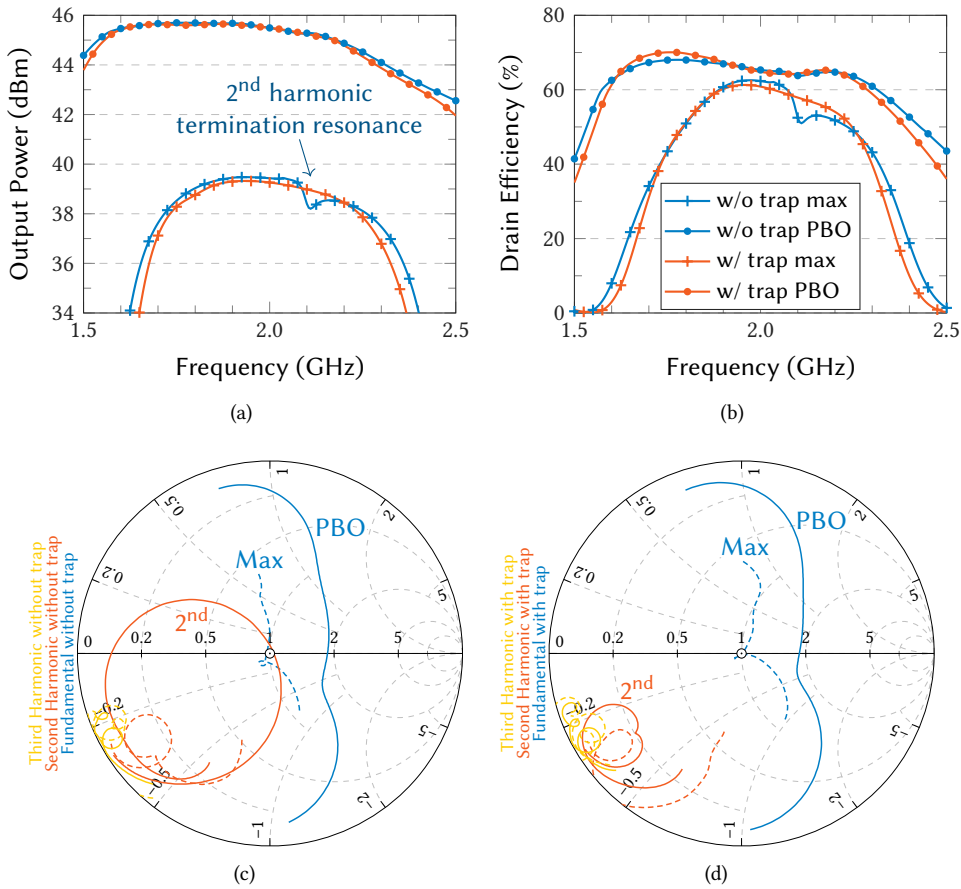
Figure 7.46: Simulated performance of the DDTX on a schematic level, showing the impact of the harmonic trap: (a) comparing the output power and (b) efficiency vs. frequency, showing the dip for PBO; (c) the impedance seen by the main DTX without the harmonic trap, for the fundamental, and second and third harmonics. The second harmonic impedance shows a resonance, causing the dip in power and efficiency; (d) the impedances with harmonic trap, showing that the harmonics now see a lower impedance.
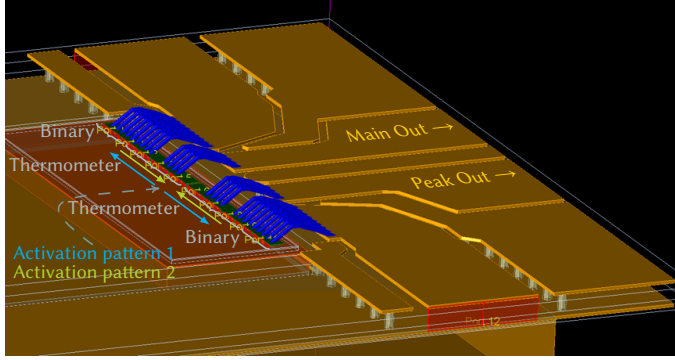
(a)
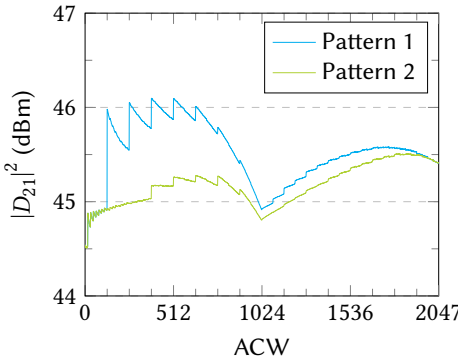
(b)
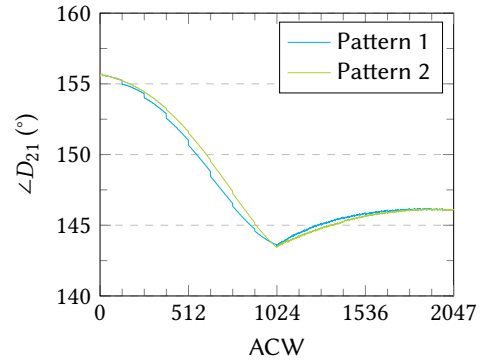
(c)

Figure 7.47: A detail of the output bond wires with the two possible activation patterns illustrated: either the inside or the outside unary-weighted cells first. Depending on physical location the LDMOS segments will see a different matching condition, and thus leading to different (simulated) transfers.

the DDTX's transfer. As such, LSB dummy LDMOS C_{GS} mismatches are not considered, and neither are input bond wire mutual inductances. Two activation patterns are possible, as illustrated in Fig. 7.47a: the unary-weighted MSBs either start activating on the inside of the die, towards the binary-weighted LSBs (pattern 1), or vice-versa (pattern 2). Pattern 2 most closely resembles the activation pattern used for the class-BE demonstrator (Section 7.5) because the MSB activations start closest to the LSBs. Simulating these patterns using a harmonic balance simulation gives us the effect of layout on the static transfers, which are given in Figs. 7.47b and 7.47c. Here, it can be seen that pattern 2 gives a smoother transfer since the current from the first MSB to be activated takes a very similar path as the current provided from the LSBs. In other words, the first MSB and the LSBs have very similar matching conditions due to their proximity. Hence, activation pattern 2 is chosen to translate the ACW to values stored in the SRAM for the measurements.

(a) (b)
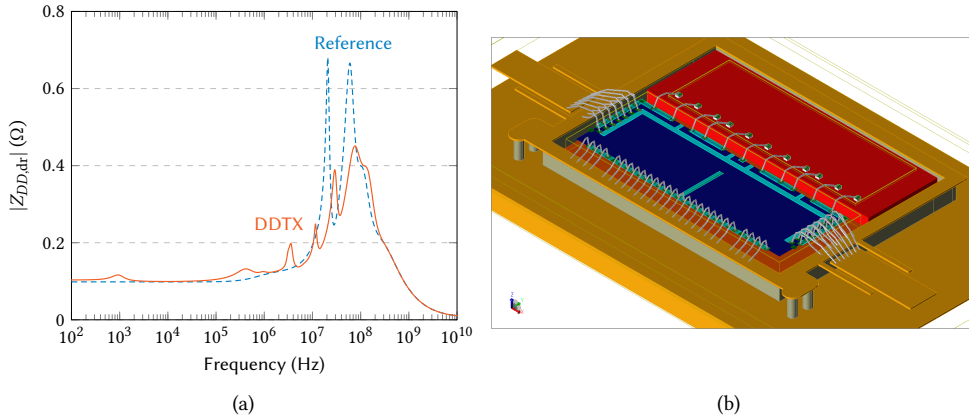
Figure 7.48: In (a) the simulated impedance $Z_{DD,\mathrm{dr}}$ seen into the $V_{DD,\mathrm{dr}}$ supply, using lumped equivalents for capacitors on the PCB and transmission line equivalents of the traces, and (b) a 3D view for the EM simulated $V_{DD,\mathrm{dr}}$ supply paths and its ground return paths used for the simulation. The reference design is Fig. 7.49a. The improved decoupling for the DDTX uses the design of Fig. 7.49b, showing lower resonance impedance magnitudes.

7.7.3 Improved Supply Decoupling

One key takeaway from the first demonstrator (Section 7.5.3) was taking the CMOS controller supply decoupling extremely seriously. Even though the CMOS controller's hardware is as-is, we can make an effort to improve the off-chip decoupling. We know from earlier measurements that a 16 MHz two-tone resulted in a large supply variation, indicating a nearby supply resonance. To improve this, first a suitable model for the DTX's supply impedance $Z_{DD,\mathrm{dr}}$ needs to be found. The DC supply path is EM simulated, of which a 3D view is shown in Fig. 7.48b. Several ports at the drivers' locations are placed inside the CMOS die, and the input ports are placed on each side of the PCB. The total on-chip capacitance (16.3 nF) is distributed over the ports on the chip, with individual time constants varying between 2.9 ps and 9 ps. The discrete surface-mounted device (SMD) capacitors on the PCB are modeled using SPICE netlists provided by their manufacturer, where the PCB traces are approximated using microstrip and conductor-backed coplanar waveguide (CBCPW) transmission line equivalents based on the PCB substrate, and trace length and width. This provides the blue impedance curve $Z_{DD,\mathrm{dr}}$ of Fig. 7.48a, which is the impedance seen into one of the on-chip ports using the approximated PCB layouts of demonstrators I and Ⅱ shown in Fig. 7.49a, and functions as a reference. A sharp impedance peak is found at 20 MHz in this model, which is the most likely cause of the supply variations measured before at 16 MHz. A second resonance can be seen at 60 MHz, which is also present when the ports at the edge of the PCB are AC shorted. This resonance seems unavoidable due to the on-chip capacitance and the bond wire structures. That means, however, that the resonance at 20 MHz can be avoided by proper PCB design.

Several improvements were made to the DDTX PCB design, which is shown in Fig. 7.49b. First, all components have now been annotated on the design to ease communication and avoid the risk of assembly errors. To improve the decoupling, the most notable are C1 and C2, which are low series inductance three-terminal feed-through capacitors of 0.1 µF. These capacitors have now been placed as close as possible to the PCB bond pads for $V_{DD,\mathrm{dr}}$, namely almost 1 mm closer, while keeping enough spacing to have some solder
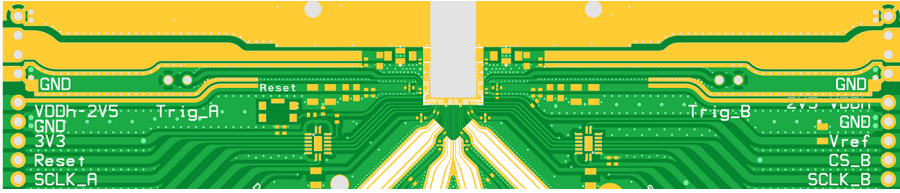
(a) Reference PCB design (as used in demonstrators I and II)
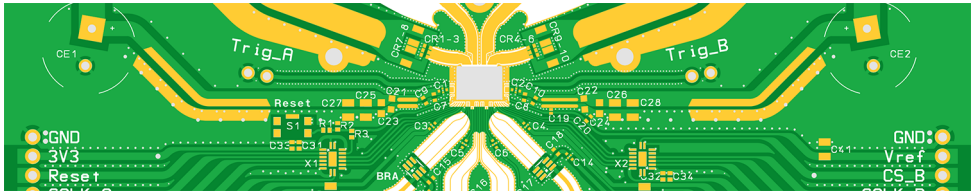
(b) DDTX PCB design

Figure 7.49: Improvements on PCB level for the DC decoupling of the DDTX's $V_{DD,\mathrm{dr}}$ domain. Most notably the three-terminal capacitors C1 and C2 have been moved closer to the bond pads, C7–C10 have been added close-by with lower characteristic impedance traces, and dedicated space is reserved for the (big) electrolytic decoupling capacitors CE1–CE2.

mask between them to avoid contamination of the bonding surfaces due to soldering. The PCB material thickness is halved with respect to the reference PCB design for layouting of the inverted Doherty combiner, which is also beneficial for lowering the characteristic impedance of the $V_{DD,\mathrm{dr}}$ feed lines. Next, the flange design has also been improved for this design, as shown in Fig. 7.50, which allows the CMOS and LDMOS dies to be placed closer together. This shifts the unavoidable resonance of this design up to 75 MHz, as seen in the orange impedance plot of Fig. 7.48a. The next capacitors (C7–C10) have been placed close by and have increasing effective capacitance values to keep the supply path impedance low. These result in new resonance peaks at 29 MHz and 11.5 MHz, but with a strongly reduced impedance magnitude compared to the reference case. From there, the capacitor sizes keep increasing until the big through-hole electrolytic capacitors CE1–CE2 and the board edge. These add some new resonant peaks but at a near-negligible magnitude. Also, the board connector itself was improved to lower DC series resistance, from pin headers to a screw terminal that could accommodate thicker wires. These effects are not explicitly included in the simulation model of Fig. 7.48a.

7.7.4 Measurement Results

A photo of the realized DDTX design is shown in Fig. 7.51. The DDTX was measured using external signal generators capable of changing the mutual phase relations for the digital sampling clock, and the RF activation clocks for the main and peak DTXs. The amplitude information programmed in the DTX controllers' memory has a time duty cycle of 10 % to prevent excessive thermal heating. The resulting output powers have been measured using a Keysight U8488A power meter, with the losses of a 30 dB high-power attenuator de-embedded.

First, the performances in terms of output power and efficiency for full power and power back-off operations are evaluated in Fig. 7.52. Comparing these measurements with
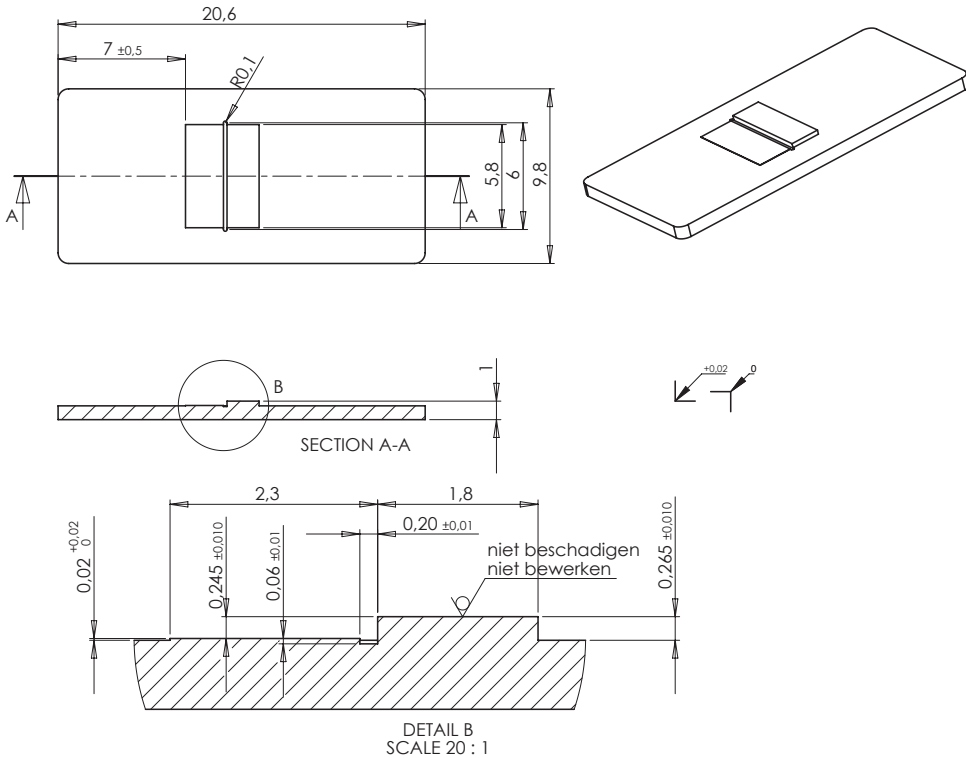
Figure 7.50: Improvement of the flange, featuring a small undercut/cavity on the edge between the CMOS and LDMOS dies. This allows the CMOS die to be placed closer to the LDMOS' plateau without lifting up due to the milling tool's radius, while providing an alternative path for the silver epoxy glue to prevent it from creeping up due the capillary formed between the copper wall and the CMOS die. This modification provides easier assembly, while also reducing the CMOS–LDMOS interconnect distance and, thus, inductance.

a 3D EM simulated version of the output network show that there has been some significant downshift due to some deviations in its actual assembly, impacting the realized efficiencies. Unfortunately, the three capacitors functioning as RF shorts cannot be moved easily for tuning. Nonetheless, at 1.66 GHz, a peak drain efficiency of 60 % was measured at an output power of 34.2 W. With only the main DTX activated, optimum operation was achieved at 1.77 GHz. The performance over the output power range for this frequency (Fig. 7.52b) shows a peak drain efficiency of 57 % with an output power of 39.1 W. At 6.6 dB power back-off this is 52 %, an 25 %pt. improvement over the case without Doherty efficiency enhancement. Relative to these levels, −1 dB-power bandwidths of 430 MHz and 440 MHz for respectively peak power and power back-off are achieved.

Similarly, for the −10 %-drain efficiency bandwidths these are 590 MHz and 370 MHz. System efficiencies are close to the drain efficiencies, as the controllers' dynamic power consumptions scale with the output stages' activation (between 0 W and 3.47 W), and the static power consumption is only 191 mW (such as SRAMs and the clock tree). In Fig. 7.53, the system efficiencies are 55 % and 48 % for respectively peak power and power back-off. In fact, over the entire measured operating range, the system efficiency remains within

(a)   (b)
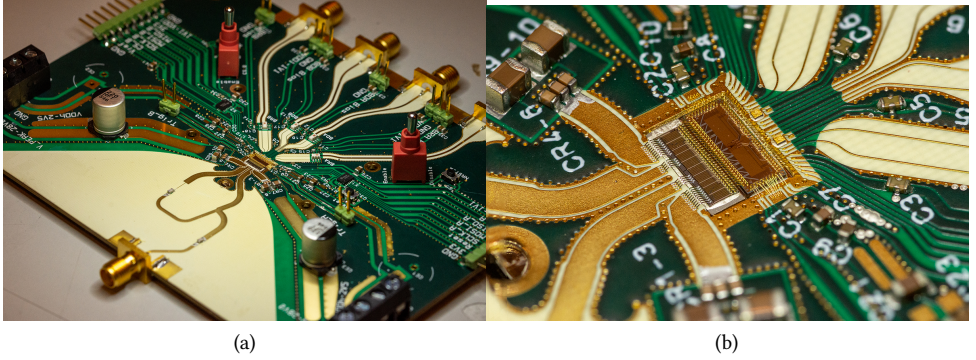
Figure 7.51: Photograph of the realized Doherty combiner on PCB and a detail photo of the die assembly.

(a)   (b)
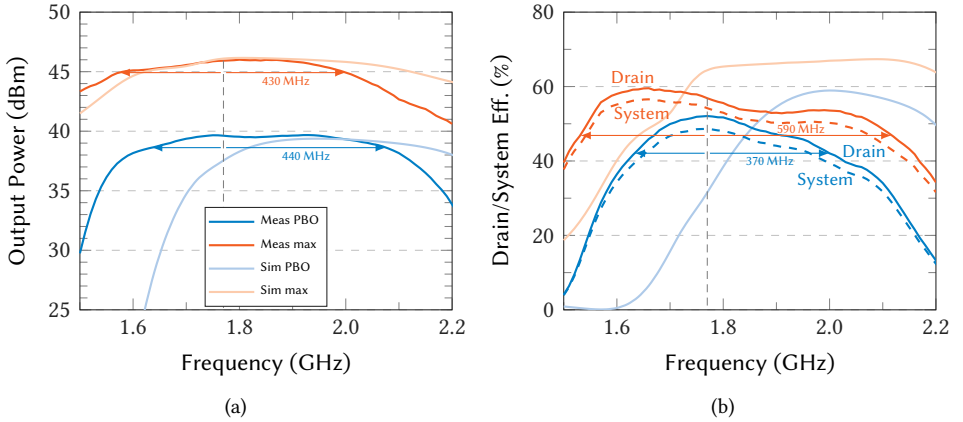
Figure 7.52: Measurement results of the DDTX compared with the 3D EM simulated design. In (a) the power and in (b) the efficiency bandwidths at peak power and power back-off of the DDTX are given. Compared to the EM simulated design, it is clear that the realized design has shifted down in frequency.

4 percentage points of the drain efficiency.

The phase relationship between the main and peaking DTX branches should nominally be 90° at the center frequency f_0, but can be freely modified during measurement thanks to the dual line-up. This yields two optima: a value for maximum peak output power or a value for maximum peak efficiency. All curves in Figs. 7.52–7.54 assume the value for maximum peak efficiency. In the measurement setup, non-phase-matched SMA cables were used, so no conclusions can be drawn from the input phase regarding the actual applied phase difference between the two branches of the DDTX. Different phase relations can also be applied when sweeping over frequency to compensate for the changing electrical length over frequency of the inverted Doherty power combiner. The applied peaking DTX phase is then provided by

$$\phi_p = \phi_{p0} + \alpha_\phi \left(f_c - f_0 \right), \tag{7.1}$$

where ϕ_0 is the phase applied at the RF carrier or center frequency f_c, and α_ϕ relates the
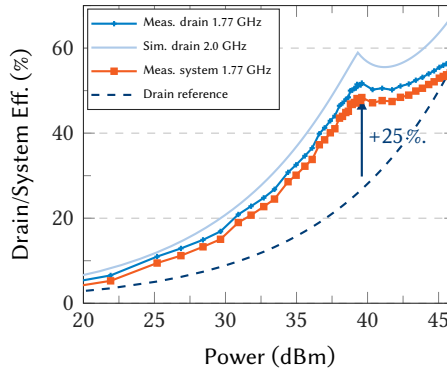
Figure 7.53: The measured and simulated efficiencies vs. output power at f_c = 1.77 GHz, showing efficiency improvement in power back-off by 25 percentage points with respect to a situation with the same peak efficiency but without efficiency enhancement.

rate of the phase change vs. frequency. The resulting power and efficiency measurements are given in Fig. 7.54 for two different samples. The first sample is similar to the sample measured in Fig. 7.52, and the second sample has been modified to accommodate moving the capacitors functioning as RF shorts closer to the die in an effort to restore operation closer to the design frequency of 2 GHz. Unfortunately, this did not result in a higher drain efficiency, most likely due to these capacitors' increased effective series resistance. However, from these measurements, it becomes clear that changing the DDTX's phase relations can enhance the efficiency over the bandwidth.

The effectiveness of the improved supply decoupling is first tested using two-tone signals without any correction or DPD. Again, the on-chip supply voltage is monitored through the TRIG pin outputs, similar to the measurements in Section 7.5.2. The two-tone spacing is swept, and the resulting supply voltage variation is analyzed statistically in Fig. 7.55. Three time-domain waveforms for different frequencies are provided in Fig. 7.55a. The relatively large swing at 1 MHz is unexpected compared to the simulated impedance magnitude of Fig. 7.48a, but it could also be an artifact of the unknown voltage transfer of the TRIG pin driver, bond wire, PCB traces, and the oscilloscope probe. Namely, at 1 MHz, nothing in the measured *IQ* diagram hints at supplying decoupling-related hysteresis. Nonetheless, note that the scale of the *y*-axes in Fig. 7.55 is 150 mV, whereas in Fig. 7.32 it was 300 mV. The measured peak-to-peak voltage swing has decreased significantly, indicating that the effort to improve the supply decoupling does seem to pay off.

The capability of the DDTX to handle modulated signals is demonstrated using a piecewice[3] polynomial fitted LUT for the AM and iterative learning control (ILC) DPD for the PM with a 7 MHz 256-QAM signal around 1.77 GHz with a PAPR of 5.5 dB. This initially achieved an ACLR of < −52.0 dBc and an EVM of 0.4 %, as shown in Fig. 7.56a. In this operation, average drain and system efficiencies of respectively 49 % and 46 % were measured. This result was published in [80], but a timing misalignment between the amplitude and phase paths became apparent upon closer inspection of the full spectrum. This yielded a modulation of the phase error due to the used DPD loop, but away from the

---

[3]One polynomial piece for the region with the peaking DTX inactive, one for when the peaking DTX is activated.
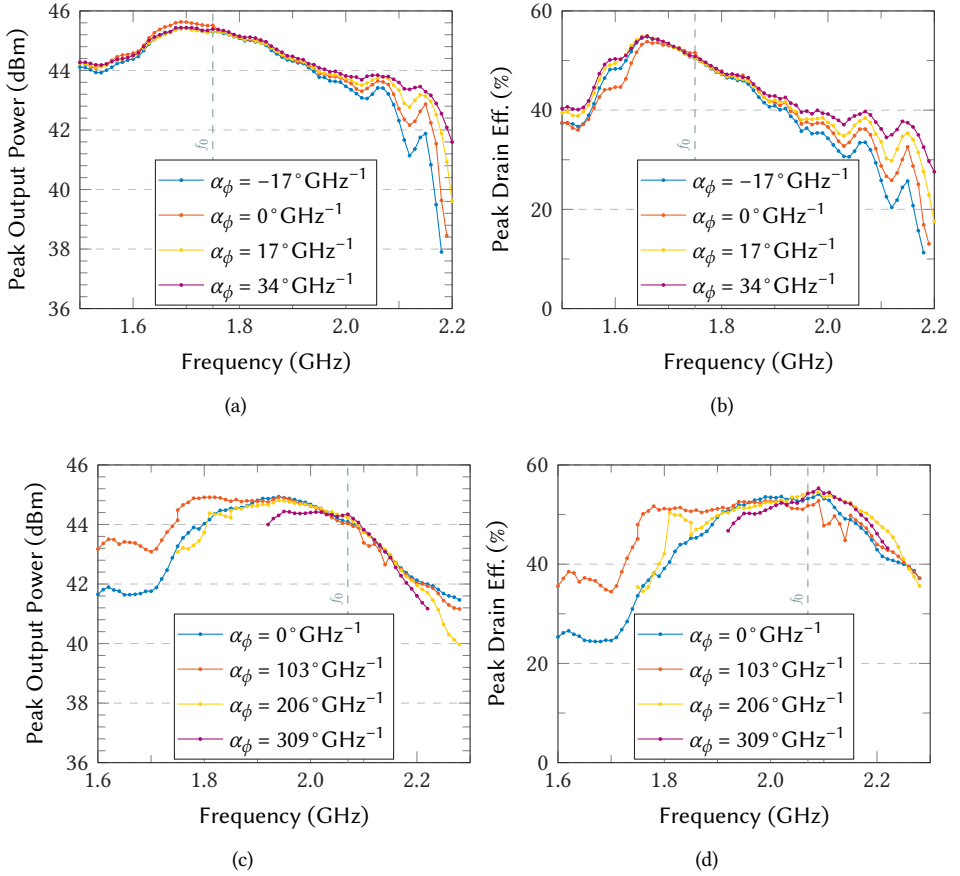
Figure 7.54: Measurements with varying the phase relation of the peaking amplifier over frequency for two different samples. The first sample (a&b) operates around $f_0 = 1.75$ GHz and the second sample (c&d) has the RF shorts moved to tune the shorted stub be closer to the design frequency, using $f_0 = 2.07$ GHz.

carrier such that the close-by spectrum and in-band were clean. In later measurements, after publication, this timing alignment was improved. This improved the ACLR and EVM to < -53.0 dBc and 0.3 %, respectively. This improvement is only slight, but the out-of-band spectrum is much cleaner, such that only the spectral dirt injected by the duty cycle adjustment loop remains. Several other modulation bandwidths have been tested, of which the measurement results are summarized in Fig. 7.57. The bandwidth of 6.9 MHz yielded the highest effective number of bits (ENOB). Above that frequency, the power supply decoupling again decreases the effective resolution.

Going to higher modulation bandwidths, it was also tested whether adjusting the peaking DTX phase vs. baseband frequency by combining digital fractional delay and a static carrier phase shift had any effect, but no significant change could be measured. This is most likely since, with the DDTX's resolution, the quantization error is more limiting than the improvement that might be gained. The highest modulation bandwidth measured
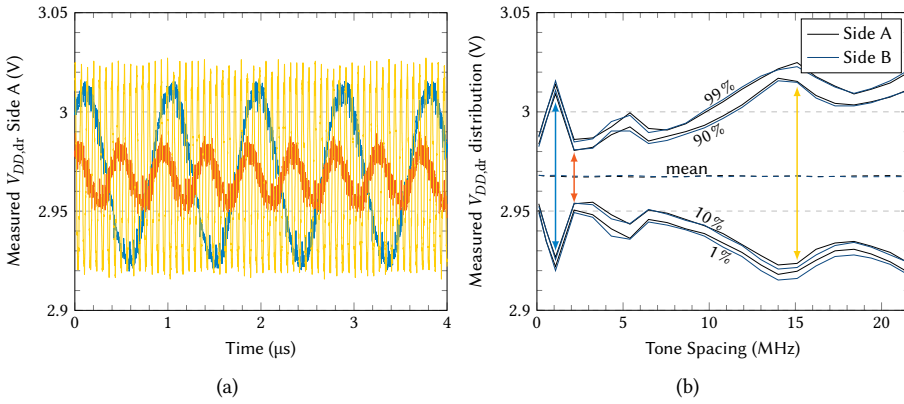
Figure 7.55: Measured $V_{DD,\mathrm{dr}}$ through the TRIG outputs in a two tone scenario with varying two tone spacing. In (a), a time domain section is shown for two tone spacings of 1.08 MHz, 2.16 MHz, and 15.1 MHz as an example, and in (b) the percentiles of the measured signals as a measure of peak-to-peak voltage variation of $V_{DD,\mathrm{dr}}$.

was 221 MHz, where the adjacent channel gets dirt from the duty cycle adjustment loop. Going beyond this bandwidth wouldn't make any sense: this dirt would then be placed in-band, and any DPD fails to correct for it as the injected signal is random, or at least uncorrelated with the baseband ACW and phase information.

7.7.5 ETSI Power Model

To estimate the power consumption in a base station scenario, it should be considered that a base station is not always at full capacity. For example, the traffic is lower in rural areas or at night. For this purpose, a static measurement method can be applied using the ETSI 24 h standard [91]. Static, in this context, means that three discrete power back-off points, but with modulated signals, are considered. Their total DC energy consumption is then considered for a given time duration for each load level, according to the weighting provided in Table 7.5.

Table 7.5: Load level durations for daily average calculation using ETSI ES 202 706-1 V1.6.0 (2020-11) [91].

	Low load	Medium load	Busy hour load
Duration/day	t_{low}	t_{med}	t_{BH}
Default value	6 h	10 h	8 h
Modulated RF power	5 %	33 %	52 %

The expected drain and system efficiencies vs. modulated power back-off can be estimated from the static ACW sweep (Fig. 7.53) using a discrete version of Eq. (6.34). Namely, the drain efficiency is estimated as

$$\eta_{D,\mathrm{est}} = \frac{\sum_{\mathrm{ACW}} P_{\mathrm{RFout}}(\mathrm{ACW})}{\sum_{\mathrm{ACW}} P_{DD,\mathrm{RF}}(\mathrm{ACW})}, \tag{7.2}$$

(a)

(b)

(c)
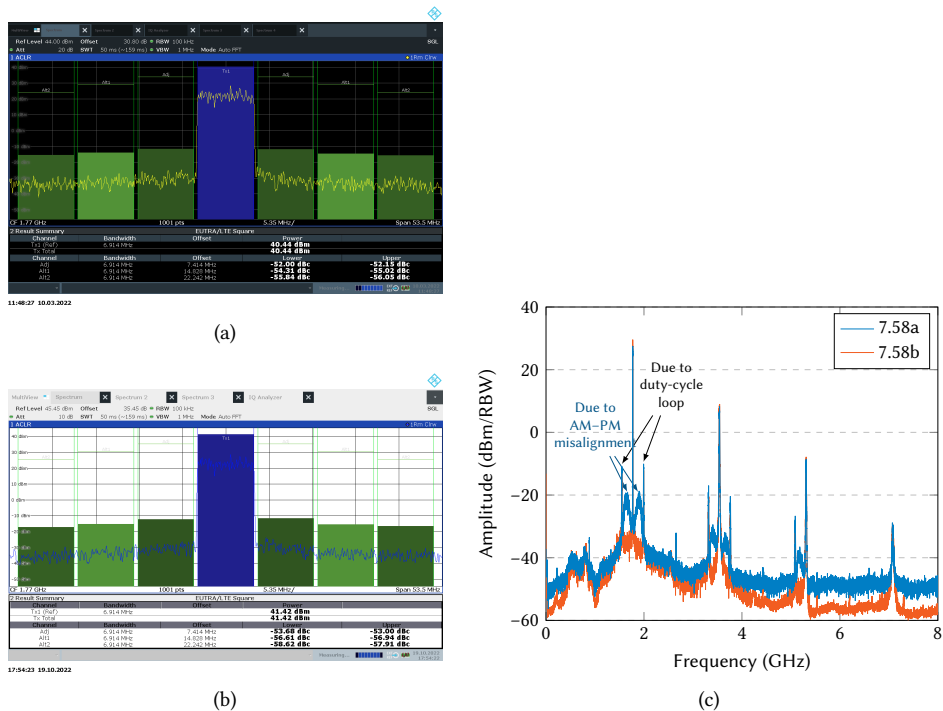
7

Figure 7.56: Measured 6.9 MHz 256-QAM signal spectra with f_c = 1.77 GHz: ACLR measurements in (a) and (b), full spectrum in (c). The results in (a) were published in [80] as ACLR ≤ −52.0 dBc, but a timing mismatch between AM and PM was present in the DPD loop. This resulted in the phase error being corrected in-band and in the adjacent and alternate channels, but pushing the error to different frequencies. In later measurements this alignment error was fixed, resulting in a slightly improved ACLR ≤ −53.0 dBc, but without the error pushed to a different frequency.
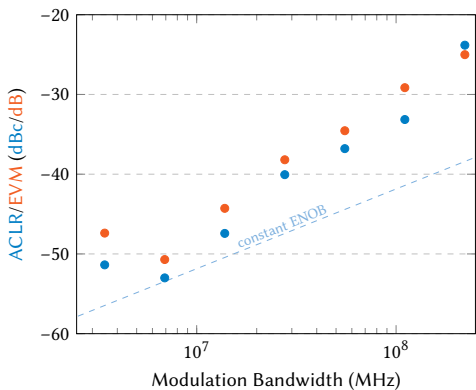
Figure 7.57: The measured ACLR and EVM vs. modulation frequency.
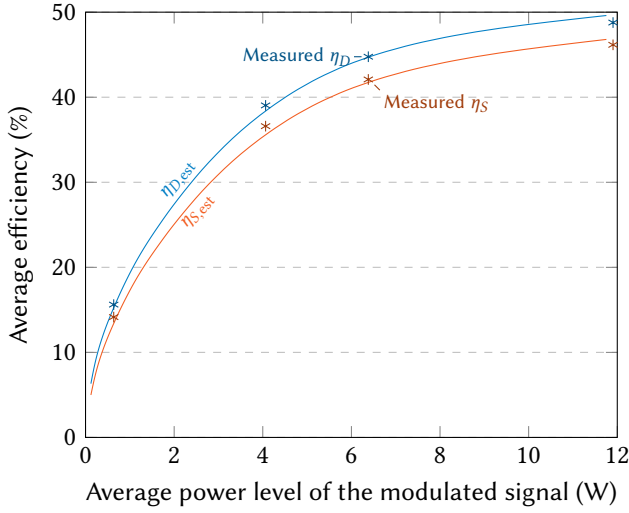
Figure 7.58: The estimated and measured drain/system efficiencies vs. modulated RF power back-off to estimate 24 h energy consumption of the DDTX.

and the system efficiency as

$$\eta_{D,\text{est}} = \frac{\sum_{\text{ACW}} P_{\text{RFout}}(\text{ACW})}{\sum_{\text{ACW}} P_{DD,\text{core}}(\text{ACW}) + P_{DD,\text{dr}}(\text{ACW}) + P_{DD,\text{RF}}(\text{ACW})}, \tag{7.3}$$

where the ACW has the same distribution as the modulated signal to be used, with the appropriate power back-off. The resulting curves using these estimations are given in Fig. 7.58 for a 256-QAM signal with PAPR = 5.5 dB. Next, the three relevant power back-off levels have been measured using the DDTX demonstrator, which are shown as discrete points in Fig. 7.58. DPD is used for all points to ensure the correct distribution for the ACW and sufficient spectral purity. The measured points are very close to the calculated estimations, which validates the correctness of Eq. (6.34).

Finally, we can calculate the ETSI 24 h energy consumption using the measured powers. The calculation is shown in Table 7.6, resulting in a system energy usage of 259.2 W h per

Table 7.6: Calculation of the DDTX's energy usage with the ETSI 24 h standard.

Case	% (dB)	P_{RFout} (W)	$P_{DD,\text{LD}}$ (W)	P_{system} (W)	t (h)	E_{RFout} (W h)	$E_{DD,\text{LD}}$ (W h)	E_{system} (W h)
Max	100 (0)	11.90	24.41	25.79	0	0	0	0
Busy	53.6 (−2.71)	6.381	14.26	15.16	8	51.05	114.1	121.3
Medium	34.1 (−4.66)	4.064	10.41	11.10	10	40.64	104.1	111.0
Low	5.3 (−12.72)	0.6359	4.072	4.494	6	3.815	24.43	26.96
Zero	0 (−∞)	0	0	0.191	0	0	0	0
Sum						95.50	242.6	259.2

24 h, or a daily average efficiency of 36.8 %. This energy consumption can be improved by assuming a (more realistic) TDD system, where 75 % of the time is used for TX and 25 % for RX. This way, the time spent in busy, medium, and low cases changes to 6 h, 7.5 h, and 4.5 h, respectively. Then 6 h is spent in the zero state, which has no useful RF output power, but also has very low DC power. In that case, the energy values change to E_{RFout} = 71.63 W h and E_{system} = 195.6 W h (36.6 % daily average efficiency). Compared to an analog system with bias currents, this DDTX can provide a massive improvement, especially when combined with more clever base station packet scheduling, e.g., (micro) discontinuous transmission [92].

7

8

Design of a High-Resolution High-Power DTX

The measurements of the fabricated DTX designs, discussed in the previous chapter, provide the proof-of-concept for high-power DTXs. However, they employ experimental hardware that introduces a great amount of uncertainty in the design. Consequently, the hardware still includes some design flaws, which, for example, limit the instantaneous video bandwidth. Nonetheless, a lot can be learned from these high-power DTX designs and their measurements.

This chapter addresses these learning points and translates them into a new set of requirements for the next generation of high-power DTX demonstrators, which must provide high integration, instantaneous video bandwidth, and power efficiency while maintaining high spectral purity. This new set of requirements is first discussed in detail in Section 8.1. Next, Section 8.2 translates the requirements to a high-resolution DTX switch-bank layout using a high-density flip-chip assembly flow. The flip-chip approach enables a finer gate segmentation of the power devices, which should be activated so that the DTX's transfer remains smooth. The activation pattern to achieve such a smooth transfer is discussed in Section 8.3. Section 8.4 addresses the advanced design and modeling needs of DTXs to allow their designs to meet the renewed specifications. The theory and implementation of improved passive DC-supply decoupling are discussed in Section 8.5. An overview of the realized CMOS controller, designed in collaboration with other designers, is provided in Section 8.6, together with a brief overview of the targeted demonstrators.

Parts of this chapter are based on published works:

[93]: D.P.N. Mul, R.J. Bootsman *et al.*, "Method of Applying an Activation Scheme to a Digitally Controlled Segmented RF Power Transmitter," US Patent US20240146346A1.

[94]: R.J. Bootsman, D.P.N. Mul *et al.*, "A Switch-Bank Approach for High-Power, High-Resolution, Fully-Digital Transmitters," *2024 54th European Microwave Conference (EuMC)*, Paris, France, 2024, pp. 23–26, doi: 10.23919/EuMC61614.2024.10732131.

[95]: D.P.N. Mul, R.J. Bootsman *et al.*, "A 20 W CMOS/LDMOS All-Digital Transmitter with Dynamic Retiming and Glitch-Free Phase Mapper, Achieving 68/63 % Peak Drain/System Efficiency," *2025 IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, 2025, pp. 104–106, doi: 10.1109/ISSCC49661.2025.10904650.

Finally, Section 8.7 discusses the measurements of the new power DTX generation, which confirm the high resolution targeted in this chapter.

8.1 Goals and Design Requirements for the Next-Generation of Base Stations

For our next-generation high-power DTX, we can now use the knowledge from the proof-of-concept DTX hardware as discussed in Chapter 7. First, the design requirements were set together with our project partners. They are given below:

- Peak P_{RFout} = 50 W, which allows an average $P_{\text{RFout}} \geq$ 5 W at the antenna interface when assuming a PAPR ≈ 8 dB.
- Instantaneous signal bandwidth > 200 MHz, with 600 MHz as a final target (ripple < 0.5 dB).
- The RF operating frequency can range from 3.3 GHz to 3.8 GHz, demanding an RF bandwidth greater than 1 GHz.
- Power amplifier PAE > 50 %, or for a DTX/RFDAC η_S > 40 %.
- Retain system efficiency in power back-off over a −12 dB range, i.e., the DTX remains reasonably efficient for RF output power levels above 0.3 W in support of transmitting complex modulated signals.
- Support 256-QAM modulated signals with an EVM < 3 %, PM error < 3 %, ACLR ≤ −50 dBc.
- A maximum flange temperature of 110 ℃.

These targets need to be translated into DTX design requirements. The technologies available for the demonstrator are, once again, TSMC 40 nm CMOS and Ampleon LDMOS. New in this design is the possibility to further optimize the LDMOS devices by using a thinner gate oxide, lowering their V_T and increasing their g_m. This enables the use of stacked core-oxide drivers (see Section 4.2.2) in the CMOS controller to switch the segments of the LDMOS devices. In this approach, the stacked driver uses $V_{DD,\text{dr}} = 2V_{DD,\text{core}}$ = 2.2 V and offers faster switching than a simpler driver with thick-oxide devices. This enables use at higher RF carrier frequencies, e.g., 3.5 GHz, and lower RF duty cycles for improved efficiency.

The modulation bandwidth and efficiency requirements dominate the upconversion architecture selection. Namely, polar operation cannot support high modulation bandwidths, while signed Cartesian operation, although supporting wideband modulation, is considered not very energy efficient (see Sections 2.1 and 6.2). So, to enable the generation of wideband complex modulated signals in an energy-efficient manner, the 8-phase multi-phase upconversion architecture is used. Digital class-C with a duty cycle of 25 % is selected as the operation class. This combination offers a theoretical upper limit of 85.4 % for the phase-averaged drain efficiency (Section 6.2). Reaching the specified ACLR using 8-phase multi-phase requires some more effort to obtain the spectral purity. This is described in detail in [26]. In this dissertation, it suffices to state that advanced data interpolation with strict retiming requirements is required to reduce the impact of the sampling errors in combination with a push-pull architecture to reject unwanted even-order upconversion products.

Achieving the desired drain and system efficiencies for the signals with a high PAPR demands efficiency enhancement techniques. More specifically, retaining the (system)

(a) Push-pull 3-way Doherty   (b) PD-LMBA

(c)
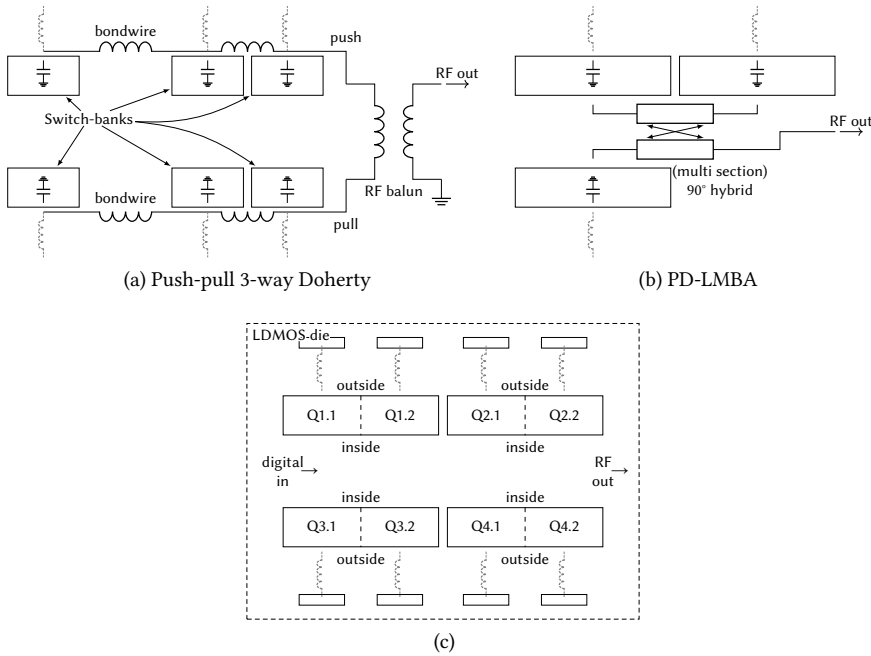
Figure 8.1: Two example switch-bank configurations in (a) and (b), leading to the generic switch bank configuration as shown in (c). This allows for the most flexible configuration for possible demonstrators, including inshin with on-chip return path.

efficiency over a large PBO range will require (at least) a 3-way Doherty DTX demonstrator, while a PD-LMBA (see Section 2.3.2) is an interesting concept to reach a large RF operating bandwidth [96]. In either case, a minimum of 3 controllable switch banks is required, possibly configured in a push-pull fashion.

Translating these DTX demonstrators to switch-bank configurations leads to the structures shown in Figs. 8.1a and 8.1b, or similar. To make the DTX design universally applicable for each demonstrator, the switch-bank layout shown in Fig. 8.1c is proposed, consisting of 4 switch-banks, Q1–Q4. To allow push-pull operation for each switch bank, each switch bank is split into two 'sub-banks,' e.g., Q1.1 and Q1.2. These sub-banks can be controlled independently in terms of activation amplitude and phase to ensure maximum flexibility, which can be useful for implementing a 3-way Doherty, for example. The configuration shown in Fig. 8.1a uses a low-pass equivalent of a quarter-wave transformer (see Section A.4.1). Space is left in the center, allowing a high-pass equivalent Doherty combiner to be placed or a hybrid coupler for the PD-LMBA configuration. Still, the switch-bank's output capacitance needs to be resonated out in the scenario of a power combiner placed at the inside edge of the switch bank. Integrated shunt inductor (inshin) wires can therefore be placed at the outside edge of the switch banks, which can be bonded to an on-chip RF-shorting capacitor such that the ground return path is formed by the lower metal layer(s). In this way, the return current does not have to pass through the (lossy) substrate, increasing the Q-factor of the inshin.

Next, the achievable ACLR should not be limited by quantization errors. To evaluate this, we can use the dynamic range equation for (oversampled) DACs to determine the quantization noise power, compensated for the PAPR[1]

$$DR_{RMS} = 2^{N_b} \cdot \sqrt{\frac{f_s}{2BW} \cdot \frac{P_{RFout,avg}}{P_{RFout,peak}}} \tag{8.1}$$

$$DR_{dB} \approx 6.02 N_b + 10 \log\left(\frac{f_s}{2BW}\right) - PAPR_{dB}. \tag{8.2}$$

In this design, we choose the sampling rate equal to the RF carrier, $f_s = f_c$, such that the sampling replicas fall into the baseband and the second harmonic band and can be suppressed by the bias and harmonic terminations. To translate the impact of the dynamic range to the ACLR, the bank implementation (including its current utilization factor) and ZOH-related sinc shaping should also be considered; however, their impact will be limited to one or two dB and is left out here for simplicity. Filling in the known values from the design specifications gives

$$N_b = \frac{-ACLR - 10\log\left(\frac{f_s}{2BW}\right) + PAPR_{dB}}{6.02} = \frac{50 - 10\log\left(\frac{3500}{2\cdot600}\right) + 8}{6.02} = 8.86, \tag{8.3}$$

thus the DTX switch bank should support at least 9 bits, which also includes the sign bit. Since some headroom is desired to get to a minimum of 8 *effective* amplitude bits, we target an 11-bit implementation for each sub-bank.

8.2 LDMOS Layout and Flip-Chip Assembly

We know from previous demonstrators (Section 7.5.3) that finer (multi-level) thermometer coding is required to increase the effectiveness of the available bits. Reaching an 11-bit power switch bank resolution was not possible using the previously introduced bond-wire-based approaches (Fig. 8.2a). Consequently, a much finer segmentation of the LDMOS devices should be used (see Section 3.1.2), which requires more advanced packaging techniques, such as fine-pitch flip-chip bonding, as illustrated in Fig. 8.2b. Implementing 11 bit as a single level of thermometer coding would yield $2^{11} - 1 = 2047$ gate segments for each sub-bank, which is very challenging to implement, even when using a high-density flip-chip approach.

Here, we adopt a multi-level thermometer coding approach (see Section 5.1.1), which uses 8 bits in the MSB layer and 3 bits in the LSB layer. The MSBs use a bank-sharing approach to improve the bank's current utilization factor (Eq. (6.10), Section 6.2), which means each unit-cell can be either activated with either the *A* or *B* phase. The LSBs cannot be executed in this manner, as each activation phase may need the full range at their disposal at any given time, giving a total of $(2^8 - 1) + 2(2^3 - 1) = 269$ LDMOS gate segments in the switch bank that need to be controlled by the CMOS drivers. Next, the layout of such a switch bank and the high-density flip-chip assembly are discussed.

---

[1]The best-known equation for the dynamic range includes a factor $\sqrt{3/2}$ (or 1.76 dB), which is only applicable in the case of a full-scale sine wave. Here it is replaced by the signal's PAPR.
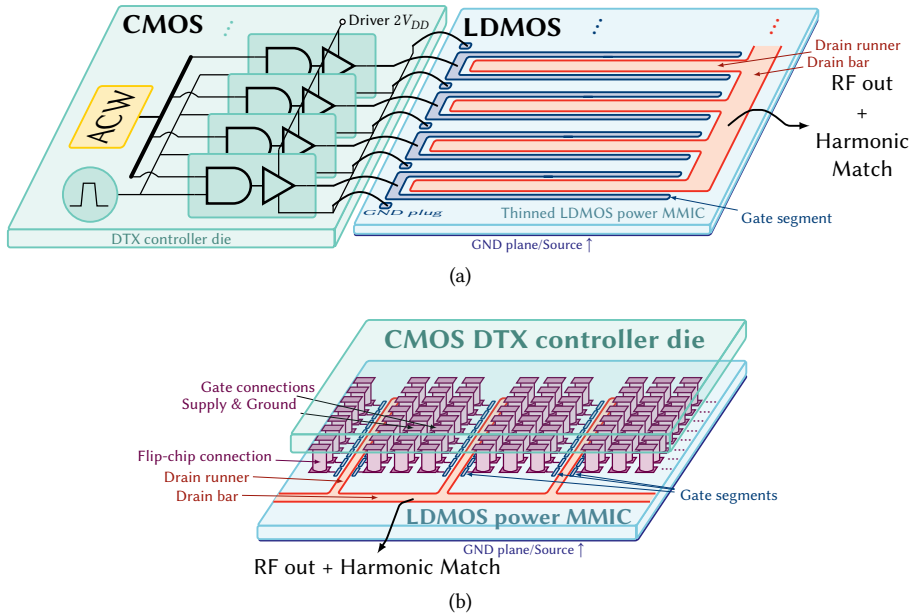
Figure 8.2: Conceptual illustrations of (a) bond-wire-based DTX and (b) using high-density flip-chip. Reaching the required resolution with as much thermometer-coding as possible necessitates many interconnections between the CMOS controller and the power die, for which flip-chip assembly is a much better solution.

8.2.1 Switch-Bank Layout using Flip-Chip

Several parameters influence the dimensions of the DTX power switch banks. First, the bump pitch of the considered high-density flip-chip process sets the minimum spacing between interconnections, which in our case is 40 μm. Next, we should identify the most dominant parasitics introduced by the layout for quick switch-bank layout iterations. The detailed parasitics will be discussed when modeling the power die for the DTX demonstrators in Section 8.4.2.

From the previous demonstrators (Chapter 7), we know that the inductive interconnect parasitics should be minimized, especially the inductive coupling between the interconnections, to avoid bit-to-bit interaction. Furthermore, complementing every driver–gate-segment with alternating $V_{DD,\mathrm{dr}}$ and V_{SS} connections provides a close-by driver return path, which minimizes both the series inductance in the gate–source path as well as the mutual inductance between gate connections. Further, the path for charging the gate–source capacitance should not be shared with the drain–source RF current, such that no additional source inductance (i.e., degeneration) is present.

For the inductance, it is preferred to have all connections as short and close together as possible; however, this does not automatically apply when aiming to minimize capacitive parasitics. The physical design rules dominate the smallest possible pitch. Namely, the pads of the LDMOS power die need to be 25 μm to accommodate the flip-chip bumps. These pads are in the topmost and thickest metal layer, which needs to be shared with metal lines for routing. The minimum width and spacing rules align quite well with the 40 μm flip-chip pitch, meaning that, between each bump pad, there is just enough space for exactly one
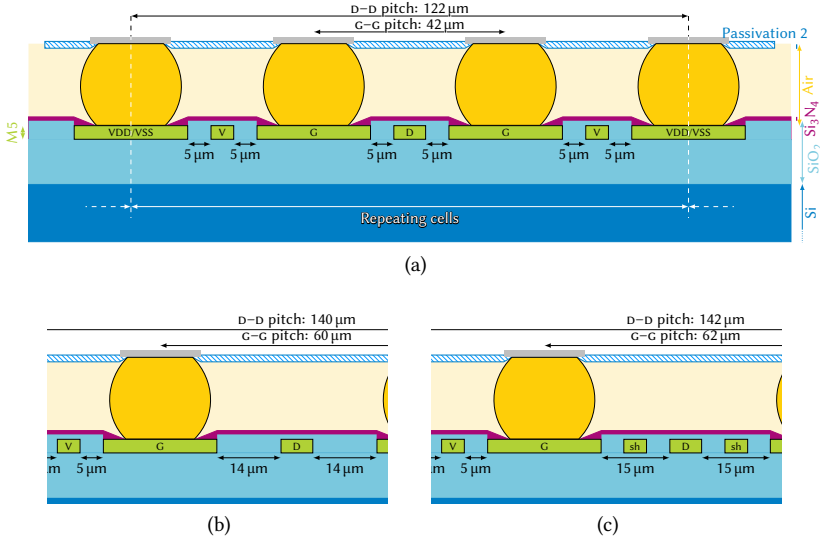
Figure 8.3: Choosing the unit-cell pitch is dominated by feedback capacitance C_{GD} layout parasitics in the LDMOS die: (a) shows the minimum possible pitch, (b) with increased G–D spacing, and (c) inserts a G–D ground shield.

line for routing purposes. We need to route $V_{DD,\mathrm{dr}}$ over the LDMOS connecting to the CMOS drivers, whereas the ground is widely available through the bottom metal layers and the LDMOS substrate by means of heavily doped source/ground plugs. We also need to route the drain runners for the RF output.An example layout sketch for minimum layout pitch is shown in Fig. 8.3a.

Increasing any of the metal spacings in Fig. 8.3b will decrease the capacitive coupling between pads and lines. Some are more impactful than others; for example, capacitance between ground and supply are no problem, as both are supposed to be AC short anyway. By far the most significant impact from a spacing point of view can be found from the gate–drain connection. The opposite polarity and much larger voltage swing of the drain causes the gate–drain, or feedback, capacitance C_{GD} to appear larger, as described in Eqs. (6.12) and (6.14). The voltage gain in the envisioned DTX operation from gate to drain is $-56/2.2$ at peak output power, resulting in C_{GD} contributing 26.5 times to C_{GG}. Simulation using Keysight Advanced Design System (ADS) controlled impedance line designer (CILD) of two M5 traces and additional verification in Ansys high-frequency structure simulator (HFSS) without bumps indicates that, when using minimum spacing, $43.3(5)\,\mathrm{pF\,m^{-1}}$ is added. This added capacitance is considered for a 25 µm pad and is compared in Table 8.1 relative to a verified LDMOS model of a conventional LDMOS finger of 38 µm gate width. We note that C_{GG} increases by 57 % using the minimum spacing, and, as such, with it will come a 57 % increase in drive power. Worse, the C_{GG} now strongly depends on the DTX's output magnitude: it increases by 44.5 % when comparing the $C_{GG,\mathrm{max}}$ at peak output power to the $C_{GG,\mathrm{min}}$ at no output power. At the drain side, C_{DD} is expected to increase by 12 % with negligible variation over output power, especially when considering the intrinsic nonlinearity of the LDMOS' output capacitance. This C_{DD} can simply be resonated out, but it may decrease the DTX's achievable bandwidth.

Table 8.1: Estimated impact using coupled microstrip calculations per unit cell (40 μm wide) of the M5 gate–drain strategies shown in Fig. 8.3 on DTX performance parameters.

	$S = 5\,\mu m$	$S = 14\,\mu m$	$S = 5\,\mu m$+shield
$+C_{GD}$	1.09 fF	0.35 fF	0 fF
$+C_{GS}$	4.53 fF	4.53 fF	5.61 fF
$+C_{DS}$	0 fF	0 fF	1.09 fF
$C_{GG,\text{max}}/C_{GG,\text{nom}}$	1.57	1.27	1.15
$C_{GG,\text{max}}/C_{GG,\text{min}}$	1.445	1.185	1.060
$C_{DD,\text{max}}/C_{DD,\text{nom}}$	1.123	1.050	1.119
$C_{DD,\text{max}}/C_{DD,\text{min}}$	1.004	1.002	1.001

Two other scenarios are also considered to estimate their impact, which are shown in Figs. 8.3b and 8.3c, the related results are also included in Table 8.1. We observe that increasing the spacing between gate and drain can significantly lower the impact of C_{GD} on both C_{GG} and C_{DD}. Inserting a grounded shield (Fig. 8.3c) minimizes both the C_{GG} increase and its variation, but the impact on C_{DD} remains. In later iterations, it was decided to add two metal layers to the LDMOS metal stack, making it possible to implement a ground shield further away from the drain runner while assuming a G–G pitch of 60 μm. Further optimization is possible by assuming an octagonal pad shape, which can still accommodate the round flip-chip bump.

All-in-all, this results in 448 flip-chip CMOS–LDMOS interconnections in a switch bank, of which the conceptual layout is shown in Fig. 8.4. Indicated are the alternating V_{SS} and $V_{DD,\text{dr}}$ connections, the positions of the MSB and LSB gate segments, and the drain connections. This gives one switch bank a 1300x640 um^2 footprint. The outside drain bar facilitates a bonding area for inshin wires and should 'stick out' underneath the CMOS die that is flip-chipped on top. Consequently, a minimum area of 5.2x1.28 mm^2 is required to implement eight of these switch banks in a configuration as shown in Fig. 8.1, not yet considering any spacing requirements between the banks due to CMOS clock or data routing, or RF matching structures on the LDMOS. The resulting LDMOS die size eventually becomes 7.0x5.3 mm^2 when these structures are eventually included. The large footprint adds to the challenges of successfully performing flip-chip assembly, which will be discussed next.

8

8.2.2 Flip-Chip Assembly Flow for Minimized Risk

We assume a minimum bump pitch of 40 μm for the flip-chip process, although even finer pitches, down to 25 μm, can be supported [68]. Multiple manufacturers are involved in manufacturing and the flip-chip assembly of the LDMOS and CMOS dies before they are placed on the PCB demonstrator. Fraunhofer IZM supports advanced flip-chip, including several pre-processing steps that may be necessary. Ampleon has the capabilities and expertise in manufacturing (LDMOS) RF power devices, including their assembly. The CMOS design is processed by TSMC in an industrial setting, taped out by IMEC as a multi-project wafer (MPW). Several challenges are present here that need to be addressed while pioneering a flip-chip assembly flow for combined digital and RF power purposes.

First, the LDMOS and CMOS designs should be compatible with the requirements
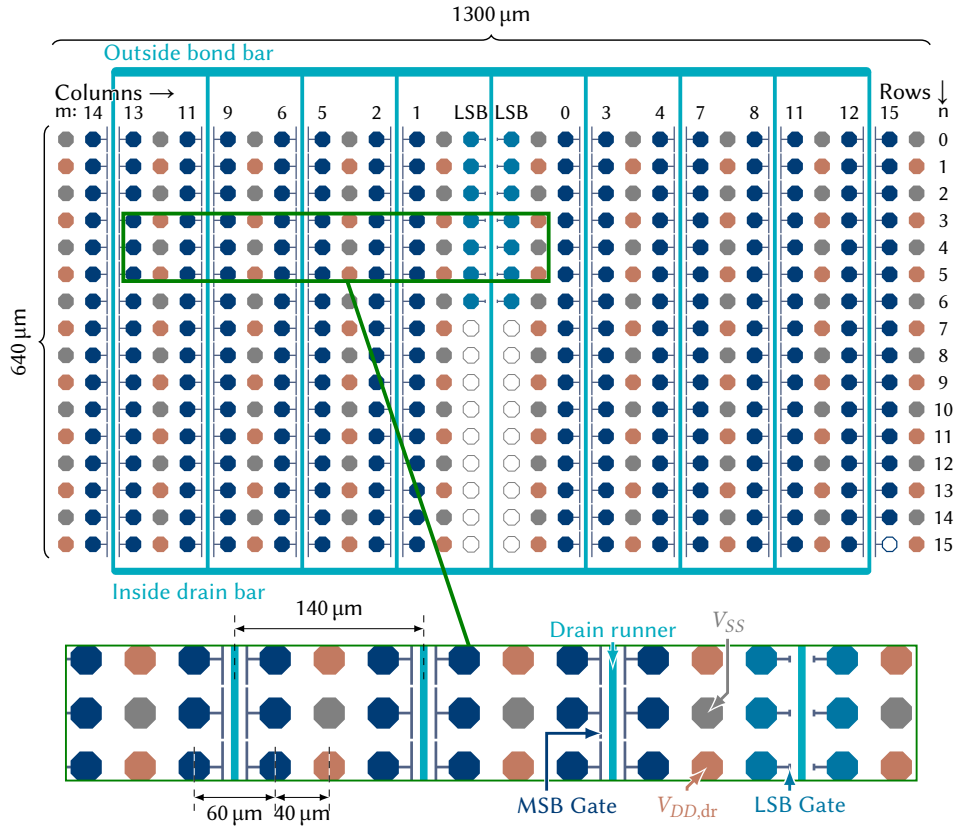
Figure 8.4: Layout for a single switch bank, also featuring driver supply connections and ground return paths.

imposed by Fraunhofer for flip-chipping. These include requirements on the used layer stack with its passivation layers to minimize topography (height differences at the die surface) and related design rules. TSMC's default packaging processes available in small-scale MPWs do not provide options for very high-density flip-chip, causing design rule violations when designing small bump pads, including their passivation openings. After inquiry, these rules are only a restriction for in-house assembly processes, not for wafer manufacturing.

The flip-chip process uses tin-silver (SnAg) micro solder bumps grown on copper under bump metallization (UBM) in a reflow process. This process has some self-aligning properties, so applying pressure during flip-chip is unnecessary. As a result, no restrictions are present on the structures underneath the pads: no additional metals are required in the CMOS pads, and there is no problem with placing active devices underneath them. Only very sensitive circuits should not be placed directly underneath, such as SRAMs, since there may be radioactive traces present in the tin, of which their decay may cause bit flips.

Table 8.2: A summary of key manufacturing capabilities.

	Unique capabilities:		
Ampleon	• LDMOS wafer fabrication • Wafer thinning (conventional & TAIKO) • Die-to-flange attach • Back-side metallization		**Shared options:** • Dicing
Fraunhofer IZM	• UBM for LDMOS • Bump growth (LDMOS & CMOS) • Flip-chipping	**Shared options:** • UBM for CMOS	
TSMC	• CMOS wafer fabrication		

A summary of the manufacturing capabilities resulting from initial discussions is presented in Table 8.2. The search space for a suitable assembly flow can be significantly decreased by defining the unique capabilities per manufacturer: the related steps simply *have* to be performed by that party. The only remaining point of discussion is then "in what order?" After several iterations of logical elimination, the flip-chip assembly flow of Fig. 8.5 was defined.

Starting on the bottom part of the assembly flow, TSMC does have capabilities to deposit UBM, but does not (officially) support the fine structures or material layers required for this particular flip-chip process. This means that CMOS UBM deposition has to be performed by Fraunhofer, for which full wafers are preferred. However, since this concerns an MPW with multiple designs, the other designs have to be considered confidential, and have to be etched or burned away. This renders the resulting wafer fragile and thus useless for further processing. As such, the UBM for CMOS has to be performed on a per-die basis.

For the LDMOS, however, full wafer access is not an issue, making depositing UBM and bumps straightforward. However, temperatures that are too high for the solder bumps are required when attaching the LDMOS power die to a metal flange. This means the solder bumps should be deposited on the CMOS dies instead.

8
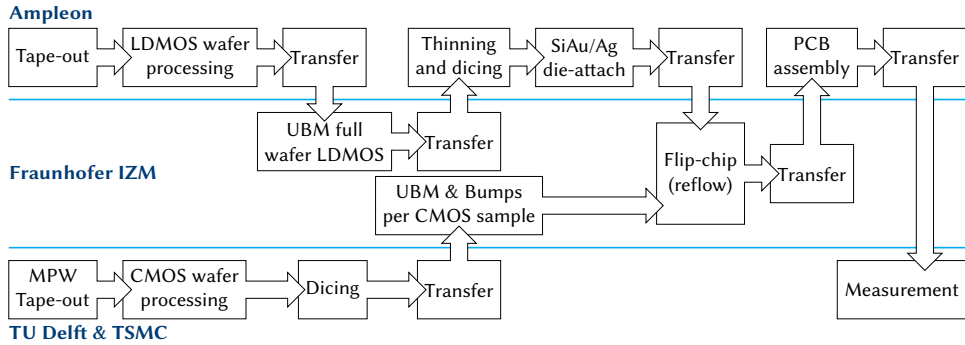
Figure 8.5: The flip-chip assembly flow used that is compatible with all the manufacturing capabilities and requirements.

The LDMOS die attach process is a major concern due to the die warpage and die shrinkage that are possible during this process. For that reason, two die-attach options are considered. The first option uses a high-temperature die-attach with an eutectic SiAu mixture under high pressure. This will result in the lowest die warpage, while the temperatures during the reflow step in the flip-chip assembly will not melt the SiAu mixture. Die shrinkage is a larger concern due to the high temperatures used in this die-attach, caused by the mismatch in expansion coefficients of the die and flange materials (see Section 3.2.2 and Table 3.1). From the center of the edge to the die is 4390 μm, meaning that a shrinkage of even 0.5 % is more than half the flip-chip pitch, risking misalignment or shorts. The second die-attach option uses a low-temperature silver sintering die attach without pressure. The lower temperature lessens the risk of unacceptable die shrinkage, but the lack of pressure increases the risk of die warping. Another concern is that this type of die-attach is performed at lower temperatures than the reflow step while flip-chipping, however, the sintered result fortunately disintegrates at a higher temperature than required to perform the sintering itself. The LDMOS die can be left thicker than usual to minimize risk further, e.g., 160 μm instead of 50 μm. However, this may affect the LDMOS' electrical performance. Preliminary die-attach tests indicated that both options result in acceptable die-shrinkage or die-warping to proceed safely.

8.3 Activation Pattern[2]

From previous demonstrators, we know that the physical location of the LDMOS gate segments combined with the order of their activation impacts the DTX's transfer (see Sections 7.5.2 and 7.7.2). Special attention should be given to such activation order to implement a high-resolution DTX using an electrically large switch bank. For that purpose, a strictly symmetrical activation pattern as shown in Fig. 8.6 is defined, targeting a maximally uniform current distribution at any given time also to minimize (dynamic) current redistribution effects.

The activation with increasing ACW starts from the middle segments and extends symmetrically to the outer ones, on a row-by-row basis from the outside bank edge towards

---

[2]Parts of this section are based on the published work [93]: D.P.N. Mul, R.J. Bootsman *et al.*, "Method of Applying an Activation Scheme to a Digitally Controlled Segmented RF Power Transmitter," US Patent US20240146346A1.

drain connnections   middle
                                    ↓ outside ↓

15	13	11	9	7	5	3	1	•	0	2	4	6	8	10	12	14
31	29	27	25	23	21	19	17		16	18	20	22	24	26	28	30
47	45	43	41	39	37	35	33		32	34	36	38	40	42	44	46
63	61	59	57	55	53	51	49		48	50	52	54	56	58	60	62
79	77	75	73	71	69	67	65		64	66	68	70	72	74	76	78
95	93	91	89	87	85	83	81		80	82	84	86	88	90	92	94
111	109	107	105	103	101	99	97		96	98	100	102	104	106	108	110
127	125	123	121	119	117	115	113		112	114	116	118	120	122	124	126
143	141	139	137	135	133	131	129		128	130	132	134	136	138	140	142
159	157	155	153	151	149	147	145		144	146	148	150	152	154	156	158
175	173	171	169	167	165	163	161		160	162	164	166	168	170	172	174
191	189	187	185	183	181	179	177		176	178	180	182	184	186	188	190
207	205	203	201	199	197	195	193		192	194	196	198	200	202	204	206
223	221	219	217	215	213	211	209		208	210	212	214	216	218	220	222
239	237	235	233	231	229	227	225		224	226	228	230	232	234	236	238
255	253	251	249	247	245	243	241	•	240	242	244	246	248	250	252	254

middle inside

Figure 8.6: Activation pattern as used for the MSBs, activating from 0 (light) to 255 (dark).

the inside edge. The outside edge first is used as the most significant currents are expected in the path with the highest Q factor: resonating between the LDMOS segments' output capacitance and the inshin. Each sub-bank follows this activation pattern, meaning that every full bank has twice this activation pattern next to each other. Some additional flexibility is present as the banks can be controlled completely independently. Using this activation prevents irregularities, yielding a smooth and monotonic ACW-to-RF output signal transfer. The remaining linearity error due to, e.g., output stage compression can be compensated using DPD, which can be significantly relaxed due to the much smoother (and monotonic) behavior.

8.3.1 Unit Cell Logic

In the requirements and design goals discussion (Section 8.1), we already selected 8-phase multi-phase upconversion and digital class-C operation with a 25 % RF duty cycle using bank sharing activations. For each unit cell to be able to activate with a 25 % duty cycle at either the A or B clock, both these A and B clocks are provided to the unit cell. The routed clocks have a 50 % duty cycle, as will be motivated in the next subsection. Combining two 50 % duty cycle clocks with a 90° relative phase offset can simply be done using a symmetrical and-gate [97]. However, a unit cell should also 'know' whether it should activate at the A or the B phase. This dynamic phase allocation must be made compatible with the activation pattern described above, while still maintaining a symmetric and maximally uniform pattern.

The MSB unit cells are divided into two groups, and each group is given a native clock phase, either phase A or phase B, as illustrated in Fig. 8.7a. The division is made such that the A and B phases are distributed uniformly over the switch bank. The total number of wanted activations is calculated first to determine which MSB cell should be activated at all. This information is encoded into row and column activation information, and decoded by a row-column-decoder in the unit cell, as shown above left in Fig. 8.8. This sets an outline of active unit cells, as shown in Fig. 8.7b, the other unit cells should remain deactivated. A second row-column-decoder (bottom right of Fig. 8.8) provides information on whether the unit cell in question should swap its activation to the other clock. The result is fed
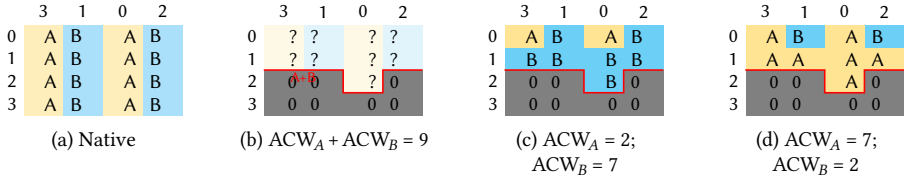
Figure 8.7: The dynamic phase allocation of the unit cells using 4×4 unit cells as simplified example: (a) each unit cell is assigned a native A or B phase; (b) first the total number of activations is determined by calculating $ACW_A + ACW_B$; (c) and (d) show example activations with the dynamic phase allocation adhering to the activation pattern of Fig. 8.6, where the minority phase activation is prioritized.
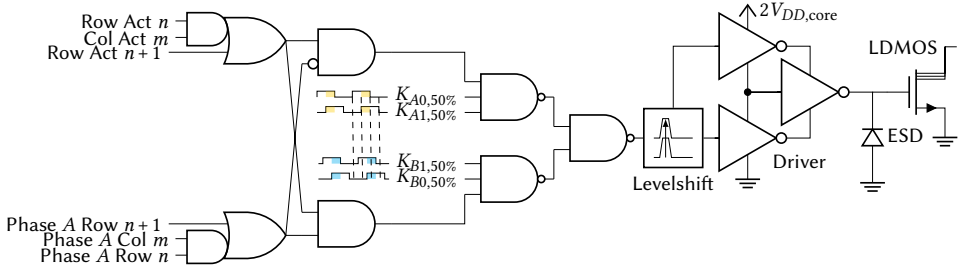
Figure 8.8: Simplified logic as applied in an unit cell (data retiming, glitch prevention, and buffering removed) for clock combining and selecting. Here a native 'A' cell is shown. The three bottom left inputs are swapped for phase B row and column data lines for a native 'B' cell, as well as the clock inputs being mirrored.

into the level shifter and driver and is transferred to the LDMOS segment. The phase with the lowest ACW has all its activations closest to the middle outside, as dictated by the activation pattern. Take Fig. 8.7c for example, where ACW_A is in the minority. The B phase has its activations evenly distributed among them, after which the remaining A unit cells have their activation swapped for the B phase. Figure 8.7d shows the opposite situation, illustrating the maximum uniformity and minimized redistribution, while each drain runner sees, on average, the same (re)active loading.

Three data lines are typically required per row and per column: one for the activation and two for swapping the phase of natively A and B cells. The rows containing LSB cells form an exception, there an additional data line for the activation of the LSB cell is required. The timing of the data should be strictly aligned with the targeted activation phase, which is described in detail in [26, 95].

8.3.2 Bank Clock Line Design

The A and B clocks dictate the moment and duration of the unit cells' activations. These clocks need to be phase-modulated to access all octants of the complex baseband's unit circle (see Section 2.1.3). It is important to ensure the signal integrity of these clocks, for which fast rise and fall times are required, especially if 25 % duty cycle clocks are routed. Namely, 25 % of the clock period at 3.5 GHz is only 71 ps; the propagation delay and the rise and fall times should be a fraction of that. Routing two clocks of 50 % instead relaxes this timing requirement by a factor of 2. The power consumed by the clock line drivers is also
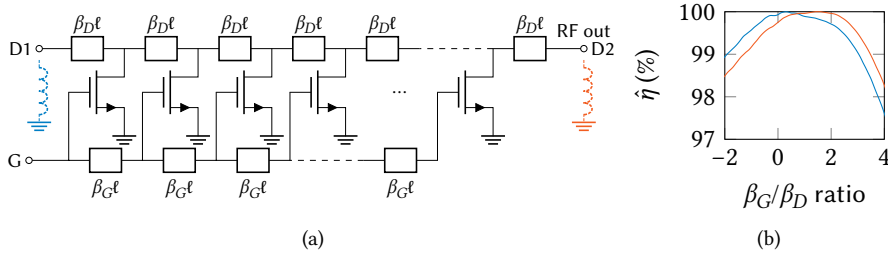
Figure 8.9: Investigating the effects of different propagation constants (β) of the driver's clock line at the gate and line at the drain side. The propagation constants should roughly match if the shunt inductance is placed at the outside. When the shunt inductance is placed at the inside, the gate line should be faster. If the propagation constants are severely mismatched, e.g., by a factor 10, the efficiency loss increases to more than 20 %.

lowered, as the power consumed in the driver itself grows asymptotically with increasing speed (see Section 4.1.3), even though it doubles the number of clock lines needed. The clock lines are executed pseudo-differentially to secure the clocks' integrities further, making it more resilient against common-mode interference, while the 50 % duty cycle makes it more resilient against even-order interference. The phase modulation of these clocks is performed by a global phase mapper for each sub-bank, after which the clocks are routed to the switch bank using a binary clock tree. Coincidentally, by routing two sets of 50 % duty-cycle clocks and creating the 25 % duty-cycle activations within the unit cell, it becomes possible to create a glitch-free phase mapper, as described in [26, 95, 97].

Here, we focus on routing the clock lines through the switch bank. These clock lines cannot be buffered within the switch bank itself, since additional buffer stages would create timing mismatches between the unit cells. Here, we should also consider the wave propagation of the RF output in the LDMOS drain runners, combined with the clock propagation delays in the CMOS. Figure 8.9a conceptually shows the situation where the clock signals at the gate are distributed by transmission lines with a different phase constant β_G compared to the drain runners' β_D. Two situations are considered: the shunt inductance meant for resonating out the LDMOS segments' output capacitance can be placed at either side of the drain, say D1 and D2, while the RF output remains always located at D2. This situation is analogous to having the inshin at the outside of the bank and placing the RF output at the inside or the outside of the switch bank, and the RF clocks are routed parallel to the drain runners. Conceptually, when the ratio $\beta_G/\beta_D = 1$, the wave propagation in the LDMOS drain runners matches the wave propagation of the clock lines in the CMOS. When this ratio is bigger than one, this would mean the CMOS clock lines are slower[3]; conversely, the CMOS clock lines are infinitely fast when the ratio is 0. A negative ratio would indicate that the CMOS wave propagation is in the opposite direction of the RF output. Varying this ratio and inspecting the resulting normalized drain efficiency, as shown in Fig. 8.9b, shows that the β_G required for optimum efficiency changes depending on the location of the shunt inductance, such that the gate wave propagation aligns with the RF wave to to the shunt inductance. The effect here is relatively little, less than 2 % efficiency loss. However, this loss quickly increases to, for example, more than 20 % if the gate lines are a factor 10 slower,

---

[3]Note that $\beta\ell$ is the 'electrical length' of the lines, where $\beta \propto v_p^{-1}$.
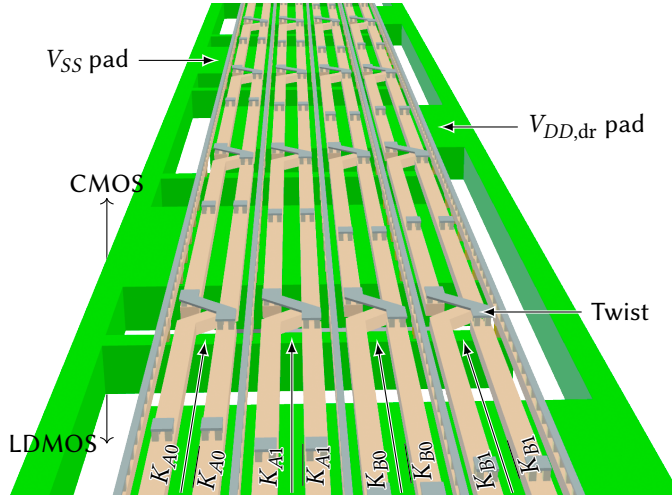
Figure 8.10: The implemented RF clock lines in CMOS, using the thicker M6 layer (pale yellow). The lines are capacitively shielded by the supply pads in M7 (green), and magnetically by twisting the clock lines. The ground shield implemented in M4 is not rendered for better visibility.

regardless of direction. We can conclude that the gate lines' propagation is preferred to be in the same direction as the RF output, and should be faster or, at worst, a factor 2 slower. This impact of different propagation constants becomes stronger when considering larger switch-bank structures or higher frequencies, and the conclusions drawn here may need to be reevaluated.

When modeling the LDMOS drain runner in M5 only (see Section 8.4.2 for more details), it results in a phase velocity of $128.1 \cdot 10^6 \, \mathrm{m \, s^{-1}}$, or $6.3°$ at $3.5 \, \mathrm{GHz}$ and $640 \, \mathrm{\mu m}$. When also considering the capacitive line loading of the LDMOS segments' drains loading, the line's phase velocity drops to $66.3 \cdot 10^6 \, \mathrm{m \, s^{-1}}$, possibly varying up to $87.7 \cdot 10^6 \, \mathrm{m \, s^{-1}}$ for different LDMOS configurations ($12.1°$ to $9.2°$).

The lower CMOS metal layers have a relatively high sheet resistance, resulting in a delay that is RC-dominated rather than LC-dominated. Instead, we have to consider the thicker M6 layer ($23 \, \mathrm{m\Omega \, \square^{-1}}$). We shield the clock lines by the alternating $V_{DD,\mathrm{dr}}$ and V_{SS} pads in M7 on the top to minimize capacitive coupling from the LDMOS drains' large voltage swing, while placing an M4 ground on the bottom. The $V_{DD,\mathrm{dr}}$ and V_{SS} pads are placed farthest away from LDMOS drains. The clock lines are periodically twisted to reject possible magnetic interference from drain currents, resulting in the layout structure as shown in Fig. 8.10. A good trade-off between size, power, and line delay was found when using a line width of $1.125 \, \mathrm{\mu m}$, spacing between the lines of $1.215 \, \mathrm{\mu m}$, and a clearance to the ground on the sides of $0.56 \, \mathrm{\mu m}$. This results in an odd-mode phase velocity of $77.5 \cdot 10^6 \, \mathrm{m \, s^{-1}}$, which is an electrical length of $10.4°$ at $3.5 \, \mathrm{GHz}$ and $640 \, \mathrm{\mu m}$. This would be equivalent to a β_G/β_D ratio of 0.85 to 1.13, depending on the LDMOS configuration, which is considered good enough not to significantly impact the drain efficiency. The odd-mode capacitance of one such line is $332 \, \mathrm{pF \, m^{-1}}$, meaning that for the switch bank under consideration, eight clock lines at $3.5 \, \mathrm{GHz}$ and $1.1 \, \mathrm{V}$ require $7.2 \, \mathrm{mW}$ to charge and discharge. Since each sub-bank requires 10 sets of these lines, every sub-bank dissipates $72 \, \mathrm{mW}$ in the clock lines, or $576 \, \mathrm{mW}$ for all

banks together. A (dynamic) disable signal is hence added to each complete bank to save power for when the bank is not in use, for example, in demonstrators requiring only one or two banks, or in the case of a Doherty, where the peaking branches are mostly inactive.

8.4 Advanced DTX Technology, Modeling and Design

With the high-power, high-resolution DTX demonstrator, we are pioneering a new assembly method and modified LDMOS technology. That means no verified models are available, not for the flip-chip interconnect or active device performance. Fortunately, some models were available for the previous designs, although they were imperfect (see Section 7.2.1). Consequently, we need to create a model that predicts the device performance, including its capacitances, to iterate on the CMOS driver design and output matching networks. This model should also reflect the impact of an LDMOS segment's physical location to evaluate its impact on the DTX's performance.

The modifications made to the LDMOS technology and the modeling of the active device are discussed next in Section 8.4.1. The surrounding passive interconnect parasitics are characterized in Section 8.4.2, which results in a segmented LDMOS model suitable for demonstrator design and simulating the predicted performance of the DTX. The input capacitance of an LDMOS gate segment can be found by combining the active and passive models, based on which a CMOS driver can be designed. Its design and further modeling for system-level simulations are discussed in Section 8.4.3.

8.4.1 A Modified LDMOS Technology

For the power DTX, we prefer an LDMOS technology that can be switched by low gate voltages. Since the DTX demonstrators of Chapter 7 showed that a high-power DTX is indeed feasible, the motivation of our industry partner was further increased to modify the LDMOS technology for the DTX application. In Chapter 7, the V_T-shift was achieved by changing the substrate doping concentrations. In this generation, the thickness of the LDMOS gate oxide is also adjusted. In a DTX scenario, there is no risk of oscillations or overvoltage conditions at the gate because the LDMOS segments are controlled by a digital driver with a very low impedance; using a thinner gate oxide (up to a certain point) does not pose a reliability risk. Initial TCAD simulations indicate that by thinning the gate oxide by 33 %, the LDMOS V_T goes down from a nominal 1.73 V (simulated) to 0.97 V, while the peak g_m rises from 129 mS mm^{-1} to 166 mS mm^{-1}. Reducing the oxide thickness by 50 % or even 67 % would yield a V_T of 0.71 V or 0.45 V, and a g_m of 208 mS mm^{-1} or 263 mS mm^{-1}, respectively, showing that pursuing such a path can be very worthwhile.

Changing the gate oxide thickness is quite a drastic measure in the device processing. While everything might be fine at the gate side, at the drain side, voltage breakdown limits still need to be respected, especially for the electric fields present from drain to gate. Hence, it was chosen to only reduce the process's gate oxide by 33 % and leave the gate length as-is. Further TCAD simulations indicate that the nominal gate–source capacitance would increase to 1.9 pF mm^{-1}.

To model the changes in the LDMOS devices, a proprietary, scalable standard LDMOS model can be modified to obtain some initial DTX performance estimates. The LDMOS process changes involve the customized oxide thickness, new processing steps, two additional metal layers, and a thicker substrate (to lower the assembly risk, see Section 8.2.2). The change in gate oxide can be modeled by scaling voltages and currents at the gate side, as

shown in Fig. 8.11. The bottom metal layer will be used as the LDMOS source connection, providing the current return path, hence little to no change in source inductance L_S is expected compared to the old LDMOS model, despite the thicker substrate. However, the effect on drain/source series resistance is still unknown.

This first exploratory LDMOS model allowed designers to do a coarse iteration on their demonstrator designs, helping them define the required bank spacing and die dimensions, as discussed in Section 8.2.

Eventual technology processing tests and DC I–V measurements indicate that the modifications made lowered the V_T from 2.086 ± 0.013 V to 0.701 ± 0.016 V, a reduction of 66 %. An LDMOS model accurate only for DC was extracted from these measurements. For the purpose of including capacitance information, the parameters of the scaled model were updated to
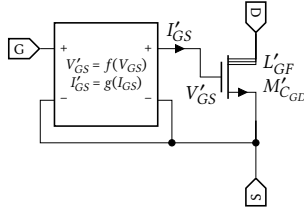
$$V_{GS}' = 1.78 V_{GS} + 0.46 \qquad (8.4)$$

$$I_{GS}' = 1.37 I_{GS} \qquad (8.5)$$

$$L_{GF}' = 1.17 L_{GF} \qquad (8.6)$$

$$M_{C_{GD}}' = 1.17 M_{C_{GD}}. \qquad (8.7)$$

Figure 8.11: LDMOS model modification using parameter conversion.

Doing so, the I–V curves of the two models match closely, while matching the gate capacitance values that were obtained by the TCAD simulation. The majority of all demonstrator design iterations were done using this modified LDMOS model.

A final proprietary RF LDMOS model for the modified technology was developed by Ampleon using pulsed S-parameter measurements from the AMCAD PIV 3200 system, targeting only the intrinsic LDMOS structures. The extrinsic layout parasitics related to, e.g., flip-chip connections can be added to this model using EM simulations. A simple I_D model with 2D capacitance would most likely suffice, as only a single (small) LDMOS segment needs to be modeled. The resulting V_{GS}–I_{DS} and $-g_m$ curves are in Fig. 8.12 compared to a typical non-modified LDMOS node. It can be seen that the V_T has shifted down significantly, while also the transconductance has increased, making this modified technology more compatible with the CMOS voltage levels and thus much better suited for DTXs. Self-heating is included in the model, but unverified. Hence, the recommendation for use in design work was to disable any self-heating effects of the model. The method of parameter conversion, such as done in Eqs. (8.4)–(8.7), was validated by Ampleon using this new RF model for segmented LDMOS. This proprietary, more accurate, RF model can now serve as a drop-in replacement of the previously developed model for the active part, while the modeled passives, discussed in the next section, can remain unchanged.

8.4.2 LDMOS Interconnect and Power Stage Modeling

The active model only delivers a part of the puzzle for modeling this DTX. The surrounding structures at the gate side add capacitive and inductive parasitics, while at the drain side, the drain series resistance as well as the substrate's dielectric loss also play an important role. The series parameters of the drain play a role in the DTX's transfer (see Section 8.3.2), so any passive model should accurately reflect the effect of the segment activation order. A

3D view of the CMOS–LDMOS layout with interconnections is shown in Fig. 8.13. The LDMOS drain runner can be found in the figure's center, surrounded by the gate segments connected through the flip-chip bumps to the CMOS driver. This 3D view is for illustration purposes only; actually simulating this exact structure turned out to be computationally infeasible. The significant parasitics of this structure should be captured in a simpler passive model instead to allow other designers to simulate and design DTX demonstrators.

Drain Parasitics

The LDMOS segments are driven using rectangular voltage pulses, which require sufficient harmonics in frequency domain simulations (also see Sections B.1 and 5.2), or fine timesteps in transient simulations. Modeling the drain only by EM simulation and using the resulting *S*-parameter model might yield simulation convergence issues, due to (numerical) instabilities at higher frequencies, especially when several of these models are placed in series or have multiple internal ports to accurately reflect the position of the segments, such that the EM-model of the drain is segmented as well.

Instead of directly using the *S*-parameter model from the EM simulation, its transmission line parameters in terms of the distributed elements (i.e., *R*, *L*, *G*, and *C*) are extracted using the procedure as described in Section A.6. A 1000 μm section of the drain runner in LDMOS metals 4&5 is simulated, while using metal 1 as the ground return path. No coupling (magnetic nor electric) between two drain runners is assumed, as GND connections and shields are placed between them. The extracted frequency-dependent transmission line parameters are then plugged into an HSPICE W-element model, which is essentially a circuit element implemented by a frequency-dependent *RLGC* matrix with a line length included, where the parameters are fitted for RF accuracy up to 20 GHz, while numerical stability of the model should be guaranteed up to 120 GHz (e.g., 31st harmonic of 4 GHz).

The accuracy of the W-element fit is checked by comparing its simulated two-port parameters in the *S* and *Y* domains to those of the EM-simulated transmission line. Up to 20 GHz, the transmission parameters (S_{21} and S_{12}) deviate less than 0.03 %, and the fit for the reflection parameters (S_{11} and S_{22}) deviate less than 1.1 %. The largest deviation below 20 GHz in *Y*-parameters is less than 1.2 %. Increasing to the full EM simulated frequency range, to 56 GHz, its worst-case deviation is still less than 3.3 %, which is deemed accurate enough for the targeted purpose.

Numerical stability and further time-domain accuracy is verified by investigating the models' voltage step response with a non-zero rise time, and the current impulse response with non-zero width, performed using a transient simulation. A few simulations are performed with varying line source and load terminations. Two voltage step response examples with different rise times are shown in Fig. 8.14, where the 1000 μm line models are capacitively terminated by 16 pF. The simulation results as shown in Fig. 8.14a show good agreement, which assumes a rise time of 0.286 ns (the targeted RF period). Decreasing the rise time to 0.143 ns in Fig. 8.14b now yields numerical instability of the EM model, while the W-element shows no problems. Better yet, the length of the W-element can be scaled down freely, without consequence on the model's stability, nor any issues arising from EM frequency grid convergence, edge effects, or port calibrations.

Gate Parasitics

To extract the relevant layout parasitics at the individual LDMOS segment level, two simplified EM simulations are performed using a layout very similar to the 3D view,
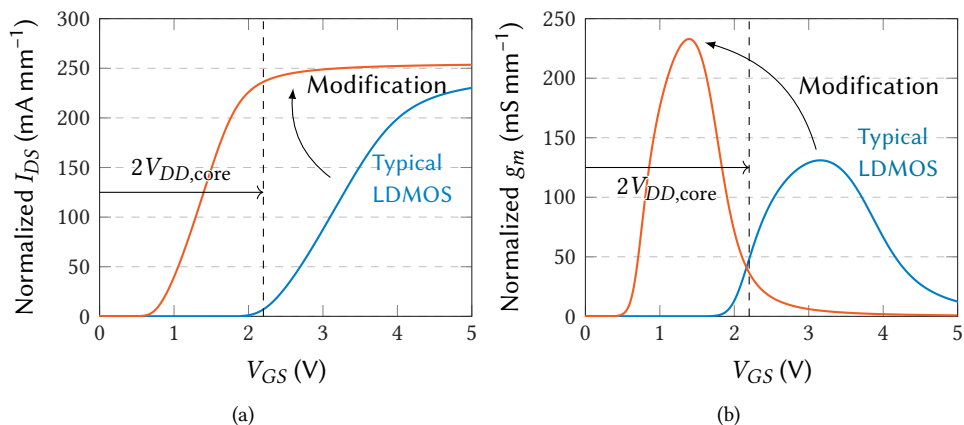
Figure 8.12: Modeled V_{GS}–I_{DS} and $-g_m$ curves based on measurement of the modified LDMOS process with thinned gate oxide, using $V_{DS} = 28\,\text{V}$.
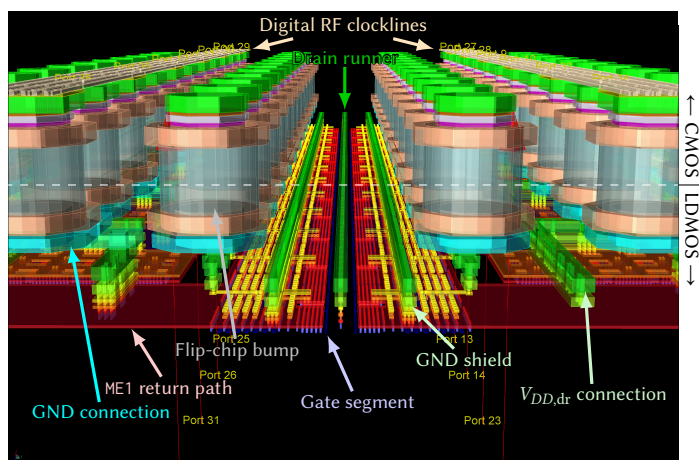
Figure 8.13: Detailed 3D view of the structure to be modeled.

(a) $t_r = 1/3.5$ ns   (b) $t_r = 1/7.0$ ns
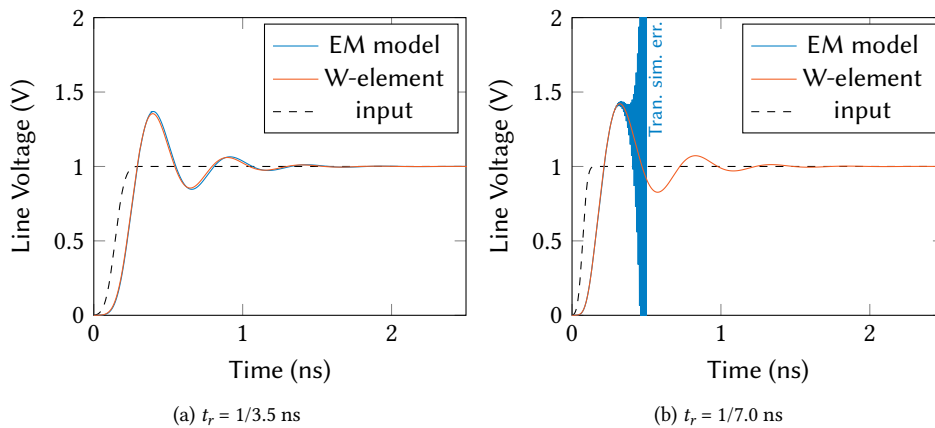
Figure 8.14: Voltage step response of two drain runner models, showing good agreement between the EM model and the fitted W-element. However, for a fast step rise time, the EM model shows numerical instability, while the W-element remains numerically stable.

as shown in Fig. 8.13. An individual segment is at most $40\,\mu m$ wide, such that lumped equivalent parasitics should suffice. The first layout focuses on extracting accurate capacitive parasitics, using accurate dimensions of the gate–drain shield metallization, drain runner, LDMOS supply and ground routing, gate and supply/ground pads and flip-chip bumps. The second focuses on extracting accurate inductive parasitics, so accurate dimensions of all current loops are used, for which the gate–drain shields can be removed to simplify, for example. Both these layouts consider a section of 4×2 unit cells, for which the parasitics of the middle ones are assumed to be representative of the total repeating structure.

Table 8.3: Capacitive gate interconnect parasitics per unit cell of $40\,\mu m$ wide.

From $\downarrow$ \ To $\rightarrow$	Gate 2	Gate 3	C_{in} weight ($V\,V^{-1}$)
Substrate	6.30 fF	6.26 fF	1
GND (ex. Shields)	1.12 fF	0.45 fF	1
Shields	2.19 fF	2.26 fF	1
$V_{DD,dr}$ (2.2 V)	1.70 fF	2.42 fF	1
Neighboring gates	3.02 fF	3.04 fF	0–0.5
Opposite gates	0.56 fF	0.56 fF	0–0.5
Drain	0.370 fF	0.388 fF	26.5
C_{in} due to interconnect (ex. substrate or ESD)	16.2 fF	16.8 fF	

Additional drain-to-shield capacitance is 0.58 fF, for $40\,\mu m$ drain runner with a ground shield at both sides.

The resulting passive interconnect capacitance values from EM simulation are summarized in Table 8.3 for the middle gate connections, gates 2 and 3. These gate connections see somewhat different capacitances, as gate 2 is closer to a V_{SS} pad, and gate 3 is closer

to a $V_{DD,\text{dr}}$ pad. Still, the driver supply can be considered a signal ground from the RF signal's perspective, so summing the capacitances to ground (including shields) and supply yields 5.01 fF for gate 2 vs. a very similar 5.13 fF for gate 3. The effective capacitance to any neighboring or opposing gates is halved, since typically those have the same activation, or worst case one of them has a different activation (inactive or the other phase). The parasitic feedback capacitance has a bigger impact, which was already discussed briefly in Section 8.2.1. The found C_{GD} is larger than was predicted using coupled microstrip lines in Table 8.1, which can be explained by the additional capacitance between the drain metal and the flip-chip bump. The interconnect metals here add 16.5 ± 0.3 fF, for which the substrate capacitance is excluded. Any substrate capacitance should already be part of the active model for the LDMOS segment, while the substrate capacitance for the CMOS will be explicitly included when extracting the parasitics of the CMOS layout. However, all LDMOS gate structures still need to be protected against ESD events, for which small diodes were placed. The capacitance values for these devices was obtained by TCAD simulation, yielding: "1.3 fF μm^{-1} (0 V) to 1.0 fF μm^{-1} (2.2 V)." A simple diode model was fitted to these values for use in simulation. The width of this protection diode is 10 μm, thus adding 10 fF to 13 fF to the LDMOS' input capacitance. To place this in perspective, the input of a nominal LDMOS MSB segment of 38 μm itself has C_{GS} = 72 fF, so all LDMOS capacitance combined results in C_{in} = 102 fF.

Next, we consider the inductive parasitics. Again, we only consider the middle gate segments, as the ones on the edge have inaccurate return paths in the EM simulation. This provides a series inductance from the CMOS driver, to the LDMOS gate, and back through the V_{SS} pad of L_{GS} = 43.58 ± 0.06 pH, as well as a series resistance of 1.92 Ω. The input capacitance we found above results in a parasitic series resonance of 75 GHz, which is a significant improvement over the 5 GHz of the bond-wire-based demonstrators of Chapter 7. We also investigate the mutual coupling between gates, which is $k_m \approx 0.20$ for neighboring gates and $k_m \approx 0.16$ for opposing gates. This yields an estimated 0.2 mV variation of the V_{GS} at the fundamental frequency, which we consider to be negligible, especially compared to other possible nonlinearities. The series inductance found in the path from the $V_{DD,\text{dr}}$ routing on LDMOS, to the CMOS supply plane, and back through the V_{SS} pad is $L_{DD,\text{dr}}$ = 27 pH. To be sure, also the inductive coupling of the gate and supply paths with the drain runner have been investigated, resulting in coupling coefficients of 0.02 and 0.002, respectively. These are very low, since their current loops are positioned in orthogonal planes, and are again considered negligible.

Segmented LDMOS Model

The active LDMOS model can be accompanied by the now-known passive layout parasitics. We target to model a single drain runner, as shown in Fig. 8.13, containing 32 LDMOS segments. Either side of the drain runner has 16 LDMOS gate segments of 38 μm wide. Several of these segment models can then be used together with (EM simulated) models of an output matching network. For example, a single sub-bank should contain 8 of these segment models, while a full bank then requires 16.

The most accurate and straightforward method to make this segment model, is to include all 32 individual gate segments together with their interconnected parasitics as a discrete simulation model (see Section 5.2.1. Two output ports are placed at either side of the drain runner, and the drain runner itself is divided into small segments modeled as the W-element discussed above. Two LDMOS unit cells are then connected between each
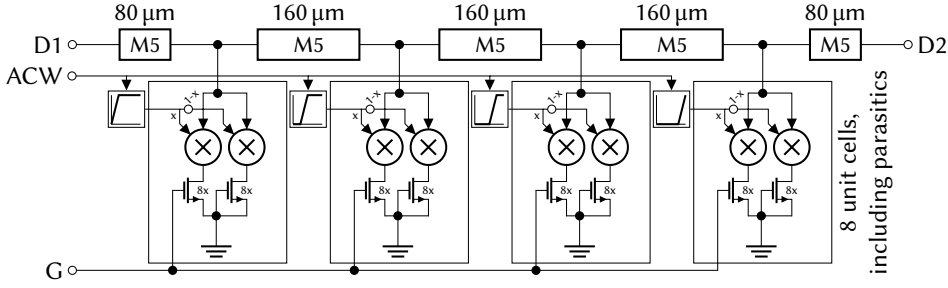
Figure 8.15: Simplified LDMOS segment model, where the 32 individual unit cells are replaced by a single continuously scalable ACW, while maintaining the layout dependent effects and activation pattern.

such segment. In addition to a ground reference port, this 'advanced' discrete model has 35 ports, allowing each gate segment to be controlled individually. This makes this advanced discrete segment model hard to use in a simulation set-up. However, it allows, for example, to accurately simulate the activation pattern and phase swaps of the individual unit cells. This requires many electrical nodes in practice, causing either long simulation times, or the simulation to not converge at all.

A more workable solution is to simplify the advanced discrete model using the current scaling simulation model (see Section 5.2.2). However, we cannot simply group all segments together into one, as the activation pattern and related layout effects would not be reflected accurately. Instead, we make four groups of 8 unit cells, where the drain current is scaled between an active and an OFF device, as illustrated in Fig. 8.15. Choosing four groups compromises simulation speed and model accuracy; doing so greatly simplifies simulation while maintaining the distributed nature of the LDMOS segments.

Both the advanced and simplified model efforts can be repeated to model the drain runner containing the LSBs. To finish up the entire flip-chip DTX model, the corresponding LSB model can then be placed in the middle of the bank. One additional point of attention is the modeling of the drain capacitance, depending on which active model is used. The parameter conversion model already contains the drain runner's parasitics of a typical LDMOS process. To not double count these parasitics, 38 μm worth of capacitance should be subtracted for every 40 μm of drain runner. This can easily be done by subtracting this value from the W-elements C-parameter. This is unnecessary when using the final proprietary RF model, as extrinsic capacitances are mostly deembedded from the model anyway.

8.4.3 CMOS: Driver and ESD

We now know the LDMOS $C_{in} = 102$ fF from modeling the LDMOS power stage, which needs to be driven by a driver in CMOS. Operating it at 3.5 GHz with an RF duty cycle of 25 % requires a pulse width of 71 ps and thus should have compatible rise and fall times. This requires a fast driver, for which a stacked topology is selected (see Section 4.2.2). Next, the various design considerations for this driver structure are discussed.

ESD Circuits

ESD protection should be in place to ensure the CMOS driver does not get damaged during the flip-chip assembly process. Standard ESD protection in this technology consists of an IO ring with supply lines and reversed diodes from the IO to these supplies. We cannot use a ring-based design here, as the switch bank IOs are placed in a grid. Furthermore, these diodes contribute a capacitance of at least 0.6 pF, not yet considering additional layout parasitics. For example, the ESD protection used in Chapter 7 together with the IO pad contributed to a load capacitance of 0.81 pF, for which the LDMOS segments were accordingly scaled (Section 7.2.1). Scaling the LDMOS segments is not an option here, so a capacitive overhead in the order of 0.6 pF is unacceptable.

We now require custom ESD protection circuitry that has a much smaller footprint, such that it can be placed in a grid, and has a lower capacitance. The point of using a ring-like design for the ESD protection is to ensure a low enough series resistance to supply ESD clamps and diodes, such that an ESD event of any polarity can be safely dissipated. Here, we have to ensure low supply line series resistance in the switch bank anyway, to ensure the supply *IR*-drop will not impact the DTX performance. We can thus assume that the supply ESD clamps and diodes can be safely placed elsewhere, not interfering with the bank design or ESD operation. Next, much smaller diodes are required, but device integrity should still be guaranteed. Fortunately, the driver only has CMOS drains connected to its output, which are more resilient against ESD events than CMOS gates.

A small test chip was taped out with decreasing customized diode sizes, ranging from medium-sized high current application (HIA) diodes (e.g., a predicted protection level of 1 kV HBM) to minimum-size regular diodes. These diodes are then connected to the IO and device drains (both thick and thin oxide devices) on one side, and to the normal IO ring on the other. These test chips were wire-bonded to a PCB to verify that they can survive manufacturing, handling, and assembly. All IOs were measured to be functional, even the thick oxide drivers with the smallest possible diodes. Still, the high current application (HIA) diodes are used for the CMOS controller design to be a bit on the safe side. These diodes do not structurally differ from a regular core voltage junction diode but are optimized in terms of their layout and are used for logic, high speed, or low capacitance ESD protection [98, Section 10.2.9.6.2]. We use small diodes with an estimated protection level of 250 V HBM for our custom ESD circuit, which has an acceptable nonlinear capacitance value (including layout parasitics) of 6.42 fF to 6.90 fF, depending on the IO voltage while assuming a supply of 2.2 V.

Driver Design

The total capacitive load for the CMOS driver is now set to a nominal value of 109 fF when including the CMOS ESD. Allowing for some design margin, we assume a nominal load value of 150 fF for the CMOS driver, which it should drive within the 25 % duty cycle. Achieving this kind of speed with a thick oxide driver chain is not feasible, so instead, we aim for a stacked driver topology in thin oxide. The resulting driver topology is shown in Fig. 8.16 [99].

Ensuring this driver reaches the full 2.2 V swing over all process corners is still challenging for this technology and speed requirement. Since we target a 25 % duty cycle, we can allow the driver's fall time to be larger, as there is more time for each RF cycle to reach 0 V than there is to reach 2.2 V. We can scale down the size of the NMOS devices to lower the output capacitance, making the PMOS relatively faster. Even though this makes the rise
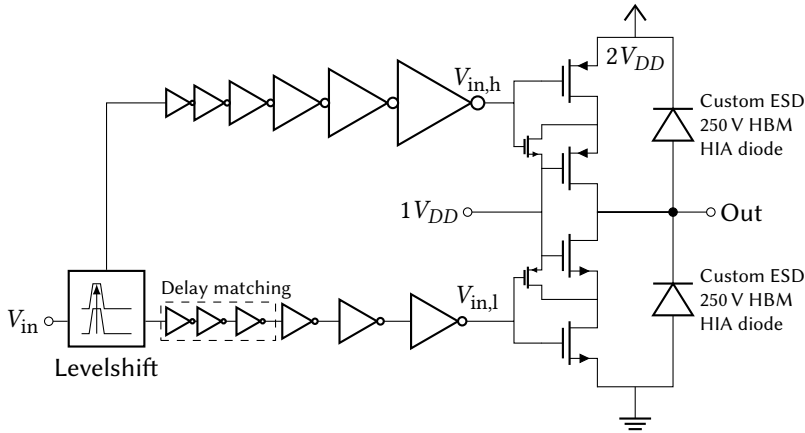
Figure 8.16: The stacked driver topology with its tapered buffer chains and ESD diodes.

and fall times asymmetrical, the result is a more power-efficient driver stage, consuming a nominal 1.60 mW at 3.5 GHz, excluding the load. With a load of 150 fF, this becomes 4.14 mW.

The driving chains are sized according to the device they are driving, resulting in a larger and longer chain for the top PMOS than for the bottom NMOS. Additional inverters are added to the bottom signal to equalize the number of inverters, and, thus, the top and bottom propagation delays. The top chain requires 2.1 mA when active, whereas the bottom chain only requires 0.17 mA due to its smaller load and size. The resulting nominal driver performance in terms of rise and fall times (10–90 %), for the ESD and an external load of 150 fF, are respectively 29 ps and 69 ps (linearly scaled to 0–100 %, these are related to a $t_{r/f}$ of 12.8 % and 30.5 %). The full driver (again including load) then requires a supply current of 3.56 mA at the 2.2 V domain, and −0.91 mA at the 1.1 V domain.

Level Shifter
The level shifter for the stacked driver needs to shift the core logic level (V_{SS} to $V_{DD,\text{core}}$) upward to $V_{DD,\text{core}}$ to $V_{DD,\text{dr}} = 2V_{DD,\text{core}}$ and is required to be DC-coupled. The level shifter should also be capable of handling the narrow voltage pulses while maintaining good rise and fall times. It uses a similar structure as used in the stacked driver design, resulting in the schematic shown in Fig. 8.17, but the top part uses a cross-coupled PMOS latch. These nodes are internal to the level shifter, and are also AC-coupled to the opposite input polarity using a capacitor. This coupling capacitor helps the output to latch faster to the new logic state than what would be the case without [72]. The settling of the node between the bottom two NMOS devices is aided by an additional PMOS device, similar to what is used in the driver [99]. While active, it only consumes 32 µA at the 2.2 V domain, which was already included in the above-mentioned numbers.

Driver Simulation Model
The driver's performance should also be captured correctly when combined with the segmented LDMOS model, for example, Fig. 8.18. Simulating the driver as a simple equivalent resistor will not suffice anymore, as the load capacitance can vary (e.g., due to the difference
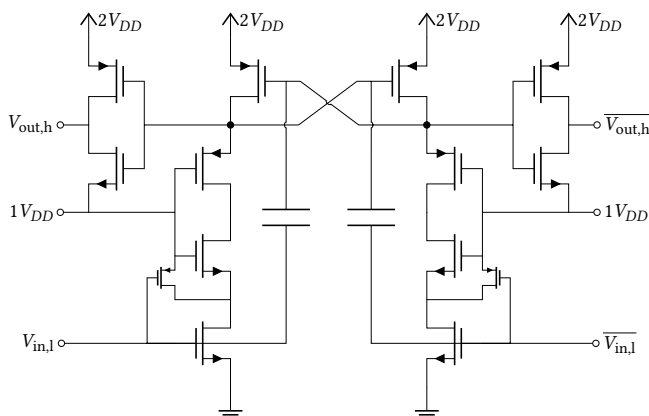
Figure 8.17: The schematic of the level shifter responsible for shifting from V_{SS}–$V_{DD,\text{core}}$ to $V_{DD,\text{core}}$–$2V_{DD,\text{core}}$ = $V_{DD,\text{dr}}$.

8
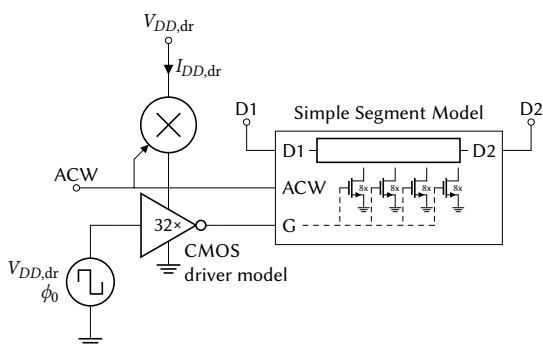
Figure 8.18: Combining the CMOS driver model with the simplified LDMOS segment model from Fig. 8.15 to accurately reflect the driver's speed and power consumption.
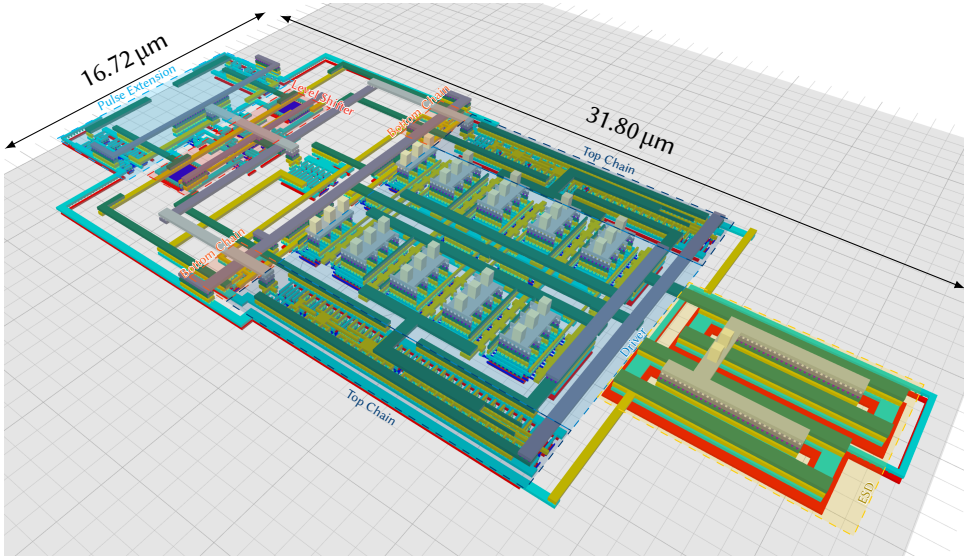
Figure 8.19: 3D view of the complete realized driver structure, including ESD diodes, buffer chains, level shifter, and pulse extension.

between MSBs and LSBs, or when using LDMOS devices for different supply voltages, see Section 8.6). Besides the CMOS design being in different EDA software than the LDMOS power stage model, using the full driver schematic is also too complicated to use in frequency domain simulations. Instead, we can use a resistive switch element with a certain saturation current, similar to how actual metal–oxide–semiconductor (MOS) devices would operate. The circuit implementation of this current-saturating resistive switch is provided in Sections B.2.2 and B.2.4.

The power consumption and speed of the CMOS driver model should be identical to the power consumption and speed of the simulated CMOS driver with parasitics. While fitting these parameters, it was observed that the resulting system efficiency was quite poor due to lower-than-expected RF output power. A contributing factor is that the LDMOS current pulse was narrower than the targeted 25 % duty cycle, which can be explained by combining the V_T of the LDMOS devices with the drive pulse's rise and fall times. Further, the relative $t_{r/f}$ (12.8 % and 30.5 %) also degrade the achievable fundamental output current. To counteract these issues, a digital pulse extension circuit was added to the CMOS driver prior to the level shifter, targeting a 32 % duty cycle after taking the LDMOS V_T into consideration.

The complete layout of the resulting driver is shown in Fig. 8.19, including the pulse extension circuit, level shifter, chains, driver, and ESD diodes. Some key performance metrics of this driver circuit, including the parasitics of this layout, are characterized over a range of external load capacitances. These key performance metrics relate to the driver's wave shape, power consumption, and delay. The output duty cycle is only a part of the wave shape, as it, by definition, only captures the width at the 50 % voltage point of the pulse. Here, we additionally define the 25 % and 75 % voltage point duty cycles to characterize the driver's performance. Figure 8.20 shows the targeted and the fitted model's
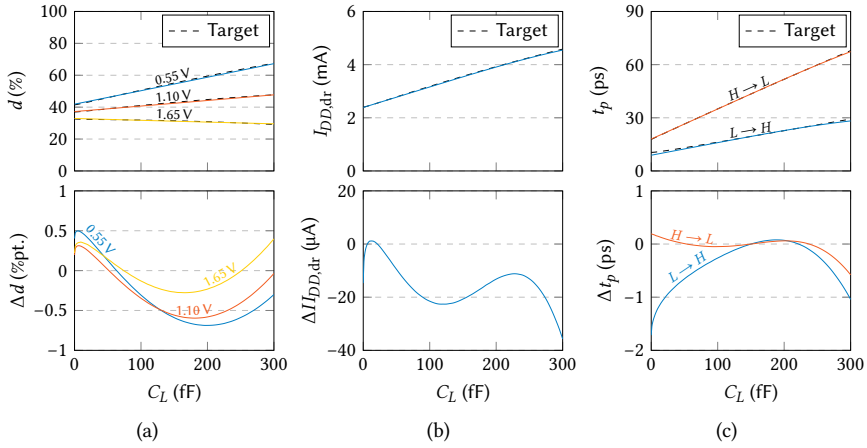
Figure 8.20: Fitting a behavioral driver model (see Section B.2.4) to the post-layout simulation results while varying the load capacitance C_L: (a) the output duty cycle for three selected voltage points; (b) the consumed supply current; (c) variation of the propagation delay normalized to $C_L = 150\,\mathrm{fF}$. The top grahps show the absolute fitting results, while the bottom graphs show the difference.

performance metrics in the top graphs, and their differences in the bottom graphs. The delay is characterized primarily for variations around the nominal load of 150 fF, since the model only captures a single driver stage and not the full chain. As such, a constant $t_{pL \to H}$ of 247.3 ps is subtracted, as well as a $t_{pH \to L}$ of 273.9 ps. A different normalization point could also have been chosen to characterize the driver, but, as can be seen in Fig. 8.20c, this will barely influence anything, as the driver's propagation delay is approximately linearly increasing with the load capacitance.

8

8.5 DC Supply Requirements for Wideband Operation

The CMOS drivers' actual supply voltage on-chip can have a big impact on a DTX's transfer, as was concluded from the measurements of earlier demonstrators (see Sections 7.5.2, 7.5.3 and 7.7.3). The flip-chip assembly has many more ground and supply connections available inside the switch bank, providing an improved starting point for our supply decoupling design. The next sections introduce a theory for improved DTX decoupling design, leading to the design requirements and implementation for the high-resolution DTX demonstrators.

8.5.1 Definition of the Relevant Frequency Regions

A successful DC decoupling means that supply voltage variations are suppressed to such an extent that they do not negatively affect the circuit performance. This can be achieved by ensuring a low impedance path from the on-chip circuits to the power supply for all relevant frequencies. A schematic illustrating this conceptually is provided in Fig. 8.21a. To quantify what constitutes as 'low enough' voltage variations and impedances, we first need to understand the drivers' current requirements at these 'relevant frequencies.' The relevant frequencies can be divided into the RF carrier and its harmonics, DC, and baseband, which are discussed next.
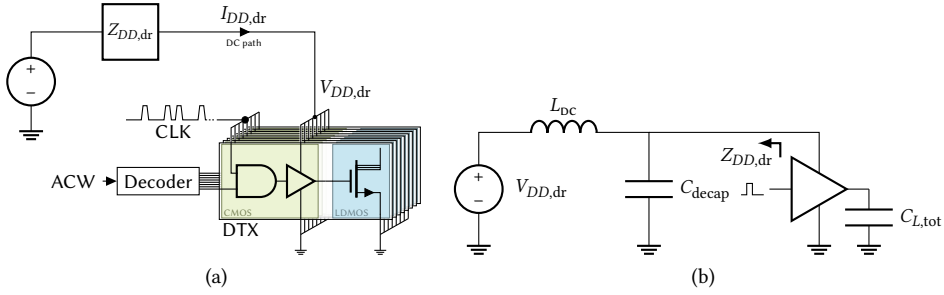
Figure 8.21: Schematic of the CMOS driver supply path: (a) conceptual supply path and its impedance; (b) simplest schematic of a supply decoupling structure.

Decoupling at RF

The situation for the RF carrier is illustrated in Fig. 8.21b, where a digital driver aims to quickly charge the load C_L to the driver's supply voltage $V_{DD,dr}$. This situation can be understood conceptually by assuming the feed inductance L_{DC} to pose an open circuit condition during the charging of C_L, such that all the driver current must be supplied by the (close-by) DC-decoupling capacitance C_{decap}. The driver also has capacitance itself (see Section 4.1.3, Eq. (4.10)), such that the total charge drawn from C_{decap} becomes

$$q_{DD,dr} = MC_{L,tot}V_{DD,dr}. \tag{8.8}$$

After this charge is drawn, the resulting change in the supply voltage is then

$$\Delta V_{DD,dr} = \frac{q_{DD,dr}}{C_{decap}}. \tag{8.9}$$

Intuitively, this means that if we aim for a maximum 10 % variation due to the RF current, we require a minimum C_{decap} of 10× larger than MC_L. Or, for example, a maximum 1 % variation will then require a 100× larger C_{decap}, and so on. The charge on C_{decap} is then replenished through the feed inductance when the driver is discharging or inactive. The C_{decap} has to be designed for the case that all drivers are active at the same time, even though the average number of active drivers is lower.

Translating this statement in terms of allowable supply impedance $Z_{DD,dr}$, the maximum RF current drawn by the drivers is the cycle charge multiplied by the number of cycles per second, i.e., $I_{DD,dr,max} = f_c q_{DD,dr}$. In this scenario, where we assume the time domain driver current to be an impulse train (i.e., $I_{DD,dr}(t) = q_{DD,dr}\text{Ш}(f_c t)$), the steady-state supply voltage becomes a sawtooth wave at the repetition rate of f_c with a peak-to-peak variation of $\Delta V_{DD,dr}$. The fundamental (i.e., at f_c) voltage magnitude of this square wave is then $\Delta V_{DD,dr}/2\pi$, such that

$$\left| Z_{DD,dr}\left(f_c\right) \right| \le \frac{\Delta V_{DD,dr}/2\pi}{f_c q_{DD,dr}} = \frac{\Delta V_{DD,dr}}{2\pi I_{DD,dr,max}}. \tag{8.10}$$

As a sanity check, we can substitute in Eq. (8.9), which gives us exactly the impedance of the decoupling capacitor

$$\left| Z_{DD,dr}\left(f_c\right) \right| \le \frac{q_{DD,dr}/2\pi C_{decap}}{f_c q_{DD,dr}} = \frac{1}{2\pi f_c C_{decap}}. \tag{8.11}$$
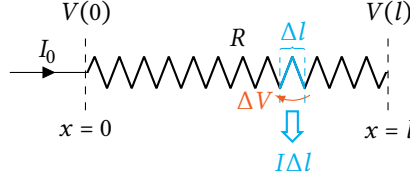
Figure 8.22: Calculating the effective resistance of a distributed resistive line with a constant uniformly distributed current drawn, resulting in an effective $R/2$ at the end of the line.

This holds for any f_c in the targeted RF bandwidth, and, since we have assumed an impulse train for the driver current, also for its harmonics kf_c.

'Decoupling' at DC

At DC, all capacitances pose an open circuit and all inductances are a short circuit, such that the only remaining factor is the series resistance found in the supply path. No dynamics are present at DC, such that 'decoupling' is not the correct term. However, we still want to minimize the (static) voltage drop across the supply path, which relates to the average power consumption. The DC power consumption of the DTX drivers is given by Eq. (6.37), which, to summarize, mainly depends on the average number of activations and the capacitance that needs to be charged, i.e.,

$$P_{DD,\text{dr}} \approx \rho_{\text{avg}} \cdot f_c M C_{L,\text{tot}} V_{DD,\text{dr}}^2 = \rho_{\text{avg}} P_{DD,\text{dr,max}}, \tag{8.12}$$

where the average magnitude ρ_{avg} for a single line-up is, by definition, equal to $\sqrt{\text{PAPR}}$. There is only a factor difference related to the bank implementation and upconversion method to provide the exact value (see Eq. (6.28)), but for simplicity we here assume single-line up polar operation, for which this factor is simply 1. This gives a DC current of

$$I_{DD,\text{dr}}(0) = \frac{\rho_{\text{avg}} P_{DD,\text{dr,max}}}{V_{DD,\text{dr}}} = \rho_{\text{avg}} \cdot f_c M C_{L,\text{tot}} V_{DD,\text{dr}}. \tag{8.13}$$

The tolerable DC resistance (i.e., *IR*-drop) then becomes

$$R_{DD,\text{dr}} = \frac{\Delta V_{DD,\text{dr}}}{V_{DD,\text{dr}}} \cdot \frac{1}{\rho_{\text{avg}} \cdot f_c M C_{L,\text{tot}}} = \frac{\Delta V_{DD,\text{dr}}}{\rho_{\text{avg}} I_{DD,\text{dr,max}}}. \tag{8.14}$$

The on-chip supply line resistance can contribute a significant amount to the total DC resistance seen from the driver. This means the *IR*-drop depends on the position of a driver and is thus not the same for each driver. The worst case is at the end of the line, for which we need to determine the effective resistance a driver sees under the influence of the other drivers along the resistive line. This situation is conceptually shown in Fig. 8.22. We can find that a small line section has a resistance of $R\Delta l/l$. The total current through the line at position x is then is then $I(x) = I_0(1 - x/l)$, such that the voltage drop across that section of line is $\Delta V = I(x)R\Delta l/l$. The voltage at position x can then be found using

$$\lim_{\Delta l \to 0} V(x) = \int_0^x I_0\left(1 - \frac{x}{l}\right)\frac{R}{l} \, \mathrm{d}l = \frac{I_0 R}{l}\left(x - \frac{x^2}{2l}\right), \tag{8.15}$$

which gives for $V(l) = I_0 R/2$. In other words, the effective distributed resistance of a line is half the line resistance as observed at the end of the line. This may seem familiar when considering a distributed RC line from RF models of FET gates [100] or the base resistance R_{bv} in bipolar junction transistors (BJTs) [101], but in these cases, the equivalent resistance is a factor 3 smaller. The difference is that here we consider the effect seen at the end of the line, rather than finding an equivalent resistance to model the transfer from a changing source. Our assumption is based on a signal ground at $x = 0$ and a uniform current drawn across the line, which is only valid as long as the voltage change over the line is small. Having a small voltage change is exactly what we aim for, and thus a valid assumption for our application.

Decoupling at Baseband

We have assumed the current at RF to be an impulse train, which is not true when considering (complex) modulated signals. Aside from the obvious fact that the current pulses will not be a Dirac delta, their magnitude is also scaled by the applied ACW, while also being phase modulated. This means that the time-domain signal envelope is multiplied with the impulse train, leading to a convolution in the frequency domain of the envelope and impulse train spectra. The impulse train also has a DC component, meaning that a copy of the current envelope will thus be present in the baseband. While a phase reference is present at RF, this is not the case for the baseband. This is explicitly captured, e.g., in Eqs. (A.26), (A.27) and (5.12), and the implications for the time domain current made in the sign/phase inequality as discussed in Section 5.1.3. This implies that the spectrum of the envelope's magnitude is present around DC, which we define here as the baseband current.

The time domain baseband current into the drivers is here proportional to the sum of all ACWs. So, in polar the baseband current $I_{\mathrm{BB}}(t) \propto \rho(t)$, in signed-Cartesian $I_{\mathrm{BB}}(t) \propto |I(t)| + |Q(t)|$, and in multi-phase $I_{\mathrm{BB}}(t) \propto |A(t)| + |B(t)|$ (see Section 2.1). The absolute values need to be used since the ACW(s) are always positive, while in baseband the RF phase modulation has no influence. Maybe more physical intuitively, the absolute value relates to the driver current flowing from the power supply through the driver into the load capacitance, while discharging the load has the current flowing into the ground. In other words, the current from the supply in baseband has always to be positive regardless of sign or phase of the RF output.

In general, it is challenging to describe the baseband current in the frequency domain for a complex modulated signal since the absolute value is a nonlinear operation. Furthermore, it will strongly depend on the type of modulation used, as well as the upconversion method. However, a two-tone scenario concentrates the baseband current into a single fundamental baseband frequency, allowing an analytical expression of the baseband current magnitude in the frequency domain, while being independent of the used upconversion method. Simulating a DTX in a two-tone scenario is discussed in Section 5.2.3, where the baseband signal is described by $A\cos(\omega_{\mathrm{TT}} t)$, which is split in activations at 0° and 180° using two half-rectifier functions to translate the baseband signal to RF activations. This means the baseband current for a two-tone can be given by a fully rectified version using the absolute value of the baseband description

$$I_{\mathrm{BB}}(t) = I_{DD,\mathrm{dr,max}} \left| A\cos\left(\frac{\omega_{\mathrm{TT}} t}{2}\right) \right|. \tag{8.16}$$

The Fourier series of this function can be easily calculated, such that its baseband harmonic components at $k\omega_{\mathrm{TT}}$ are given by

$$I_{\mathrm{BB}}[k] = -A\, I_{DD,\mathrm{dr,max}} \frac{4}{\pi(4k^2 - 1)} \tag{8.17}$$

and for DC

$$I_{\mathrm{BB}}[0] = I_{DD,\mathrm{dr,max}} \frac{2A}{\pi}. \tag{8.18}$$

So, the baseband current is concentrated at f_{TT}, even though one might initially expect it at $f_{\mathrm{TT}}/2$, as that is the spacing of each tone to the carrier signal. This is due to the absolute value used, which is also observed in Section 5.2.3. The Fourier series of the two half-rectifiers is an infinite series, showing the bandwidth expansion of using the absolute value. The Fourier series of other rectifying activation functions are provided in Section A.7.1.

We can use the two-tone scenario as a proxy for a complex modulated signal. We can specify the tolerable supply impedance again by dividing the tolerable voltage variation at the harmonic baseband components of the rectified two-tone signal. Setting $A = 1$ gives

$$\left| Z_{DD,\mathrm{dr}}\left(kf_{\mathrm{TT}}\right) \right| \le \frac{\Delta V_{DD,\mathrm{dr}}}{I_{DD,\mathrm{dr,max}}} \cdot \frac{\pi(4k^2 - 1)}{4}. \tag{8.19}$$

This represents the worst-case scenario, since the baseband currents are spread out for a complex modulated signal over the baseband spectrum. Still, assuming a certain channel bandwidth means that, due to the rectifying activation function, the majority of its current spectral density components occupy the full channel's bandwidth from DC, rather than half of it.

Impedance Mask

With the three relevant frequency regions defined, their information can be combined to set a maximum tolerable supply path impedance. We assume the requirements from Section 8.1 as example: a carrier frequency of 3.5 GHz, and a signal bandwidth of 600 MHz. We assume a normalized $\Delta V_{DD,\mathrm{dr}} = 1\,\mathrm{V}$ and $I_{DD,\mathrm{dr,max}} = f_c q_{DD,\mathrm{dr}} = 1\,\mathrm{A}$. Using the two-tone scenario, the maximum allowed DC resistance becomes related to the average magnitude of a two-tone, following from Eq. (8.18) is $2/\pi$ (3.92 dB PAPR), becomes $R_{DD,\mathrm{dr}} = \pi/2$ Ohm. Similarly, we can sweep the two-tone spacing from 0 Hz to 600 MHz while keeping track of the tolerable impedances for all baseband frequency and RF components and fill in the maximum currents to get the maximum tolerable impedances, as shown in Fig. 8.23. The red shaded area forms an impedance mask that the supply impedance should be lower than.

8.5.2 Sensitivity Analysis

To motivate the maximum tolerable supply variation, we should look at its impact on the DTX performance in terms of its transfer (Section 5.1.3). For that, we can assume the modulated carrier to be the only relevant transfer, i.e., the outgoing wave $b_{2[1]}$ of the DTX. An additional port for the driver's supply path can be added to the DTX black box model, whose supply (small-signal deviation) modifies the digital forward transfer D_{21}.

For DC, only the average driver supply voltage changes depending on the type of output signal. This yields a static change of the DTX transfer, which can be easily solved in
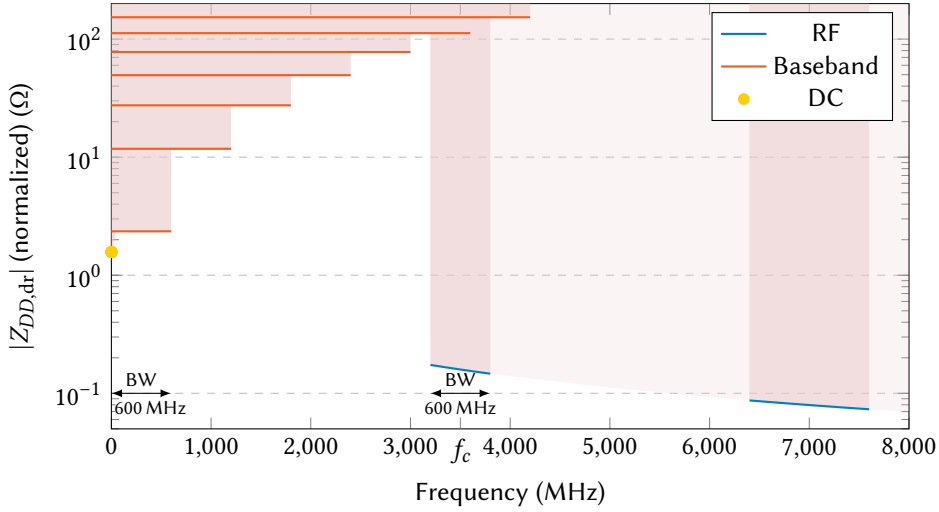
Figure 8.23: The normalized impedance mask when assuming $I_{DD,\mathrm{dr,max}} = 1\,\mathrm{A}$ and the same maximum $\Delta V_{DD,\mathrm{dr}} = 1\,\mathrm{V}$ for the three relevant frequency regions: RF, baseband, and DC. The supply impedance $Z_{DD,\mathrm{dr}}$ seen from the drivers should remain below this impedance mask, so out of the red shaded areas.

DPD, provided the *IR*-drop is limited enough not to impair the driver's functionality. Other parameters also change with the signal type (i.e., PAPR) that may be more impactful, such as the LDMOS junction temperature, which impacts the transfer and should also be addressed by the DPD.

At RF, the peak-to-peak voltage variation of the driver's supply depends only on the driver's instantaneous activation level, provided the supply voltage reaches its original value after one RF cycle. This also results in a static change in the DTX's transfer, which, however, does not depend on signal type. Thus, this can also be compensated using DPD, as long as the minimum peak voltage does not interfere with the driver's functionality and the maximum peak voltage does not impact device reliability. To evaluate the impact (e.g., in simulation), the actual current waveform of the drivers should be used rather than the impulse train used for initial analysis. For example, the current waveform is much more spread out over the RF cycle due to the driver chain delay. This can also be utilized by design, for example, by implementing a push-pull architecture where the drivers activate twice per RF cycle, causing the driver current to be concentrated around the second harmonic of the RF carrier. This increases the effectiveness of capacitive decoupling, provided the supply impedance between the decoupling capacitors of the push and the pull drivers is low enough to be effective. Special attention should be paid to the instantaneous voltage variation at RF when operating with multiple activation phases during an RF cycle, as is the case in signed Cartesian and multi-phase upconversion. The earlier activation may change the instantaneous voltage available for the driver for the next activation, impacting its magnitude, and necessitating a 2D LUT to correct it. This can be an argument to minimize $\Delta V_{DD,\mathrm{dr}}$ in the RF region, and can be helped by having an ultra-local low impedance path, e.g., by placing decoupling capacitors connected as directly as possible to a single driver.

Any supply variation in the baseband region can have a much greater impact on the

DTX performance. Namely, it should be regarded as a memory effect, changing the DTX's transfer depending on past activations. We here only consider the influence of the baseband $\Delta V_{DD,\mathrm{dr}}$, so, without the RF-related peak-to-peak variations superimposed on it. We can then evaluate what effect $\Delta V_{DD,\mathrm{dr}}$ has on the steady-state transfer. The sensitivity of phase variation will depend on the change in the driver's delay

$$\frac{\partial(\angle D_{21})}{\partial V_{DD,\mathrm{dr}}} = \frac{\partial t_p \cdot \omega_c}{\partial V_{DD}} \quad (\mathrm{rad}). \tag{8.20}$$

We know the delay change from Eq. (4.18), providing the sensitivity in $\frac{\Delta V_{DD,\mathrm{dr}}}{V_{DD,\mathrm{dr,nom}}} = 0$

$$\frac{\partial(\angle D_{21})}{\partial V_{DD,\mathrm{dr}}} = -\alpha t_{p,\mathrm{nom}}\omega_c \quad (\mathrm{rad}). \tag{8.21}$$

The sensitivity of the amplitude will vary with the change in the driver's output voltage, causing a change in the current provided by the LDMOS, as well as a change in the fundamental current due to changes in rise and fall time t_{rf}. If we assume the current change to be dominant, we can write the amplitude sensitivity as

$$\frac{\partial|D_{21}|}{\partial V_{DD,\mathrm{dr}}} \approx \frac{\partial I_{DS,\mathrm{LD}}}{\partial V_{GS,\mathrm{LD}}} = g_{m,\mathrm{LD}}, \tag{8.22}$$

which provides another motivation to minimize the LDMOS g_m. Linearizing around the nominal point provides the changes in AM and PM by

$$\Delta\mathrm{AM} = \frac{\partial|D_{21}|}{\partial V_{DD,\mathrm{dr}}}\Delta V_{DD,\mathrm{dr}} \qquad \Delta\mathrm{PM} = \frac{\partial(\angle D_{21})}{\partial V_{DD,\mathrm{dr}}}\Delta V_{DD,\mathrm{dr}} \tag{8.23}$$

which in a two-tone scenario gives by first-order approximation

$$\mathrm{IM}_3(\mathrm{dBc}) \approx 10\log\left(\frac{1}{4}\Delta\mathrm{PM}^2 + \frac{8}{35}\Delta\mathrm{AM}^2\right). \tag{8.24}$$

Note that this is not an actual IM_3 product originating from odd-order distortion of the two tones at RF, but the result from the second baseband harmonic upconverted to RF.

Depending on DTX architecture and requirements, it might be preferable to define different tolerable $\Delta V_{DD,\mathrm{dr}}$ values for the different frequency regions. This can relax the impedance requirements on one frequency region while necessitating more strict ones on another.

8.5.3 Implementation

Now that the frequency regions relevant for DTX operation and their associated tolerable supply impedances are defined, we can fill in the parameters for the targeted high-resolution DTX demonstrators. As discussed in Section 8.4.3, one CMOS driver requires a nominal supply current of 3.56 mA when switching at 3.5 GHz (i.e., $q = 1.02\,\mathrm{pC}$) and is loaded by 150 fF (i.e., $q_L = 330\,\mathrm{fC}$). For a full switch bank containing 512 MSB cells this means $I_{DD,\mathrm{dr,max}} = 1.82\,\mathrm{A}$. Setting a maximum supply variation of 10 % yields $\Delta V_{DD,\mathrm{dr}} \leq 0.220\,\mathrm{V}$.
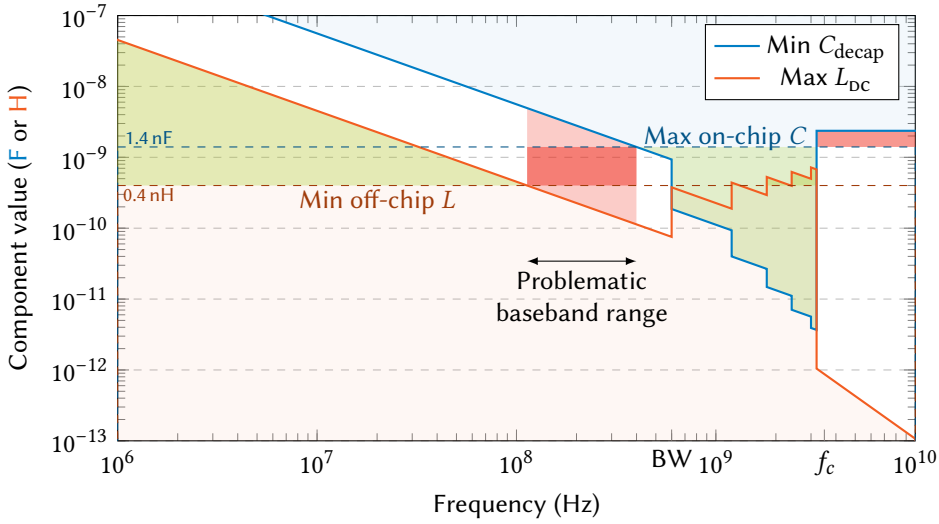
Figure 8.24: The impedance mask translated to component values. If the maximum tolerable inductance is higher than the minimum implementable inductance, the impedance mask requirements can be met, and vice versa for the capacitance. These areas are shaded green. The frequency ranges are shaded red where neither requirement is met.

Translating the Impedance Mask to Component Values

Filling in the requirements for decoupling at RF gives a maximum impedance $\left|Z_{DD,\mathrm{dr}}\left(f_c\right)\right| \leq$ $\frac{0.220\,V}{2\pi \cdot 1.82\,A} = 19\,\mathrm{m\Omega}$, or directly as a minimum decoupling capacitor (per bank) of $C_{\mathrm{decap}} \geq$ $\frac{1.82\,\mathrm{A}}{0.220\,\mathrm{V} \cdot 3.50\,\mathrm{GHz}} = 2.37\,\mathrm{nF}$. The total driver-related capacitance charged is $\frac{1.02\,\mathrm{pC}}{2.20\,\mathrm{V}} \cdot 512 = 237\,\mathrm{pF}$, and setting a maximum 10 % variation at RF sensibly yields this 2.37 nF. However, initial calculations show it is only possible to implement 1.4 nF on-chip, for which we need to consider other measures later.

We can repeat this procedure for the entire impedance mask, that is, the normalized mask of Fig. 8.23, scaled by 0.220/1.82 = 0.121. Assuming the simplest decoupling structure available, a feed inductor and a decoupling capacitor, as shown in Fig. 8.21b, we can calculate the maximum allowed inductance or the minimum required capacitance that fulfils the required impedance levels. The calculated component values vs. frequency are plotted in Fig. 8.24; the blue line indicates the minimum required C_{decap} and has the area above it shaded blue to indicate these are the values that result in a low enough impedance. Similarly, the orange line indicates the maximum allowed L_{DC}, with the area below it shaded orange. Next, initial calculations indicated that it is possible to implement 1.4 nF of decoupling capacitance on-chip, per bank, which is a maximum indicated in the figure as a dashed line, and a 0.4 nH effective series inductance due to bond wires and supply routing over the LDMOS die, as a minimum also indicated by a dashed line. As long as the maximum tolerated inductance is higher than the minimum implementable inductance, we can meet the requirements set by the impedance mask. For the capacitance, this is the other way around; as long as the minimum required capacitance is lower than the maximum implementable capacitance, the impedance mask requirement can be met. The
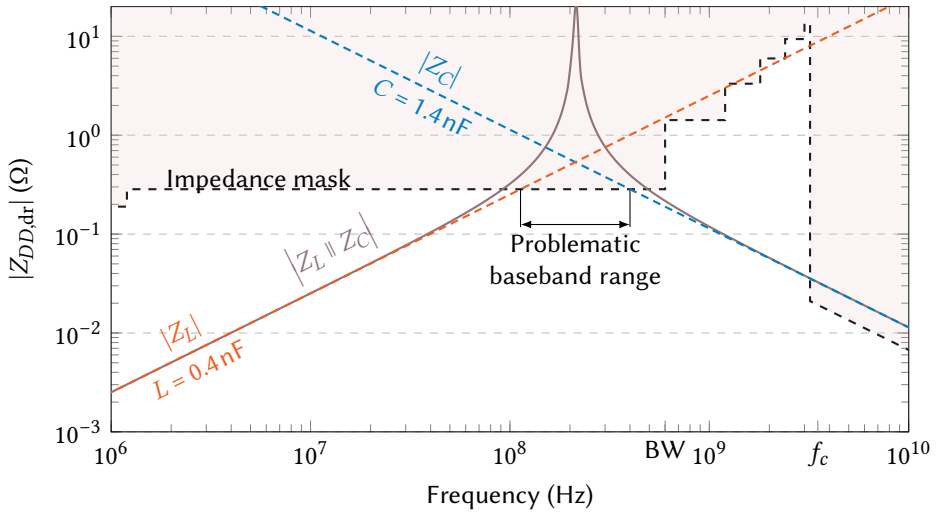
Figure 8.25: The actual impedance as seen by the driver when using only an (ideal) feed inductance and decoupling capacitance, while assuming realistic implementable values.

green shaded areas indicate these regions. The regions where neither required value is met are shaded red, indicating a problem. A problem arises for the decoupling at RF, as already recognized above, but also for the baseband region between 113 MHz and 400 MHz. This means there are four options: increase the capacitance available on-chip, reduce the inductance to the chip, reduce our considered bandwidth, or relax our requirements on the tolerated voltage variation.

Passive Decoupling for Minimized Resonance

Let us ignore the voltage variation requirements for now, since another problem occurs when only a feed inductance and a decoupling capacitance are considered. The inductor and capacitor are a perfect parallel resonator with impedance $Z_{DD,\mathrm{dr}} = Z_L \parallel Z_C$ from the CMOS driver's perspective. Its impedance magnitude is shown in Fig. 8.25, where the parallel resonance shows an infinite impedance peak. The location of this peak could already be predicted from the component value mask of Fig. 8.24, namely in the (geometric) middle of the problematic baseband range.

In a practical scenario, this resonance will never be this sharp. There is always some resistance present in the supply path, especially relative to the low impedance levels we target. This means we can also add resistance to this system on purpose in an attempt to dampen the supply's parallel resonance. This resistance can be added in two locations: in series with the decoupling capacitance or in series with the feed inductance, as shown in Figs. 8.26a and 8.26b, respectively. Damping resonance at the capacitor degrades the supply impedance towards the RF region, where the impedance requirements are the most strict. The tolerable resistance values for the RF region hardly dampen the resonance, and vice versa. Damping the resonance at the inductor now degrades DC performance, increasing the *IR*-drop, which neither provides any satisfactory resistance value, nor does balancing the resistance between the two shunt paths.

(a) series-*RC* parallel to *L*

(b) series-*RL* parallel to *C*
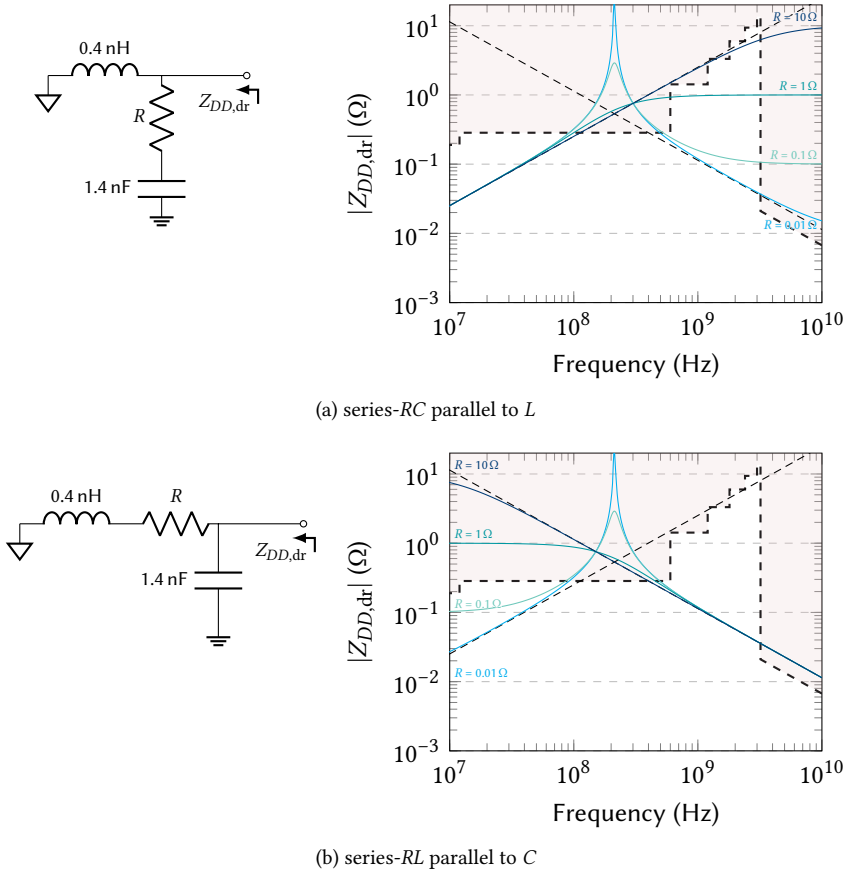
Figure 8.26: Damping the inevitably occurring resonance peak, here within the unwanted baseband region. Here the resonance is dampened by placing a resistance in series with (a) the decoupling capacitor, or (b) the feed inductance. Neither method meets the impedance mask requirements.

8
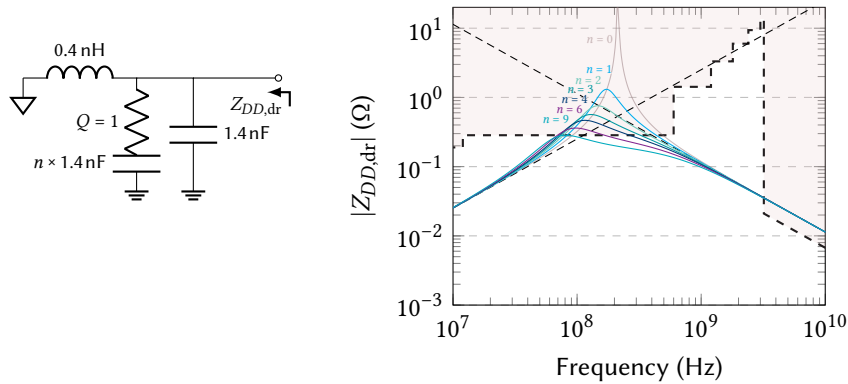
Figure 8.27: Placing a capacitor with $Q = 1$ parallel to the nominal (high-Q) decoupling capacitor, damping the resonance peak effectively.

Clearly, a different strategy is required. One very effective strategy is to add another damped *RC* in parallel to the required C_{decap}, as shown in Fig. 8.27. Adding a capacitor with a *Q* of 1 at the theoretical undamped *LC*-resonant frequency is always the best solution in this strategy. To be of any effect, the added capacitance in a damped shunt *RC* combination should provide a low impedance to have any significant effect. For example, adding a 1 pF capacitance with a series resistance of 800 Ω to dampen a 200 MHz resonance caused by a 1 nF capacitance (−16j mΩ) will have no effect whatsoever. The structure as shown in Fig. 8.27 presents near identical impedances for DC and RF, while effectively damping the parallel resonance. In this case, adding nine times the 'primary' decoupling capacitance allows to meet the requirements of the impedance mask in baseband. Using this as an example, the total capacitance present after the 0.4 nH feed inductance is now 14 nF, which in theory resonates at 67.3 MHz. As such, the added capacitance (12.6 nF) should have *Q* = 1 at this frequency, necessitating a series resistance of 188 mΩ. Obviously, stating that we now require ten times the achievable capacitance on-chip to meet the mask is quite sudden. However, there is no specific requirement for it to be on the CMOS chip, as long as it is after the most significant feed inductance, which in this case are the bond wires from the PCB to the LDMOS die.

From a theoretical standpoint, this design philosophy—using damped shunt capacitors—can be repeated several times, leading to the theoretical values as provided in Section A.7.2. This widens the bandwidth for which *Q* ≈ 1, which can help deal with uncertainty in the actual inductance of the supply path. It could potentially even be used for the (on-chip) RF decoupling provided the bandwidth is large enough. However, as can be seen from Table A.2, widening the dampened frequency range of the decoupling structure quickly grows the total capacitance required, which may not be feasible to realize (on-chip).

Additional Implementation Measures

The lines for routing the supply on the LDMOS are characterized using EM simulation, from which the line's distributed element equivalent values are extracted (see Section A.6). This procedure is like what was done in Section 8.4.2, but here the emphasis for fitting is placed at the lower frequency region, e.g., DC to 1 GHz. A supply routing model of the CMOS and LDMOS is made using the extracted line parameters and mimicking the layout. This way, all 640 $V_{DD,\text{dr}}$ connections of the entire CMOS die can be assigned their own port. The distributed impedance as seen from the drivers can be simulated using AC simulation by connecting an AC current source to each port that injects 1/160 A. The voltage at each port is then equal to the $Z_{DD,\text{dr}}$ since exactly 1 A is injected for one switch bank.

The effective feed inductance per switch bank due to the bond wires from the PCB to the LDMOS power die is 0.2 nH. However, the supply lines on the LDMOS also add inductance. High-density capacitors (HDC) are added along these lines in an effort to lower their characteristic impedance, and thus their inductance. Unfortunately, this does not have the desired effect, as only the line's *C* parameter is increased, while the *L* parameter remains largely unchanged. Put differently, the line's characteristic impedance and propagation velocity drop, such that the electrical length of the line does not change (by first-order approximation for low frequencies).

Nonetheless, we can use these HDCs to implement additional dampened shunt capacitance, as shown in Fig. 8.27. Each HDC element is dampened by a 100 Ω series resistance, which can be easily modified by a mask update to be 50 Ω or 150 Ω in later fabrication iterations if deemed necessary. All lines and HDCs can provide an accumulated shunt
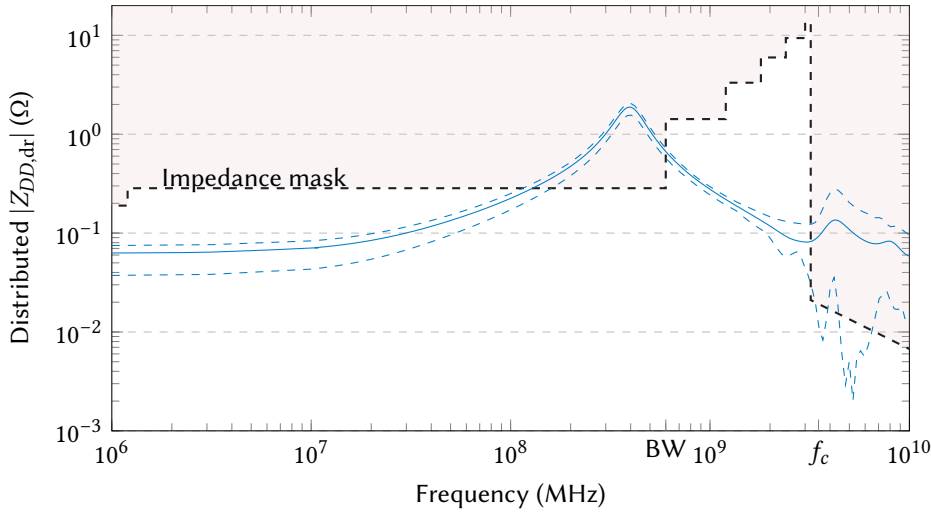
Figure 8.28: The simulated distributed supply impedance per bank, assuming all drivers are active and an AC short at the PCB reference plane. Since the distributed impedance simulation contains 640 ports, all the impedances seen over the bank are averaged to estimate the effective impedance. Also the maximum and minimum impedances found are shown dashed.

capacitance of 2.33 nF, so 0.58 nF per bank, which is still not enough to meet our specifications. Still, it helps dampen the unavoidable resonance. The resulting distributed supply impedance is shown in Fig. 8.28.

To achieve more capacitance on the LDMOS die, pads at the chip corners are placed, which can host on-chip multi-layer ceramic capacitors (MLCCs). While unconventional, the CMOS die effectively gets soldered on the LDMOS die, so placing an SMD on the LDMOS die should be possible. Since this is a risk during assembly, the first demonstrators will not use this until its feasibility is proven and whether it interferes with the rest of the assembly process is checked.

8.5.4 PCB Design Considerations

So far, we have assumed the feed inductance to be connected to signal ground, but, in reality, it is connected to a PCB design. Here we can try to offer a very low impedance by placing discrete capacitor components. However, we can also attempt to absorb the feed inductance by placing a component that has the opposite impedance. Using 1.4 nF and 0.4 nF again as an example, they resonate at 212 MHz, for which the 0.4 nF inductance has an impedance of 533j mΩ. Placing a component on the PCB that provides an effective −533j mΩ then cancels the inductance present. Since this resonance is caused by the 1.4 nF on-chip, we know that placing a 1.4 nF capacitor on the PCB will provide the required impedance. This effectively makes a π-shaped impedance inverter at the resonance frequency (e.g., Section A.4.1, Fig. A.2c), which needs to be considered when deciding on component values placed further on the PCB.

We have more flexibility in choosing capacitor components and materials at the PCB level. Close by, we need to place small package SMDs in order to place them close to the

bond pad and to have a capacitor with a low ESL. Large capacitances in a small package typically use class II ceramics, of which the capacitance drops significantly with applied DC bias. When selecting actual component values, attention must be paid to the SRF and ESR. If the ESR is much more than 250 mOhm, the decoupling capacitor will hardly do anything to lower the peak impedance as seen by the bank. If the SRF is lower than the resonance frequency we try to mitigate, the incorrect impedance is provided at the reference plane. In other words, the effective component value should be used, including the effect of any series inductance of the component and the applied DC bias of 2.2 V.

8.6 Demonstrator Overview

The most important design aspects for the high-resolution DTX demonstrator have been covered in the sections above, which are the flip-chip assembly flow, the switch bank layout and its activation pattern, design and modeling of the CMOS driver and the modified LDMOS technology, and the DC supply requirements. There are more aspects to the design that do not require a full section of their own or are implemented by other designers. A broader overview of the demonstrators is provided next, highlighting the global functionality and a few details.

8.6.1 CMOS Overview

Figure 8.29 shows the micrograph of the realized CMOS controller. The design measures $5937.0 \times 2664.9\,\mu m^2$ and features over 4000 bump pads. It contains 4 banks, one of which is highlighted in green. Each bank consists of two sub-banks (see Sections 8.1 and 8.2), also one is highlighted. The digital IOs, including their ESD protection and supply-related ESD diodes and clamps, are on the right side. Note that the CMOS top side is mirrored when compared to a 'through' view after the flip-chip assembly, such that these IOs appear on the left side with respect to the LDMOS power die. To ensure a low resistance path to supply ESD diodes, the supply ESD diodes are repeated on the left side of the CMOS die, while the ESD clamps are repeated in every bank. In the middle, on top of the synthesized digital blocks with the SRAMs, a ground plane is placed in the topmost metal layer to function as a shield against any matching structures internal to the LDMOS die, most significantly, the hybrid coupler for the planned fully integrated PD-LMBA demonstrator.

A micrograph detail is shown in Fig. 8.30 with annotations of the functional blocks. The baseband data is stored in 56-bit SRAMs, and eight are placed in parallel to support sampling rates up to $3.5\,\mathrm{GSa\,s^{-1}}$. This data is then to be deserialized by an 8:1 multiplexer and fed into a DSP block running at 3.5 GHz for calculating the sum of the A and B activations for determining the activation pattern (see Section 8.3), as well as mapping the three multi-phase sign bits for the octants to two times two bits to be used by the phase mapper for the A and B clock phases. This phase-mapper is made glitch-free with respect to the wanted phase-modulated RF clock and the (phase static) data clock [95]. After the phase mapper, the clocks are routed through a binary clock tree to the bank, utilizing the in-bank clock lines described in Section 8.3.2. The amplitude activation data is routed to the corners of the sub-banks, then to the row and column encoders, where it is dynamically retimed depending on the wanted activation phase to avoid glitches [95].

The digital architecture of several parallel SRAMs is very similar to that of the CMOS controller of previous demonstrators, as described in Section 7.4.4. Each of the four banks has its own digital block with eight parallel SRAMs that are 56-bit wide and 1024 addresses
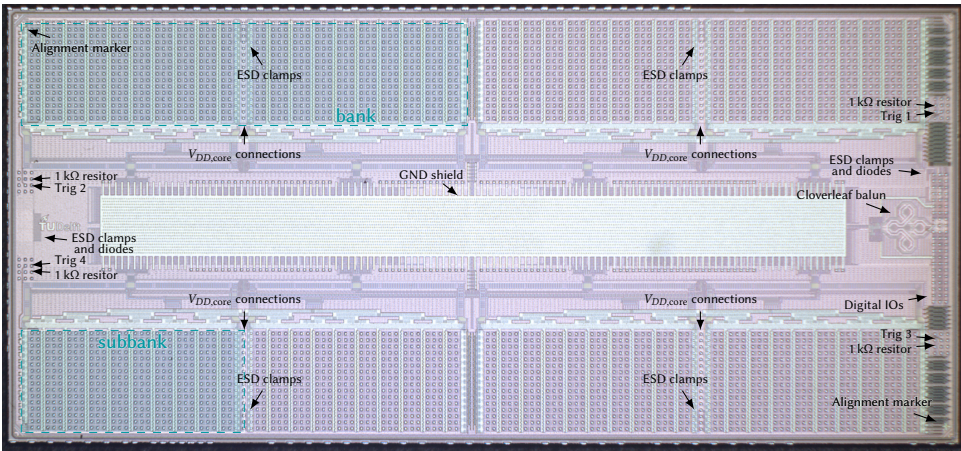
Figure 8.29: Full chip micrograph of the realized CMOS controller.
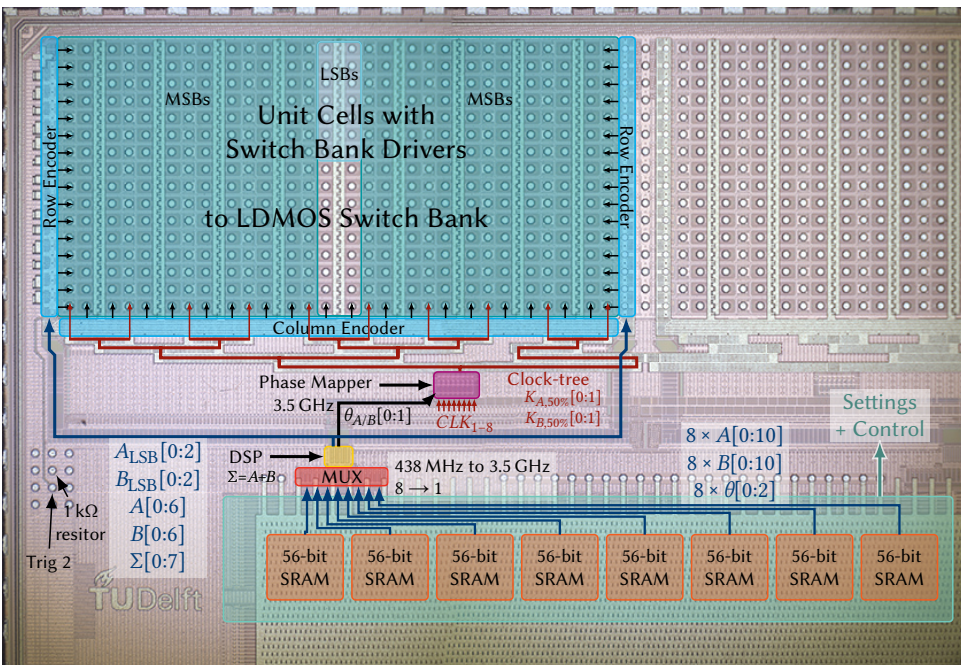
Figure 8.30: Micrograph of the top left chip corner with the functional block diagram of one sub-bank.
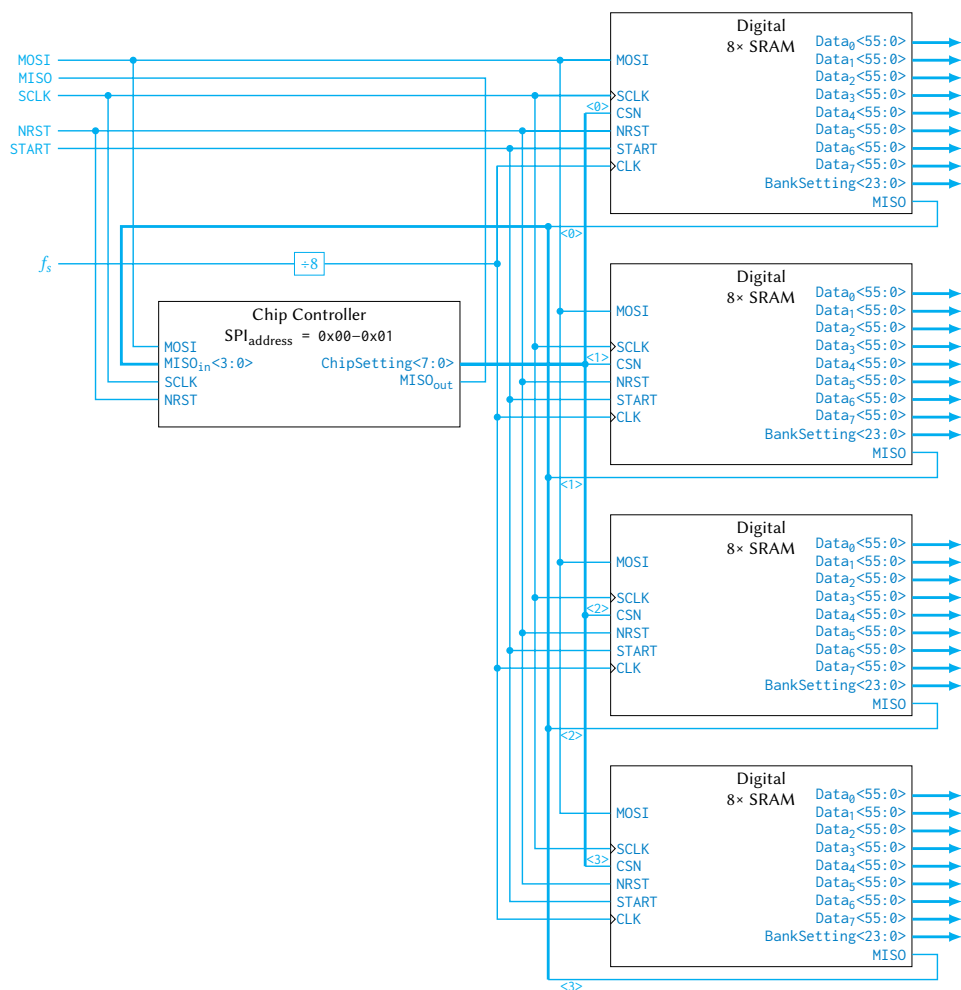
Figure 8.31: Simplified digital block diagram, showing the chip's external interfacing.

deep. Approximately one half is allocated to one sub-bank, and the other half to the other, which each has its own 8:1 multiplexer (3-bit Gray counter), DSP block, and phase mapper, and the outermost sub-bank also has a TRIG output for debugging purposes, routed to the 1 kΩ resistor bump pads used for assembly verification (see next section). A ring of ground pads then surrounds these two pads.

This CMOS controller has four of these digital blocks, each controlled through an SPI. It was already challenging enough to communicate over two individual SPIs in the demonstrators of Chapter 7, let alone synchronizing the two. Here, a single SPI for the entire chip was chosen, and a chip controller SPI block directing which digital block should be 'listening' to the external commands. The digital block diagram is shown in Fig. 8.31. This chip controller always listens to the addresses associated with it (0x00–0x01) while ignoring the rest. The chip controller's ChipSetting first register outputs are then routed
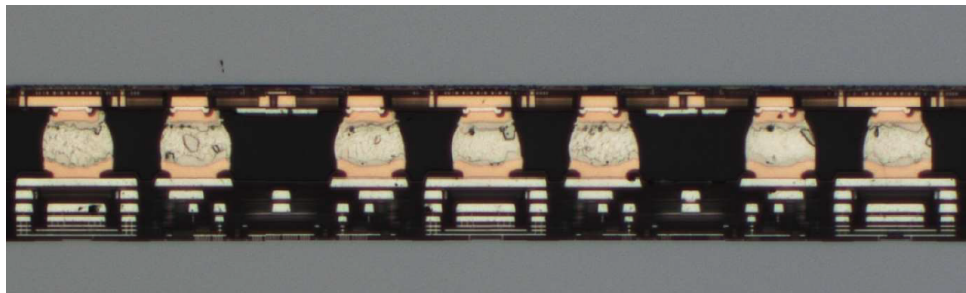
Figure 8.32: Cross section for for verifying the flip-chip assembly.

to a chip select[4] pin on each digital block, directing them to listen or ignore the external communication. The second register is used to multiplex the MISO signals from the digital blocks, as only one MISO signal can be routed to the SPI bus. In the digital blocks, as used in Chapter 7, the only way to reset the SRAM address counter was to entirely reset the digital block using the NRST signal, including its chip settings. To avoid this situation, an additional START signal has been added. The START signal only resets the SRAM address counters and the 8:1 mux clock dividers while leaving the BankSetting registers untouched. Unfortunately, in the implementation, the sampling clock divider for the digital blocks was, by accident, implemented as a divide by 4 rather than a divide by 8 block, while the mux does have the correct divide by 8 clock, causing every odd SRAM address to be skipped, effectively halving the memory depth. This also limits the maximum sampling clock, as the digital blocks now run twice as fast as initially intended.

8.6.2 Flip-Chip Assembly Verification

Special care was taken in the design of the CMOS to have a density that is as uniform as possible for the top two metal layers underneath the aluminum pad layer. Namely, the chip surface should be planar within the tolerance of the flip-chip bumps. Custom DRC rules were written to visualize the densities and density gradients of metals 6 and 7 and sequentially fix potential issues. A plus mark is placed in two opposite corners of the CMOS die (Fig. 8.29), which aligns with four dots on the LDMOS die for optical alignment in the flip-chip process. As checking the flip-chip assembly with X-ray verification may impact electrical parameters, four $1\,\text{k}\Omega$ resistor contacts are placed along the chip edges to allow for quick probe-based testing of whether the flip-chip process was successful. The first flip-chip trial run was performed after passing tests for the topography of the CMOS and LDMOS dies using a confocal microscope. This run was destructively verified using X-rays and by cutting the assembly in half to get a cross-section, which is shown in Fig. 8.32. Here, the successful connection between the two dies can be clearly observed.

Checking for shorts between ground and supply is another method to confirm correct assembly, since V_{SS} and $V_{DD,\text{dr}}$ connections are the most abundant on the chips. The capacitance between ground and the supplies is also measured at low frequency for two assemblies. First, a measurement before flip-chip assembly and then again with the CMOS die attached. The first measurement (LDMOS only) provided 2.337 nF for the $V_{DD,\text{dr}}$ domain

---

[4]Chip select (or slave select) is the terminology commonly used in context of SPI, while here a more appropriate naming would be 'bank select.'

and 0.63 nF for the $V_{DD,core}$ domain, for both samples. After flip-chip assembly, these values increased to 8.09 nF to 8.20 nF for the $V_{DD,dr}$ domain, and 15.52 nF to 16.40 nF for the $V_{DD,core}$ domain. The uncertainty is most likely caused by the other CMOS inputs left floating. It still provides valuable information to refine the DC decoupling structures further

8.6.3 LDMOS Variants and Demonstrators

For the LDMOS power die, a total of 11 layout variants have been designed. These LDMOS dies can be attached to a flange using Ag-sintering or Au-eutectic bonding, using different die thicknesses (see Section 8.2.2). Especially, designs with passive RF structures on-chip may have different performances depending on substrate thickness. Since the number of CMOS dies provided with flip-chip bumps is limited to 50, some of which are sacrificed for assembly tests, not all combinations are tested. Twelve different demonstrator variants can be defined, some of which have multiple variations due to slight variations in the used PCB layout and LDMOS substrate thicknesses.

Demonstrator I targets quick functionality verification, specifically verifying the high resolution and the DTX operation at 3.5 GHz. Demonstrator II has an embedded multi-section 90° hybrid coupler, which makes a fully-integrated PD-LMBA, enabling a very large RF bandwidth. Demonstrators III–VII target 3-way Doherty, with variants using low-pass or high-pass equivalents for the $\lambda/4$ impedance inverters, different driving profiles, single-ended or push-pull variations, and using different supply voltages for the different DTX branches. Demonstrators VIII–X and XII target more experimental applications, e.g., load-insensitive TXs, GaN drivers, and low-frequency over-sampled DTX with more control over the wave shape. An additional demonstrator XI was added after the measurements of demonstrator I, which contains some fixes to measure modulated signals better. The following section discusses only the measurements of demonstrators I and XI, as different designers designed the other demonstrators.

8.7 Measurements

The next sections discuss two different demonstrators and their measurement results. More demonstrators were assembled using the models discussed in Section 8.4, but they were designed by others and thus are not covered in this dissertation.

Demonstrator I is a general-purpose demonstrator, discussed in Section 8.7.1. It uses the two sub-banks furthest from the digital inputs (Fig. 8.1, Q2.2, and Q4.2), which are matched single-ended for 3.5 GHz. Its main purpose is to perform initial assembly testing and measurement debugging (e.g., writing code for controlling the different banks over SPI, downloading data to the SRAMs, and more) while the other demonstrators are still being assembled and give measurement results for initial publication.

During the demonstrator I measurement process, we discovered several design issues that complicated measuring (wideband) modulated signals. For that reason, another general-purpose demonstrator, designated demonstrator XI, was designed. This demonstrator uses a push-pull topology, now matched at 1.8 GHz to avoid the design issues. This demonstrator is discussed in Section 8.7.2.

8.7.1 Demonstrator I: Single-Ended 3.5 GHz Operation

This is the very first demonstrator of the design discussed in this chapter. It uses a general-purpose LDMOS design, unlike the specialized designs for (integrated) Doherty or PD-LMBA operation. It serves as a quick verification of operation, so its performance may not be optimized. Next, a brief overview of its design is discussed, followed by its measurements.

Design

A single sub-bank is used for this design. Its only target is to provide a sample for assembly tests and DTX functionality verification, so no specific efficiency (e.g., Doherty) or linearity (e.g., push-pull) enhancements are necessary. The LDMOS output is matched at the drain for digital class-C operation: resonate out the drain capacitance with shunt bond wires to the on-die RF short and out to the ohmic match by a $\lambda/4$ line on the PCB to the 50 Ω output. This matching structure is shown in Fig. 8.33a and, below it, its simulated performance for peak output power over frequency.

During assembly, the bond wire loop height was infeasible to implement, as the CMOS die on top of the LDMOS interfered with the path of the bonding tool. The two tall wires at the sides of the bond bar were replaced by a single lower wire with the same inductance value in the middle, as shown in Fig. 8.33b. Replacing two wires with a single wire results in higher DC and AC resistances, causing the simulated output power and efficiency to degrade, as shown in the graphs below the figure. The expected system efficiency degrades by 2.3 %pt.—more than the drain efficiency—caused by decreased RF output power over the controller's power consumption.

Another unexpected issue arose during the assembly. One of the ovens used during the die-attach for the silver sintering appeared to leak, causing oxygen to leak into the oven. This corroded the copper UBM on the LDMOS die. These bump pads needed plasma cleaning prior to flip-chip reflow. A different oven was used for subsequent demonstrators until the other one was repaired.

The PCB design surrounding the flip-chip assembly is shown in Fig. 8.34. The input clock lines and the digital SPI are shown on the left side. The primary clock input uses the LO_{high} line for the RF upconversion clock at $4f_0 = 14$ GHz. On the right side of the LDMOS die, routed to the top and bottom, are the $\lambda/4$ RF matching lines implemented as a CBCPW line. Further, decoupling capacitors of various sizes surround the die for the 1.1 V, 2.2 V, and 28 V supply domains, some using an optional series resistor. For example, the DC feed for the LDMOS drain uses a baseband decoupling capacitor of 47 nF with a series resistance of 2.2 Ω, which is sufficiently low for the impedance requirements of the power die while high enough to dampen the resonance peak. The impedance requirements for the CMOS decoupling are much more strict, but somewhat more relaxed than was presented in Section 8.5.3 due to using only one or two sub-banks. The PCB design was EM simulated for the CMOS supply feeds to choose appropriate capacitor and resistor values. Manufacturer models, including ESL and ESR, are used for the capacitors. The resulting impedance seen from the drivers is shown in Fig. 8.35. Here, the impedance and required impedance mask are calculated per sub-bank, assuming only sub-banks Q2.2 and Q4.2 are active. The impedance without the PCB, as shown in Fig. 8.28, is dashed for reference. It can be seen that, with the additional decoupling on the PCB, the main resonance peak has now shifted down in frequency and with a smaller magnitude. The impedance curve is somewhat more irregular due to the various component values (including package and

8
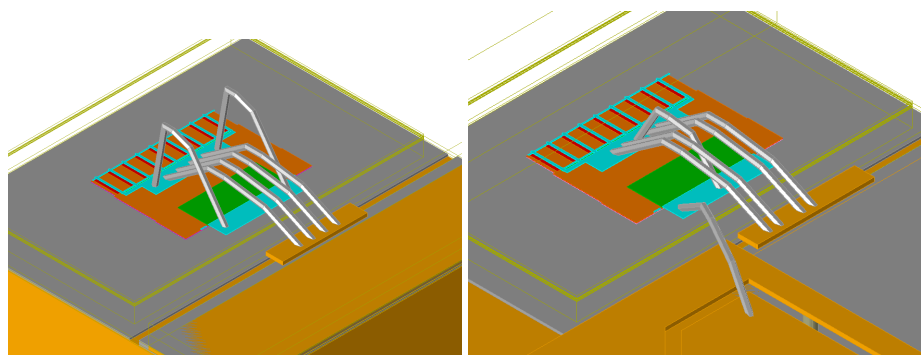
(a) Original design

(b) Implementation
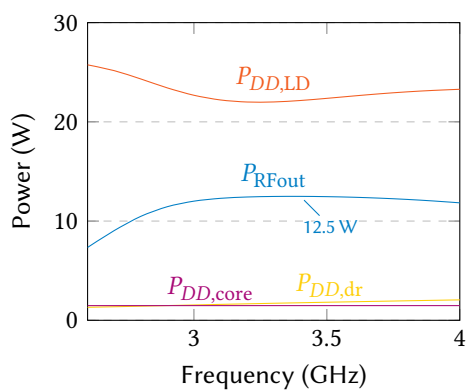
(c)
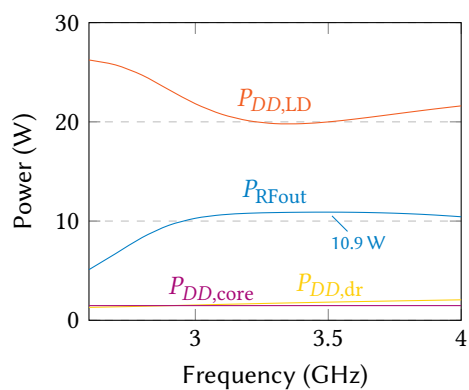
(d)
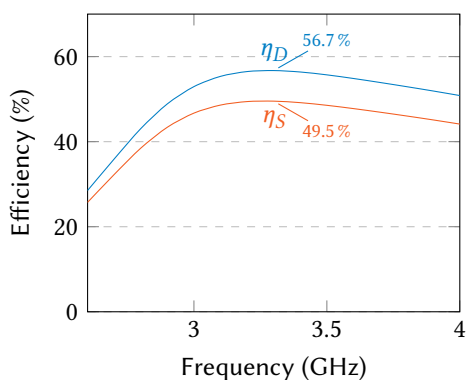
(e)
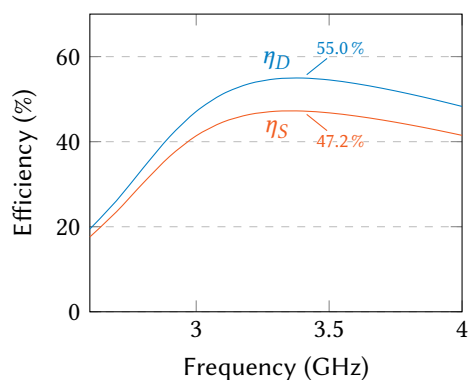
(f)

Figure 8.33: Expected performance (simulated) of demonstrator I. The initial bond wire design was not possible (a) to realize in assembly due to bond wire height, requiring a smaller single shunt wire in the implemented version (b) that has a lower Q, negatively impacting output power and efficiency.

(a) Top layer (M1)   (b) Inner layer (M3)
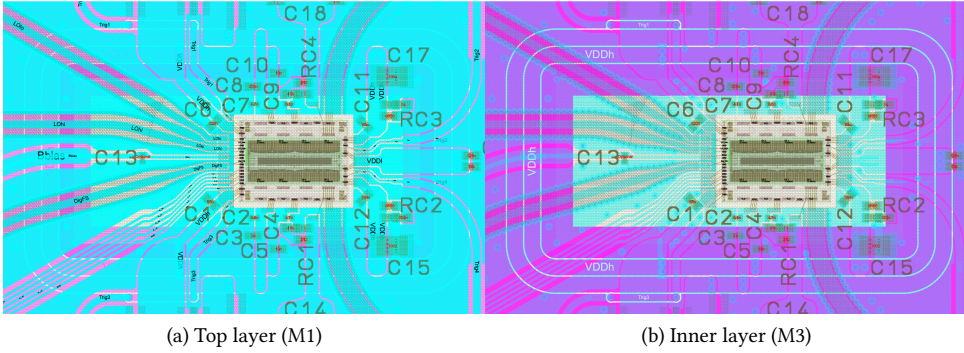
Figure 8.34: A close-in view of the PCB design used for demonstrator I, showing the first and third metal layers surrounding the die assembly. The inner structure reveals the ring structure used for DC routing of the 1.1 V and 2.2 V supplies, the top layer shows the decoupling structures present, as well as the various input and output lines. The second metal layer is a ground plane (not shown).

return path inductance) placed in parallel. Figure 8.36 shows a photo of the fully realized demonstrator.

Measurements

The DTX measurements are performed by uploading multi-phase test data in the on-chip memory and providing the external RF up-conversion clocks to the DTX. After debugging, an issue appeared with the stability of the high-frequency LO input due to a common-mode resonance at the center tap of the cloverleaf balun. This issue is caused by the inherently asymmetric nature of a balun, namely, at the input, one side is the single-ended signal while the other is ground, combined by the relatively low coupling factor of cloverleaf transformers. Further, the digital blocks ran twice the expected frequency, which the SRAMs did not support and corrupted the data (see Section 8.6.1). Instead the additional f_s (data sampling) input was used to lower the sampling rate to a supported frequency. Then, the DTX demonstrator was measured at the frequencies that did not cause problems with the high-frequency LO input.

First, pulsed CW test signals with a 12.5 % time duty cycle are uploaded to the SRAMs to avoid excessive self-heating of the DTX at high RF-output power conditions while mimicking a PAPR of 9 dB. The related measured RF-output power, drain, and system efficiencies are given in Fig. 8.37 for $V_{DD,RF}$ = 28 V. At 3.525 GHz, this DTX sample reaches 45 %/40 % peak-drain/system efficiency, with an output power of 10.5 W, providing a close to 1 W mm^{-1} power density. Figure 8.37a gives the measured DTX switch-bank efficiencies and (output) powers at 3.525 GHz vs. amplitude code word (ACW), while Fig. 8.37b gives the measured peak-efficiency and peak-RF-output power vs. frequency. In Fig. 8.37c, the power breakdown of $P_{DD,core}$ (the 1.1 V domain) is shown, which is a continuous 1.48 W. Almost one watt is continuous for the whole chip, irrespective of how many banks are used. Here, we only use one-eighth of the controller's capacity.

Second, to characterize the realized (effective) DTX resolution, a spectrum analyzer in VSA mode is used, and the DTX is programmed with a slow linear up-and-down ACW ramp. Due to the limited SRAMs' depth available, first, a coarse up-and-down sweep is

Figure 8.35: The simulated distributed supply impedance of demonstrator I per sub-bank, assuming only the drivers of sub-banks Q2.2 and Q4.2 are active. Here the impedance includes an EM simulated model of the PCB and the placed decoupling capacitors. Since the distributed impedance simulation contains 160 ports, all the impedances seen over the sub-bank are averaged to estimate the effective impedance. Also the maximum and minimum impedances found are shown dashed, and the initial simulated average impedance without the PCB (see Fig. 8.28) is shown dotted.

(a) Overview   (b) Detail

Figure 8.36: Photo of the realized demonstrator I.

Figure 8.37: Pulsed CW measurements with 12.5 % time duty-cycle pulses, showing RF output power, drain and system efficiencies (η_D resp. η_S) for (a) powers and efficiencies vs. ACW at 3.525 GHz; (b) peak output power (triangle) and efficiencies vs. frequency; (c) continuous $P_{DD,\text{core}}$ power breakdown.
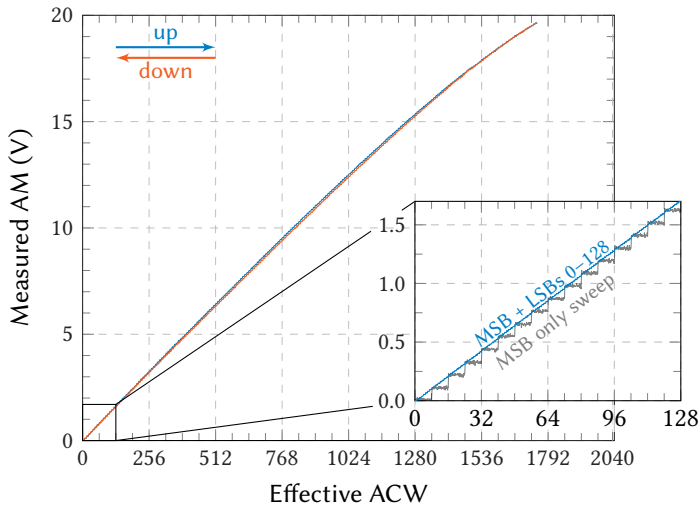
Figure 8.38: Measured ACW–AM curve of the high-resolution DTX. In the main graph only the MSB segments are used, while in the zoomed graph both the MSBs and the second layer of 7 thermometer coded LSBs are used.
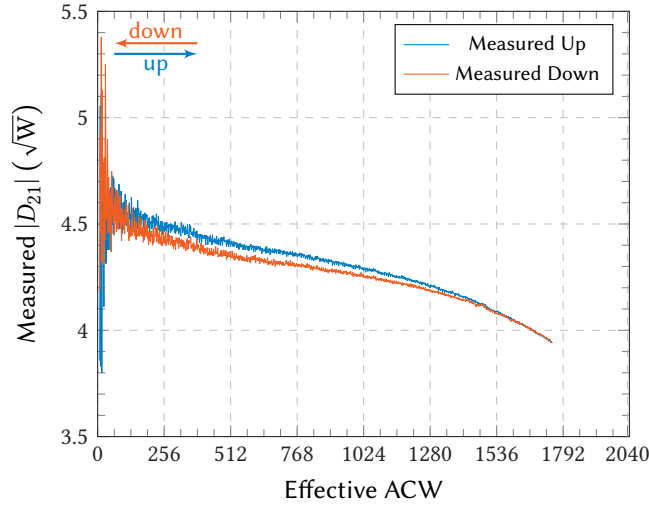
Figure 8.39: Measured dynamic transfer of the high-resolution DTX only using the MSB segments.

performed using only the MSB cells. It was found that 36 MSB cells were nonresponsive, most likely due to the abovementioned corrosion of the copper UBM. Note that this issue was not encountered in any of the next samples. The measured ACW–AM is shown in Fig. 8.38, where the nonresponsive MSBs were skipped, but no DPD has been applied. Minimal compression can be observed, as the response is mostly linear, which can be explained by the 36 MSB cells that cannot provide output current. This likely affected the achievable output power and drain efficiency as well. Next, both the MSB cells and LSB cells are swept over the 0–128 ACW range (see the zoomed graph of Fig. 8.38), showing the very fine ACW–AM control of the high-resolution DTX.

The measured MSB-only ramp can also be used to determine this demonstrator's full-scale transfer, which is shown in Fig. 8.39. The amplitude transfer shows little irregularity, especially compared to the proof-of-concept demonstrators, such as in Figs. 7.29 and 7.30. This also shows, by the lack of undershoots and overshoots, that the applied activation pattern (Section 8.3) is effective in preventing current redistribution effects. Also, the increased thermometer coding and the lowered coupling of the flip-chip gate connections are effective in preventing bit-to-bit interactions.

Conclusion

Unfortunately, some design imperfections sneaked in this first demonstrator, making it unsuitable for measuring modulated signals. However, it achieved its purpose as a path-finder for assembly and quick functionality verification. More attention should be paid to decoupling the core supply domain for newer designs. Ideally, all functional blocks with amplitude-dependent current consumption should have been isolated in their own supply domain so as not to influence the long chains for clock routing over the full CMOS die.

Still, the aimed high-resolution operation is verified and a smooth transfer is achieved, while low layout parasitics enable RF output powers above 10 W with drain/system efficiencies of 45 %/40 %, respectively. The proposed DTX configuration offers an excellent

(a) Overview

(b) Detail
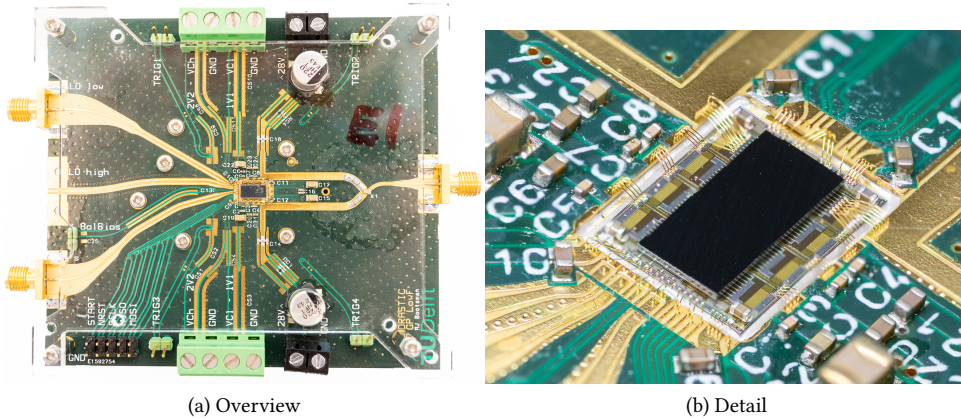
Figure 8.40: Photo of the realized demonstrator XI.

starting point for developing highly integrated, energy-efficient, wideband DTX solutions for mMIMO.

8.7.2 Demonstrator XI: Push-Pull 1.8 GHz Operation

This additional demonstrator addresses some of the issues found during the measurement of demonstrator I. Most notably, the high-frequency LO input is now avoided by targeting a 1.8 GHz carrier frequency at the low-frequency LO input. This CMOS input uses an active balun, avoiding the common-mode issues of the cloverleaf balun, while the lower frequency also allows the SRAMs to run at 450 MHz, such that the sampling rate is now equal to the RF carrier after the 4:1 (effectively) serializer. Next, a brief overview of its design is discussed, followed by its measurements.

Design

This design aims to demonstrate modulated signals with good linearity to showcase the improved resolution. From the discussion on the design targets in Section 8.1, we have selected push-pull operation in combination with 8-phase multi-phase upconversion to increase wideband linearity. For this design, we use two matched sub-banks connected to DC-blocking capacitors and a commercial of-the-shelf balun in the RF output to minimize design time and uncertainty. Unfortunately, commercial baluns are either compact ($1.6 \times 0.8 \text{ mm}^2$) but low power (< 0.5 W typical) and lossy (> 1 dB), or high power (> 100 W) with low loss (< 0.5 dB) but very bulky ($28 \times 30 \text{ mm}^2$) and typically low frequency (at best 2.5 GHz, commonly below 1 GHz), with no options in between. A small package was chosen supporting an average RF power of 2 W (thermal limitation, the breakdown is > 1 kV), with a typical insertion loss of 1 dB [102]. From demonstrator I, we know that the expected peak RF power of two sub-banks is around 20 W. The DTX is to be characterized with modulated signals with a PAPR > 8 dB in a lab setting (short measurements rather than continuous operation), so the average power should never exceed thermal limitations. An additional calibration board is manufactured to deembed the losses from the DC-blocking capacitors and the balun, not to include their effects in characterizing the DTX performance.
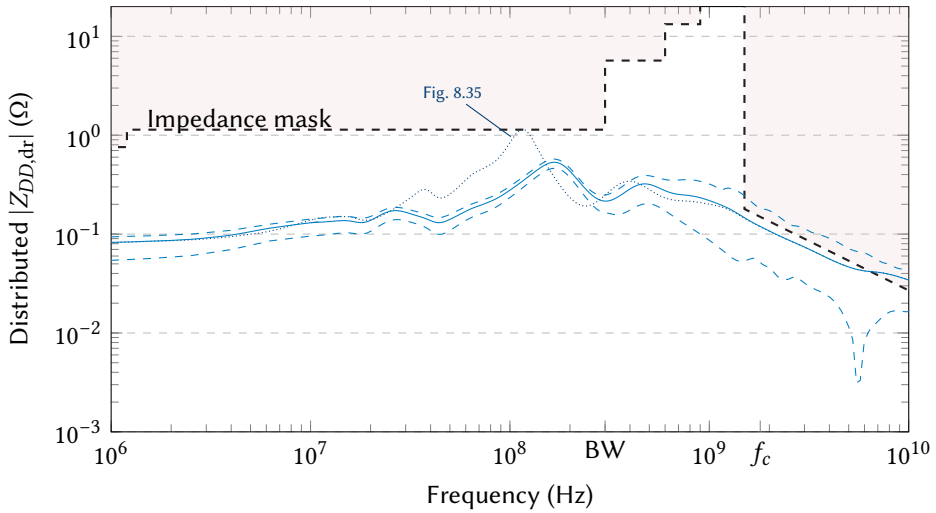
Figure 8.41: The simulated distributed supply impedance of demonstrator XI per sub-bank, assuming all drivers of sub-banks Q2.2 and Q4.2 are active, and with added on-chip SMD MLCCs. Here the impedance mask is adjusted for f_c = 1.8 GHz while keeping the fractional bandwidth the same. Since the distributed impedance simulation contains 160 ports, all the impedances seen over the sub-bank are averaged to estimate the effective impedance. Also the maximum and minimum impedances found are shown dashed, and average impedance of demonstrator I (see Fig. 8.35) is shown dotted.

The DC-decoupling of the core voltage domain on the PCB level has been improved by allowing large multi-layer ceramic capacitors (MLCCs) to be placed closer to the die, which are C19 and C22 in the photo of the demonstrator in Fig. 8.40a. Assembly tests have proved that additional on-chip SMD capacitors have no interference with the rest of the assembly process. The driver supply domain DC-decoupling can be improved using these additional SMDs, shown in the four corners of the LDMOS die in Fig. 8.40b. Figure 8.41 shows the resulting distributed supply impedance with these capacitors added, which now meets the set impedance mask requirement. Since the carrier frequency is halved, the impedance mask is relaxed further by a factor of 2. Even without this relaxation, this approach would have met the baseband impedance mask.

Measurements

First, the calibration board is characterized for deembedding purposes, of which the results are shown in Fig. 8.42. An insertion loss of 1.083 dB was measured at 1.8 GHz, which is largely caused by the balun. Next, the resulting power-DTX test configuration was characterized for maximum output power and drain/system efficiency directly at the RF-output connector (Fig. 8.43), yielding a simultaneous 16 W and 54 %/49 %, respectively. Using the measurement of the calibration board, the losses of the balun and the connecting traces can be deembedded, yielding a simultaneous 20 W and 68 %/63 % at the DTX output plane, respectively.

Using static calibration only, the power DTX was characterized with modulated signals with a bandwidth of 13 MHz (Fig. 8.44a) and 53 MHz (Fig. 8.44b), providing an ACLR better than −43.9 dBc and −37.0 dBc, respectively. EVM levels are 1.6 % and 3.0 % for 13 MHz

(a)   (b)
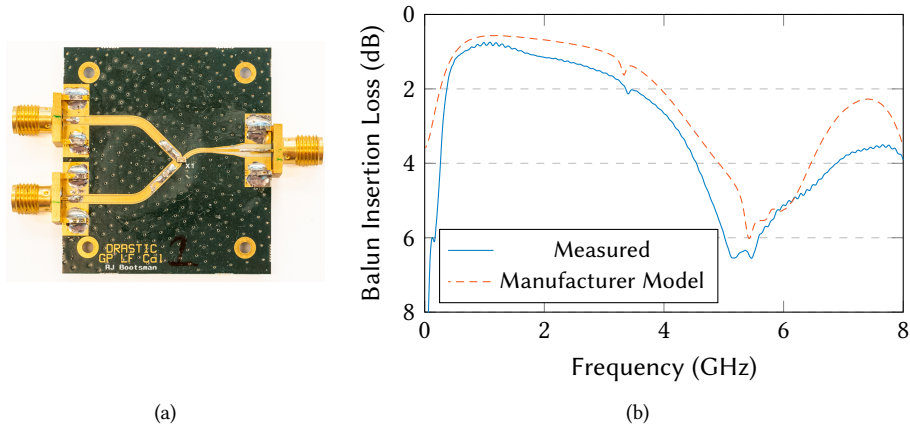
Figure 8.42: Measurements of the calibration board for deembedding the commercial of-the-shelf balun and DC-blocking capacitors from the DTX measurement. Shown dashed is the insertion loss from the *S*-parameter model provided by the manufacturer, which does not include the PCB lines or DC-blocking capacitors.

8

(a)   (b)
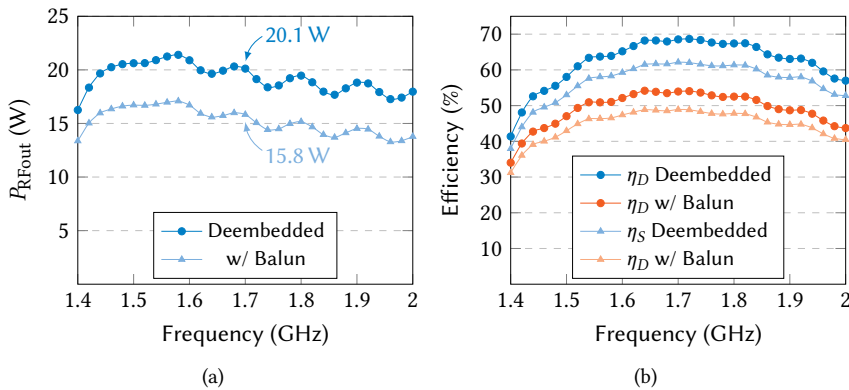
Figure 8.43: Measured output power and efficiencies of the DTX at peak RF output power vs. frequency, with and without deembedding of the balun.

(a)

(b)

(c)

(d)

(e)

(f)
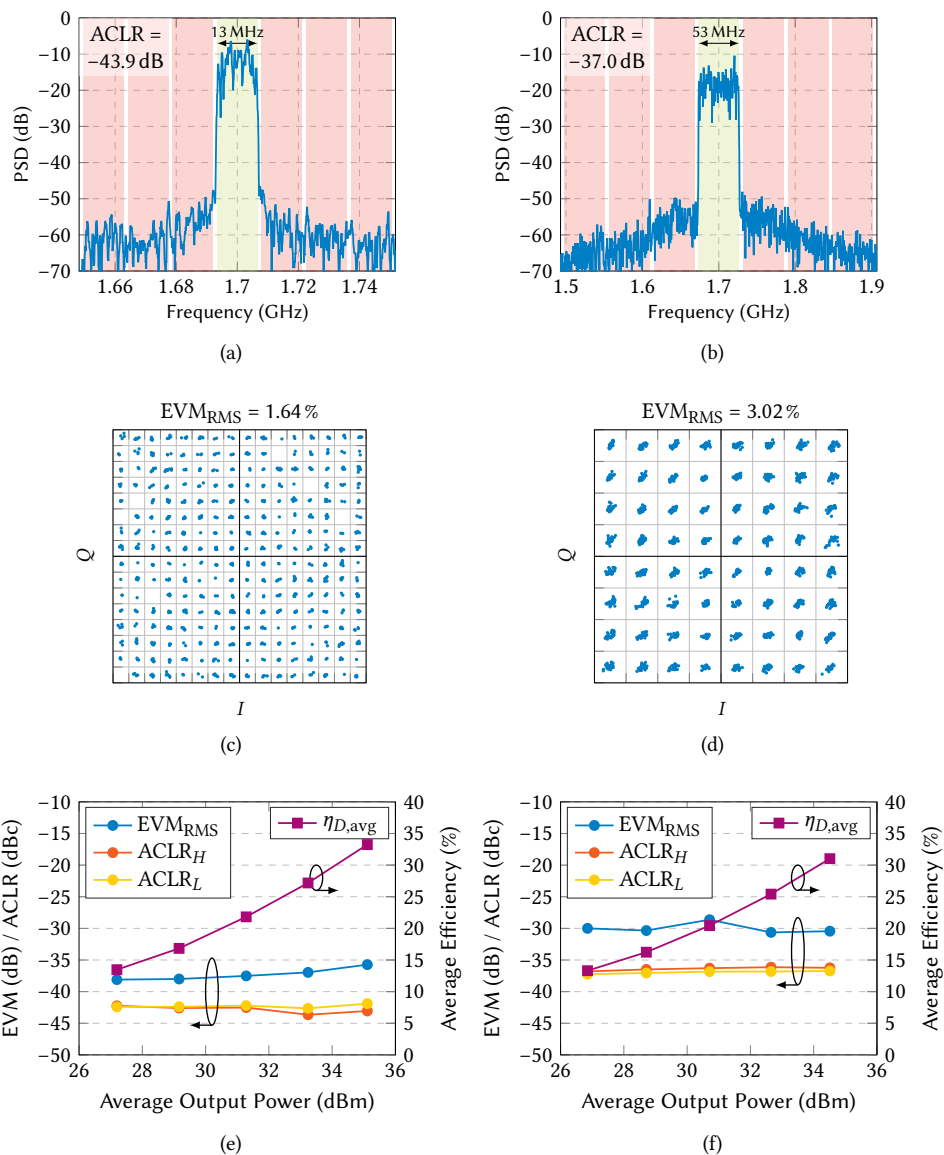
Figure 8.44: Measurements of modulated signals, using 13 MHz 256-QAM and 53 MHz 64-QAM, at maximum modulated output power, as well as in additional power back-off. The EVM and ACLR remain constant vs. power back-off, illustrating the realized resolution of the demonstrator.

256-QAM (Fig. 8.44c) and 53 MHz 64-QAM signals (Fig. 8.44d), respectively. The ACLR and EVM are limited mainly by imperfections of the DC-decoupling of the 1.1 V supply domain in this implementation, rather than by the resolution of the switch banks. This can be seen in Figs. 8.44e and 8.44f, which show constant ACLR and EVM levels vs. power back-off. No additional efficiency enhancement is applied, causing the drain efficiency to drop off proportional to the back-off level.

Conclusion

The proposed new approach for realizing a high-power, high-resolution DTX uses a CMOS controller that is interconnected using a high-density flip-chip to a low-V_T LDMOS MMIC containing gate-segmented switch bank with low layout losses. High drain and system efficiency is achieved by adopting 8-phase multi-phase upconversion with 25 % duty cycle current scaling digital class-C operation, reaching peak drain/system efficiencies of 68 % and 63 %. High effective resolution is confirmed using modulated measurements over varying average output power, showing constant ACLR and EVM levels.

8

9

Conclusion and Outlook

The feasibility of high-power digital transmitters (DTXs) using a combination of digital-low-power CMOS with high-power RF technology has been proven by several demonstrators, ranging from a bond-wire-based approach as a proof-of-concept to advanced high-density flip-chip packaging resulting in a high-resolution high-power DTX. Next, in Section 9.1, we provide this dissertation's key conclusions using the knowledge gained from developing the design techniques and implementation technology behind these power DTX demonstrators, with their subsequent measurements. Section 9.2 projects the gained knowledge to possible future implementations of high-power DTX, including promising technology trends that can be very beneficial for future DTX implementations, as well as other DTX development directions.

9.1 Dissertation Conclusions

Chapter 2 provided the background theory on (digital) transmitter architectures and power amplifier operating classes, including the theory on digital current scaling classes. The core of the theory could be found in the embryonic state in the literature [30], but its practical implementation is verified for the first time in the measurements in Section 7.6. The theoretical trade-off in drain efficiency and output power always favors the rectangular current waveforms as present in the digital current scaling classes over the analog transconductance classes, while it benefits from similar extremely simple output matching conditions: a resistive fundamental with all harmonic currents shorted, similar to the analog transconductance classes, such as class-B. Practical implementations have non-zero rise and fall times, which may limit the DTX output power and efficiency. Still, no quiescent or bias currents are required for the digital case, which is a significant benefit at lower output amplitudes over analog implementations.

Chapters 3 through 5 address several practical background aspects of high-power DTXs, which resulted from the accumulated experience of designing several demonstrators. Namely, Chapter 3 highlighted the high-level technology and packaging aspects that arise from the heterogeneous integration required between (relatively) low-voltage CMOS and high-power RF technologies for implementing high-power DTXs. The threshold voltage of the power devices should be made compatible with the supply voltage available in digital CMOS technologies, while assembly methods that enable a fine gate-level segmentation

of the power device are preferred. Chapter 4 discussed the design aspects for high-speed digital drivers, primarily the power–speed trade-off in scaling the driver and its preceding tapered buffer chain(s). The highest driver speeds relative to their power consumption can be achieved using the most advanced devices available in that technology, which tend to become faster with each generation. Chapter 5 proposes a mathematical definition of a DTX's transfer, which is primarily a direct relation from numerical value(s) in baseband to an RF amplitude and phase. Power gain—as used in analog power amplifiers—is not a useful linearity metric in DTXs due to the numerical (digital) input quantity and a transmitter system's modulating/upconverting action.

A power model capable of estimating DTX performance in terms of power and efficiency is proposed in Chapter 6. This power model explicitly combines the theory of digital current-scaling classes (Section 2.2.2), the electrical compatibility of the two technologies (CMOS V_{DD} and the V_T of the power technology) discussed in Chapter 3, and the formulation of the driver's power consumption as a function of its speed and capacitive load from Chapter 4. The combined theories result in a handful of equations that describe the power relations in a DTX by first-order approximation, which are useful for hand calculations and can help conceptual understanding of the underlying relations. These relations can be used to optimize DTX designs, from both the digital CMOS and power technology perspectives.

The subsequent chapters discuss the design and measurements of five different demonstrators. Three are used as a proof-of-concept, as discussed in Chapter 7. The first demonstrator uses a single line-up polar class-BE matching at 2.1 GHz, measuring 18.5 W peak RF output power with a drain efficiency of 67 % and a system efficiency of 60 %, proving that high-power DTX is indeed feasible. Modulated signals result in an ACLR of −46.1 dBc and an EVM of 1.2 % using 10 MHz 256-QAM. The next demonstrator introduces digital class-C operation and is designed at 1.0 GHz with reduced RF duty cycle, measuring 25.9 W peak RF output power with a drain efficiency of 76 % and a system efficiency of 73 %. Modulated signals result in an ACLR of −48.3 dBc with an EVM of 1.0 %. The third demonstrator applies the digital class-C matching in a Doherty configuration for enhanced modulated (average) efficiency targeting 2.0 GHz operation. It measures a peak RF output power of 39.1 W with a drain efficiency of 57 %, while only consuming 0.19 W in standby. More importantly, using a 7 MHz 256-QAM modulated signal with a PAPR of 5.5 dB, the average drain and system efficiency were measured at 49 % and 46 %, respectively. It also achieved an ACLR of −53.0 dBc and EVM of 0.3 % at that same bandwidth. At larger modulation bandwidths, these demonstrators' performance degraded due to resonances in the driver's supply path. Still, at, for example, 26 MHz modulation bandwidth, the Doherty demonstrator measures −40 dBc ACLR.

The final two measured demonstrators use a high-density flip-chip assembly, which targets higher resolutions by incorporating more thermometer coding. Eight individually controllable LDMOS switch-banks can enable a range of operating conditions, such as push-pull operation, 3-way Doherty, or PD-LMBA, in which several designers were involved. Chapter 8 describes the design of the switch bank and its drivers, and the modelling efforts of these switch banks that enabled these designers, while pioneering the flip-chip flow for combined digital and RF power purposes. The demonstrators of the other designers are not discussed in this dissertation; instead, the measurements of two (simple) demonstrators that do not include additional efficiency enhancements are discussed. The first high-resolution DTX demonstrator is used as an assembly path-finder and was designed for 3.5 GHz single-

ended 8-phase multi-phase operation with digital class-C matching using an RF duty cycle of 25 %. For a single (sub)bank, it measured a peak RF output power of 10.5W with a drain efficiency of 45 % and a system efficiency of 40 %. A dynamic ramp was used to show the smooth transfer, proving the high resolution of the design. The second demonstrator targeted modulated signals since the assembly flow has matured. It is similarly designed for 8-phase multi-phase operation with digital class-C matching using an RF duty cycle of 25 %, but now at 1.8 GHz in a push-pull operation using a commercial off-the-shelf balun. With the balun deembedded from the measurements, the design measured 20 W of peak RF output power with a drain efficiency of 68 % and a system efficiency of 63 %. Modulated performance resulted in an ACLR of −43.9 dBc and an EVM of 1.6 % using a 13 MHz 256-QAM signal. Using a larger bandwidth, 53 MHz 64-QAM, provided an ACLR of −37.0 dBc and an EVM of 3.0 %. These ACLR and EVM levels are retained over a more than 10 dB power back-off range, illustrating the resolution of this demonstrator.

In summary, the theory provided in this dissertation shows how digital-oriented low-power CMOS can be combined with high-power RF technologies by showing how they can be made electrically and physically compatible. The introduced power model for DTXs explicitly defines the achievable power efficiency. Measurements on several demonstrators have verified the foregoing theories, proving the feasibility of our proposed high-power fully digital transmitter concept. The digital concept allows low standby power by eliminating quiescent and bias currents, high integration, frequency agile operation, and flexible output matching conditions. In the author's view, no fundamental limitations stand in the way of implementing a power-efficient transmitter system for the next generation of sub-7 GHz mMIMO base stations using DTXs, enabling significant energy savings over traditional analog-based transmitter systems.

9.2 Outlook on the Future of High-Power DTX

The feasibility of high-resolution efficient power-DTXs has been demonstrated, but that does not mean that nothing is left to do. Still, some engineering challenges remain, and continuous technological developments may positively impact future DTX designs.

9.2.1 Technology Trends

Technology scaling of digital CMOS nodes is still happening. The most advanced node in production is 2 nm using gate-all-around (GAA) at the time of writing, and the IRDS roadmap projects an "1.4 nm" node for 2027, continuing down to an "1.8 Å" equivalent node using stacked N and P devices (CFET-3D) by 2039 [103]. The main motivation is to keep on minimizing switch energy and increasing logic density. For DTX, this means more digital functionality can be cointegrated, expanding the range of 'digital tricks' that can be performed to optimize DTX linearity and energy efficiency, while the driver's speed–power trade-off (Chapter 4) becomes better with the lowered switch energy consumption. The nominal supply voltage of these nodes is a metric to be wary of; as the supply voltage decreases, close attention should be paid to the electrical compatibility with the RF power devices. An interesting example that could even enable monolithic medium power DTX is by using fully integrated LDMOS devices in SOI, which can have a breakdown voltage $BV_{DSS} > 10$ V with an $f_t = 77$ GHz [104].

Not only is the CMOS device technology improving, but more advanced packaging solutions are also being pursued. More often, system-in-package (SiP) solutions are used

Normalized g_m (mS mm^{-1})
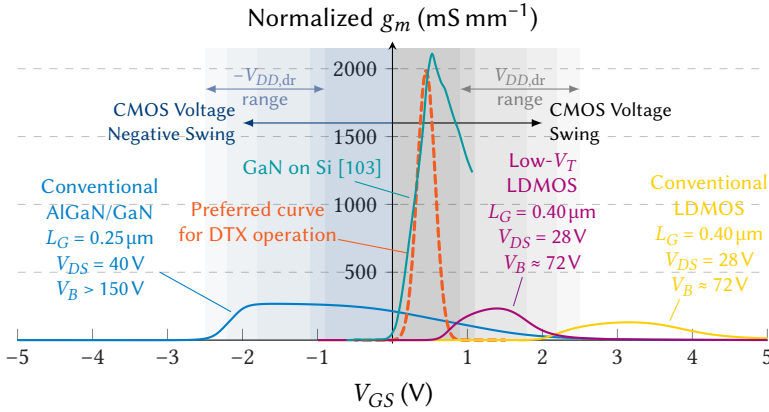
Figure 9.1: Example V_{GS}–g_m curves for some RF power technologies, with the shaded areas for possible $\pm V_{DD,\mathrm{dr}}$ ranges. Preferably the $g_m = 0$ for $V_{GS} = 0\,\mathrm{V}$ and $V_{GS} = V_{DD,\mathrm{dr}}$ and peaks in between to ensure complete 'OFF' and 'ON' switching. The E-mode GaN MOSHEMT from [106] has a characteristic that comes close to the preferred curve for DTX operation.

that can provide heterogeneous integration, compared to system-on-chip (SoC) solutions. This is driven by advanced packaging solutions, such as using advanced (silicon) interposers in, e.g., chip-on-wafer-on-substrate (CoWoS) packaging. The IRDS roadmap targets a 3 μm solder bump pitch by 2039 [103]. DTX can benefit from these developments to enable further integration and finer segmentation of the power device for increased resolution.

Another potentially useful development in CMOS technology is the use of backside power delivery networks (PDN) and buried power rails (BPR), which enable lower resistance supply routing [105]. Integrating high-density passives (e.g., deep trench capacitors) on-chip or on the interposer can enable more effective passive supply decoupling, as well as the use of advanced PCB technology with buried SMD components.

9.2.2 Segmenting III-V Semiconductors

When discussing the electrical compatibility of RF power technologies for their use in DTX in Chapter 3, a preferred curve for such technology was given. Namely, a sharp rise in device g_m at a low V_T is preferred and, after a high peak value, it should drop again quickly when reaching the device's $I_{\max}$. Figure 9.1 shows this preferred curve again. Out of context, it may seem a crazy requirement; however, experiments using GaN devices have shown that such a curve is actually possible [106]. This publication realizes an enhancement-mode (E-mode) GaN MOSHEMT on silicon, allowing it to be integrated with other silicon-based devices (e.g., CMOS FinFETs). This would greatly interest future DTX development if such a process becomes commercially available.

Alternatively, we could look toward GaN-based RF power technologies, typically GaN-on-SiC. Using the blue curve in Fig. 9.1 as an example, this technology has a negative V_T that is on the edge of a negative CMOS supply voltage. The electrical compatibility of CMOS with GaN can be improved if similar V_T engineering of GaN devices is performed as was done with the custom LDMOS technology (Chapter 8), or possibly modify the devices to be E-mode similar to [106].
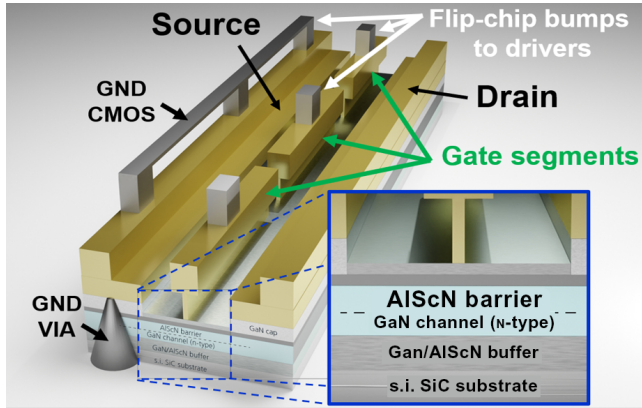
Figure 9.2: Artist's impression of a segmented GaN technology (courtesy of Fraunhofer IAF).

The benefits of using GaN as a semiconductor material are its higher bandgap and electron mobility, which increase its breakdown voltage and achievable current density, both contributing to a higher power density. GaN typically has lower intrinsic gate and drain-related capacitances for the same RF output power and less substrate losses. This can increase the achievable gain for a given operating frequency and bandwidth in analog applications. For DTX, it would mean that the required drive power lowers and the output bandwidth can be increased.

In the analog domain, there are many challenges related to using GaN, like stability issues related to the high device gain and the long-term memory effects caused by charge trapping. Since the digital drivers used in DTX provide a very low source impedance, stability is no longer a concern, while the consequences of gate-side trapping effects can be mitigated if the driver voltage range is sufficiently large to switch the GaN devices completely 'ON' and 'OFF', irrespective of trapping-related V_T changes. Drain-side trapping may still be present, but it would have a lower impact than in analog implementations. More challenging would be to define a GaN technology that can be finely segmented and is compatible with a CMOS BEOL stack. Namely, typical GaN technologies use up to two gold interconnect layers with airbridges. To be suited for (high-density) flip-chip assembly, copper or aluminum interconnect with little topography needs to be present. Figure 9.2 shows an artist's impression of how such a GaN technology stack would look.

9.2.3 Future Work

One topic for future work is further optimizing the heterogeneous integration of CMOS with RF power technologies through more advanced CMOS nodes and better RF device optimization, which may include using different semiconductor technologies as discussed for GaN above. This also involves using advanced packaging, such as die-to-die high-density flip-chip (Chapter 8). Alternatively, dedicated DTX MMICs could be developed if suitable RF power devices can be monolithically integrated, such as the LDMOS in SOI from [104] or GaN-on-Si from [106].

Using more advanced CMOS nodes enables more digital processing and higher driver speeds, which can implement more functionality. An interesting option could be to imple-

ment an oversampled DTX, which would allow for explicitly defining the DTX transfer for other frequencies, i.e., specifying a wanted $D_{2[k]1[0]}$ for k more than just the fundamental. For example, the drain current waveform can be shaped by defining a non-zero transfer to the (odd) harmonics, which lowers the effective RF duty cycle while band-limiting the spectral content around these harmonics and avoiding intermodulation [107]. Another function could be to integrate an observation receiver with the DTX controller, which can reuse the RF clock phases available in the DTX controller [108, 109]. Doing so can fully integrate a DPD loop, calibrating itself online. Energy-efficient DPD algorithms have been introduced recently that use AI techniques [110–112]. These have shown better linearization performance at a lower power consumption than conventional generalized memory polynomial (GMP) based approaches, and are a prime candidate for integration with a DTX controller.

In our experiments, the CMOS controller's supply voltage stability has proved critical to the DTX's performance. The passive decoupling can be improved further through the technology trends discussed above. Specifically interesting would be the integration of deep trench capacitors with backside PDNs. Alternatively, we can look towards active measures to control the supply voltage by using DC regulators. A topic of further reseach could be the modeling of a DTX using D-parameters that explicitly includes the driver supply as additional port and its (small-signal) influence on the DTX transfer (similar to X-parameters). This may eventually be beneficial for system integration using DTXs.

Although the primary targeted application of the power DTX developments in this dissertation is wireless communication, with the primary objective of realizing enormous energy savings, DTX also offers higher integration and a close-to-frequency agile system functionality. These are features crucial to other RF application fields, like MRI or radar systems, thus giving rise to new, inspiring ideas and implementations in these application domains. As such, although this is the last chapter of this dissertation, in the author's opinion, we are only at the very beginning of a new chapter in RF and wireless design.

9

A

Definitions and Derivations

In literature sometimes conflicting definitions are used. For completeness and to avoid any potential ambiguity, several definitions common across chapters of this dissertations are given here. Also some full derivations omitted in the main text are provided here for completeness.

A.1 Mathematics

The Fourier transform is defined as

$$F(\omega) = \mathscr{F}_t\{f(t)\}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\, \mathrm{d}t. \tag{A.1}$$

The two argument inverse tangent provides the angle in four quadrants, namely its range is $(-\pi, \pi]$, whereas the normal inverse tangent is only defined for the range $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. The two argument inverse tangent provides unambiguous angles for the full Cartesian plane and can be given by

$$\mathrm{atan2}\,(y, x) = \mathrm{Arg}\,(x + jy) = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right) & \text{if } x > 0, \\ -\tan^{-1}\left(\frac{y}{x}\right) + \frac{\pi}{2} & \text{if } y > 0, \\ -\tan^{-1}\left(\frac{y}{x}\right) - \frac{\pi}{2} & \text{if } y < 0, \\ \tan^{-1}\left(\frac{y}{x}\right) \pm \pi & \text{if } x < 0. \end{cases} \tag{A.2}$$

The sinc function is here defined normalized by

$$\mathrm{sinc}\,(x) = \begin{cases} \dfrac{\sin(\pi x)}{\pi x} & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases} \tag{A.3}$$

Shorthand notations of logarithms are defined here for base e and 10, namely $\ln x \equiv \log_e x$ and $\log x \equiv \log_{10} x$.

The generalized mean of an array of numbers is

$$M_p\,(x_1, ..., x_n) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^p\right)^{1/p}, \tag{A.4}$$

A

with a names defined for $p = -1$ being called the harmonic mean, $p = 1$ is the "normal" geometric mean and $p = 2$ the root-mean-square (RMS). A special case is for $p = 0$, which is called the geometric mean. In that case

$$\lim_{p \to 0} M_p(x_1, ..., x_n) = M_0(x_1, ..., x_n) = \left(\prod_{i=1}^{n} x_i\right)^{1/n} = \exp\left(\frac{1}{n} \sum_{i=1}^{n} \ln x_i\right). \tag{A.5}$$

The parallel operator is then $1/n^{\text{th}}$ of the harmonic mean. For two values the parallel operator is defined as

$$a \parallel b = \left(a^{-1} + b^{-1}\right)^{-1}. \tag{A.6}$$

A.2 Power Amplifiers

A.2.1 Power, Gain, and Efficiencies

The drain efficiency η_D of a PA can be given by its output power divided by the DC power consumption at the drain, whereas the total efficiency η_T also includes the input power of a PA

$$\eta_D = \frac{P_{\text{RFout}}}{P_{\text{DC}}} = 1 - \frac{P_{\text{diss}}}{P_{\text{DC}}} \qquad \eta_T = \frac{P_{\text{RFout}}}{P_{\text{in}} + P_{\text{DC}}}. \tag{A.7}$$

The normalized efficiency $\hat{\eta}$ is a scaling step to map the efficiency from some interval $[0, \eta_{\text{pk}}]$ to $[0\,\%, 100\,\%]$. This is, for example, useful for making comparisons, or expressing the stand-alone efficiency modification of an efficiency enhancement technique or upconversion technique, without considering the underlying efficiency of the used operating class

$$\hat{\eta}(\cdot) = \frac{\eta(\cdot)}{\eta_{\text{pk}}}. \tag{A.8}$$

The gain of a PA is relating input and output powers, where 2 common power gains can be defined: ratio gain G_P and slope gain G_{sP}

$$G_P = \frac{P_{\text{RFout}}}{P_{\text{in}}} = \left.|S_{21}|^2\right|_{\Gamma_L=0,\ \Gamma_{\text{in}}=0} \qquad G_{sP} = \frac{\partial P_{\text{RFout}}}{\partial P_{\text{in}}}. \tag{A.9}$$

Sometimes the power-added efficiency (PAE) is used, which can be linked to the gain of a device

$$PAE = \eta_D \frac{G_P - 1}{G_P} \iff \eta_D = PAE \frac{G_P}{G_P - 1}. \tag{A.10}$$

Also the total efficiency can be related to gain and PAE

$$\eta_T = \eta_D \parallel G_P \iff \eta_D = \eta_T \parallel -G_P, \tag{A.11}$$

$$\eta_T = \frac{PAE \cdot G_P}{G_P - 1 + PAE} \iff PAE = \frac{G_P - 1}{G_P \eta_T^{-1} - 1}. \tag{A.12}$$

The system efficiency η_S includes all DC powers consumed in the system that are required to generate the wanted output signal. For example, for a transmitter system that may consist out of a PA and one or more (pre-)driver(s), the (DC) power consumed by the driver(s) also needs to be considered

$$\eta_S = \frac{P_{\text{RFout}}}{\sum_k P_{\text{DC},k}}. \tag{A.13}$$

A.2.2 Equations for the Analog Transconductance Classes

The conduction angle α is a parameter that unifies the operation of all analog transconductance classes A, AB, B, and C, which is set by the voltage bias V_{bias} at the input of the PA. Normalizing all relevant voltages and currents here gives for the transistor threshold voltage $V_T = 0\,\text{V}$, the maximum drain current $I_{DS,\text{max}} = 1\,\text{A}$, and the drain supply voltage is $V_{DD} = 1\,\text{V}$. The input sinusoid can then be given by

$$V_{GS}(\theta, \alpha) = \left(1 - V_{\text{bias}}(\alpha)\right)\cos(\theta) + V_{\text{bias}}(\alpha), \tag{A.14}$$

where

$$V_{\text{bias}}(\alpha) = \frac{\cos\left(\frac{\alpha}{2}\right)}{\cos\left(\frac{\alpha}{2}\right) - 1} \tag{A.15}$$

such that its instantaneous peak voltage is always exactly 1 V. The resulting current waveform then is

$$I_{DS}(\theta, \alpha) = \begin{cases} \left(1 - V_{\text{bias}}(\alpha)\right)\cos(\theta) + V_{\text{bias}}(\alpha), & |\theta - 2n\pi| \le \alpha/2 \\ 0, & \text{elsewhere.} \end{cases} \tag{A.16}$$

Fourier decomposition for DC ($k = 0$) gives

$$I_{DS}(\alpha)[0] = \frac{2}{2\pi}\int_0^{\frac{\alpha}{2}} I_{DS}(\theta, \alpha)\,\mathrm{d}\theta = \pi^{-1}\left[1 - V_{\text{bias}}(\alpha)\right]\sin\left(\frac{\alpha}{2}\right) + V_{\text{bias}}(\alpha)\frac{\alpha}{2\pi}$$

$$= \frac{2\sin\left(\frac{\alpha}{2}\right) - \alpha\cos\left(\frac{\alpha}{2}\right)}{4\pi\sin^2\left(\frac{\alpha}{4}\right)}, \tag{A.17}$$

and for all harmonics k gives

$$I_{DS}(\alpha)[k] = \frac{4}{2\pi}\int_0^{\frac{\alpha}{2}} I_{DS}(\theta, \alpha)\cos(k\theta)\,\mathrm{d}\theta. \tag{A.18}$$

For $k = 1$ (the fundamental), it is

$$I_{DS}(\alpha)[1] = \frac{2}{\pi}\left[1 - V_{\text{bias}}(\alpha)\right]\cdot\frac{\frac{\alpha}{2} + \sin\left(\frac{\alpha}{2}\right)\cos\left(\frac{\alpha}{2}\right)}{2} + \frac{2}{\pi}V_{\text{bias}}(\alpha)\sin\left(\frac{\alpha}{2}\right)$$

$$= \frac{\alpha - \sin(\alpha)}{4\pi\sin^2\left(\frac{\alpha}{4}\right)} \tag{A.19}$$

and for all harmonics $k \ge 2$

$$I_{DS}(\alpha)[k] = \frac{2}{\pi}\left[1 - V_{\text{bias}}(\alpha)\right]\cdot\frac{k\cos\left(\frac{\alpha}{2}\right)\sin\left(\frac{k\alpha}{2}\right) - \sin\left(\frac{\alpha}{2}\right)\cos\left(\frac{k\alpha}{2}\right)}{k^2 - 1} + \frac{2}{\pi}V_{\text{bias}}(\alpha)\cdot\frac{\sin\left(\frac{\alpha k}{2}\right)}{k}. \tag{A.20}$$

Perfect harmonic shorts are assumed, no power is generated or lost at the harmonics. That allows expressing the drain efficiency in terms of the DC and fundamental

$$\eta_D(\alpha) = \frac{I_{DS}(\alpha)[1]}{2I_{DS}(\alpha)[0]} = \frac{\alpha - \sin(\alpha)}{4\sin\left(\frac{\alpha}{2}\right) - 2\alpha\cos\left(\frac{\alpha}{2}\right)}, \tag{A.21}$$

as well as to formulate the (normalized) optimum load resistance

$$R_{L,\text{opt}}(\alpha) = \frac{1}{I_{DS}(\alpha)[1]} = \frac{4\pi \sin^2\left(\frac{\alpha}{4}\right)}{\alpha - \sin(\alpha)}. \tag{A.22}$$

A.2.3 Linearity

The ACLR (formerly also called adjacent channel power ratio (ACPR)) is defined as

$$\text{ACLR} = \frac{\text{adjacent channel power}}{\text{main channel power}} \tag{A.23}$$

commonly expressed in dBc.

The RMS EVM is the RMS magnitude of the error vector over the RMS magnitude of the targetted constellation points [113]

$$\text{EVM}_{\text{RMS}} = \sqrt{\frac{N^{-1}\sum_{i=1}^{N}\left|V_{\text{meas},i} - V_{\text{ideal},i}\right|^2}{N^{-1}\sum_{i=1}^{N}\left|V_{\text{ideal},i}\right|^2}}. \tag{A.24}$$

It is commonly represented as a percentage or in dB (using $\text{EVM}_{\text{dB}} = 20\log\text{EVM}_{\text{RMS}}$).

A.3 Further Derivations on D-Parameters for DTX

In Section 5.1.2 the basic definitions for the D-parameters are provided for the idealized case. In the nonideal case, however, the value of $D_{2[1];1[0]}$ will be a function of $da_{1[0]}$ and the clock will have harmonics. Also, a_2 might not equal zero. Consequently, a more generalized version of Eq. (5.14) using the harmonic superposition principle [114], which includes the above dependencies, is given as

$$B_{p[k]} = \sum_{q}\sum_{l} D_{p[k];q[l]}\left(da_{1[0]}\right)A_{q[l]}P^{k-l}\Big|_{a_3=P,\ \text{Arg}(a_2)=\text{Arg}(P)} \tag{A.25}$$

where A and B are matrices containing all incident and outgoing waves, respectively, for all harmonics considered. Furthermore the phase of a_2 can (strongly) influence the large-signal operation, which is still not captured in the this generalized equation. This dependency is included next when considering multiple phase reference clocks for a DTX, as is the case in Cartesian or multi-phase DTXs.

A.3.1 Higher-Order Multi-Port D-Parameters

Depending on the upconversion architecture of a DTX (see Section 2.1), more (phase) reference clocks might be present. For example, a polar DTX has only one continuously phase-modulated reference clock, in contrast a Cartesian or multi-phase DTX will have two or more phase reference clocks. To handle also these DTX concepts, additional data input ports and reference clock ports must be added. For example, as illustrated in Fig. A.1, port 1 and port 2 can be configured for the in-phase and quadrature digital baseband data, and port 4 and port 5 are used for the 90° phase-shifted reference clocks to support the
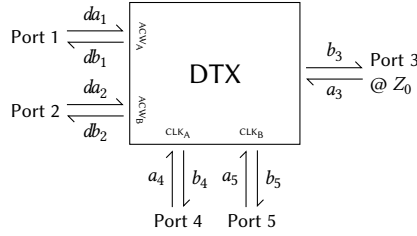
Figure A.1: Full 5-port DTX representation when using 2 phase references.

Cartesian operation of a DTX. Similar reasoning works for multi-phase operation. Without loss of generality, we can choose port 4 here to be the phase reference P. However, with two variables and two phases present, the influence of the phase of the incident wave a_3 on the large signal operation at the output port can no longer be covered by simply assuming it to be in phase with the reference phase. To describe this influence, the D-parameters have to be split into two variants, S-type and T-type, as is practice in polyharmonic distortion modeling [115]. An S-type D-parameter, $D^{(S)}$, only describes the contribution to a reflected wave dependent on the amplitude of any incident wave. This makes $D^{(S)}$ a mixed-signal equivalent to large-signal S-parameters, as it is dependent on the digital input signal's magnitude. Large signal S-parameters are also referred to as X-parameters, so in fact $D^{(S)}$ is the mixed signal equivalent of $X^{(S)}$. The T-type D-parameter also describes the contribution of the phase difference between the incident wave and the reference phase, as well as the incident wave's amplitude. This makes $D^{(T)}$ the mixed-signal equivalent of $X^{(T)}$. Also, $D^{(S)}$ and $D^{(T)}$ now become a function of both $da_{1[0]}$ and $da_{2[0]}$

$$B_{p[k]} = \sum_q \sum_l D^{(S)}_{p[k];q[l]}\left(da_{1[0]}, da_{2[0]}\right) P^{k-l} A_{q[l]} + \sum_q \sum_l D^{(T)}_{p[k];q[l]}\left(da_{1[0]}, da_{2[0]}\right) P^{k+l}\overline{A_{q[l]}},$$

(A.26)

where $\overline{A}$ denotes the complex conjugate of A.

Alternatively, a multi-port DTX could be described by two parallel DTXs with an LTI combining network. Rather than capturing their interaction explicitly in the D-parameters, it becomes implicit by the incident wave at each DTXs' output port.

A.3.2 Multi-Rate *D*-Parameters

Only steady-state signals have been considered thus far, where the digital inputs were only considered as a DC bias. It is also possible to describe a large-signal operation using a sinusoid at the digital input. As there are now two frequency regions present, a harmonic index (or mixing index) is used in the subscript to specify which (mixing) term results from which signal, which is common practice in the modeling of mixers [116]. Here we use $[k, l]$ as the additional subscript, where k describes the harmonic index of ports 4 and 5, and l the harmonic index of ports 1 and 2. For example, $a_{4[1,0]}$ then contains the fundamental of the used LO clock A, and $da_{1[0,1]}$ a tone of the ACW$_A$ in the baseband. Two phase references are now also required, one for the fundamental tone of the clocks and one for the baseband input. As such, port 4 and port 1 are chosen as references, without the loss of generality: $P_{[1,0]} = e^{\text{j}\cdot\text{Arg}(a_{4[1,0]})}$ and $P_{[0,1]} = e^{\text{j}\cdot\text{Arg}(da_{1[0,1]})}$. The wanted output of a DTX is the outgoing

A

wave $b_{3[1,0]}$. The full multi-rate black-box description of a DTX then becomes

$$B_{p[k,l]} = D^{(F)}_{p[k,l]}\left(\left|da_{1[0,1]}\right|, \left|da_{2[0,1]}\right|\right) P^k_{[1,0]} P^l_{[0,1]}$$

$$+ \sum\sum\sum_{(q,m,n)\neq\{(1,0,1)(2,0,1)\}} D^{(S)}_{p[k,l];q[m,n]}\left(\left|da_{1[0,1]}\right|, \left|da_{2[0,1]}\right|\right) P^{k-m}_{[1,0]} P^{l-n}_{[0,1]} A_{q[m,n]}$$

$$+ \sum\sum\sum_{(q,m,n)\neq\{(1,0,1)(2,0,1)\}} D^{(T)}_{p[k,l];q[m,n]}\left(\left|da_{1[0,1]}\right|, \left|da_{2[0,1]}\right|\right) P^{k+m}_{[1,0]} P^{l+n}_{[0,1]} \overline{A_{q[m,n]}}. \quad (A.27)$$

Here $D^{(F)}$ sets the large-signal operating point based on the magnitudes of $da_{1[0,1]}$ and $da_{2[0,1]}$, hence these shouldn't be considered a second time in the summation of $D^{(S)}$ and $D^{(T)}$. The F-type D-parameter is then the mixed-signal equivalent of $X^{(F)}$ [74, 116].

This description still has some shortcomings. For example, all incident waves other than $da_{1[0,1]}$ and $da_{2[0,1]}$ are considered to be small signal deviations, which might not be accurate. Additionally, there's a third frequency involved, namely the sampling frequency of the digital input. This goes beyond the scope of this dissertation; as such, the sampling rate is simply considered as a periodic spectrum shaped by a sinc function due to a zero-order-hold (ZOH), and can be captured in the higher harmonics and negative harmonics of $da_{1[0,n]}$, while assuming the sampling frequency and the fundamental frequency have an integer relation. Nonetheless, this harmonic description helps understand frequency domain simulations on a DTX; see Section 5.2.3 and Fig. 5.15 for an example.

A.4 Circuits
A.4.1 Impedance Inverters

(a) Using $\lambda/4$ transmission line   (b) Lumped low-pass equivalent   (c) Lumped high-pass equivalent
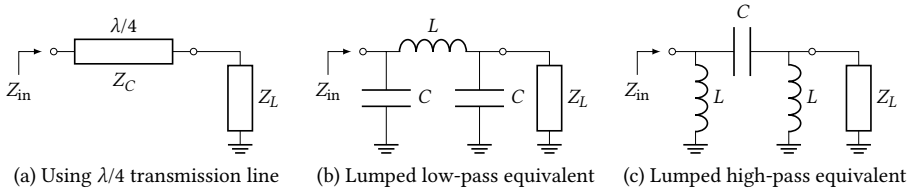
Figure A.2: Quarter wave transmission lines and their lumped equivalents at the design frequency

$$Z_{\text{in}} = \frac{Z_C^2}{Z_L} \qquad\qquad Z_C = \sqrt{Z_{\text{in}} Z_L} \qquad (A.28)$$

$$L = \frac{Z_C}{\omega} \qquad\qquad C = \frac{1}{\omega Z_C} \qquad (A.29)$$

$$\theta' = \sin^{-1}\left(\frac{Z_C}{Z_{C'}}\right) \quad \forall Z_{C'} > Z_C \qquad (A.30)$$

$$C' = \frac{\cos\theta'}{\omega Z_C} \qquad (A.31)$$
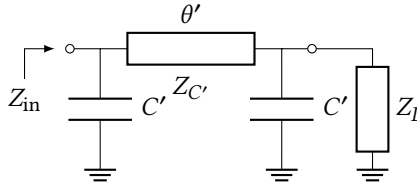
Figure A.3: Generalized semi-lumped equivalents of a quarter wave transmission line at the design frequency.

A.4.2 Coupled Inductors with a Common Node
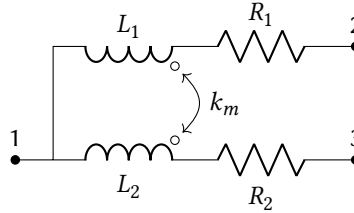
Figure A.4: Two coupled inductors with a common node.

Formulas for inductance may also work in the presence of shunt capacitors to ground. Just in case of problems, these could be de-embedded by simulating or measuring separately without bondwires.

$$A = 2\pi f \cdot (\Im\{Y_{12}\} \cdot \Im\{Y_{13}\} + \Im\{Y_{12}\} \cdot \Im\{Y_{23}\} + \Im\{Y_{13}\} \cdot \Im\{Y_{23}\}) \tag{A.32}$$

$$M = k_m \sqrt{L_1 L_2} \tag{A.33}$$

$$L_1 = \frac{\Im\{Y_{13} + Y_{23}\}}{A} \tag{A.34}$$

$$L_2 = \frac{\Im\{Y_{12} + Y_{23}\}}{A} \tag{A.35}$$

$$M = \frac{\Im\{Y_{23}\}}{A} \tag{A.36}$$

$$k_m = \frac{\Im\{Y_{23}\}}{\sqrt{\Im\{Y_{13} + Y_{23}\} \cdot \Im\{Y_{12} + Y_{23}\})}} \tag{A.37}$$

$$\tag{A.38}$$

$$R_{\mathrm{DC1}} = \lim_{f \to 0} -\Re\left\{\frac{1}{Y_{12}}\right\} \qquad R_{\mathrm{DC2}} = \lim_{f \to 0} -\Re\left\{\frac{1}{Y_{13}}\right\} \tag{A.39}$$

$$Q_1 = -\frac{\Im\{Y_{12}\}}{\Re\{Y_{12}\}} = \frac{\Im\left\{\frac{1}{Y_{12}}\right\}}{\Re\left\{\frac{1}{Y_{12}}\right\}} \qquad Q_2 = -\frac{\Im\{Y_{13}\}}{\Re\{Y_{13}\}} = \frac{\Im\left\{\frac{1}{Y_{13}}\right\}}{\Re\left\{\frac{1}{Y_{13}}\right\}} \tag{A.40}$$

$$R_{\mathrm{AC1}} = -2\pi f L_1 \frac{\Re\{Y_{12}\}}{\Im\{Y_{12}\}} \qquad R_{\mathrm{AC2}} = -2\pi f L_2 \frac{\Re\{Y_{13}\}}{\Im\{Y_{13}\}} \tag{A.41}$$

A.5 Optimization of Driver Chains

A.5.1 For Inverter Based Drivers

From the definition in [71], the equivalent switch resistance R_{dr} is defined as the resistance that discharges the capacitive load C_L from $V_{DD,dr}$ to $V_{DD,dr}/2$ in the same time as a fully modeled NMOS device would, or charges from 0 to $V_{DD,dr}/2$ in case of a PMOS device. The related propagation delay for a step input is $t_p = \ln(2)R_{dr}\left(C_{dr,o_1} + C_L\right)$. To have the same delay for low-to-high $t_{pH \to L}$ and high-to-low $t_{pL \to H}$, the NMOS and PMOS should have the same equivalent resistance $R_{dr_1} = R_{dr,p_1} = R_{dr,n_1}$ (Fig. 4.2a). The effective input and output capacitances of the kth CMOS inverter in the chain (numbering starting from the load) are then given by $C_{dr,i_k} = C_{dr,ip_k} + C_{dr,in_k}$ and $C_{dr,o_k} = C_{dr,op_k} + C_{dr,on_k}$ respectively (Fig. 4.2). Further, the technology-dependent self-loading factor is defined as $\gamma = \frac{C_{dr,o_k}}{C_{dr,i_k}}$.

Then, from [71, Eqs. (5.30) and (5.37)] $t_{p,int_k} = t_{p0}\left(1 + \frac{f_k}{\gamma}\right)$ and $t_{p_k} = t_{p,int_k} + \varsigma t_{p_{k+1}}$. Defining the tapered buffer chain such that the propagation delays and effective fan-outs are constant (i.e., $t_{p_k} = t_p \forall k$ and $f_k = f \forall k$) results in

$$t_p(1 - \varsigma) = t_{p,int} \tag{A.42}$$

and therefore

$$f(t_p) = \gamma\left(\frac{t_p}{t_{p0}}\left[1 - \varsigma\right] - 1\right), \tag{A.43}$$

or alternatively in terms of 0–100 linear rise and fall time

$$f(t_{rf}) = \gamma\left(\frac{t_{rf}}{t_{p0}r_{rf/p1,0\to100}}\left[1 - \varsigma\right] - 1\right). \tag{A.44}$$

Note that $1 - \varsigma = t_{p,int}/t_p$, which can be empirically determined for a buffer chain with $f = 1$, resulting in t_{p1} and

$$f(t_p) = \gamma\left(\frac{t_p}{t_{p0}}\frac{t_{p,int_1}}{t_{p1}} - 1\right) \tag{A.45}$$

Then filling in $t_p = \ln(2)R_{dr}(C_{dr,o_1} + C_L) + \varsigma t_p = \ln(2)R_{dr}(C_{dr,o_1} + C_L)/(1 - \varsigma)$ and $C_{dr,o_1} = t_{p0}/\left(\ln(2)R_{dr_1}\right)$ which results in Eq. (4.13).

The total segment capacitance is the sum of the load capacitance and all the effective (including intrinsic) inverter input and output capacitances.

$$C_{seg} = C_L + C_{dr,o_1} + C_{dr,i_1} + \cdots + C_{dr,o_N} + C_{dr,i_N} \tag{A.46}$$

$$= C_L + \frac{C_{dr,o_1} + C_{dr,i_1}}{f^0} + \cdots + \frac{C_{dr,o_1} + C_{dr,i_1}}{f^{N-1}} \tag{A.47}$$

$$= C_L + C_{dr,i_1}(1 + \gamma)\sum_{k=0}^{N-1} f^{-k} \tag{A.48}$$

giving Eq. (4.9) by the sum of a geometric series and that $C_{dr,i_1} = C_L/f$. Then by the definition of M (Eq. (4.10)), and substituting in $N = \log_f F + 1$, gives

$$M = 1 + \frac{1 + \gamma}{f}\left(\frac{fC_L - C_{dr,i_N}}{(f - 1)C_L}\right). \tag{A.49}$$

By assuming $C_L(f-1) \gg C_{\mathrm{dr},i_N}$ the last factor can be approximated by $f/(f-1)$, giving Eq. (4.12).

An additional assumption made here is that the total number of stages N is allowed to be fractional. Logically, the latter cannot be implemented and should be rounded up. However, this estimation is adequate since the smallest inverter stage will have a negligible capacitance compared to the final stage.

A.5.2 For Stacked Drivers

Split $t_{p,s}$ and $t_{p,c}$:

$$t_{p,\mathrm{tot}} = \varsigma_s t_{p,c} + t_{p,s}. \tag{A.50}$$

Using $t_{p,c}$ as input value gives the total multiplication factor

$$M_{\mathrm{tot}} = \frac{C_{\mathrm{tot},c} + C_{\mathrm{in},s}}{C_L} + \frac{C_{\mathrm{out},s} + C_L}{C_L}$$

$$= \frac{M_c \gamma_s^{-1} C_{\mathrm{out},s} + C_{\mathrm{out},s} + C_L}{C_L}. \tag{A.51}$$

$$\frac{C_{\mathrm{out},s}}{C_L} = \frac{t_{peq,s} - t_{p0,s}}{t_{p,\mathrm{tot}} - \varsigma_s t_{p,c} - t_{p0,s}} \tag{A.52}$$

$$M_c \approx \frac{f_c + \gamma_c}{f_c - 1} \tag{A.53}$$

$$f_c = \gamma_c \left(\frac{t_{p,c}}{t_{p0,c}} \left[1 - \varsigma_c \right] - 1 \right). \tag{A.54}$$

Providing

$$M_{\mathrm{tot}} = \left(\frac{\gamma_c + 1}{\gamma_c \frac{t_{p,c}}{t_{p0,c}} \left[1 - \varsigma_c \right] - \gamma_c - 1} + 1 \right) \frac{(t_{peq,s} - t_{p0,s})/\gamma_s}{t_{p,\mathrm{tot}} - \varsigma_s t_{p,c} - t_{p0,s}} + \frac{t_{peq,s} - t_{p0,s}}{t_{p,\mathrm{tot}} - \varsigma_s t_{p,c} - t_{p0,s}} + 1. \tag{A.55}$$

Optimizing for minimum M_{tot} for a given $t_{p,\mathrm{tot}}$ gives

$$\frac{\partial M_{\mathrm{tot}}}{\partial t_{p,c}} = 0 \tag{A.56}$$

giving

$$t_{p,c}^{(\min M)} = \frac{\gamma_c + 1}{\gamma_s + 1} \cdot \frac{\varsigma_s \gamma_s + \sqrt{\varsigma_s^2 \gamma_s^2 - \varsigma_s (\gamma_s + 1) \left(\varsigma_s \gamma_s - \frac{\gamma_c (1 - \varsigma_c)(t_{p,\mathrm{tot}} - t_{p0,s})}{(\gamma_c + 1) t_{p0,c}} \right)}}{\gamma_c \frac{1 - \varsigma_c}{t_{p0,c}} \varsigma_s} \tag{A.57}$$

However, the supply voltages of stack and chain are different, also resulting in a different power consumption. When assuming a stack of two, the chains' supplies are half the stack's, thus consuming a quarter times the power with respect to their capacitances

$$M_{P,\mathrm{tot}} = \left(\frac{\gamma_c + 1}{\gamma_c \frac{t_{p,c}}{t_{p0,c}} \left[1 - \varsigma_c \right] - \gamma_c - 1} + 1 \right) \frac{(t_{peq,s} - t_{p0,s})/4\gamma_s}{t_{p,\mathrm{tot}} - \varsigma_s t_{p,c} - t_{p0,s}} + \frac{t_{peq,s} - t_{p0,s}}{t_{p,\mathrm{tot}} - \varsigma_s t_{p,c} - t_{p0,s}} + 1. \tag{A.58}$$

A

Then minimizing the power consumption typically results in optimizing for a faster chain, as

$$t_{p,c}^{(\min P)} = \frac{\gamma_c + 1}{4\gamma_s + 1} \cdot \frac{4\varsigma_s\gamma_s + \sqrt{16\varsigma_s^2\gamma_s^2 - \varsigma_s(4\gamma_s + 1)\left(4\varsigma_s\gamma_s - \frac{\gamma_c(1-\varsigma_c)(t_{p,\text{tot}}-t_{p0,s})}{(\gamma_c+1)t_{p0,c}}\right)}}{\gamma_c\frac{1-\varsigma_c}{t_{p0,c}}\varsigma_s}. \tag{A.59}$$

A.6 Conversion from S-Parameters to Distributed Elements

It is possible to convert the simulated or measured S-parameters of a (quasi-)TEM transmission line to distributed (frequency dependent) series resistance R, series inductance L, shunt conductance G, and shunt capacitance C, as used in the telegrapher's equations. For that, first the complex propagation constant γ is extracted from the S-parameters

$$e^{-\gamma\ell} = \frac{1 - S_{11}^2 + S_{21}^2 \pm K}{2S_{21}} \tag{A.60}$$

where ℓ is the line's physical length and

$$K^2 = \left((1 + S_{11})^2 - S_{21}^2\right)\left((1 - S_{11})^2 - S_{21}^2\right). \tag{A.61}$$

This then gives

$$\gamma = \ell^{-1}\ln\left(\frac{1 - S_{11}^2 + S_{21}^2 + K}{2S_{21}}\right) \quad \vee \quad \ell^{-1}\ln\left(\frac{1 - S_{11}^2 + S_{21}^2 - K}{2S_{21}}\right) \tag{A.62}$$

as well as the line's characteristic impedance using Z_0 as the S-parameter port impedance

$$Z_c^2 = Z_0^2\frac{(1 + S_{11})^2 - S_{21}^2}{(1 - S_{11})^2 - S_{21}^2}. \tag{A.63}$$

Since

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{A.64}$$

$$Z_c = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \tag{A.65}$$

calculating the distributed elements' values is straightforward:

$$R = \Re\{\gamma Z_c\} \qquad\qquad G = \Re\left\{\frac{\gamma}{Z_c}\right\} \tag{A.66}$$

$$L = \frac{\Im\{\gamma Z_c\}}{\omega} \qquad\qquad C = \frac{\Im\left\{\frac{\gamma}{Z_c}\right\}}{\omega}. \tag{A.67}$$

All parameters are provided for a normalized transmission line length of 1 m. A software implementation using ADS AEL is provided in Section B.3.

A.7 Baseband considerations of a DTX
A.7.1 Baseband Current Magnitude Calculation
Using a 'sharp' function (as opposite to smooth) yields many harmonics, which can result in a circuit simulator to not converge to a solution. For example, the half-rectifier function $f_r(x) = \max(0, x)$ is only C^0 continuous and yields infinite harmonics. Reducing the number of harmonics also modifies the DC component, and strongly depends on the used rectifying activation function f_r.

Any rectifying activation function maps a real domain to a positive range, specifically $f_r : [-1, 1] \rightarrow [0, 1]$. Since the wanted activation after upconversion should remain unmodified, the rectifying function is constrained by

$$f_r(x) + f_r(-x) = x. \tag{A.68}$$

This can be rewritten to

$$f_r(x) - x/2 = -(f_r(-x) - x/2) \tag{A.69}$$

which means $f_r(x)$ is inherently an even function added to $x/2$. Using $f_r(x) = \max(0, x)$ as an example, $f_r(x) - x/2 = |x|/2$, which is even. Assuming that $f_r(x)$ is a polynomial, i.e., $f_r(x) = \sum_n p_n x^n$, solutions can be found by solving for $f_r(-1) = 0$ and $f_r(1) = 1$, for the first derivative $f_r'(-1) = 0$ and $f_r'(1) = 1$, and all higher order derivatives $f_r^{(n)}(-1) = f_r^{(n)}(1) = 0 \forall n >= 2$, as illustrated in Fig. A.5. By including more derivatives as boundary condition,
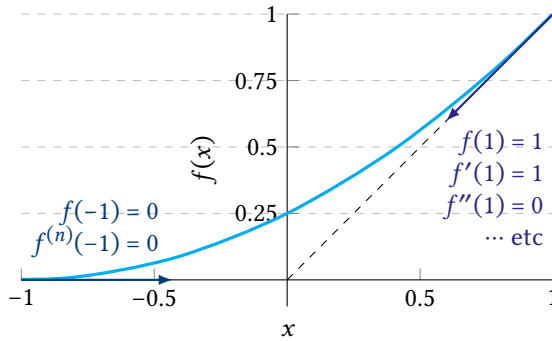
Figure A.5: Setting the boundary conditions for finding analytical smooth (continuously differentiable C^∞) rectifying functions.

the curve will be pulled more towards the limit of the half-rectifier function, making the 'corner' at $x = 0$ sharper. The first few solutions then are

```
Poly1  = (X)/2
Poly2  = (1*X.^2+2*X+1)/4;
Poly4  = (-1*X.^4+6*X.^2+8*X+3)/16;
Poly6  = (1*X.^6-5*X.^4+15*X.^2+16*X+5)/32;
Poly8  = (-5*X.^8+28 *X.^6-70*X.^4+140*X.^2+128*X+35)/256;
Poly10 = (7*X.^10-45*X.^8+126*X.^6-210*X.^4+315*X.^2+256*X+63)/512;
Poly12 = (-21*X.^12+154*X.^10-495*X.^8+924*X.^6-1155*X.^4+1386*X.^2+1024*X+231)/2048;
Poly14 = (33*X.^14-273*X.^12+1001*X.^10-2145*X.^8+3003*X.^6-3003*X.^4
  +3003*X.^2+2048*X+429)/4096;
Poly16 = (-429*X.^16+3960*X.^14-16380*X.^12+40040*X.^10-64350*X.^8+72072*X.^6
  -60060*X.^4+51480*X.^2+32768*X+6435)/65536;
```

```
Poly18 = (715*X.^18-7293*X.^16+33660*X.^14-92820*X.^12+170170*X.^10-218790*X.^8
  +204204*X.^6-145860*X.^4+109395*X.^2+65536*X+12155)/131072;
Poly20 = (-2431*X.^20+27170*X.^18-138567*X.^16+426360*X.^14-881790*X.^12
  +1293292*X.^10-1385670*X.^8+1108536*X.^6-692835*X.^4+461890*X.^2+262144*X+46189)/524288.
```

Using a higher order means the simulation result is more accurate, in the sense that it is closer to the actual operation of a DTX. These rectifying activation functions, when composed with a (co)sinusoidal function, actually yield identical results to the harmonic series as in [88, 90]. As such, the results in Table A.1 correspond to the results in [88, Table 1.1] and [90, Appendix C], where [88] refers to `poly2` as a 'square-law' and the half-rectifier as the 'dog-leg' function.

Table A.1: Fourier components of $f_r(\sin(x)) + f_r(-\sin(x))$ (baseband), normalized by a factor $\pi/2$ such that the half-rectifier has a DC component of 1 for comparison.

n^{th} harmonic	Poly1	Poly2	Poly4	$\cdots$	half-rect
DC	1.570 796	1.178 097	1.104 466	$\cdots$	1
2nd	0	0.392 699	0.490 874	$\cdots$	0.666 667
4th	0	0	0.245 437	$\cdots$	0.133 333
6th	0	0	0	$\ddots$	0.057 143
8th	0	0	0		0.031 746

Smoothed activation functions suitable for Doherty branches can be found by setting $g^{(n)}(0) = 0\,\forall n$ and modifying the first derivative $g'(1)$, for example to $g'(1) = 2$ for the driving profile of a symmetrical 2-way Doherty peaking DTX. This now inherently yields an odd function, for which the first viable solution is $g(x) = (-x^5 + 3x^3)/2$. Such activation should then still be passed through a rectifying function, such that the composed Doherty activation becomes $(f_r \circ g)(x)$.

Other approximations are possible by letting go of of the strict boundary conditions. A softplus function can be used, for example, the LogSumExp: $f_{LSE}(x) = \ln(1 + e^x)$. This function is also C^∞ continuous and can be freely scaled, for example introducing a 'smoothness' factor $a > 0$ and again ensuring $f_r : [-1, 1] \rightarrow [0, 1]$ gives

$$f_r(x, a) = \frac{f_{LSE}(ax) - f_{LSE}(-a)}{a} = \frac{\ln(1 + e^{ax}) - \ln(1 + e^{-a})}{a}. \tag{A.70}$$

Here, a smaller value of a results in a more 'smooth' function, in the sense of smaller harmonics magnitudes, and a larger value of a is closer to the DTX half-rectifier operation. In contrast to the polynomial based rectifying functions, the f_{LSE} will introduce infinite Fourier components, having a magnitude somewhere between the linear (Poly1, equal to using $a = 0$) function, and the half-rectifier function (e.g., using $a \rightarrow \infty$). A more computationally favorable look-alike would be the 'squareplus' function $f_{sqp}(x, b) = \left(x + \sqrt{x^2 + b^2}\right)/2$ with $b \geq 0$ [117]. Again ensuring $f_r : [-1, 1] \rightarrow [0, 1]$ gives

$$f_r(x, b) = f_{sqp}(x, b) - f_{sqp}(-1, b) = \frac{x + 1 + \sqrt{x^2 + b^2} - \sqrt{1 + b^2}}{2}. \tag{A.71}$$

Here, a larger value of b results in a more 'smooth' function, and gives very similar shaping results compared to f_{LSE} by using $b \approx 2/a$. Both these functions allow easy parameter

modification to serve different function domains and ranges, e.g., to implement smoothed Doherty activations, and increase simulation speed.

For the ADS implementation of f_{LSE} to be used as rectifying activation function see Section B.2.7.

A.7.2 Distributed Decoupling Capacitors

As discussed in Section 8.5.3, we know that optimal damping of a resonance occurs when the shunt capacitor has $Q = 1$ at the frequency of the (undamped) resonance. Even more optimal would be a capacitor with a constant Q, which would be a theoretical constant phase element [118]. Using $Q = 1$, the impedance becomes

$$Z_{\text{CPE}}(\omega) = \frac{1}{\sqrt{\text{j}\omega T}} \tag{A.72}$$

providing a constant phase of $-45°$, and an impedance decreasing proportional to the square root of the frequency.

Even though the theoretical capacitance with a constant $Q = 1$ does not exist, we can approximate its impedance around a nominal frequency by placing several series RC combinations with different time constants in parallel. Scaling all component values down by powers of two, we get the schematic as shown in Fig. A.6a. A single RC combination has $Q = 1$ for $\omega_1 = 1/RC$, such that its impedance is $Z_C(\omega_1) = (1 - \text{j})R = \frac{1-\text{j}}{\omega_1 C}$. We assume this to be the nominal impedance and frequency that our approximated constant phase element should be defined around, and its (normalized) impedance is shown in orange in Fig. A.6b. Increasing the order increases the total capacitance required to be able to reach the nominal value, but increases the total bandwidth for which $Q \approx 1$. The required normalized component values, as well as the total capacitance required, are provided in Table A.2.

A different way to approximate the constant phase element is by using an RC ladder configuration in shunt, as shown in Fig. A.7. This type of implementation is more compatible with structures on an IC, primarily keeping the $Q \approx 1$ for the higher frequencies. This is reflected in the f_H/f_0 values, as can be found in Table A.3. In the limit, the line becomes a fully distributed RC line, for which the total capacitance and resistance can be solved algebraically. Setting the distributed line impedance $Z_\infty = \sqrt{\frac{R}{\text{j}\omega C}} \frac{1}{\tanh\left(\sqrt{\text{j}\omega RC}\right)}$ equal to that of a single RC combination at ω_1 results in many solutions, of which the first one is $R_{\text{tot}} = \frac{\pi}{\tanh(\pi/2)}$ and $C_{\text{tot}} = \frac{\pi}{2}\tanh\left(\frac{\pi}{2}\right)$. The approximate value of this solution is also provided in Table A.3. In the implementation, the value of C_{tot} does not increase much, in the limit it 'only' becomes 44 % larger. However, attention should be paid to the R_{tot}, as it more than triples. If this is not considered, the damping will be less than targeted, possibly resulting in sharper resonance peaking than intended.

The second solution is $R_{\text{tot}} = 2\pi\tanh(\pi)$ and $C_{\text{tot}} = \frac{\pi}{\tanh(\pi)}$. Further solutions become increasingly large and would be similar to choosing a lower value for ω_1, which also requires a larger value of C_{tot}.
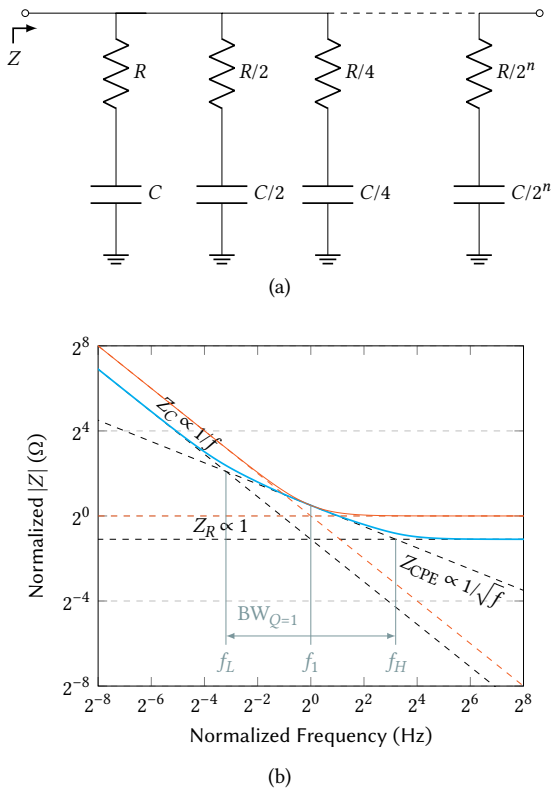
(a)

(b)

Figure A.6: Parallel binarily scaled series RC combinations and their impedance.

Table A.2: Required nominal component values for parallel binarily scaled series RC combinations for different orders.

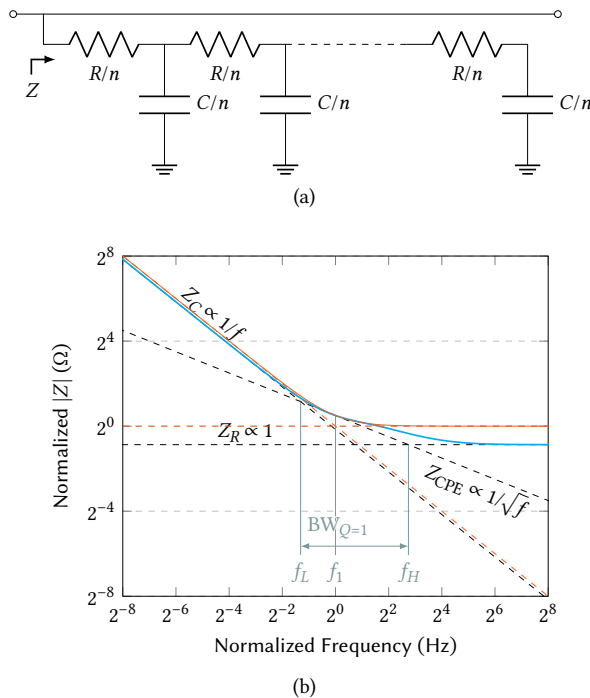| Order	R	C	$\sum_n C/2^n$	$f_1/f_L = f_H/f_1$
1	1	1	1	2
2	2.4	0.833 333	1.25	3.125
3	4.352 94	0.918 919	1.608	5.169
4	7.015 38	1.140 35	2.138	9.139
5	10.8226	1.478 39	2.864	16.40
6	16.0913	1.988 66	3.915	30.64
7	23.6760	2.703 16	5.364	57.52
8	34.1980	3.742 91	7.457	111.2
9	49.3597	5.186 41	10.35	214.2
10	70.3999	7.272 73	14.53	422.0
11	100.721	10.1667	20.32	825.5
12	133.734	13.4243	26.84	1652

(a)

(b)

Figure A.7: *RC* ladder combinations and their impedance.

Table A.3: Required nominal component values for *RC* ladder combinations for different orders.

Order	R	C	$\sum_n R$	$\sum_n C$	f_1/f_L	f_H/f_1
1	1	1	1	1	2	2
2	0.806 831	0.521 166	1.613 66	1.042 33	2.166	3.078
3	0.654 088	0.360 725	1.962 26	1.082 18	2.333	4.686
4	0.547 037	0.278 656	2.188 15	1.114 63	2.481	6.700
5	0.469 562	0.228 233	2.347 81	1.141 16	2.595	9.083
6	0.411 246	0.193 876	2.467 48	1.163 26	2.696	11.86
7	0.365 857	0.168 854	2.561 00	1.181 98	2.791	14.98
8	0.329 550	0.149 759	2.636 40	1.198 07	2.865	18.46
⋮			⋮	⋮	⋮	⋮
∞			3.425 38	1.440 66	4.151	∞

B

Simulation Models

B.1 General DTX Simulation Remarks

B.1.1 Harmonic Balance (HB) Simulations

In Harmonic Balance simulations, 'enough' harmonics should be added. Although an order of 5 or 7 harmonics can already be enough in analog simulations, it should be around 30 harmonics or beyond in digital transmitter simulations. Also, some fundamental oversampling must be included for simulation convergence and accuracy. To have time-domain waveforms that are visibly clean (free of Gibbs ripples/ringing artifacts), even higher orders can be needed, though not required for simulation accuracy itself. Consequently, one should ensure that the EM/S-parameter models used in these simulations behave well for these high frequencies, including enforced passivity when extrapolating. Convergence issues may arise when many harmonics are included when working with complex (matching) networks. These typically involve many nodes which are not directly connected together, resulting in a sparse matrix to be solved. Here, a direct solver might not be able to find a solution. Instead, a Krylov subspace solver might be needed to find a solution that can be (re)used to speed up subsequent simulations (possibly using the direct solver again) by including it as the initial state.

It is also possible to limit the harmonics present in the simulation. Using the simulation example of Section 5.2.3, the half-sine sources or the sine wave signal splitting of Fig. 5.12a are the reason for the harmonics of the baseband signal. This input splitting can be made smoother than a hard knee or 'dog-leg' half-rectifier function as used by the LinearActivation component (Section B.2.5). The half-rectifier has a discontinuous derivative, i.e., it is only C^0 continuous. By choosing a C^1 continuous function, gradient methods have a better change of converging. Even better is to use C^2 continuous functions, allowing the use of Newton's method. This yields less harmonic expansion helping simulator convergence, which can reduce simulation time of the example in Section 5.2.3 from 28 seconds to 18 seconds with otherwise identical simulator settings. For this smooth signal splitting, the LinearActivation_Smooth or LinearActivation_SmoothExp components can be considered (Sections B.2.6 and B.2.7, see also Section A.7.1).

B.1.2 Envelope Simulations

For envelope simulations, the same holds true for HB simulations regarding the number of harmonics. The envelope timestep should be smaller than the actual sampling rate to properly include the effects of sampling. The envelope simulation provides successive frequency domain (harmonic balance) simulations with a time-varying input. From timestep to timestep, it keeps the transient operating point information of reactive components (e.g., current through an inductor and charge stored in a capacitor), such that dynamic memory effects can also be included. This results in a discrete time-varying spectrum in the simulator output, which can be 'stitched' together to a time domain waveform using the built-in ADS function `ts()`, which is implemented as

$$v_x(t) = \Re\left\{\sum_{k=0}^{N} V_{x[k]}(t)e^{\mathrm{j}2\pi f_k t}\right\}. \tag{B.1}$$

Here $v_x(t)$ is the voltage at node x, and $V_{x[k]}(t)$ the time-varying Fourier coefficient of the voltage at that node. To get a complete time domain waveform without aliasing, the available timestamps and the Fourier coefficients will be linearly interpolated. This provides a 'free' first-order-hold (FOH) on the input data, which might not be there in reality. As such, sampling errors might be represented too optimistically. To properly include sampling errors in simulation, first make sure the interpolation mode for timed sources using data access components is a (floor) value lookup, i.e., not interpolated. Next, ensure the envelope timestep is small enough to accurately display the effects of a zero-order-hold. For example, in a signed Cartesian operation, a timestep smaller than $\frac{1}{4f_0}$ is proper, while for 8-phase multi-phase $\frac{1}{8f_0}$ is required. The simulator will provide the warning: "There are analysis frequencies inside Envelope bandwidths of other frequencies. A simpler frequency grid may be possible and faster." This is normal. When examining the output spectrum, only looking at the spectrum of the fundamental (for example, `fs(V_out[::,1])`, or similar) might not show the full picture. Instead, use the full harmonic content and transform to the time domain using enough points, e.g., `ts(V_out,,,numtpts)`, before creating the spectrum. A good choice for numtpts is the total number of envelope cycles (simulation time over timestep) with an additional oversampling factor (2~8, depending on wanted accuracy and highest frequency available in the resulting spectrum) with 1 added. This +1 is necessary to avoid the 'fencepost error' when interpolating the data points. This signal can be plugged directly into the `fs()` function when using ADS. In MatLab, the last sample has to be removed again. Be wary of FFT-windowing effects in general; cyclic data together with a rectangular window works best: if any spectral leakage occurs, it is a clear sign that the FFT settings are incorrect.

B.2 ADS Components

B.2.1 Imult

This component serves as a current multiplier block, where the current of port 1 is a scaled version of the current of port 2 depending on the applied control voltage (CV) at port 3. The port 1 and 2 are directly connected in terms of voltage, or scaled/shifted depending on constants `MV` or `Vconst`. Port 3 is terminated by a parameterized resistor to provide a

DC path to ground to aid convergence, as well as to enable the use of impedance matched components in the CV path.
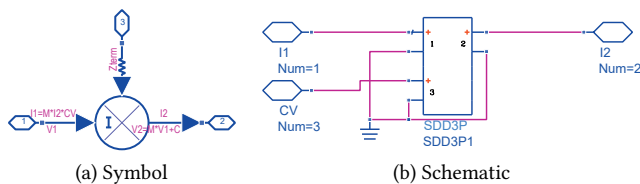

(a) Symbol   (b) Schematic

Figure B.1: Imult component definition.

Cell Parameters Imult

Parameter Name	Value Type	Default Value	Type/Unit	Parameter Description
MI	Real	1	Unitless	Current Multiplier (static)
MV	Real	1	Unitless	Voltage Multiplier (static)
Vconst	Real	0	Voltage	Voltage Shift (DC) (static)
Zterm	Real	50	Resistance	Control Voltage termination impedance

3-Port Symbolically Defined Device (SDD3P) parameters

```
F[1,0]=_i1+MI*_i2*_v3
F[2,0]=MV*_v1-_v2+Vconst
I[3,0]=(_v3)/Zterm
C[1]=
Cport[1]=
```

B.2.2 Switch_ISAT_Ron

Based on an ideal switch model with added current saturation to behaviorally implement a MOSFET switch, as its load line travels through the saturation region before reaching the R_{ON} related to the triode region.

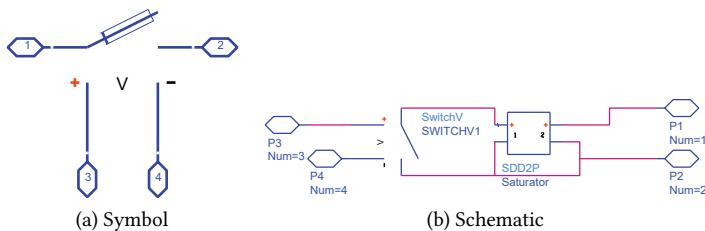(a) Symbol   (b) Schematic

Figure B.2: Switch_ISAT_Ron component definition.

Switch (SWITCHV1) parameters

```
Model=
R1=R1 Ohm
V1=V1 V
R2=R2 Ohm
V2=V2 V
```

Cell Parameters Switch_ISAT_Ron

Parameter Name	Value Type	Default Value	Type/Unit	Parameter Description
R1	Real	1.0	Resistance	Resistance at voltage 1
V1	Real	0.0	Voltage	Voltage 1
R2	Real	1 M	Resistance	Off resistance
V2	Real	1.0	Voltage	Voltage 2
ISAT	Real	1.0	Current	Saturation Current
Sharpness	Real	1	Unitless	Sharpness of saturation

2-Port Symbolically Defined Device (Saturator) parameters

```
F[1,0]=_v1-_v2
F[2,0]=_i2+(atan((abs( _i1 )/ISAT)**Sharpness*pi/2)*2/pi*(ISAT)**Sharpness)**(1/Sharpness)
 * sgn(_i1)
C[1]=
Cport[1]=
```

B.2.3 SPDT_Dynamic_ADJcmosVDD

Modeled after the ADS SPDT_Dynamic (Single Pole Double Throw Switch, Dynamic) component, but adjusted for user defined switch voltages, namely the V_{DD} of a CMOS inverter (driver). Note that this is (intentionally) an inverting component, but can be made non-inverting by swapping the SWITCHV1 and SWITCHV2 parameter values (e.g., SPDT_Dynamic_ADJcmosVDD_noInv).

Convergence problems may occur if VDD is too small or the difference between Ron and Roff is too large. Also, enough harmonics of the control voltage input (port 4) should be included in harmonic balance or envelope simulations, also see the general simulation remarks (Section B.1).
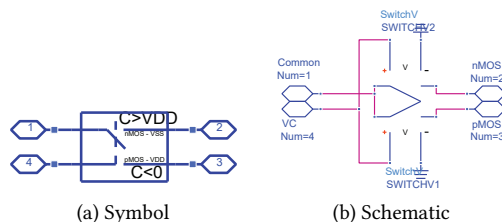

(a) Symbol   (b) Schematic

Figure B.3: SPDT_Dynamic_ADJcmosVDD component definition.

Cell Parameters SPDT_Dynamic_ADJcmosVDD

Parameter Name	Value Type	Default Value	Type/Unit	Parameter Description
VDD	Real	2.5	Voltage	Supply voltage
Ron	Real	1	Resistance	On resistance
Roff	Real	1 M	Resistance	Off resistance
Overlap	Real	0.5	Voltage	Switching overlap
M_pMOS	Real	1	Unitless	Relative pMOS size (higher means lower Ron)
M_nMOS	Real	1	Unitless	Relative nMOS size (higher means lower Ron)

Switch 1 (SWITCHV1) parameters

```
Model=
R1=Ron*(1/M_pmos)
V1=0 V
R2=Roff
V2=(VDD+Overlap) V
```

Switch 2 (SWITCHV2) parameters

```
Model=
R1=Ron*(1/M_nmos)
V1=VDD V
R2=Roff
V2=(0-Overlap) V
```

B.2.4 SPDT_Dynamic_ADJcmosVDD_noInv_Sat

Similar to SPDT_Dynamic_ADJcmosVDD (Section B.2.3), but not inverting and resistive switches with current saturation (Section B.2.2). This way, it is a behavioral model of a CMOS inverter, allowing for simple but accurate modeling of a CMOS driver (see Section 8.4.3).
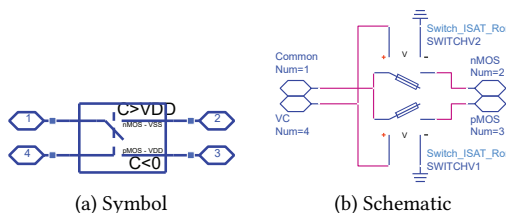

(a) Symbol   (b) Schematic

Figure B.4: SPDT_Dynamic_ADJcmosVDD_noInv_Sat component definition.

Cell Parameters SPDT_Dynamic_ADJcmosVDD_noInv_Sat

Parameter Name	Value Type	Default Value	Type/Unit	Parameter Description
VDD	Real	2.5	Voltage	Supply voltage
Ron	Real	1	Resistance	On resistance
Roff	Real	1 M	Resistance	Off resistance
Overlap	Real	0.5	Voltage	Switching overlap
M_pMOS	Real	1	Unitless	Relative pMOS size (higher means lower Ron)
M_nMOS	Real	1	Unitless	Relative nMOS size (higher means lower Ron)
ISAT	Real	1.0	Current	Saturation Current
Sharpness	Real	1	Unitless	Sharpness of saturation

Switch_ISAT_Ron 1 (SWITCHV2) parameters

```
Model=
R1=Ron*(1/M_pmos)
V1=VDD V
R2=Roff
V2=(0-Overlap) V
ISAT=ISAT*M_pMOS
Sharpness=Sharpness
```

Switch_ISAT_Ron 2 (SWITCHV1) parameters

```
Model=
R1=Ron*(1/M_nmos)
V1=0 V
R2=Roff
V2=(VDD+Overlap) V
ISAT=ISAT*M_nMOS
Sharpness=Sharpness
```

B.2.5 LinearActivation

This block maps the domain [TH1 : TH2] to the range [0 : 1]. All input voltages below TH1 result in 0 output, while all above TH2 result in 1. Only C^0 continuous.

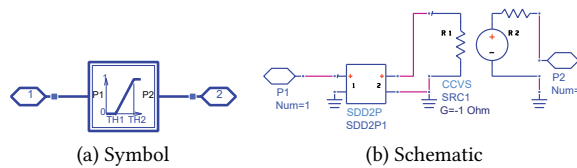(a) Symbol   (b) Schematic

Figure B.5: LinearActivation component definition.

Cell Parameters LinearActivation

Parameter Name	Value Type	Default Value	Type/Unit	Parameter Description
TH1	Real	0	Voltage	Threshold voltage
TH2	Real	1	Voltage	Limit voltage (threshold 2)

2-Port Symbolically Defined Device (SDD2P) parameters

```
I[1,0]=0
I[2,0]=if (_v1<=TH1) then 0 else if (_v1>=TH2) then 1 else _v1*(1-TH1/_v1)/(TH2-TH1) endif
  endif
C[1]=
Cport[1]=
```

Current Controlled Voltage Source (CCVS) parameters

```
G=-1 Ohm
T=0.0 nsec
R1=0 Ohm
R2=0 Ohm
F=0.0 GHz
```
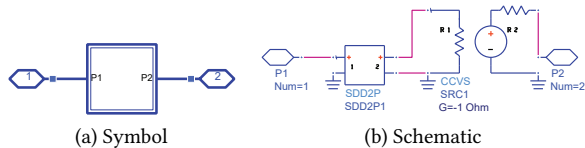
B.2.6 LinearActivation_Smooth

Same function as LinearActivation (Section B.2.5), but with smoothed curve around TH1. This block is C^1 continuous around TH1, only C^0 continuous at TH2.

2-Port Symbolically Defined Device (SDD2P) parameters

```
I[1,0]=0
I[2,0]=if (_v1<=TH1-TH1_smooth) then 0 else if (_v1>=TH2) then 1 else
  if(_v1<=TH1+TH1_smooth) then (1*(_v1-TH1+TH1_smooth)**2/(4*TH1_smooth*(TH2-TH1))) else
  _v1*(1-TH1/_v1)/(TH2-TH1) endif endif endif
C[1]=
Cport[1]=
```

(a) Symbol   (b) Schematic

Figure B.6: LinearActivation_Smooth component definition.

Cell Parameters LinearActivation_Smooth

Parameter Name	Value Type	Default Value	Type/Unit	Parameter Description
TH1	Real	0	Voltage	Threshold voltage
TH1_smooth	Real	0	Voltage	Smoothing range around TH1
TH2	Real	1	Voltage	Limit voltage (threshold 2)

Current Controlled Voltage Source (CCVS) parameters

```
G=-1 Ohm
T=0.0 nsec
R1=0 Ohm
R2=0 Ohm
F=0.0 GHz
```

B.2.7 LinearActivation_SmoothExp

This block maps the domain [DomainStart : ∞) to the range [0 : ∞) using LogSumExp functions, making it C^{∞} continuous, while also the smoothness can be controlled by a continuous parameter > 0. These use ln(exp(..)) functions, note that the argument should not exceed floating point limits (-708.39,709.78). Consider replacing LogSumExp by SquarePlus function for computational efficiency in the future.
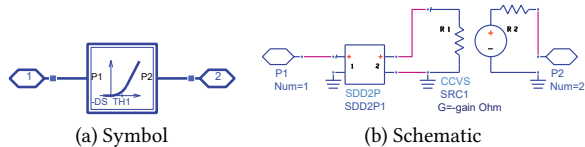

(a) Symbol   (b) Schematic

Figure B.7: LinearActivation_SmoothExp component definition.

2-Port Symbolically Defined Device (SDD2P) parameters

```
I[1,0]=0
I[2,0]=if(abs(sUse)<=1e-9)then((_v1-TH1+1)/2)else((LSE((_v1-TH1)*sUse,1)-y0)/sUse)endif
C[1]=
Cport[1]=
```

Note: abs(sUse)<=1e-9 can also be 1e-15, but anything below 1e-8 will start to impact quantization noise from floating point limit.

Cell variable equations (VAR)

```
LSE(x,c)=ln(c+exp(x,709))
sUse=smooth*sourceLevel
y0=if(smooth*(DomainStart-TH1)<=-36.75)then(0)else(LSE(sUse*(DomainStart-TH1),1))endif
```

Cell Parameters LinearActivation_SmoothExp

Parameter Name	Value Type	Default Value	Type/Unit	Parameter Description
TH1	Real	0	Voltage	Threshold voltage
smooth	Real	1	Unitless	Smoothness, smaller is smoother
gain	Real	1	Unitless	First derivative for large input
DomainStart	Real	−100	Voltage	Input value for which the output should be 0. Note, using smooth*DomainStart<-37 will not modify the transfer.

Notes: Use sourceLevel for convergence. If DomainStart is small enough, use 0 for y0. 1+exp(-36.74)-1 = 0 in double precision floating point, while 1-1+exp(-36.74) ≈ 1.1e-16.

Current Controlled Voltage Source (CCVS) parameters

```
G=-gain Ohm
T=0.0 nsec
R1=0 Ohm
R2=0 Ohm
F=0.0 GHz
```

B.3 ADS AEL

These application extension language (AEL) functions implement the *S*-parameter conversion of (quasi-)TEM transmission lines to distributed elements, as used in telegrapher's equations. See Section A.6.

B.3.1 stogammaz

```
// 2-port S parameters to propagation constant and characteristic impedance
// Rob Bootsman, April 2022
defun stogammaz(s, linelength, zRef)
{
  decl z0 = if (zRef == NULL) then 50.0 else zRef;
  decl ll = if (linelength == NULL) then 1 else linelength;
  decl aSMatSize = size(s);
  if (aSMatSize(1) == 2 && aSMatSize(2) == 2) {
    decl s11_2 = s(1,1).**2;
    decl s21_2 = s(2,1).**2;
    decl Z2_num = (1 + s(1,1)).**2 - s21_2;
    decl Z2_den = (1 - s(1,1)).**2 - s21_2;
    decl knum = sqrt(Z2_num.*Z2_den);
    decl expGammaLenPos = (1 - s11_2 + s21_2 + knum) ./ (2*s(2,1));
    decl logexp = ln(expGammaLenPos);
    if (sum(real(logexp)<0) != 0){
      expGammaLenPos = (1 - s11_2 + s21_2 - knum) ./ (2*s(2,1));
      logexp = ln(expGammaLenPos);
    }
    decl gamma = (real(logexp) + 1i*unwrap(imag(logexp)))/ll;
    decl Zc = sqrt(z0**2 * Z2_num./Z2_den);
    return {gamma,Zc};
  }
  print_function_error("stogammaz","Transformation is only available for 2-port scattering
    parameters, for now and probably forever");
}
```

B.3.2 storlgc

```
// 2-port S parameters to distributed RLGC parameters
```

```
// Rob Bootsman, April 2022
defun storlgc(s, linelength, zRef)
{
  decl z0 = if (zRef == NULL) then 50.0 else zRef;
  decl ll = if (linelength == NULL) then 1 else linelength;
  decl aSMatSize = size(s);
  decl freq = indep(s);
  if (aSMatSize(1) == 2 && aSMatSize(2) == 2) {
    decl gammazc = stogammaz(s, linelength, zRef);
    decl R = real(gammazc(1).*gammazc(2));
    decl L = imag(gammazc(1).*gammazc(2))./(2*pi*freq);
    decl G = real(gammazc(1)./gammazc(2));
    decl C = imag(gammazc(1)./gammazc(2))./(2*pi*freq);
    return {R,L,G,C};
  }
  print_function_error("storlgc","Transformation is only available for 2-port scattering
    parameters, for now and probably forever");
}
```

B.4 Cadence

B.4.1 Imult

Same core functionality as the Imult block in ADS (Section B.2.1), though without the possibility to scale the node voltages.
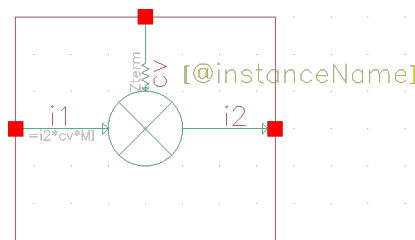
Figure B.8: Imult component symbol.

Listing B.1: Imult VerilogA code

```
// VerilogA for DesignLib, Imult, veriloga
// Rob Bootsman, January 2025
// with help of Tariq Ibrahim

`include "constants.vams"
`include "disciplines.vams"

module Imult(i1, i2, cv);
electrical i1, i2, cv; //cv: control voltage.
parameter real Zterm=50 from (0:inf);
parameter real MI=1;

analog
        begin
                V(cv) <+ Zterm*I(cv);
                I(i1) <+ I(i1,i2)*(V(cv)-1)*MI;
        end

endmodule
```

C

Chip Gallery

The research on the topic of (high-power) DTX involved designing multiple chips, often requiring a broad area of expertise. To realize these chips single-handedly would take forever, or might even be impossible due to the unavailability of technology PDKs. The multidisciplinary nature of DTX resulted in multiple designers contributing to my chips, as well as me contributing in one way or another in those of others. This varies between only simulating parts of a design and providing recommendations, all the way to providing the full details to be implemented on the chip. Below are the chip designs I have made or have contributed to, in chronological order and to scale.
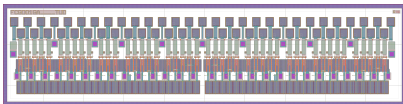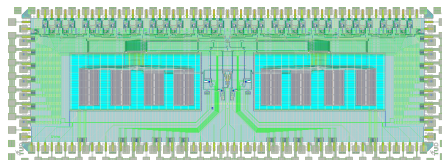

Design Lead:	Rob Bootsman
Technology:	Fraunhofer IAF GaN25
Code name:	FCG002GD____TUD
Size:	$5000 \times 1250\,\mu m^2$
Tape-out:	2018 March
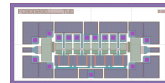
Design Lead:	Marco Pelk
Technology:	Fraunhofer IAF GaN25
Code name:	FCG001GA____TUD
Size:	$2000 \times 1000\,\mu m^2$
	(2 variants)
Tape-out:	2018 March
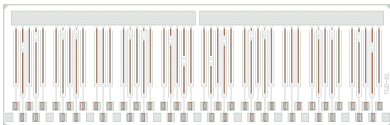
Design Lead:	Rob Heeres
Technology:	Ampleon LDMOS LM8
Size:	$1530 \times 4830\,\mu m^2$
	(3 variants)
Tape-out:	2018 April

Design Lead:	Rob Bootsman
Technology:	TSMC CLN40LP
Code name:	Thoth
Size:	$5445 \times 1985.4\,\mu m^2$
Tape-out:	2018 August
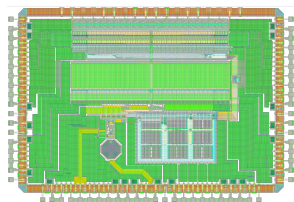
Design Lead:	Yiyu Shen
Technology:	TSMC CLN40LP
Code name:	Vermeer
Size:	3582 × 2457 µm^2
Tape-out:	2019 March
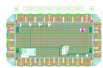
Design Lead:	Rob Bootsman
Technology:	TSMC CLN40LP
Code name:	Escher p1
Size:	839.9 × 1192.7 µm^2
Tape-out:	2019 December

Design Lead:	Rob Bootsman
Technology:	TSMC CLN40LP
Code name:	Turner p4
Size:	1215 × 569.8 µm^2
Tape-out:	2020 April
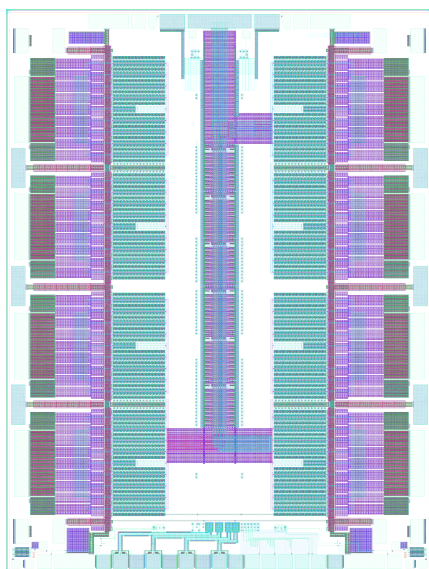
Design Lead:	Daniel Maassen
Technology:	Ampleon LDMOS LM9
Size:	5300 × 7000 µm^2 (11 variants + 1 PCM)
Tape-out:	2022 September
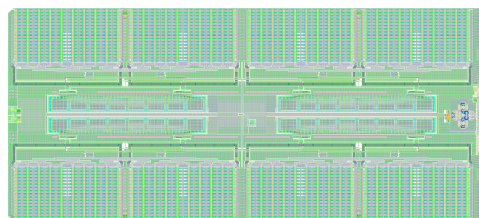
Design Leads:	Dieuwert Mul Rob Bootsman Mohammadreza Beikmirza
Technology:	TSMC CLN40LP
Code name:	Vosmaer
Size:	5937.0 × 2664.9 µm^2
Tape-out:	2022 October
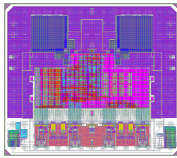
Design Leads:	Mohammadreza Beikmirza
	Tariq Ibrahim
	Ang Li
Technology:	GlobalFoundries 22FDX+
Code name:	DRASTIIC V2A
Size:	$2185 \times 1905.6\,\mu\text{m}^2$
Tape-out:	2024 October

C

Bibliography

References

[1] M. Roser, H. Ritchie, and E. Mathieu, "Technological change," *Our World in Data*, 2023, https://ourworldindata.org/technological-change.

[2] Ericsson Mobility Reports, "Mobile data traffic outlook," online, August 2023, https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast.

[3] 5G Infrastructure Public Private Partnership, https://web.archive.org/web/20231219064000/https://5g-ppp.eu/.

[4] G. E. Moore, "Cramming more components onto integrated circuits, reprinted from Electronics, volume 38, number 8, april 19, 1965, pp. 114 ff." *IEEE Solid-State Circuits Society Newsletter*, vol. 11, no. 5, pp. 33–35, Sept 2006.

[5] M. Roser, H. Ritchie, and E. Mathieu, "What is Moore's Law?" March 2023, https://ourworldindata.org/moores-law.

[6] Wikipedia contributors, "Transistor count — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Transistor_count&oldid=1185987635, 2023, [Online; accessed 21-November-2023].

[7] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, 1974.

[8] A. Andrae, "Projecting the chiaroscuro of the electricity use of communication and computing from 2018 to 2030," *ResearchGate*, 02 2019.

[9] H. Ritchie, M. Roser, and P. Rosado, "Energy," *Our World in Data*, 2023, https://ourworldindata.org/energy.

[10] H. T. Friis, "A note on a simple transmission formula," *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, 1946.

[11] H. Holma, H. Viswanathan, and P. Mogensen, "Extreme massive MIMO for macro cell capacity boost in 5G-advanced and 6G," Nokia Bell Labs, Tech. Rep., 2021.

[12] S. Wesemann, J. Du, and H. Viswanathan, "Energy efficient extreme MIMO: Design goals and directions," *IEEE Communications Magazine*, vol. 61, no. 10, pp. 132–138, 2023.

[13] Next Generation Mobile Networks Alliance, "Green future networks: Network energy efficiency v1.1," online, Dec 2021, www.ngmn.org.

[14] AD9161/AD9162, *11-Bit/16-Bit, 12 GSPS, RF Digital-to-Analog Converters*, Analog Devices, 2019, rev. D.

[15] DS926, *Zynq UltraScale+ RFSoC Data Sheet: DC and AC Switching Characteristics*, AMD Xilinx, 2023, v1.12.

[16] ADRV904x, *ADI RadioVerse® SoC Series Drives 5G Radio Efficiency and Performance*, Analog Devices, 2022, pre-release.

[17] R. J. Bootsman, D. P. N. Mul, Y. Shen, M. Hashemi, R. M. Heeres, F. van Rijs, M. S. Alavi, and L. C. N. de Vreede, "High-power digital transmitters for wireless infrastructure applications (a feasibility study)," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 5, pp. 2835–2850, 2022.

[18] M. P. van der Heijden, M. Acar, J. S. Vromans, and D. A. Calvillo-Cortes, "A 19W high-efficiency wide-band CMOS-GaN class-E Chireix RF outphasing power amplifier," in *2011 IEEE MTT-S International Microwave Symposium*, June 2011, pp. 1–4.

[19] E. McCune, Q. Diduck, W. Godycki, R. Booth, and D. Kirkpatrick, "A fully polar transmitter for efficient software-defined radios," in *2017 IEEE MTT-S International Microwave Symposium (IMS)*, June 2017, pp. 1946–1949.

[20] D. P. N. Mul, R. J. Bootsman, Q. Bruinsma, Y. Shen, S. Krause, R. Quay, M. J. Pelk, F. van Rijs, R. M. Heeres, S. Pires, M. Alavi, and L. C. N. de Vreede, "Efficiency and linearity of digital "class-C like" transmitters," in *2020 50th European Microwave Conference (EuMC)*, 2021, pp. 1–4.

[21] D. P. N. Mul, R. J. Bootsman, M. Beikmirza, M. S. Alavi, and L. C. N. de Vreede, "The efficiency and power utilization of current-scaling digital transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 72, no. 7, pp. 4350–4366, 2024.

[22] L. Ding, Z. Ma, D. R. Morgan, M. Zierdt, and G. T. Zhou, "Compensation of frequency-dependent gain/phase imbalance in predistortion linearization systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 1, pp. 390–397, 2008.

[23] M. Mehrpoo, M. Hashemi, Y. Shen, L. C. N. de Vreede, and M. S. Alavi, "A wideband linear I/Q-interleaving DDRM," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 5, pp. 1361–1373, 2018.

[24] T. Matsuura, W. S. Lee, T. Urushihara, and T. Nakatani, "High efficiency transmitter," US Patent US20 130 058 435A1, 2013. [Online]. Available: https://patents.google.com/patent/US20130058435A1

[25] M. Beikmirza, Y. Shen, L. C. N. de Vreede, and M. S. Alavi, "A wideband four-way Doherty bits-in RF-out CMOS transmitter," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 12, pp. 3768–3783, 2021.

[26] D. P. N. Mul, "Digital transmitter architectures – A signal processing perspective," Ph.D. dissertation, Delft University of Technology, 2026.

[27] D. C. Prince, "Vacuum tubes as power oscillators," *Proceedings of the Institute of Radio Engineers*, vol. 11, no. 3, pp. 275–313, 1923, part I.

[28] I. J. Kaar and C. J. Burnside, "Some developments in broadcast transmitters," *Proceedings of the Institute of Radio Engineers*, vol. 18, no. 10, pp. 1621–1660, 1930.

[29] W. L. Everitt, "Optimum operating conditions for class C amplifiers," *Proceedings of the Institute of Radio Engineers*, vol. 22, no. 2, pp. 152–176, 1934.

[30] S. C. Cripps, *RF power amplifiers for wireless communications*, 2nd ed.   Artech House Norwood, MA, 2006.

[31] B. Razavi, *RF microelectronics*, 2nd ed.   Pearson Education, Inc, 2012.

[32] D. Chowdhury, S. V. Thyagarajan, L. Ye, E. Alon, and A. M. Niknejad, "A fully-integrated efficient CMOS inverse class-D power amplifier for digital polar transmitters," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 5, pp. 1113–1122, 2012.

[33] F. Raab, "Idealized operation of the class E tuned power amplifier," *IEEE Transactions on Circuits and Systems*, vol. 24, no. 12, pp. 725–735, 1977.

[34] M. Acar, A. J. Annema, and B. Nauta, "Analytical design equations for class-E power amplifiers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 12, pp. 2706–2717, 2007.

[35] ——, "Analytical design equations for class-E power amplifiers with finite DC-feed inductance and switch on-resistance," in *2007 IEEE International Symposium on Circuits and Systems*, 2007, pp. 2818–2821.

[36] M. K. Kazimierczuk and W. A. Tabisz, "Class C-E high-efficiency tuned power amplifier," *IEEE Transactions on Circuits and Systems*, vol. 36, no. 3, pp. 421–428, March 1989.

[37] S. D. Kee, "The class E/F family of harmonic-tuned switching power amplifiers," Ph.D. dissertation, California Institute of Technology, 2002.

[38] M. Hashemi, Y. Shen, M. Mehrpoo, M. S. Alavi, and L. C. N. de Vreede, "An intrinsically linear wideband polar digital power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 12, pp. 3312–3328, 2017.

[39] F. Raab, "Efficiency of outphasing RF power-amplifier systems," *IEEE Transactions on Communications*, vol. 33, no. 10, pp. 1094–1099, 1985.

[40] J. H. Qureshi, M. J. Pelk, M. Marchetti, W. C. E. Neo, J. R. Gajadharsing, M. P. van der Heijden, and L. C. N. de Vreede, "A 90-W peak power GaN outphasing amplifier with optimum input signal conditioning," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 8, pp. 1925–1935, 2009.

[41] L. C. N. de Vreede, R. Gajadharsing, and W. C. E. Neo, "On the bandwidth performance of Doherty amplifiers," in *2013 IEEE International Wireless Symposium (IWS)*, 2013, pp. 1–4.

[42] J. H. Qureshi, W. Sneijers, R. Keenan, L. C. N. de Vreede, and F. van Rijs, "A 700-W peak ultra-wideband broadcast Doherty amplifier," in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, 2014, pp. 1–4.

[43] J. Pang, Y. Li, C. Chu, J. Peng, X. Y. Zhou, and A. Zhu, "Extend high efficiency range of Doherty power amplifier by modifying characteristic impedance of transmission lines in load modulation network," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 707–710.

[44] L. Zhou, L. Liu, M. Pelk, A. R. Qureshi, and L. C. N. de Vreede, "A 90W high-efficiency four-way Doherty power amplifier with 37.8% fractional bandwidth over a 15 dB power back-off range," in *2025 IEEE/MTT-S International Microwave Symposium - IMS 2025*, 2025.

[45] R. Quaglia and S. Cripps, "A load modulated balanced amplifier for telecom applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 3, pp. 1328–1338, 2018.

[46] Y. Cao and K. Chen, "Pseudo-Doherty load-modulated balanced amplifier with wide bandwidth and extended power back-off range," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 7, pp. 3172–3183, 2020.

[47] P. Eloranta, P. Seppinen, S. Kallioinen, T. Saarela, and A. Parssinen, "A multimode transmitter in 0.13 μm CMOS using direct-digital RF modulator," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 12, pp. 2774–2784, 2007.

[48] B. Schafferer and R. Adams, "A 3V CMOS 400mW 14b 1.4GS/s DAC for multi-carrier applications," in *2004 IEEE International Solid-State Circuits Conference (IEEE Cat. No.04CH37519)*, 2004, pp. 360–532 Vol.1.

[49] Y. Shen, M. Hoogelander, R. Bootsman, M. S. Alavi, and L. C. N. de Vreede, "A wideband digital-intensive current-mode transmitter line-up," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 9, pp. 2489–2500, 2023.

[50] S.-M. Yoo, J. S. Walling, E. C. Woo, B. Jann, and D. J. Allstot, "A switched-capacitor RF power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 12, pp. 2977–2987, 2011.

[51] S.-W. Yoo, S.-C. Hung, J. S. Walling, D. J. Allstot, and S.-M. Yoo, "10.7 a 0.26mm2 DPD-less quadrature digital transmitter with <−40dB EVM over >30dB Pout range in 65nm CMOS," in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2020, pp. 184–186.

[52] J. Zanen, E. Klumperink, and B. Nauta, "A predistortion-less digital MIMO transmitter with DTC-based quadrature imbalance compensation," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 8, pp. 2214–2225, 2023.

[53] L. C. N. de Vreede, S. M. Alavi, R. J. Bootsman, M. R. Beikmirza, D. P. N. Mul, R. Heeres, and F. van Rijs, "Digital transmitter with high power output," US Patent US12 294 360B2, May, 2025, US Patent 12,294,360. [Online]. Available: https://patents.google.com/patent/US12294360B2

[54] Z.-Y. Cui, J.-W. Park, C.-S. Lee, and N.-S. Kim, "Integration of CMOS logic circuits with lateral power MOSFET," in *2013 4th International Conference on Intelligent Systems, Modelling and Simulation*, 2013, pp. 615–618.

[55] K. Hoo Teo, Y. Zhang, N. Chowdhury, S. Rakheja, R. Ma, Q. Xie, E. Yagyu, K. Yamanaka, K. Li, and T. Palacios, "Emerging GaN technologies for power, RF, digital, and quantum computing applications: Recent advances and prospects," *Journal of Applied Physics*, vol. 130, no. 16, 2021.

[56] M. Acar, M. P. van der Heijden, and D. M. W. Leenaerts, "0.75 Watt and 5 Watt drivers in standard 65nm CMOS technology for high power RF applications," in *2012 IEEE Radio Frequency Integrated Circuits Symposium*, June 2012, pp. 283–286.

[57] S. J. C. H. Theeuwen, J. A. M. de Boet, V. J. Bloem, and W. J. A. M. Sneijers, "LDMOS ruggedness reliability," *Microw. J*, vol. 5, pp. 96–104, 2009.

[58] Y. Shen, M. Mehrpoo, M. Hashemi, M. Polushkin, L. Zhou, M. Acar, R. van Leuken, M. S. Alavi, and L. de Vreede, "A fully-integrated digital-intensive polar Doherty transmitter," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2017, pp. 196–199.

[59] D. Chowdhury, L. Ye, E. Alon, and A. M. Niknejad, "An efficient mixed-signal 2.4-GHz polar power amplifier in 65-nm CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 8, pp. 1796–1809, 2011.

[60] M. S. Alavi, R. B. Staszewski, L. C. N. de Vreede, A. Visweswaran, and J. R. Long, "All-digital RF I/Q modulator," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 11, pp. 3513–3526, 2012.

[61] A. A. M. Saleh and J. Salz, "Adaptive linearization of power amplifiers in digital radio systems," *The Bell System Technical Journal*, vol. 62, no. 4, pp. 1019–1033, 1983.

[62] M. Hashemi, M. S. Alavi, and L. C. N. de Vreede, "Pushing the linearity limits of a digital polar transmitter," in *2018 13th European Microwave Integrated Circuits Conference (EuMIC)*, 2018, pp. 174–177.

[63] ESDA/JEDEC Joint Standard, ANSI/ESDA/JEDEC JS-001-2014, *For Electrostatic Discharge Sensitivity Testing; Human Body Model (HBM) - Component Level*, Aug. 2014.

[64] ——, ANSI/ESDA/JEDEC JS-002-2014, *For Electrostatic Discharge Sensitivity Testing; Charged Device Model (CDM) - Device Level*, Apr. 2015.

[65] GlobalFoundries, *22FDX-Plus Process Design Kit; ESD Reference guide*, Jun. 2023, V1.0_2.0.

[66] JEDEC EIA, JESD22-A115C, *Electrostatic Discharge (ESD) Sensitivity Testing Machine Model (MM)*, Nov. 2010, inactive as of Sep. 2016.

[67] JEDEC Standard, JESD47L, *Stress-Test-Driven Qualification of Integrated Circuits*, Aug. 2014.

[68] P. Svihra, J. Braach, E. Buschmann, D. Dannheim, K. Dort, T. Fritzsch, H. Kristiansen, M. Rothermund, J. V. Schmidt, M. V. B. Pinto *et al.*, "Development of novel single-die hybridisation processes for small-pitch pixel detectors," *Journal of Instrumentation*, vol. 18, no. 03, p. C03008, 2023.

[69] Ampleon, BLC10G22XS-401AVT, *Power LDMOS transistor - Product data sheet*, May 2022, rev. 1.

[70] J. W. Arblaster, *Selected values of the crystallographic properties of elements.* ASM International, 2018.

[71] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolić, *Digital integrated circuits: a design perspective.* Pearson Education Upper Saddle River, NJ, 2003, vol. 7.

[72] L. G. Salem, J. F. Buckwalter, and P. P. Mercier, "A recursive switched-capacitor house-of-cards power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 7, pp. 1719–1738, 2017.

[73] N. R. Shay, E. Solomon, L. Zohar, A. Ben-Bassat, E. Socher, and O. Degani, "A watt level, 5-7GHz all digital polar TX based on 3.3V switched capacitor digital PA in 16nm fin-FET for Wi-Fi7 applications," in *2024 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2024, pp. 255–258.

[74] D. C. Ribeiro, A. Prata, P. M. Cruz, and N. B. Carvalho, "*D*-Parameters: A novel framework for characterization and behavioral modeling of mixed-signal systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 10, pp. 3277–3287, 2015.

[75] F. van Rijs and S. J. C. H. Theeuwen, "Efficiency improvement of LDMOS transistors for base stations: towards the theoretical limit," in *2006 International Electron Devices Meeting*, 2006, pp. 1–4.

[76] R. M. Fano, "Theoretical limitations on the broadband matching of arbitrary impedances," *Journal of the Franklin Institute*, vol. 249, no. 1, pp. 57–83, 1950.

[77] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 814–826, 2002.

[78] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. McGraw-hill, 2002.

[79] R. J. Bootsman, D. P. N. Mul, Y. Shen, R. M. Heeres, F. van Rijs, M. S. Alavi, and L. C. N. de Vreede, "An 18.5 W fully-digital transmitter with 60.4 % peak system efficiency," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 1113–1116.

[80] R. Bootsman, Y. Shen, D. Mul, M. Rousstia, R. Heeres, F. van Rijs, J. Gajadharsing, M. S. Alavi, and L. C. N. de Vreede, "A 39 W fully digital wideband inverted Doherty transmitter," in *2022 IEEE/MTT-S International Microwave Symposium - IMS 2022*, 2022, pp. 979–982.

[81] M. Hashemi, L. Zhou, Y. Shen, and L. C. N. de Vreede, "A highly linear wideband polar class-E CMOS digital Doherty power amplifier," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 10, pp. 4232–4245, 2019.

[82] E. McCune, "A technical foundation for RF CMOS power amplifiers: Part 5: Making a switch-mode power amplifier," *IEEE Solid-State Circuits Magazine*, vol. 8, no. 3, pp. 57–62, 2016.

[83] V. Diddi, S. Sakata, S. Shinjo, V. Vorapipat, R. Eden, and P. Asbeck, "Broadband digitally-controlled power amplifier based on CMOS / GaN combination," in *2016 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, May 2016, pp. 258–261.

[84] V. Diddi, H. Gheidi, J. Buckwalter, and P. Asbeck, "High-power, high-efficiency digital polar Doherty power amplifier for cellular applications in SOI CMOS," in *2016 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR)*, Jan 2016, pp. 18–20.

[85] W. Yuan and J. S. Walling, "A multiphase switched capacitor power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 5, pp. 1320–1330, May 2017.

[86] D. A. Calvillo-Cortes, M. Acar, M. P. van der Heijden, M. Apostolidou, L. C. N. de Vreede, D. Leenaerts, and J. Sonsky, "A 65nm CMOS pulse-width-controlled driver with $8V_{pp}$ output voltage for switch-mode RF PAs up to 3.6GHz," in *2011 IEEE International Solid-State Circuits Conference*, 2011, pp. 58–60.

[87] Q. Bruinsma, "Wideband and energy efficient digital transmitter," Master's thesis, Delft University of Technology, Delft, The Netherlands, 2020. [Online]. Available: http://resolver.tudelft.nl/uuid:2cb1c1a7-51c3-4368-824c-41921e37be94

[88] S. C. Cripps, *Advanced techniques in RF power amplifier design.* Artech House Norwood, MA, 2002.

[89] M. P. van der Heijden, H. C. de Graaff, L. C. N. de Vreede, J. R. Gajadharsing, and J. N. Burghartz, "Theory and design of an ultra-linear square-law approximated LDMOS power amplifier in class-AB operation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 9, pp. 2176–2184, 2002.

[90] M. P. van der Heijden, "RF amplifier design techniques for linearity and dynamic range," Ph.D. dissertation, Delft University of Technology, 2005.

[91] ETSI Standard, ETSI ES 202 706-1 V1.6.0, *Environmental Engineering (EE); Metrics and measurement method for energy efficiency of wireless access network equipment; Part 1: Power consumption - static measurement method*, Nov. 2021, final draft.

[92] S. Wesemann, "Energy-efficient radio unit design for the next generation of MIMO systems," in *27th International Workshop on Smart Antennas (WSA 2024)*, Mar. 2024, invited talk.

[93] D. P. N. Mul, R. J. Bootsman, M. R. Beikmirza, S. M. Alavi, and L. C. N. de Vreede, "Method of applying an activation scheme to a digitally controlled segmented RF power transmitter," US Patent US20 240 146 346A1, May, 2024, US Patent App. 18/263,896. [Online]. Available: https://patents.google.com/patent/US20240146346A1

[94] R. J. Bootsman, D. P. N. Mul, M. Beikmirza, O. El Boustani, D. Maassen, B. van Velzen, M. Rousstia, R. Koster, J. R. Gajadharsing, T. Fritzsch, Y. Shen, M. S. Alavi, and L. C. N. de Vreede, "A switch-bank approach for high-power, high-resolution, fully-digital transmitters," in *2024 54th European Microwave Conference (EuMC)*, 2024, pp. 23–26.

[95] D. P. N. Mul, R. J. Bootsman, M. Beikmirza, O. El Boustani, Y. Shen, D. Maassen, B. van Velzen, M. Rousstia, R. Koster, J. R. Gajadharsing, T. Fritzsch, M. Alavi, and L. C. N. de Vreede, "5.8 a 20W CMOS/LDMOS all-digital transmitter with dynamic retiming and glitch-free phase mapper, achieving 68%/63% peak drain/system efficiency," in *2025 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 68, 2025, pp. 104–106.

[96] F. Bagdonas, "Wideband digital intensive Doherty concepts," Master's thesis, Delft University of Technology, Delft, The Netherlands, 2021. [Online]. Available: https://resolver.tudelft.nl/uuid:99cd71df-1b11-4878-80d2-05635824104c

[97] M. R. Beikmirza, L. C. N. de Vreede, R. J. Bootsman, D. P. N. Mul, S. M. Alavi, and Y. Shen, "Digital transmitter featuring a 50%-LO signed phase mapper," US Patent US20 240 146 503A1, May, 2024, US Patent App. 18/263,895. [Online]. Available: https://patents.google.com/patent/US20240146503A1

[98] TSMC, *TSMC 45/40 nm CMOS Logic and MS_RF Design Rule (CLN45LP/LPG, CLN40LP/LPG/LP+, CLN40G)*, Mar. 2015, T-N45-CL-DR-001 V2.3.

[99] O. El Boustani, "CMOS drivers for RF-DACs," Master's thesis, Delft University of Technology, Delft, The Netherlands, 2023. [Online]. Available: https://resolver.tudelft.nl/uuid:e7ae220e-3554-4435-8ce6-ec9679f57e45

[100] B. Razavi, R.-H. Yan, and K. F. Lee, "Impact of distributed gate resistance on the performance of mos devices," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 11, pp. 750–754, 1994.

[101] H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, S. Selberherr, Ed.   Springer-Verlag/Wien, 1990.

[102] X4B20L1-5050G, *Xinger®IV Ultra Small Low Profile 0603 Balun 50Ω to 50Ω Balanced*, TTM Technologies, 2022, rev. D.

[103] IEEE International Roadmap for Devices and Systems, *Lithography and Patterning*, Institute of Electrical and Electronics Engineers, 2024.

[104] A. Balasubramaniyan, X. Hui, A. Bellaouar, M. M. Campos, A. Bharadwaj, E. Veeramani, and S. Syed, "A 22FDX® Wi-Fi PA demonstrating a new LDMOS device with 10V breakdown achieving output power of 29.5dBm at 40% PAE," in *2024 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2024, pp. 35–38.

[105] B. Cline, D. Prasad, E. Beyne, and O. Zografos, "Power from below: Buried interconnects will help save Moore's law," *IEEE Spectrum*, vol. 58, no. 9, pp. 46–51, 2021.

[106] H. W. Then, M. Radosavljevic, P. Koirala, N. Thomas, N. Nair, I. Ban, T. Talukdar, P. Nordeen, S. Ghosh, S. Bader, T. Hoff, T. Michaelos, R. Nahm, M. Beumer, N. Desai, P. Wallace, V. Hadagali, H. Vora, A. Oni, X. Weng, K. Joshi, I. Meric, C. Nieva, S. Rami, and P. Fischer, "Advanced scaling of enhancement mode high-k gallium nitride-on-300mm-Si(111) transistor and 3D layer transfer GaN-silicon finfet CMOS integration," in *2021 IEEE International Electron Devices Meeting (IEDM)*, 2021, pp. 11.1.1–11.1.4.

[107] L. C. N. de Vreede, D. P. N. Mul, R. J. Bootsman, M. R. Beikmirza, S. M. Alavi, and T. Ibrahim, "Digitally controlled segmented RF power transmitter," NL Patent NL2 036 038B1, April, 2025, WIPO WO2025080139A1 PCT/NL2024/050563. [Online]. Available: https://patents.google.com/patent/WO2025080139A1

[108] T. Ibrahim, M. R. Beikmirza, M. S. Alavi, and L. C. N. de Vreede, "A quadrature harmonic rejection voltage-domain mixer with 20 dBm OIP3 and 800 MHz IF bandwidth," in *2024 19th European Microwave Integrated Circuits Conference (EuMIC)*, 2024, pp. 443–446.

[109] T. Ibrahim, M. Beikmirza, M. S. Alavi, and L. de Vreede, "A fully passive harmonic rejection quadrature mixer for TX observation with 20 dBm OIP3 and 800 MHz IF bandwidth," *International Journal of Microwave and Wireless Technologies*, p. 1–12, 2025.

[110] Y. Wu, G. D. Singh, M. Beikmirza, L. C. N. de Vreede, M. Alavi, and C. Gao, "OpenDPD: An open-source end-to-end learning and benchmarking framework for wideband power amplifier modeling and digital pre-distortion," in *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2024, pp. 1–5.

[111] Y. Wu, A. Li, M. Beikmirza, G. D. Singh, Q. Chen, L. C. N. de Vreede, M. Alavi, and C. Gao, "MP-DPD: Low-complexity mixed-precision neural networks for energy-efficient digital predistortion of wideband power amplifiers," *IEEE Microwave and Wireless Technology Letters*, vol. 34, no. 6, pp. 817–820, 2024.

[112] Y. Wu, Y. Zhu, K. Qian, Q. Chen, A. Zhu, J. Gajadharsing, L. C. N. de Vreede, and C. Gao, "DeltaDPD: Exploiting dynamic temporal sparsity in recurrent neural networks for energy-efficient wideband digital predistortion," *IEEE Microwave and Wireless Technology Letters*, vol. 35, no. 6, pp. 772–775, 2025.

[113] M. Vigilante, E. McCune, and P. Reynaert, "To EVM or two EVMs?: An answer to the question," *IEEE Solid-State Circuits Magazine*, vol. 9, no. 3, pp. 36–39, 2017.

[114] J. Verspecht and P. Van Esch, "Accurately characterizing hard nonlinear behavior of microwave components with the nonlinear network measurement system: Introducing 'nonlinear scattering functions'," in *Proceedings of the 5th International Workshop on Integrated Nonlinear Microwave and Millimeterwave Circuits*, Oct 1998, pp. 17–26.

[115] J. Verspecht and D. E. Root, "Polyharmonic distortion modeling," *IEEE Microwave Magazine*, vol. 7, no. 3, pp. 44–57, 2006.

[116] C. Xie, T. Zhang, and D. Liu, "Using *X*-parameters to model mixers," in *2012 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, vol. 3, 2012, pp. 1–3.

[117] J. T. Barron, "Squareplus: A softplus-like algebraic rectifier," *CoRR*, vol. abs/2112.11687, 2021. [Online]. Available: https://arxiv.org/abs/2112.11687

[118] A. Lasia, "The origin of the constant phase element," *The Journal of Physical Chemistry Letters*, vol. 13, no. 2, pp. 580–589, 2022.

Acknowledgments

During my PhD journey, I have had the support of a broad range of people to whom I owe many thanks. On the one hand I've met many new people to collaborate with, as well as my family, friends, and colleagues whom I already knew.

First and foremost, I express my gratitude to my promotor, prof. dr. ing. Leo de Vreede, for the opportunity to pursue my PhD degree. Even though the research has spanned almost seven years, I could always count on your support and enthusiasm. Where a typical PhD student would have somewhere between weekly to monthly meetings with their supervisor, I typically could (and still can) encounter you in my office several times a day, which is a testament to your never-ending enthusiasm for digital transmitters. I also thank dr. Morteza Alavi as my (co)promotor for his useful input and feedback on all my (our) publications.

Next, I want to thank my paranymphs Dieuwert Mul and Moritz Fieback. Dieuwert, I am very grateful for our research collaboration. We have worked together closely, first within the DIPLOMAT project and both continued with the DRASTIC project. It may have yielded significant overlap in our research, even though we had a different focus, but I do feel that, by working together, we have produced a result that is greater than the sum of its parts. Moritz, we have shared a significant portion of our academic lives together. For example, we started our PhD's roughly at the same time while being flatmates. We've been up to all kinds of shenanigans, from our Mario Kart adventures or late night chess, to visiting random bars, even if only to shelter from the rain on our way home.

I would like to thank the committee members for reading my dissertation, namely, dr. ir. Fred van Rijs, prof. dr. ir. Willem van Driel, prof. dr. ir. Bart Smolders, prof. dr. ir. Bram Nauta, and prof. dr. Piet Wambaqc. Here I'd like also to mention prof. dr. Earl McCune, who was the first to 'volunteer' to be on my promotion committee, but is unfortunately no longer with us. He always pressed for using clear definitions, to avoid possible ambiguity to the reader, as he used to say: "All communication happens at the receiver." This was one of the reasons for me to attempt to clearly define the transfer of a DTX, which has resulted in a part of Chapter 5 of this dissertation.

Many thanks are due to the Dutch government (through NWO STW and EZ TKI HTSM) and the industry partners who were involved in the projects and contributed to this work, as well as to the other companies that made this work possible. Starting with the project partners, this work was definitely not possible without the support of (many people of) Ampleon. Here, I'd like to thank the contributions of John Gajadharsing, Rob Heeres, Fred van Rijs, Daniel Maassen, Mohadig Rousstia, Bart van Velzen, Ronald Koster, Vittorio Cuoco, Jos Klappe, Michel de Langen, Vincent Gerritsma, Dave Hartskeerl, Nick Pulsford, Sergio Pires, and André Prata. I would also like to thank the people of the Ampleon PTL who assembled the majority of the demonstrators discussed in this dissertation, including Antoine van Dijk, Geert Arts, Marcel Hendriks, Bang Ong, Frans van Elk, and Alex Wijdeveld. I'm also grateful for the support from the people of Nokia Networks and Bell Labs: Wolfgang Templ, Stefan Wesemann, Peter Vetter, Eric Wantiez, Thomas Bohn,

biz Nanbakhsh, Can Akgun, and Jingchu He.

Besides the groups of BE and ELCA, the community of PhD and MSc students within the entirety of the extended microelectronics family is active and strong. This follows from several MEST activities, borrels (Tripel Karmeliet!), bouldering together, conferencing, or simply enjoying having a coffee. For that I'm thankful to Guilherme Medeiros, Pascal 't Hart, Roger Zamparette, Joost van Ginkel, Haji Akhundov, Sven van Berkel, Mojtaba Jahangiri, Martijn Hoogelander, Nina Beschoor-Plug, Daniël Kraak, Troya Köylü, Jun Fen, Simon Verkleij, Armin Šabanović, and Ferry Musters.

Next, I'm grateful for the diverse group of friends in my life with whom I can enjoy going on skiing trips, sailing trips, random borrels, parties, making cities unsafe (both abroad and domestic), going to movies, dinners, beer tastings, wine tastings, helping each other move, housewarmings, online games, concerts, band evenings, and so on. The first group that I explicitly like to mention is mostly defined by our yearly skiing trip, the tendency to just order the plat du jour six times independently, and to check how everybody is doing (maar dan echt): Niels, Joseph, Koen, Swier, and Tom. Thanks boys, a year is not complete without our annual snowy mountainous holiday. Since seeing each other just one holiday a year is not enough, more excuses can be thought of, such as sailing weekends and BBQs, for which I also thank Jennie. The next group started during the COVID times, which is unmistakably the VriAvBo crew: Stefanie, Alex, Julia, and Reinier. Thank you for, well, the VriAvBo's, as well as any other AvBo and the weekends away. It never ceases to amaze me how it is near-impossible to plan a drink in three months' time, but when asked who's available the day after tomorrow: sure, no problem. For the majority of my student life, I have lived in the Lizstflat, where I had the honor of sharing the apartment, in chronological order, with Lieuwe, Moritz, Reinier, and Koen. Thanks for a good time and a place I could call home all those years. Trying to mention more groups would only result in an utter mess, and if I'm to avoid spending the next several pages mentioning examples, I should speed this up. So, in alphabetical order, Bart, Ben, Carlijn, Conchita, Connor, Daniel, Danny, Dennis, Derk-Jan, Dominic, Eray, Erik, Erwin, Eva, Gabriele, Guido, Jan, Janneke, Jasper, Jasper, Jeroen, Joost, Kevin, Leon, Lotte, Luc, Ludo, Maaike, Maima, Marc, Marcella, Marije, Martijn, Martin, Matthijs, Menno, Nienke, Pascal, Paul, Ralph, Reji, Richelle, Rik, Roy, Sam, Sander, Sanne, Sjors, Thomas, Tim, Timon, Tworit, you know who you are, thanks for having been a part of this journey of mine, and for whatever is to come.

This last section is for the people I can count on unconditionally. That undoubtedly includes my parents, Annelies and Peter. You have always shown interest in what I was up to (even if I thought of myself as being too busy to respond, sorry!), which usually fell into one of two categories: designing a chip, or a "vogelhuisje" for that chip. Thank you for believing in me also the years prior to starting a PhD. And also thanks to my brother Chris, for the occasional nice talks. We don't see each other often, but when we do, it's always nice to have you around. Last but foremost, my love Merel. Thank you for my 'cultural rehab' after the COVID-19 lockdowns, by going to museums and the movies with me. I am very grateful for your support and understanding when times were stressful for me, and, in general, also for providing relaxation through board games or simply having a beer together on the couch. May our future be cute; I'm excited to see what our future together may hold.

Rob
Delft, August 2025

Curriculum Vitæ

Robert Jan Bootsman

15-07-1992	Born in Nieuw-Vennep, The Netherlands
2004–2010	Pre-University Education (VWO), NT profile (Nature en Technology) Haarlemmermeerlyceum, Hoofddorp, The Netherlands
2010–2014	B.Sc. degree in Electrical Engineering Delft University of Technology, Delft, The Netherlands
2014–2018	M.Sc. degree in Electrical Engineering, track Micro-Electronics Delft University of Technology, Delft, The Netherlands

2014–2018   M.Sc. degree in Electrical Engineering, track Micro-Electronics
Delft University of Technology, Delft, The Netherlands
Thesis: Power RFDAC:
The Design of a LDMOS Class-E SMPA DRAC with a CMOS Driver
Supervisor: Prof. dr. ing. L.C.N. de Vreede

2018   Researcher
Delft University of Technology, Delft, The Netherlands

2018–2025   PhD degree in Electrical Engineering
Delft University of Technology, Delft, The Netherlands
Dissertation: High-Power Digital Transmitters for Wireless Networks
Promotor: Prof. dr. ing. L.C.N. de Vreede
Copromotor: Dr. S.M. Alavi

List of Publications

Journal Papers

4. D.P.N. Mul, **R.J. Bootsman**, M. Beikmirza, M.S. Alavi, and L.C.N. de Vreede, "The Efficiency and Power Utilization of Current-Scaling Digital Transmitters," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 72, no. 7, pp. 4350–4366, July 2024, doi: 10.1109/TMTT.2023.3336984.

3. Y. Shen, M. Hoogelander, **R.J. Bootsman**, M.S. Alavi, and L.C.N. de Vreede, "A Wideband Digital-Intensive Current-Mode Transmitter Line-Up," in *IEEE Journal of Solid-State Circuits*, vol. 58, no. 9, pp. 2489–2500, Sept. 2023, doi: 10.1109/JSSC.2023.3279235.

2. **R.J. Bootsman**, D.P.N. Mul, Y. Shen, M. Hashemi, R.M. Heeres, F. van Rijs, M.S. Alavi, and L.C.N. de Vreede, "High-Power Digital Transmitters for Wireless Infrastructure Applications (A Feasibility Study)," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 5, pp. 2835–2850, May 2022, doi: 10.1109/TMTT.2022.3153000.
   ✉ This article is featured in IEEE MTT-S Monthy Newsletter, April 2022 Issue (`https://conta.cc/3JbV2K7` or `https://mtt.org/news/e-newsletter-archive/`).

1. Y. Shen, **R.J. Bootsman**, M.S. Alavi, and L.C.N. de Vreede, "A Wideband IQ-Mapping Direct-Digital RF Modulator for 5G Transmitters," in *IEEE Journal of Solid-State Circuits*, vol. 57, no. 5, pp. 1446–1456, May 2022, doi: 10.1109/JSSC.2022.3144362.

Conference Papers

7. D.P.N. Mul*, **R.J. Bootsman***, M. Beikmirza, D. Maassen, B. van Velzen, M. Rousstia, R. Koster, J. Klappe, J.R. Gajadharsing, T. Fritzsch, Y. Shen, M.S. Alavi, and L.C.N. de Vreede, "A 20 W CMOS/LDMOS All-Digital Transmitter with Dynamic Retiming and Glitch-Free Phase Mapper, Achieving 68/63 % Peak Drain/System Efficiency," *2025 IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, USA, 2025, pp. 104–106, doi: 10.1109/ISSCC49661.2025.10904650.

6. **R.J. Bootsman***, D.P.N. Mul*, M. Beikmirza, O. El Boustani, D. Maassen, B. van Velzen, M. Rousstia, R. Koster, J.R. Gajadharsing, T. Fritzsch, Y. Shen, M.S. Alavi, and L.C.N. de Vreede, "A Switch-Bank Approach for High-Power, High-Resolution, Fully-Digital Transmitters," *2024 54th European Microwave Conference (EuMC)*, Paris, France, 2024, pp. 23–26, doi: 10.23919/EuMC61614.2024.10732131.
   🏆 This paper was awarded with the EuMC Microwave Prize for the best paper.

5. **R.J. Bootsman**, Y. Shen, D.P.N. Mul, M. Rousstia, R.M. Heeres, F. van Rijs, J.R. Gajadharsing, M.S. Alavi, and L.C.N. de Vreede, "A 39 W Fully Digital Wideband Inverted Doherty Transmitter," *2022 IEEE/MTT-S International Microwave Symposium (IMS)*, Denver, CO, USA, 2022, pp. 979–982, doi: 10.1109/IMS37962.2022.9865405.

4. D.P.N. Mul, **R.J. Bootsman**, Q. Bruinsma, Y. Shen, S. Krause, R. Quay, M.J. Pelk, F. van Rijs, R.M. Heeres, S. Pires, M.S. Alavi, and L.C.N. de Vreede, "Efficiency and Linearity of Digital

'Class-C Like' Transmitters," *2020 50th European Microwave Conference (EuMC)*, 2021, pp. 1–4, doi: 10.23919/EuMC48046.2021.9338122.

★ This paper has been nominated in the short-list of (best 25) papers for the EuMC-2020 conference best paper awards.

3. Y. Shen, **R.J. Bootsman**, M.S. Alavi, and L.C.N. de Vreede, "A 0.5–3 GHz I/Q Interleaved Direct-Digital RF Modulator with up to 320 MHz Modulation Bandwidth in 40 nm CMOS," *2020 IEEE Custom Integrated Circuits Conference (CICC)*, 2020, pp. 1–4, doi: 10.1109/CICC48029.2020.9075949.

2. **R.J. Bootsman**, D.P.N. Mul, Y. Shen, R.M. Heeres, F. van Rijs, M.S. Alavi, and L.C.N. de Vreede, "An 18.5 W Fully-Digital Transmitter with 60.4 % Peak System Efficiency," *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 1113–1116, doi: 10.1109/IMS30576.2020.9223942.

1. Y. Shen, **R.J. Bootsman**, M.S. Alavi, and L.C.N. de Vreede, "A 1–3 GHz I/Q Interleaved Direct-Digital RF Modulator as a Driver for a Common-Gate PA in 40 nm CMOS," *2020 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2020, pp. 287–290, doi: 10.1109/RFIC49505.2020.9218324.

*Equally-Credited Authors

Patents

1. L.C.N. de Vreede, M.S. Alavi, **R.J. Bootsman**, M. Beikmirza, D.P.N. Mul, R.M. Heeres, and F. van Rijs, "Digital transmitter with high power output," Pub. No. WO/2021/162545 / NL2024903 / EP4104290 / CN115136491 / PCT/NL2021/050081 / US12294360, Filed 05.02.2021, Published 19.08.2021.

2. D.P.N. Mul, **R.J. Bootsman**, M. Beikmirza, M.S. Alavi, and L.C.N. de Vreede, "Method of applying an activation scheme to a digitally controlled segmented RF power transmitter," Pub. No. WO/2022/169362 / NL2027510 / US20240146346, Filed 05.02.2021, Published 06.09.2022.

3. M. Beikmirza, L.C.N. de Vreede, **R.J. Bootsman**, D.P.N. Mul, M.S. Alavi, and Y. Shen, "Digital Transmitter Featuring a 50 %-LO Signed Phase Mapper," Pub. No. WO/2022/169361 / NL2027509 / US20240146503, Filed 05.02.2021, Published 06.09.2022.

4. L.C.N. de Vreede, M.S. Alavi, D.P.N. Mul, M. Beikmirza, **R.J. Bootsman**, and T. Ibrahim, "Digitally controlled segmented RF power transmitter," Pub. No. WO/2025/080139 / NL2036038, Filed 13.10.2023, Published 30.04.2025.