

UC Davis

UC Davis Electronic Theses and Dissertations

Title

A Millimeter-wave Molecular Clock in Silicon

Permalink

<https://escholarship.org/uc/item/79k1r1k8>

Author

Chen, Jingjun

Publication Date

2022

Peer reviewed|Thesis/dissertation

A Millimeter-wave Molecular Clock in Silicon

By

JINGJUN CHEN

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical and Computer Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Neville C. Luhmann Jr., Chair

Anh-Vu Pham

Xiaoguang Liu

Committee in Charge

2022

Copyright © 2022 by

Jingjun Chen

All rights reserved.

To my mother.

CONTENTS

List of Figures	vi
List of Tables	xii
Abstract	xiii
Acknowledgments	xiv
1 Molecular Clock Background	1
1.1 Clock Introduction	1
1.1.1 Precision Clock Applications	3
1.1.2 Existing Solutions	8
1.2 Molecular Rotational Resonance	12
1.2.1 The Rigid Rotor	12
1.2.2 Spectral Line Width	15
1.2.3 Allan Deviation	16
2 System Analysis of the Molecular Clock	18
2.1 Introduction	18
2.2 Dispersion Curve and Baseline	21
2.2.1 Dispersion Curve	21
2.2.2 Baseline Tilting	24
2.2.3 FSK Modulation	27
2.3 Transmitter	31
2.3.1 PLL Introduction	31
2.3.2 TX specifications	35
2.3.3 Low phase noise TX	39
2.4 Detector	41
2.5 Lock-in Amplifier	43
2.5.1 Analog mixers	44
2.5.2 Digital mixers	48

2.5.3	SNR	49
3	Design of a High Power Millimeter Wave Source	52
3.1	Introduction	52
3.2	Theory of Stacked-FET Oscillator	54
3.2.1	Scaling of Passives	54
3.2.2	Small-signal Analysis of Pi-Embedded Oscillators	57
3.2.3	Interpretation of the Small-signal Analysis	61
3.2.4	Stacked-FET	65
3.2.5	Discussion of the limitation of linear models	68
3.3	Implementation	69
3.3.1	Process and Transistor	69
3.3.2	Oscillator	69
3.3.3	Combining	72
3.3.4	Layout and Passives	73
3.4	Measurement of the Stacked-FET Oscillator	76
3.4.1	Calibration	77
3.4.2	Measured Power and Efficiency	79
4	Molecular Clock Implementation and Experiment	82
4.1	Implementation of a 120 GHz molecular clock system	84
4.1.1	Top-level block diagram	84
4.1.2	The LO generation	86
4.1.3	The Receiver	90
4.1.4	The Lock-in	93
4.2	A Low Phase Noise Transmitter	97
4.2.1	The Transmitter	97
4.2.2	The LO Generation and the Reference	100
4.2.3	The DDS	105
4.3	Mechanical Setup	105

4.4	Measurements	109
4.4.1	Measured Absorption Profile	109
4.4.2	Phase Noise	110
4.4.3	Stability	115
5	Summary	123
5.1	Future Work	123

LIST OF FIGURES

1.1	The simulated absorption profile of OCS gas at ~ 121.6 GHz at various pressures.	15
2.1	Two types of gas cells. (a) The cylindrical gas cell with horn antenna probing [53]. (b) The meandering waveguide gas cell where the EM wave travels a distance much longer than L [50,53].	19
2.2	Simplified system architecture including a molecular gas cell, a transmitter which is a mm-Wave PLL referenced to a low frequency VCXO, a receiver that demodulates AM signal and a feedback control circuitry that consists of a modulation generator, a lock-in amplifier and an integrator.	20
2.3	Envelope of the received signal from an FM modulated transmitter with various carrier frequencies: $f_c < f_0$, $f_c = f_0$ and $f_c > f_0$	23
2.4	(a-c) First, third and fifth order derivative of Lorentzian profile with FWHM of 1 MHz and peak absorption of 1%. Simulated odd harmonic dispersion curves with a modulation depth of 0.1·FWHM (d-f) and a modulation depth equal to FWHM (h-j), respectively.	24
2.5	(a) The measured transmission profile of OCS gas around the resonance frequency of about 145.947 GHz, where a linearly tilted baseline is observed (a). The measured transmission profile may be decomposed into a symmetric Lorentzian profile (b) plus a linear baseline (c).	25
2.6	(a) Measured transmission profiles at around 109.463 GHz for various antenna positions. (b) Measurement setup using a Keysight PNA-X network analyzer and a pair of VDI mm-Wave TRx heads.	26
2.7	Envelope of the received signal from an FSK modulated transmitter with various carrier frequencies: $f_c < f_0$, $f_c = f_0$ and $f_c > f_0$	28

2.8	(a) Frequency tones defined by Eqn. 2.12 on the transmission profile with a tilted baseline. When f_c equal to f_0 , the error difference from main FSK and auxiliary FSK comes from the linear baseline and is proportional to $\Delta f_2 - \Delta f_1$. (b) Timing diagrams of the dual FSK scheme for instantaneous frequency, detected envelope and two reference signals Ref_1 and Ref_2 . (c) Block diagram of the dual FSK lock-in.	30
2.9	(a) PLL block diagram and (b) Circuit for phase frequency detector. (c) The transfer function of the PFD combined with a charge pump. (d) First order low pass filter, and (e) second order low pass filter with an additional zero.	32
2.10	Transfer function of the PLL when it is in lock. The model also includes major noise sources, including reference phase noise, charge pump current noise, low pass filter noise and VCO phase noise, which are denoted as $\phi_{n,ref}$, $I_{n,pd}$, $V_{n,lpf}$ and $\phi_{n,vco}$, respectively.	34
2.11	Three locations in a PLL where modulation could be applied: the modulated reference, modulated frequency tuning word and direct injection at the VCO input	39
2.12	Comparison of phase noise between a multiplier based (a) and a PLL based (b) signal source.	40
2.13	(a) A non-coherent envelope detector. (b) Homodyne envelope detector. (c) Heterodyne detector. Filters are not shown for simplicity.	42
2.14	The generic block diagram of a lock-in amplifier consists of a pair of mixers followed by narrow bandwidth low-pass filters. The in-phase and quadrature outputs are further processed to obtain the magnitude and phase of the input signal.	43
2.15	(a) A simple active mixer and (b) passive mixer.	45
2.16	Illustration of the current noise in an active mixer when one of the LO switch M_2 is turned off.	46

2.17	Simplified block diagram of lock-in amplifiers when the input and reference are in phase, with (a) sine wave input and (b) square wave input	49
3.1	Power and inductance scaling of a Pi-embedding oscillator	55
3.2	Simulated Q versus L for various inductor geometries (a) without (b) with via stack.	55
3.3	Proposed scaling consisting of parallel and counteractive scaling	56
3.4	The small signal model of the transistor (a) and a generic Pi-embedding network (b)	57
3.5	Normalized inductance \tilde{L} by doubling $g_m r_o$ (3.19) versus $g_m r_o$ for various C_{gd}	64
3.6	Normalized inductance \tilde{L} by halving C_{gd} versus C_{gd0} for various $g_m r_{o0}$	65
3.7	Schematic and simplified two-port representation of a two-stack amplifier	67
3.8	Simulated and calculated effective C_{gd} of various number of stacked transistors	67
3.9	(a)Metal cross-sectional profile. (b)Layout of a 41.6 μ m transistor. (c)3-D portrait of the transistor.	70
3.10	Schematic of the stacked-FET oscillator core.	71
3.11	(a)2-way combined oscillator with capacitive matching circuit, (b) common-mode half circuit (c) differential-mode half circuit	72
3.12	Complete schematic of the proposed stacked-FET oscillator	74
3.13	(a) layout sketch of the output combiner including capacitive matching network and (b) simplified schematic of alternative output combining.	75
3.14	(a) 3-D profile and (b) simulated performance of the proposed integrated inductor	77
3.15	Simulated voltage waveform of drain and gate-to-drain swing	78
3.16	Measurement setup for (a) frequency and (b) power.	79
3.17	Calibration procedure of the total loss from s-bend and RF probe.	79
3.18	Die photography of the fabricated chip.	80
3.19	Measured output power and dc-to-RF efficiency of the proposed oscillator.	81

4.1	(a) Block diagram of the millimeter wave transceiver TRX_120_001 from Silicon radar. (b) The chip photograph [71].	83
4.2	The simplified block diagram of the molecular clock.	84
4.3	The block diagram of the transmitter and LO generation for the receiver. The PLL LTC6947 is programmed in the integer mode that multiplies the reference by 32. (a) An initial design where the PLL only controls the LSB of the VCO tuning port, while the higher bits are connected to a potentiometer. (b) All 4 bits from the VCO tuning port are connected together and controlled by the PLL. (3) The VCO tuning characteristic of TRX_120_001 reproduced from the datasheet [71].The frequency labels in the figure are only approximations.	87
4.4	Block diagram for fractional frequency synthesis using AD9913.	89
4.5	Conceptual illustration of intermodulation generation from the unfiltered DDS output.	90
4.6	Receiver IF amplification and envelope detection.	91
4.7	The setup of DAQ system with the circuit boards and LabVIEW. The transmitter and gas cell are not shown.	93
4.8	The analog waveform of the envelope and the reference, each sampled by an ADC with a sample rate four times the analog fundamental frequency. (a) ADC timing where square wave transitions are avoided. (b) ADC timing where two samples are located at the rising and falling edge of the envelope. (c) The resulting lock-in output for an unsynchronized ADC clock over time, when the analog input is constant.	94
4.9	(a) The top-level architecture for the LabVIEW program, which is a classical producer-consumer model. (b) Detailed block diagram for the lock-in and integration operation.	96
4.10	(a) System block diagram of LO generation of the improved transceiver. (b) The design considerations of the active low pass filter. (c) The interface between the comparator LTC6957-1 and the PFD HMC3716.	98

4.11	(a) A multiplier chain from 100 MHz VCXO to 1 GHz. The 100 MHz input is first multiplied by 2 and then by 5. Amplifiers are inserted between multiplication stages to boost the signal up, and bandpass filters remove the spurious tones. (b) The divider chain from 1 GHz to obtain the 10 MHz signal for stability measurement and the 100 kHz ADC sampling clock.	101
4.12	Simplified block diagram of the low phase noise transceiver and the illustration of LO leakage induced intermodulation.	103
4.13	LO buffer for high reverse isolation.	104
4.14	Two iterations of the mechanical setup: (a) an early attempt: System 1 and (b) a modified version: System 2.	105
4.15	3-D view of the acrylic fixtures: (a) gas tube holder, (b) front view of the transceiver PCB holder and (c) back view of the transceiver PCB holder.	106
4.16	Illustration of the improved setup, where the transceiver and the lenses are mounted on translation stages through metal fixtures. The transmitter and receiver move in both x and y directions, and the lenses are adjustable in the x direction.	107
4.17	Measured gas absorption profile at about 121.6 GHz with (a) synchronized Tx and Rx LO and (b) fixed Rx LO frequency. (c) The test bench for measuring the gas absorption with System 2. (d) Tx and Rx frequency chirps when they are aligned results in a fixed IF frequency. (e) Time varying IF frequency with a Rx frequency. (f) and (g) are the associated IF spectrum of (d) and (e), respectively.	108
4.18	(a) Microsemi 3120A phase noise probe. (b) Architecture of the 3120A, where each input channel is power-split before sampled by two identical analog-to-digital converters [78]. (c) A typical setup for measuring phase noise and stability. A Rubidium atomic clock PRS10 from thinkSRS serves as the reference to the 3120A.	110

4.19	(a) Frequency extension of the 3120A using two uncorrelated LOs, leveraging the cross-correlation technique of the instrument. (b) Two identical DUTs with one acting the RF and the other as the LO.	112
4.20	Phase noise measurement of the 120 GHz signal. (a) The setup of System 1 where the frequency synthesizer is a simple PLL. (b) The setup of System 2, which differs from System 1 by the low phase noise synthesizer.	114
4.21	Measured phase noise of the 120 GHz transceiver. Actual phase noise is 3-dB lower for equal contribution of the oscillator in the Tx and the Rx.	115
4.22	(a) Raw measured frequency and temperature data in 4 hours. (b) Measured frequency deviation from 10 MHz with temperature drift cancelled. (c) Measured Allan deviation with temperature drift cancelled.	116
4.23	(a) The bench setup for measuring temperature dependence, where the gas tube is enclosed in a styrofoam box. (b) The inside of the styrofoam box where the heater is placed beneath the glass tube. (c) The heater circuit. (d) The temperature inside the styrofoam box and frequency drift in one period of the power cycle of 7000s. The measurement was conducted for 15 hours with multiple power cycles.	118
4.24	Scatter plot of the measured temperature drift with respect to temperature. Each dot is the measured frequency and temperature at each second. The dashed line indicates the most probable frequency-temperature relationship.	119
4.25	(a) Measured clock frequency and received IF power as a function of the distance between the Tx and the lens. (b) Measured clock frequency and received IF power as a function of the distance between the Rx and the lens.	120
4.26	A measurement of the frequency and temperature of the molecular clock, where the positions of the transceivers were adjusted arbitrarily.	121

LIST OF TABLES

3.1	Comparison between stack-FET and common-source amplifier	70
4.1	A list of devices used in the 1 GHz frequency multiplier chain.	102

ABSTRACT

A Millimeter-wave Molecular Clock in Silicon

An atomic clock is a type of precision timekeeping device that achieves superior stability by referencing its frequency to the transition between energy states of atoms, which are the same everywhere and do not change over time. Relentless effort has been devoted to atomic clocks since their advent in the early 1950s. The most accurate clock today achieves an uncertainty on the order of 10^{-18} based on optical transitions, over 7 orders of magnitude lower than the first generation atomic clock. A number of applications, such as scientific experiments, navigation and communications, have thus benefited from the development of precision atomic clocks.

While advanced atomic clocks meet or even surpass the frequency stability requirement for most applications, they are often bulky, power-hungry and expensive. Few solutions exist that combine the atomic grade accuracy with the size, power and cost of quartz oscillators. These portable precision clocks find application in, for example, seismic data acquisition and underwater navigation. Atomic clocks based on gaseous molecular rotational resonance, which typically falls into the millimeter-wave spectrum, are also known as molecular clocks. They have the potential to fill the gap in highly stable and low-cost frequency standards, due to advancements in the silicon process and circuit design at mm-Wave frequencies. Locking to the rotational resonance in molecular clocks could be achieved exclusively with a set of mm-Wave transceivers. Without the need for microwave cavities, lasers or discharge lamps used in other types of atomic clocks, the size, cost and power of molecular clocks are expected to be lower.

In this work, a molecular clock based on the $10 \leftarrow 9$ transition of the carbonyl sulfide gas is implemented. Techniques to remove the linear baseline in the absorption profile and to reduce the transmitter phase noise are presented. The sources that affect the frequency stability are discussed as well. In addition, a high-power and high-efficiency millimeter-wave oscillator has been designed in CMOS, to address the signal generation difficulty in millimeter-wave transmitters.

ACKNOWLEDGMENTS

I am truly grateful for the guidance and support from Professor Liu and Luhmann that lead to the completion of my graduate study. Prof. Liu was my major professor while he was with UC Davis, who introduced me to the field of RF engineering. I appreciate Prof. Liu's research taste, which allowed me to explore the field in a way that is driven by intellectual curiosities. I was fortunate to have a mentor like him at the beginning of my career. After Prof. Liu left, Prof. Luhmann took the role of the major professor, who provided tremendous resources during my last year. I would like to express special thanks to Dr. Calvin Domier, who is a staff with Prof. Luhmann, shared a lot of insightful opinions to my research. I also want to thank another group staff, Lynette Lombardo, for helping communicate with the department.

I would like to recognize the assistance of members from Prof. Liu's group; it would be impossible for me to complete my work without their help. Hao Wang shared with me hands-on experience of IC design and simulation when I first joined the group; Li Zhang was the one who involved in virtually all my designs, and it was a pleasure to be his roommate for the last three years; James Do taught me a variety of techniques for mm-Wave measurements; Xiaonan Jiang and I had lots of discussions over my molecular clock design, many of which were very inspiring; Xiaohu Wu, our postdoc, is an expert in electromagnetic, who gave me valuable advices for designing passive circuits; Zhigang Peng shared his ideas in oscillator designs. In addition, I would like to thank other group members for collaboration: Daniel Kuzmenko, Mahmoud Nafe, Naimul Hasan, Hind Reggad and Saleh Hasanzade Yamchi.

A number of people from other groups helped me with my research. When I started working on IC design, Yu Ye from Prof. Jane Gu, were especially supportive to our group, who shared both design know-hows and measurement techniques. When I was working on the molecular clock, Hai Yu, Xuan Ding and I had numerous technical discussions, which led to deeper understanding of my system and significant improvements.

Apart from technical support from advisors and peers, I received genuine help from my friends, most notably, Shasha Qiu, who enlightened and encouraged me when I was

feeling down. I enjoyed hiking and BBQ with Ying Chen, Xianzi Liu, Yuan Zheng and other friends mentioned previously. I thank my former and current roommates, Zhiyuan Zhang, Mengzhe Guo and Minyue Chen for all the fun stories we had.

Last but not least, it was possible for me to study abroad because of the vision and financial support of my mother. Words do not express how I much love her, and I owe everything to this great woman.

Chapter 1

Molecular Clock Background

1.1 Clock Introduction

The history of timekeeping devices dates back thousands of years. Developed by the ancient Egyptians, the sundial is one of the earliest types of time keeping devices, which allows people to keep local time from the position of the sun [1]. Since then and until today, relentless effort has been put into the innovation and improvement of clocks. Early designs of clocks that are capable of measuring time at night rely on the continuous flow of a medium [2]. Water clocks, hourglasses, etc., share the similar idea to keep time by measuring a regulated flow into or out of the basin. Modified water clocks that repeat and totalize over endless cycles found extensive use until the late 13th century, when the verge escapement mechanism was invented [3]. One famous application of such mechanism was the clock designed by Henry De Vick in 1360 [4]. It was a revolutionary design, where time, for the first time, was derived from the rate of controlled oscillation. However, the dependence of the oscillation period upon the applied force and friction of the mechanical parts ultimately limits the accuracy.

Three hundred years later, a new type of mechanism was applied to the clock: the resonance [2]. Despite the resonant element used, the same concept is behind all precision clocks today. The pioneer of a resonance type clock is the pendulum clock invented by Christiaan Huygens in 1656 [5]. The resonance frequency of a pendulum is dependent on local acceleration of gravity g and length of the rod L connecting the weight to the pivot.

Under small perturbations, the period of a pendulum is given by

$$T \approx 2\pi \sqrt{\frac{L}{g}} \quad (1.1)$$

Pendulum clocks suffered from temperature induced variation in the rod length. Various techniques were introduced later to compensate the material thermal expansion. For example, the first device that compensates the temperature was invented by George Graham in 1726, in which the rod is made of a container filled with mercury. The thermal expansion of the mercury compensates the thermal extension of the rod, making the effective rod length relatively constant [6]. The Shortt–Synchronome free pendulum, invented in 1921 [7], was the most accurate pendulum at the time. It achieved a quality factor of about 100,000 by reducing the friction of the sustaining mechanism to the pendulum itself, on top of many other careful precautions [2].

The trouble of designing a precision pendulum clock, which is complex and bulky, was soon easily addressed by a quartz crystal clock. The physical and chemical stability and low elastic hysteresis of quartz crystals are the crucial properties for a stable oscillator. The loss of the quartz crystal is so low that in practice, it is the mounting fixture and surrounding atmosphere that dissipates the most energy [2]. The accuracy of the quartz oscillators reached 1×10^{-10} by the early 1950s [8]. However, similar to pendular, the frequency of a quartz oscillator depends on the physical dimensions of the material. Consequently, no two quartz oscillators tick at the same frequency. The best quartz crystal oscillators so far are built with a temperature controlled oven (OCXO) to isolate the ambient temperature variations. It is possible to achieve a temperature coefficient less than 50 parts-per-trillion (ppt) in today's state of the art OCXO [9]. However, the performance of the OCXO is eventually limited by so-called aging, which is the gradual frequency drift over time, usually caused by the packaging. The quartz crystal oscillators served as the frequency standard between the 1920s and 1950s, before the advent of a more accurate type of clock, the atomic clock [10]. Nevertheless, the crystal oscillators are sufficiently accurate in most of the consumer applications today. Virtually all modern microcontrollers, RF and microwave synthesizers are clocked by the quartz crystal.

Atomic clocks are by far the most accurate type of clock [11–13]. The fundamental principle, that the frequency is derived from some kind of resonance, remains the same as centuries ago. The resonator in the popular Cesium or Rubidium atomic clocks arises from the emission or absorption of photons across two hyperfine ground state energy levels of alkali atoms. The emitted or absorbed photon frequency is linked to the energy level by the Planck relation

$$\Delta E = h\nu \tag{1.2}$$

Because the atoms are identical wherever they are, the energy states must be stable over time in theory. The idea of referencing frequency to atomic energy states completely eliminates the dependence of the mechanical property of materials, which was the main source of uncertainty in quartz clocks. Ideally, we wish the absorption or emission to occur exactly at the frequency from two energy states. In reality, atoms absorb or emit energy at the surrounding of the resonance frequency f_0 . The spreading of the frequency is usually characterized by the FWHM, a width at which absorption or transmission is reduced by one half. A number of effects contribute to broadening the line width, including finite radiative lifetime involved in the transition, collisions between atoms, thermal Doppler shift, relativistic effects [14] and also noise from electronic measurement devices. Fortunately, advancements in science and engineering have eliminated most of the obstacles of frequency instability and turned the atomic clock into the most accurate timekeeping device ever.

1.1.1 Precision Clock Applications

Development of precision timekeeping techniques benefits a variety of applications. It changes the definition of second over the past century, which evolved from fraction of mean solar day to ephemeris year and finally to atomic hyperfine transitions. Accurate measurement of time enables precision science experiments, including exploring fundamental laws of physics [15, 16], astronomy [17, 18], etc. Global position system (GPS) provides another great example: the high accurate navigation services are only possible thanks to the precision atomic clocks onboard the satellites. Underwater navigation and synchronization of submerged sensor networks pose new challenges to precision clocks. In

these ocean explorations, the need is ever-increasing for precision clocks that are compact and power efficient for use in battery powered devices [19].

1.1.1.1 Time Standard

For decades, atomic clocks have been the technique behind the most accurate frequency standards. The first Cesium-beam atomic clock was built in 1955, in the National Physical Laboratory in England [20], only 10 years after the proposal by Isidor Rabi, that a clock could be made from atomic beam magnetic resonance. Successful development of Cesium clocks led to the new definition of the second. The precision of the atomic clocks provides a much more stable timekeeping, replacing the tropical year based approach adopted during the 1950s. In 1967, at the 13th Conférence Générale des Poids et Mesures, the second was redefined as the “duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the Cesium 133 atom” [21]. Several addenda have been subsequently approved, yet the body of the definition remains unchanged until now. The current primary time standard by which global time is regulated is called Coordinated Universal Time or UTC. UTC is based on the International Atomic Time (TAI), which is a continuous timescale from the weighted average of an ensemble of over 400 atomic clocks around the world [22]. The main contributors to the ensemble are the most accurate clocks available, including ten Cs fountains, one Rb fountain, one Sr optical lattice and two Yb optical lattice clocks, as well as two legacy Cs beams operated by the PTB, as of November 2021 [23]. While TAI strives to pursue the accuracy of SI definition of time, it must be adjusted regularly to accommodate the Universal Time (UT) for use as a universal time standard. The troublesome aspect simply arises from the irregular and gradual slower rotation of Earth that defines UT. UTC solves this problem, somewhat inelegantly, by applying a leap second from TAI regularly to keep a maximum 1-second difference to UT.

1.1.1.2 Physical Experiment

The twin paradox, a famous thought (gedanken) experiment, postulates that one of the twin travels at high speed ages less when he returns to Earth compared to the twin who stays on Earth all the time. It was well-explained from Einstein’s special relativity and

experimentally verified [24]. In 1971, an experiment that resembles the twin paradox was carried out by Joseph C. Hafele, a physicist, and Richard E. Keating, an astronomer. Four Cesium beam clocks took a flight around the world twice, first eastbound and then westbound. The flying clocks were compared to the clocks which remained at the U.S. Naval Observatory when they returned. It turned out that “the flying clocks lost 59 ± 10 nanoseconds during the eastward trip and gained 273 ± 7 nanoseconds during the westward trip” [25]. The result is not surprising at all, because it is no more than another proof of the already proven special relativity theory. What is really fascinating about this experiment is how the precision of the atomic clock is able to distinguish such small differences.

In recent years, development of atomic clocks enables measurement of relativity theory in the microscale. Einstein’s general relativity predicts that the clock ticks slower in a gravitation field, a phenomenon known as gravitational redshift. Prior experiments verified the theory in a much larger scale using Hydrogen masers onboard a satellite [26]. Although the theoretical frequency change due to vertical elevation is as low as $1.09 \times 10^{-19}/\text{mm}$, improved fractional frequency measurement uncertainty of 7.6×10^{-21} permits the resolution into sub millimeter regime [27].

1.1.1.3 Astronomy

In 2019, the first image of a black hole was created by a globally collaborated telescope array - Event Horizon Telescope (EHT) [28]. It was made possible by a powerful space-geodetic measurement tool, known as very-long-baseline interferometry (VLBI). The Rayleigh criterion suggests that the resolution of a diffraction-limited imaging system (which is the case for radio telescopes) is inversely proportional to the aperture. The VLBI synchronizes multiple telescopes around the globe and records the signal from a distant radio source simultaneously. The angular resolution is reduced to λ/L , where λ is the observation wavelength and L is the maximum projected baseline length of two telescopes. Effectively, the VLBI creates a virtual telescope with the diameter of the Earth. One of the keys to successful operation is to maintain coherence of all the telescopes during observation. Early VLBI attempts were hindered due to the limited performance of the

Rubidium frequency standard [18]. Hydrogen masers were adopted in the EHT program, where the high stability of the frequency standard ensures the coherence across the entire array during the integration window [29].

1.1.1.4 Ocean-Bottom Seismic

Ocean-bottom seismic (OBS) is a marine seismic acquisition technique that provides geological information of the upper crust using earthquake waves [30]. Researchers were able to better estimate the water cycle at subduction zones by the images obtained from OBS data [31]. Seismic reservoir monitoring is a common practice to survey the oil and gas field before or during production [32]. In a conventional marine seismic survey, one or more seismic sources followed by an array of marine seismic streamers are towed behind a survey vessel. The source emits acoustic waves that propagate and reflect at the seabed, where the reflected signals are received by the hydrophone sensors in the streamer array. In a 2D arrangement, one streamer is used, whereas 3D acquisition requires multiple streamers. The OBS technique has become more popular for the following benefits compared to the conventional streamer approach: 1) ocean bottom acquisition with geophones records elastic wave field while the hydrophone in the streamer only records pressure wave field, 2) higher signal-to-noise ratio with OBS, 3) more adapted to acquisition in a specific area for extended period of time, 4) ability to acquire in areas obstructed by oil platforms in the oil field. In the application of oil field development and monitoring, ocean-bottom seismic has been demonstrated with great potential. The technique has evolved from the early ocean-bottom cable (OBC) method to ocean-bottom node (OBN) and more recently, semipermanent ocean-bottom node (SPN). In the OBC system, an assembly of geophones and hydrophones are distributed on the seafloor, connected by electrical wires from which the measured data are routed to the recording vessel. The high cost for deployment usually limits its application. With OBN operation, a remotely operated vehicle (ROV) places and recovers multiple seismic recording devices on the seabed. Permanent or semipermanent installation of the sensor nodes in SPN saves cost by improving the efficiency in the deployment and recovery process.

One of the challenges in OBNs and SPNs is to minimize timing error across multiple

units in large scale surveys. Since the GPS signal cannot penetrate the seawater, clock synchronization has to rely on the low drift internal clock in each node. Existing solutions incorporates chip-scale atomic clocks [32, 33]. High stability and low temperature drift from the atomic clock could easily meet the timing requirement in these applications, leaving the higher power consumption of CSAC to be the major drawback. In these battery power systems, it is expected to operate for a longer period of time if a lower power precision atomic clock is available.

1.1.1.5 Underwater navigation

Atomic clocks also find their application in acoustic navigation for autonomous underwater vehicles (AUVs). While parameters such as depth, altitude, heading and roll/pitch could be easily measured with the internal sensors, obtaining the XY position of submerged underwater vehicles remains a challenging task. The high attenuation of GPS signals from the seawater precludes the application to AUVs, although it is an efficient navigation technique for surface and air vehicles. Acoustic positioning systems are widely used in underwater navigation. Experiments of underwater AUV were conducted [19, 34] employing a one-way-travel-time (OWTT) acoustic navigation technique. Most acoustic navigation systems are based on two-way time-of-flight measurements, where a vehicle initiates an interrogation pulse and receives acknowledgment signals from all passively listening transponders. Each vehicle communicates with all replying nodes to obtain two-way time-of-flight data. No absolute time base is required in such arrangement, but the total bandwidth increases with the number of vehicles in the network [35]. The narrow bandwidth of the acoustic communication usually limits the maximum number of vehicles in the network. In contrast, the one-way time-of-flight is calculated by comparing the transmitted and received time without sending replies. A surface ship with known position from GPS sends encoded time of launch and navigation data to passively listening submerged vehicles. The underwater vehicles carry precision clocks synchronized with the surface ship, from which time-of-flight is calculated. Without a stable time base, the underwater vehicles must synchronize with the GPS at the surface before the mission and the maximum operating period will be determined by the drift of the clock. Integrat-

ing chip-scale atomic clocks into the acoustic modem drastically improves the positioning accuracy of the achieving centimeter accuracy during a half-hour mission AUV [19].

1.1.2 Existing Solutions

1.1.2.1 State-of-the-art

Since the advent of atomic clocks, Cesium atomic clocks have been the most stable type of timekeeping device until recently. Cesium is a soft, shiny, silver-white ductile metal. It has a fairly low melting point of 28.5°C , which could turn into liquid by the heat of one's hand. Cesium atoms are relatively heavy (atomic mass $\sim 132.9\text{u}$), and as a result, they move slower than atoms with lower atomic mass. It allows them to stay in the interaction zone for longer time, reducing the spectral line width. Cesium also has higher hyperfine frequency of 9.2GHz compared to Rubidium of 6.8GHz and Hydrogen of 1.4GHz . Higher resonance frequency leads to a higher quality factor and thus better stability, with all others being equal. For about 40 years since 1959, the U.S. national primary frequency standard (NFS) was served by generations of Cesium beam atomic clocks, from NBS-1 to NBS-6 and NIST7. These clocks were named by the design in which a hot beam of Cesium atomic travels through a microwave cavity and changes their quantum states if the microwave and atomic hyperfine frequency match. The fraction of atoms that changed their states is measured by the transmitted light intensity, which serves as a feedback signal to control a local crystal oscillator [36].

The continuous effort contributed to thermal beam devices has improved the accuracy from 8.5×10^{-11} in 1959 to 5×10^{-15} in 1998. The performance of these thermal beam frequency standards were eventually limited by the short interaction time because of the fast moving atoms. First introduced by Zacharias in the 1950s and successfully demonstrated by Steven Chu in the late 1980s, the idea of long interaction time by cooled atoms was made possible by laser cooling technique. In a fountain clock, a small volume of Cesium atoms are cooled down by beams of lasers in 6 directions. A fraction of atoms whose states were preselected are slowly tossed up into a microwave cavity, with a speed slower than bicycles, flying above and then falling into the cavity. The states of the atoms are detected in the same way as with the thermal beam clocks. NIST-F1 and NIST-F2 are

the exemplary designs conducted by NIST, with NIST-F2 currently serving as NPFS [36].

Due to the high optical frequency ($\sim 10^{15}$ Hz), the quality factor of the spectral line of optical clocks could be much higher than the microwave predecessors. The accuracy of Cesium fountains have been surpassed by about two orders of magnitude by a new generation of device: the optical lattice clock. This groundbreaking performance has driven the Consultative Committee for Time and Frequency (CCTF) to initiate work towards a redefinition of the second [37]. A recent study compares the fractional uncertainty of optical lattice clocks based on $^{27}\text{Al}^+$, ^{87}Sr and ^{171}Yb [13]. These clocks feature narrow linewidths less than 10mHz and systematic uncertainty below 10^{-18} .

1.1.2.2 Chip-scale atomic clock (CSAC)

The size, weight and power consumption (~ 1 cubic feet, 30 kg and 50W for Microsemi 5071A [38], even more in Hydrogen masers and Cesium fountains) of the advanced atomic clocks essentially prohibits the use in mobile and portable applications. The military applications of secure communication and jam-resistant GPS receivers stimulated the first CSAC project funded by US Defense Advanced Research Project Agency (DARPA). The first commercial CSAC, the SA.45s, was launched in 2011 by Symmetricom, known as Microsemi today. The clock weighs 35 grams, measures less than 17cm^3 and consumes less than 120mW of power [39].

The ambitious goals of CSAC were made possible by several inventions in the clock design. The CSAC makes use of so-called coherent population trapping (CPT) to interrogate the atoms with a laser, in place of an RF discharge lamp [40]. In conventional Rubidium vapor cell clocks, the discharge lamp creates a population imbalance between the two hyperfine states of ^{87}Rb atoms. Two fluorescence lines emit from the ^{87}Rb discharge lamp, which are separated by approximately 0.014 nm. A ^{85}Rb gas cell filters one of the spectral lines from the lamp, while the other line is unaffected. The filtered light then passes through a ^{87}Rb vapor cell, where a population imbalance of two states $F1$ and $F2$ is created. When the microwave applied to the vapor cell is equal to the hyperfine frequency, there is a net transfer of atoms from $F = 1$ into $F = 2$, which decreases the transmitted light intensity [41].

Replacing the discharge lamp with the laser is motivated by the huge dc power consumption of the lamp and fulfilled by the development of CPT interrogation and semiconductor lasers. In a three-state quantum system, if the optical frequency difference matches with the two-photon microwave resonance, the atoms are placed in a superposition of two hyperfine states. This phenomenon is named the dark state because atoms do not absorb light in this state. Perfect dark state happens when the optical frequency difference is exactly equal to the hyperfine resonance, at which the optical transmission is maximum. When the optical frequency is detuned from resonance, optical transmission decreases. Contrary to the transmission dip at resonance in conventional atomic interrogation schemes, there is a transmission peak in CPT. The VCSEL laser is a crucial part of the low power atomic clock due to its high efficiency and low power compared to DFB and Fabry Perot lasers [40]. The laser also ensures coherent superposition of the dark states.

A compact atomic clock solution dictates an integrated physics package that contains the gas cell and the optics. Based on an early VCSEL clock by the Northrop Grumman, NIST developed a vertical assembly that includes a photo diode, Rubidium gas cell, optics and laser. The complete physics package only measures $1.5 \times 1.5 \times 3.2 \text{mm}^2$. Later, in the DARPA CSAC program, an improved physics package was designed. Two key technologies were involved, including the anodic bonding to seal the Rubidium gas and use of polyimide tethers to suspend the VCSEL, resonance cell and photodetector [40].

The frequency servo loop is no different from other Rubidium or Cesium clocks. In addition, there is a laser servo loop that locks the laser wavelength to optical absorption, a temperature servo to control the gas cell temperature, a microwave signal power servo that optimizes the CPT signal amplitude. The electronics, as well as the optics design, are optimized for lower power operation. It should be noted that low-power oriented design comes at a price of worse stability performance compared to dedicated Rubidium standards. For example, the Allan deviation of SA.45s (CSAC) [39] is more than one order of magnitude higher than that of PRS10 (a popular Rubidium standard) from thinkSRS [42]. Recent research advancement pushes the boundary of low power chip

scale atomic clocks. A Cesium ultra-low power atomic clock (ULPAC) achieved a long term Allan deviation of 2.2×10^{-12} at $\tau = 10^5$ s, while consuming 59.9mW total power [43].

1.1.2.3 Molecular clock

The world's first atomic clock was not based on Cesium hyperfine transitions, but instead on the ammonia rotational spectral line at about 23.87 GHz [44]. Frequency stabilization using this technique was first proposed in 1947 [44] and the first prototype of the atomic clock was developed by Harold Lyons and his colleagues at the National Bureau of Standards (NBS). The stability of this standard achieved about 2×10^{-8} , comparable to the fine quartz oscillators at that time. However, the frequency calibration continued to be performed by quartz standards until the invention of Cesium beam oscillators [45]. An improved design of the ammonia clock was investigated in 1979 [46], with about 2 orders of magnitude better stability than the first prototype. It was hoped to fill the gap between the crystal oscillator and expensive atomic clocks, but the development of high performance quartz oscillators soon outperformed the ammonia clock. A new piezo-electric resonator design, known as B.V.A., was invented in 1977, in which the ordinary electrodes bonded to the active part were replaced by tiny bridges. It significantly reduces the discontinuity in the resonator caused by the fixture and improves the aging [47]. The commercialization of the B.V.A. concept produces the BVA8600 series high performance oscillators. The short term stability reached as low as 2.5×10^{-13} with τ from 0.2 to 30 seconds [48]. Virtually all high performance quartz oscillators are housed in a temperature controlled oven to minimize drift by maintaining the crystal turnover temperature, which shows zero temperature sensitivity of quartz to the first order.

Emerging low power and portable applications precludes the use of OCXOs, due to the excessive size and power consumption of the oven. On the other hand, technology advancement has paved the way for high frequency electronics that was one of the predominant obstacles in the early exploration of molecular clocks. The rotational spectral lines of many simple molecules fall into the millimeter wave spectrum. Passive and active circuits in CMOS have been demonstrated with good performance at these frequencies. Without the need of expensive optical components or bulky microwave cavities, the molec-

ular clock only consists of a gas cell and a set of mm-Wave transceivers. Prototypes of molecular clocks based on rotational spectral line at G-band have been proposed [49, 50]. These clocks were based on the rotational resonance of carbonyl sulfide (OCS), as opposed to ammonia in the early attempts. One of the major benefits of OCS gas is the higher resonance frequency, resulting in a higher quality factor for the same linewidth. The OCS rotational spectrum covers from millimeter waves to sub-millimeter waves, where each resonance peak is about 12 GHz apart. Molecular clocks do not require optical interrogation and detection, thus full CMOS integration could be possible. The frequency, being within the millimeter wave range that CMOS process is capable of handling, opens up new opportunities for low cost and low power precision clocks. Recent work [50] on molecular clock employs an integrated transmitter and receiver, third order dispersion curve probing, achieving Allan deviation of 4.3×10^{-11} at $\tau = 1000$ s, while consuming 70.4mW dc power (excluding off-chip integrator and microcontroller).

1.2 Molecular Rotational Resonance

Molecules such as oxygen, water, and ammonia absorb distinct microwave or millimeter wave frequencies due to rotational resonance. The absorption selectivity improves as the molecular pressure is decreased, up to a point limited by so-called Doppler-shift broadening. At very low pressure of about 1-10mm Hg, the absorption only occurs within a small vicinity of the resonance. Sweeping the microwave frequency and observing the received signal power, a narrow dip shows up on a spectral plot. The unique absorption of each molecule is equivalent to the fingerprint of molecules, which plays an important role in spectroscopy. Given a spectrum of a certain gas sample and a known database of molecular spectral lines, the identity of each molecule in the gas could be determined. The spectrum of a molecule could be computed from quantum physics, and in the next section, a simplified model is analyzed.

1.2.1 The Rigid Rotor

A diatomic molecule with two atoms connected by a covalent bond may be modeled as a rigid rotor with bond length R and two point masses from the nuclei m_1 and m_2 .

The molecule rotates round the center mass, whose distance to the atoms is inversely proportional to the atom mass. The problem may be simplified to a single point mass rotating around a fixed vertex with reduced mass

$$\mu = \frac{m_1 \cdot m_2}{m_1 + m_2} \quad (1.3)$$

and the moment of inertia of the system is

$$I = \mu R^2 \quad (1.4)$$

The energy of the rigid rotor is obtained by solving the time-independent Schrödinger equation

$$\hat{H}\psi = E\psi \quad (1.5)$$

where ψ is the wave function and E is the energy eigenvalue of the system. \hat{H} is the Hamiltonian operator and since there is no potential energy in the rigid rotor system

$$\hat{H} = -\frac{\hbar^2}{2\mu}\nabla^2 \quad (1.6)$$

with \hbar being the reduced Planck's constant and ∇^2 the Laplacian operator

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (1.7)$$

Transforming into the spherical coordinates with the following relationship

$$\begin{aligned} x &= r \sin\theta \cos\varphi \\ y &= r \sin\theta \sin\varphi \\ z &= r \cos\theta \end{aligned} \quad (1.8)$$

and noticing that the bond length is constant, the Hamiltonian is simplified as

$$\hat{H} = -\frac{\hbar^2}{2I} \left[\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\varphi^2} \right] \quad (1.9)$$

The Hamiltonian must also be equal to one half of the squared angular momentum to the moment of inertia

$$\hat{H} = \frac{\hat{L}^2}{2I} \quad (1.10)$$

where \hat{L}^2 is the angular momentum squared operator

$$\hat{L}^2 = -\hbar^2 \left[\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\varphi^2} \right] \quad (1.11)$$

The solutions to the Schrödinger equation Eqn. 1.5 and Eqn. 1.9 are spherical harmonics denoted as $Y_l^m(\theta, \varphi)$. The variables m and l are angular momentum quantum number and magnetic quantum number, respectively. The $Y_J^m(\theta, \varphi)$ are a set of eigenfunctions to the Schrödinger equation of rigid rotors. From the eigenvalue equation

$$\frac{1}{2I} \hat{L}^2 Y_J^m(\theta, \varphi) = \frac{1}{2I} \hbar^2 J(J+1) Y_J^m(\theta, \varphi) \quad (1.12)$$

the energy is observed as

$$E_J = \frac{\hbar^2}{2I} J(J+1) \quad J = 0, 1, 2, \dots \quad (1.13)$$

which only assumes discrete values. The magnetic quantum number m takes the value of $0, \pm 1, \pm 2, \dots, \pm J$, but it does not appear in the energy. Thus, the energy of the rigid rotor is $2J+1$ -fold degenerate: for a fixed angular momentum quantum number J , a total of $2J+1$ states have the same energy.

The energy state of the molecular transitions up or down by absorbing or emitting energy equal to the difference of the energy states. The selection rule dictates that the change in energy states ΔJ is only allowed to assume values of ± 1 . The frequency is found from the Planck relation

$$f = \frac{\Delta E}{h} = \frac{E_{J+1} - E_J}{h} = \frac{h}{4\pi^2 I} (J+1) \quad (1.14)$$

The moment of inertia of the diatomic molecules is concentrated on the nuclei, thus for molecules with light-weight atoms, the rotational resonance frequency usually lies in the millimeter to sub-millimeter wave range. Notice that each state J has $2J+1$ degenerate states; the total absorption intensity is proportional to J up to about 500GHz when the number of higher energy molecules decreases, as a result of the Boltzmann distribution [51].

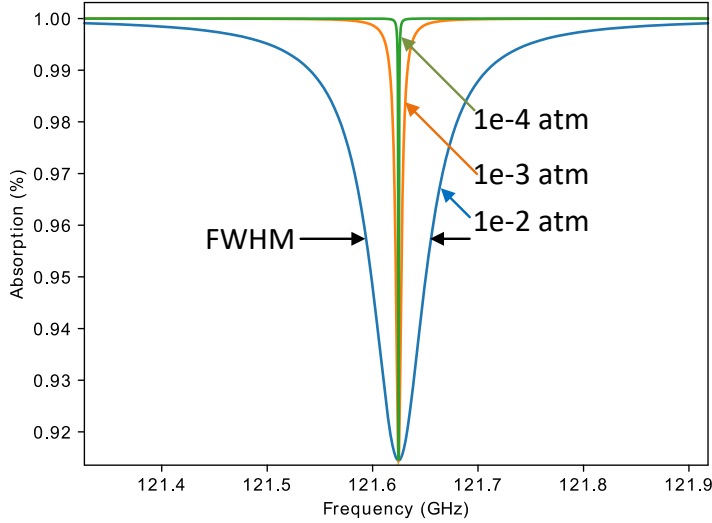


Figure 1.1: The simulated absorption profile of OCS gas at ~ 121.6 GHz at various pressures.

1.2.2 Spectral Line Width

Since the solution to the Schrödinger equation is discrete-valued, the absorption is expected to occur only at one single frequency, *i.e.*, the absorption profile is a delta function. In practice, however, the absorption profile takes a Lorentzian shape. An important parameter of the absorption profile is the full-width-half-maximum (FWHM), indicating how close the physical profile is to the ideal delta function. It is defined as the frequency difference where the absorption drops to half of its peak. A smaller FWHM is preferred, which shows better frequency selectivity. The quality factor of the spectral line is defined as the resonance frequency f_0 over the FWHM

$$Q = \frac{f_0}{FWHM}. \quad (1.15)$$

The clock stability is proportional to the quality factor, as is the case in all other resonant type of clocks. The FWHM is affected by the gas pressure, a phenomenon called pressure broadening. Figure. 1.1 simulates the OCS gas absorption profile during the $10 \leftarrow 9$ transition under different pressures. The data is sourced from HITRAN database, which provides spectroscopic parameters of a variety of gas molecules [52]. The FWHM is reduced linearly with the gas pressure, but in fact it stops decreasing at approximately

5×10^{-5} atm. Although HITRAN fails to model this effect, it was verified by measurement from literatures [51, 53]. The linewidth at very low pressure is limited by a mechanism called Doppler broadening. The OCS molecules move at the speed of about 325 m/s at room temperature, causing each molecule to see a different frequency due to the Doppler effect. The random direction and velocity eventually give rise to a FWHM of approximately 0.75 MHz. A decrease in the peak absorption at low pressure is also observed in the literature [51, 53]. An optimum pressure setting would fall in the range of 5 to 10 pa for OCS molecules.

1.2.3 Allan Deviation

The stability of oscillators may be characterized by several metrics, depending upon the observation time. For a relatively short observation period, the oscillator output frequency fluctuates due to the random noise from the resonator and active circuits. During this period, the nominal frequency is kept constant, and the instability of the oscillator could be conveniently modeled as random phase modulation around a nominal frequency. Frequency domain representation is usually favored, leading to the phase noise definition as the Fourier transform of the autocorrelation function of the random phase. In a typical crystal oscillator, the phase noise degrades with decreasing offset frequency. The white far out phase noise is usually dominated by the thermal noise of the amplifier, resistor, etc. As the offset frequency reduces, the phase noise exhibits higher order slopes. For example, the amplifier flicker noise contributes to the first order $1/f$ phase noise, while noise contribution in the resonator could vary from $1/f^2$ to $1/f^4$. Eventually, for an extended period of observation time, the nominal frequency of the oscillator drifts, violating the small phase perturbation approximation. This is when the phase noise metric fails and Allan deviation, named after David Allan, comes into play. Modern Allan variance is based on the 2-sample variance

$$\sigma_y(\tau) = \left\langle \frac{1}{2} (y_{n+1} - y_n)^2 \right\rangle^{\frac{1}{2}} \quad (1.16)$$

where $\langle \cdot \rangle$ denotes the expectation operation and y_n is the n^{th} fractional frequency average over sample time τ . Here, y_n could be calculated from the fractional reading $x(t)$ at

discrete time $t = n \cdot \tau$ and $(n + 1) \cdot \tau$

$$y_n = \frac{x((n + 1) \cdot \tau) - x(n \cdot \tau)}{\tau} \quad (1.17)$$

In practice, the ADEV from a finite dataset with M samples is calculated from

$$\sigma_y(\tau) = \left[\frac{1}{2(M - 1)} \sum_{i=1}^{M-1} (y_{i+1} - y_i)^2 \right]^{\frac{1}{2}} \quad (1.18)$$

If the measured frequency only contains zero mean random fluctuation, the ADEV falls with a slope of $\sqrt{\tau}$. It implies that for every $100\times$ longer integration time, the ADEV decreases by 10 times. It is therefore more convenient to construct a log-log plot, where time τ and ADEV are both in log scale. In this case, the $\sqrt{\tau}$ slope becomes a straight line.

Chapter 2

System Analysis of the Molecular Clock

2.1 Introduction

A molecular clock locked to gaseous rotational spectra provides exceptional stability performance by keeping track of the center of the Lorentzian resonance profile. This is accomplished through a feedback loop that detects the frequency difference between an unstable millimeter-wave signal to the rotational resonance and fine-tunes the mm-Wave frequency so that the long term frequency error is reduced to zero. The output of the system is usually a low frequency signal derived from the stable mm-Wave frequency, 10 MHz for example, as a reference to other precision timing circuitry.

One of the major components in a molecular clock system is a sealed gas container. Because low pressure reduces the line width of the rotational spectra, the molecular gas is hermetically sealed in a vessel at desired pressure for use in the clock. The vessel could be either a sealed cylindrical cell with dielectric openings for unguided mm-Wave probing or a sealed waveguide filled with gas. As the electromagnetic (EM) wave travels through the gas from the one end of the opening to the other or through the waveguide, its energy gets absorbed if the frequency of the EM wave is close to the rotational resonance. This leads to a convenient two-port setup shown in Figure. 2.1 [53], with the transmitter and receiver separated by the molecular gas with a distance L . The waveguide may be manufactured along a meandering path [50,53], making the EM wave traveling a distance L much larger

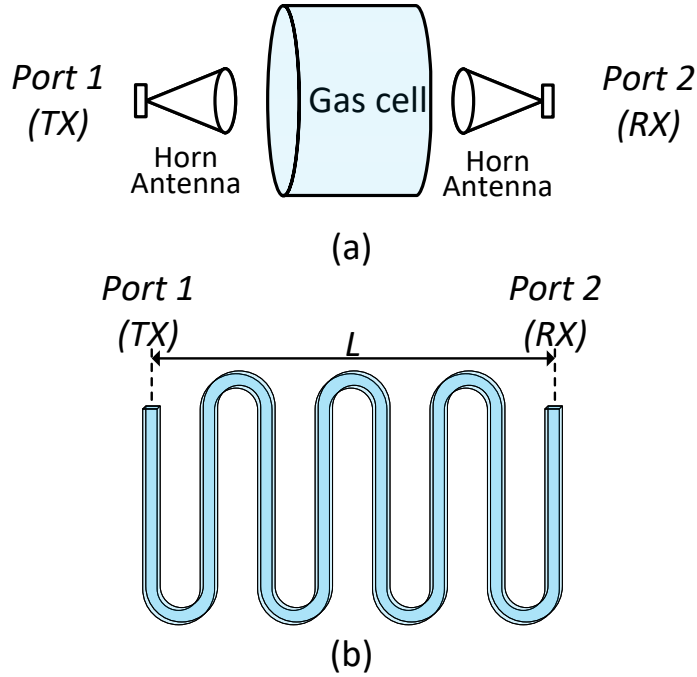


Figure 2.1: Two types of gas cells. (a) The cylindrical gas cell with horn antenna probing [53]. (b) The meandering waveguide gas cell where the EM wave travels a distance much longer than L [50, 53].

than the physical spacing of the transceiver, in contrast to the unguided scenario.

The electronic circuitry of measuring the absorption profile bears considerable resemblance to a scalar network analyzer. By sweeping the transmitter frequency around the resonance and measuring the received power, which is equivalent to $|S_{21}|$, the Lorentzian frequency response is obtained. In the simplest form, the transmitter could be a mm-Wave voltage controlled oscillator (VCO), but significant drawbacks exist. Due to the low quality factor of the resonator in a mm-Wave VCO, the frequency of the VCO can be very unstable and fail to lock to a narrow absorption line. The tuning range of the mm-wave VCO is typically around a few percent, which is much larger than the locking range of the feedback system. Therefore, the transmitter is more commonly made of a frequency synthesizer that locks the mm-Wave VCO to a more stable voltage controlled crystal oscillator (VCXO) at lower frequency. Instead of controlling the mm-Wave VCO directly, voltage tuning is applied to the VCXO. The receiver can be as simple as a diode

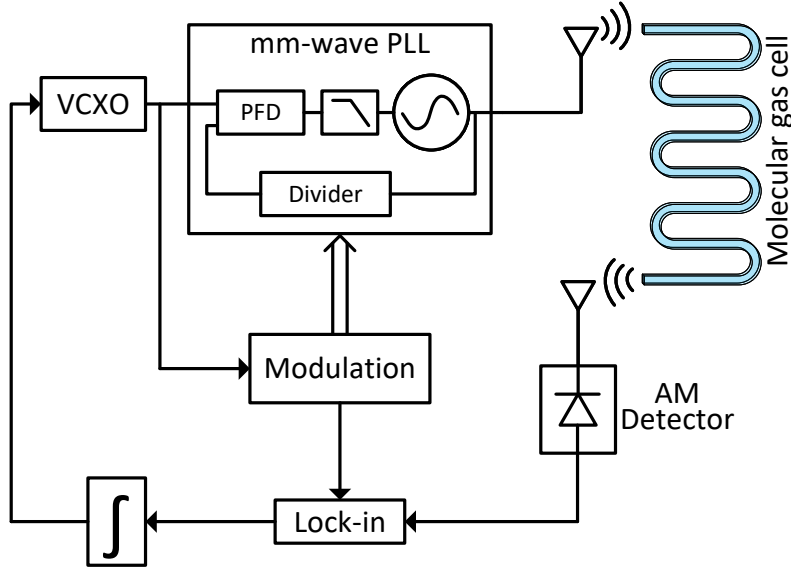


Figure 2.2: Simplified system architecture including a molecular gas cell, a transmitter which is a mm-Wave PLL referenced to a low frequency VCXO, a receiver that demodulates AM signal and a feedback control circuitry that consists of a modulation generator, a lock-in amplifier and an integrator.

power detector, although other receiver architectures exist. The receiver will be discussed in more detail in Section. 2.4.

An open loop measurement of the Lorentzian profile is useful to gain some insight about the system performance qualitatively. In order for the clock to operate in a closed loop, it is necessary to introduce modulation to the transmitter. Since the frequency difference between the mm-Wave frequency f_c and the resonance frequency f_0 is of interest rather than the shape of the profile, the transmitter should be designed such that the received signal reflects the frequency difference.

$$ERROR \propto \Delta f = f_c - f_0$$

In practice, this process is done by frequency modulation in the transmitter. When the FM signal passes through the gas, the envelope of the transmitted FM signal is imprinted by the frequency response of the spectral line. The receiver demodulates the envelope, which is converted to an error signal by a lock-in amplifier. The lock-in amplifier is

essentially a synchronous detector that compares the phase of the detected AM component to the FM baseband. If $f_c < f_0$, the lock-in output would be negative, and vice versa. The error signal is then integrated before feeding back to the VCXO to adjust f_c (because the mm-Wave VCO is locked to the VCXO). The use of an integrator ensures zero average frequency error, due to the infinite gain of an ideal integrator at dc. The simplified block diagram of the molecular clock is shown in Figure. 2.2.

2.2 Dispersion Curve and Baseline

2.2.1 Dispersion Curve

Modeling the total transmission along the distance L as a Lorentzian function:

$$\beta(f) = 1 - \alpha_p \frac{f_h^2}{f_h^2 + (f - f_0)^2} \quad (2.1)$$

where α_p is peak absorption at resonance frequency f_0 and f_h is the half-width at half maximum, *i.e.*, the half frequency bandwidth over which the transmitted signal power drops by half. The slope of the absorption profile is odd symmetric around the resonant frequency, making it a good indication of the error signal. The slope, or sometimes called dispersion curve, may be obtained through FM modulation. Assume a sinusoidal modulated signal, whose instantaneous frequency $f_c(t)$ is expressed as:

$$f_c(t) = f_c + f_\Delta \cos(2\pi f_m t) \quad (2.2)$$

where f_c is the center frequency of the carrier and f_m and f_Δ are modulation frequency and modulation depth, respectively. When the modulated signal interacts with gas molecules, the instantaneous power of the mm-Wave signal is amplitude modulated because of the frequency response of the absorption profile. The instantaneous intensity of the mm-Wave signal can be found by substituting Eqn. 2.2 into Eqn. 2.1:

$$I(t) = \beta(f_c(t)) \quad (2.3)$$

Taylor expansion of $\beta(f)$ around f_0 yields:

$$\beta(f) \approx 1 - \alpha_p + \alpha_p \left(\frac{(f - f_0)^2}{f_h^2} - \frac{(f - f_0)^4}{f_h^4} + \dots \right) \quad (2.4)$$

and assuming only the first harmonic is of concern and f_m and Δf are small, $I(t)$ can be approximated by only retaining the quadratic term in Eqn. 2.4:

$$I(t) \approx 1 - \alpha_p + \alpha_p \left(\frac{f_{\Delta}^2 \cos(4\pi f_m t) + 4f_{\Delta}(f_c - f_0) \cos(2\pi f_m t) + 2(f_c - f_0)^2 + f_{\Delta}^2}{2f_h^2} + \dots \right) \quad (2.5)$$

Realize that $I(t)$ represents the envelope of the mm-Wave signal through the gas. Using an ideal envelope detector as the receiver, the amplitude of the first harmonic from the detector would be:

$$V_1 = \frac{2\alpha_p f_{\Delta}(f_c - f_0)}{f_h^2} \quad (2.6)$$

Following the same procedure, the amplitude of the third harmonic can also be found:

$$V_3 = -\frac{\alpha_p f_{\Delta}^3 (f_c - f_0)}{f_h^4} \quad (2.7)$$

Notice that Eqn. 2.6 and Eqn. 2.11 are only valid for small Δf , because the Taylor series is evaluated for f with small deviation from f_0 . As expected, both the fundamental and third harmonic of the detector output are proportional to Δf , confirming the validity of an indication as frequency error. For large frequency deviations, the response of the dispersion curve ought to be obtained from numerical simulations.

Figure. 2.3 overlays the frequency-time relationship of an FM modulated signal on top of a transmission plot. When the carrier frequency of the mm-Wave lies exactly at the resonance, the envelope of the transmitted signal contains only even harmonics of the modulation frequency. If the mm-Wave carrier frequency is lower or higher than the resonance, there are nonzero odd harmonics in the envelope. The sign of the odd harmonic is determined by the phase relative to the modulation signal. When $\omega_c < \omega_0$, the odd harmonics are out of phase to the modulation frequency, giving a negative output, and vice versa. In principle, any of the odd harmonics could be an indication of $\Delta\omega$. In practice, the magnitude of higher order harmonics is smaller than that of lower harmonics with the same frequency difference. Thus, it suffers from lower SNR in a noisy environment. Figure. 2.4 compares the dispersion curve of various odd harmonics compared to the derivative of a Lorentzian profile. Here, a FWHM of 1 MHz and peak absorption of 1 percent are assumed. Figure. 2.4 (a-c) plots the normalized derivative of the Lorentzian

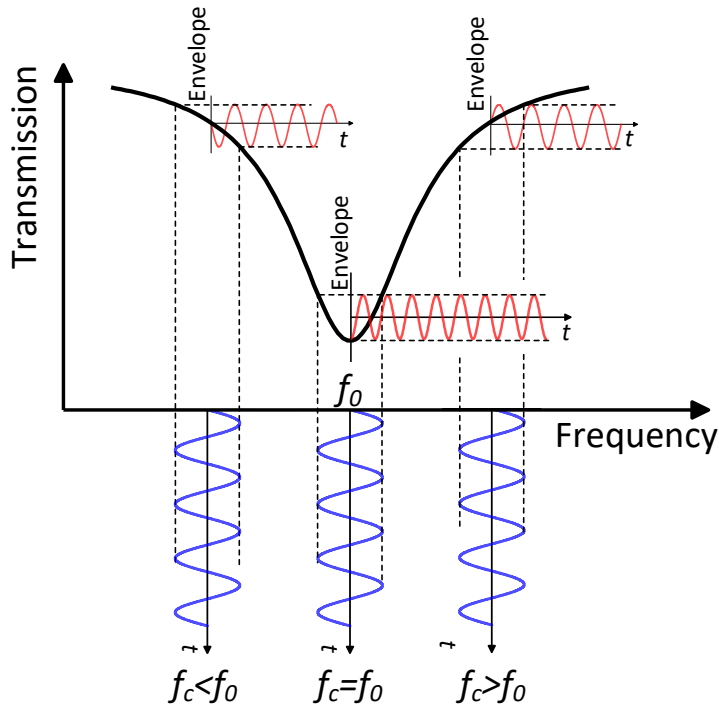


Figure 2.3: Envelope of the received signal from an FM modulated transmitter with various carrier frequencies: $f_c < f_0$, $f_c = f_0$ and $f_c > f_0$.

profile, whereas (d-f) and (h-j) are simulated dispersion curves with modulation depth of 0.1 and 1 times FWHM, respectively. The magnitude of all the harmonics is normalized to that of the fundamental. As the order of extracted harmonic increases, the output magnitude at the same offset frequency decreases compared to the fundamental. The linear region where the output is monotonically increasing with frequency error also shrinks. The benefit of using a higher order dispersion curve will be discussed later, but it is clear so far that higher order dispersion curve degrades SNR by reducing the signal amplitude. The modulation depth is another factor that affects the shape of the dispersion curve. With modulation depth small compared to the FWHM of the absorption profile, the curve is indistinguishable to the derivative of the Lorentzian function. Increasing modulation depth to FWHM results in a dispersion curve with lower slope at zero crossing. The difference may be explained by an analogy with the tangent or secant line of a function. The FM modulation and demodulation process works similarly as joining two points on

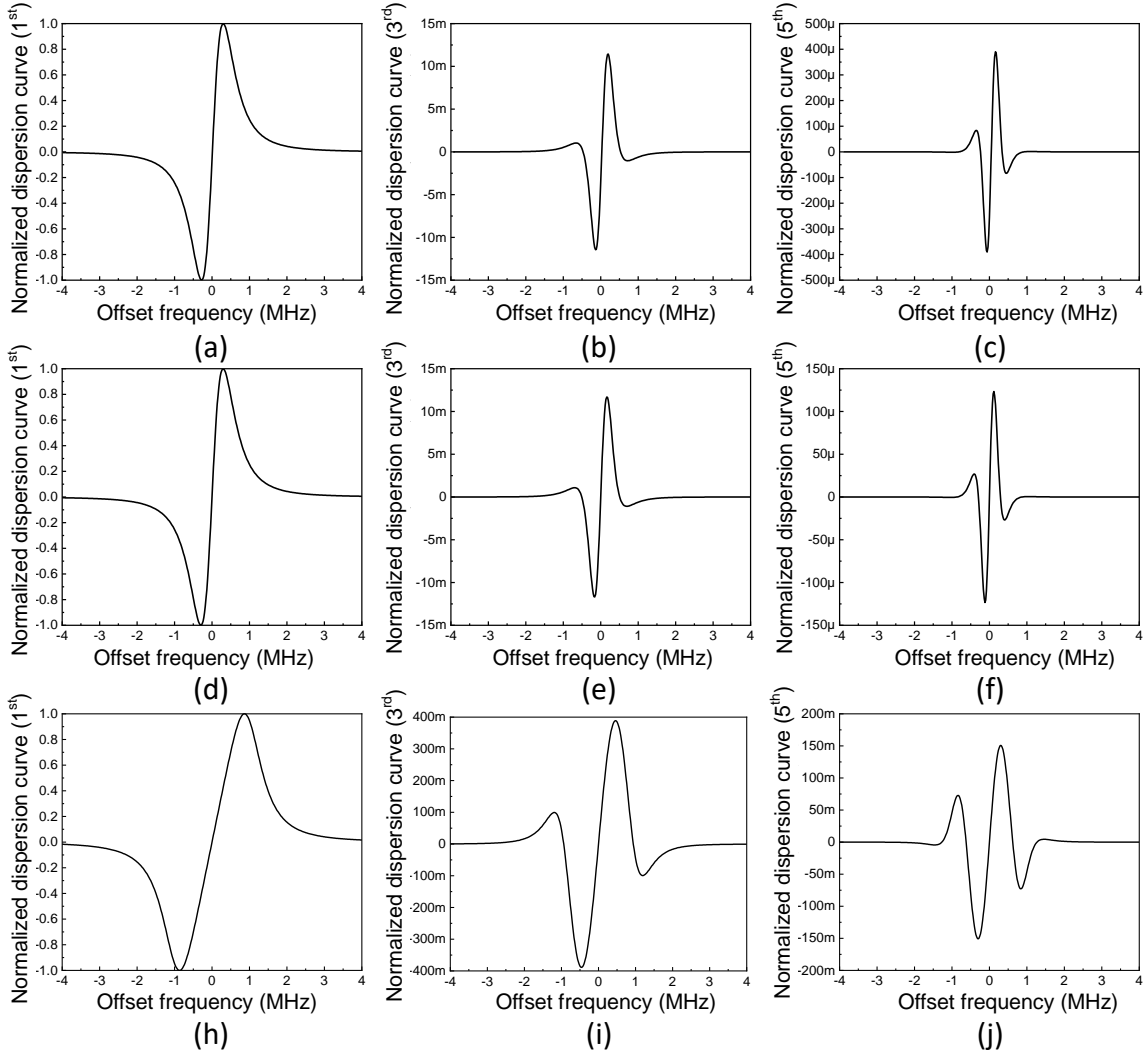


Figure 2.4: (a-c) First, third and fifth order derivative of Lorentzian profile with FWHM of 1 MHz and peak absorption of 1%. Simulated odd harmonic dispersion curves with a modulation depth of 0.1·FWHM (d-f) and a modulation depth equal to FWHM (h-j), respectively.

a function and finding the slope. As the two points get close to each other, the slope of the secant line becomes the derivative of that point.

2.2.2 Baseline Tilting

So far, it has been assumed that the absorption is in a symmetric Lorentzian shape. In fact, the measured absorption profile is generally a superposition of a slowly varying baseline and a Lorentzian function, as illustrated in Figure. 2.5. The baseline arises from the aggregated frequency response of all components, including the transceiver, free

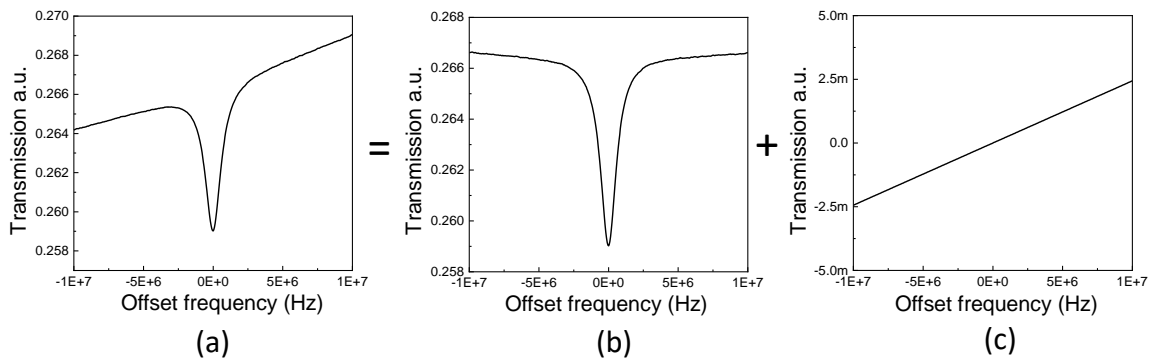


Figure 2.5: (a) The measured transmission profile of OCS gas around the resonance frequency of about 145.947 GHz, where a linearly tilted baseline is observed (a). The measured transmission profile may be decomposed into a symmetric Lorentzian profile (b) plus a linear baseline (c).

space, etc., and also from the standing wave between the transmitter and receiver. It is approximately linear around the resonance because the rotational resonance must be the only high Q resonance at this frequency; otherwise, the clock could be made more stable by locking to a higher Q resonance. The low Q baseline thus can be approximated as a straight line around the resonance (to first order, of course). The tilted baseline affects both frequency accuracy and stability. The transmission profile in Eqn. 2.1 would have to include a linear term to model the linear baseline:

$$\tilde{\beta}(f) = 1 - \alpha_p \frac{f_h^2}{f_h^2 + (f - f_0)^2} + k \cdot f \quad (2.8)$$

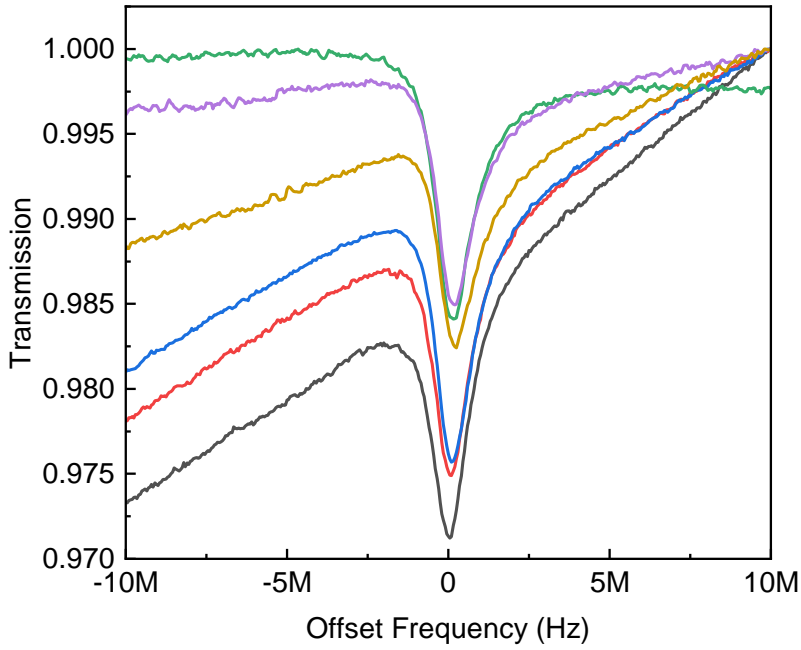
where k is the slope of the linear baseline. The fundamental component from the envelope detector of the receiver is given by:

$$\tilde{V}_1 = \frac{2\alpha_p f_\Delta (f_c - f_0)}{f_h^2} + k \quad (2.9)$$

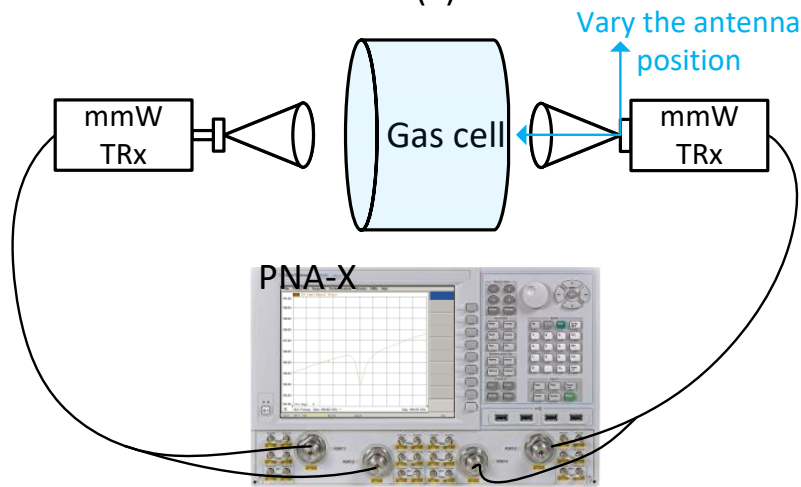
When the whole system locks, the feedback loop forces the error signal \tilde{V}_1 equal to zero. The carrier frequency must shift from f_0 by an amount of:

$$f_{offset} = \frac{k f_h^2}{2\alpha_p} \quad (2.10)$$

The dependence of offset frequency on the baseline slope and absorption peak adds to the uncertainty of the carrier frequency in the lock state. One of the sources of baseline tilting



(a)



(b)

Figure 2.6: (a) Measured transmission profiles at around 109.463 GHz for various antenna positions. (b) Measurement setup using a Keysight PNA-X network analyzer and a pair of VDI mm-Wave TRx heads.

is the standing wave, caused by the reflection at various interfaces along the propagation path. For example, reflection occurs at the window that seals the low pressure gas from air. The standing wave depends on the mechanical assembly of the gas cell and antenna, which not only varies from part to part but is hardly stable over time. In Figure. 2.6,

various rotational spectral lines of OCS gas are measured at 109.463 GHz with slightly different position of TX and RX antenna corresponding to the gas cell. The setup is illustrated in Figure. 2.6(b). The transmission is derived from $|S_{21}|$ and normalized to the maximum absorption. Note that it is difficult to study the effect of standing waves with the setup shown in Figure. 2.6(b), since the measurement is a collective result of standing wave as well as TX to RX leakage and more.

For the linearly tilted baseline, the third harmonic remains the same as a flat baseline:

$$\tilde{V}_3 = -\frac{\alpha_p f_\Delta^3 (f_c - f_0)}{f_h^4} \quad (2.11)$$

It becomes evident by looking at the higher order dispersion curve from the perspective of the derivative of the transmission profile. The n^{th} order dispersion curve is equivalent to taking an n^{th} order derivative. Any $(n - 1)^{th}$ order term in the baseline would be eliminated by the derivative operation. In the case of strong baseline environment, it is instructive to lock to a higher order dispersion curve. There is a trade-off between choosing a 3^{rd} order or 5^{th} order dispersion curve. Although the 5^{th} curve cancels the 3^{rd} baseline, it suffers from lower SNR. In practice, after canceling the linear baseline, factors such as temperature drift instead of higher order baselines usually limit the frequency stability.

2.2.3 FSK Modulation

Alternatively, the frequency difference between the carrier f_c to the resonance f_0 could be obtained with an FSK modulated transmitter. It is a special case of an FM modulation with two discrete frequency states. In the time domain, the output of FSK modulation switches between ϕ_1 and ϕ_2 , corresponding to $f_c - f_\Delta$ and $f_c + f_\Delta$. The modulation depth f_Δ , is defined the same way as in the FM modulation. Figure. 2.7 shows different f_c with respect to f_0 and the corresponding signal envelope at the receiver. When f_c is equal to f_0 , the attenuation due to wave and molecule interaction is equal for ϕ_1 and ϕ_2 . A signal with constant amplitude passing the envelope detector yields only a dc component. The dc signal is then fed to the lock-in amplifier to calculate the frequency error. The reference signal to the lock-in amplifier is a square wave of zero offset, synchronized with

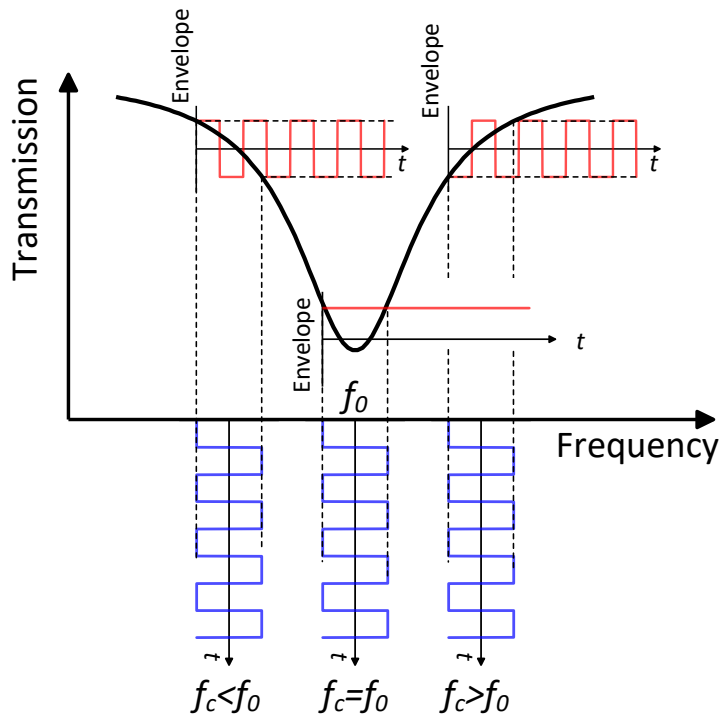


Figure 2.7: Envelope of the received signal from an FSK modulated transmitter with various carrier frequencies: $f_c < f_0$, $f_c = f_0$ and $f_c > f_0$.

the FSK modulation frequency. Since the lock-in amplifier is equivalent to a multiplier, any dc value multiplies a zero-mean square wave is also zero. When f_c is lower than f_0 , the amplitude during ϕ_1 is larger than ϕ_2 . The envelope of the FSK signal is therefore modulated by the molecular absorption at the rate of the FSK modulation frequency. Suppose the lock-in reference takes a value of $-A$ at ϕ_1 and A at ϕ_2 , the envelope is out of phase with respect to the reference. The output of the lock-in amplifier would thus produce a negative dc voltage. Similarly, when f_c is larger than f_0 , the dc output is positive. Overall, the FSK modulation scheme provides an output proportional to frequency difference around the resonance. Note that whether the lock-in reference is in-phase or out of phase with respect to ϕ_1 and ϕ_2 is determined such that the loop is negative feedback. It depends on the polarity of both the dispersion curve and the circuit that follows the lock-in amplifier.

2.2.3.1 Baseline cancellation

A simple FSK modulation also suffers from the baseline tilting problem, as in the case of locking to the fundamental dispersion curve using a sine wave. An auxiliary pair of FSK modulation may be inserted in between each period of normal FSK operation. The frequency tones with respect to the transmission profile is illustrated in Figure. 2.8(a) and the timing diagram is sketched in Figure. 2.8(b). Here, f_1 and f_2 are the two tones from the main FSK, whereas f_3 and f_4 are the additional tones to measure the baseline, with the relationship to carrier frequency f_c :

$$\begin{aligned}
 f_1 &= f_c - \Delta f_1 \\
 f_2 &= f_c + \Delta f_1 \\
 f_3 &= f_c - \Delta f_2 \\
 f_4 &= f_c + \Delta f_2
 \end{aligned} \tag{2.12}$$

where Δf_1 is smaller than Δf_2 . Suppose the transmission profile is a superposition of a linear baseline on top of an ideal Lorentzian function shown in Figure. 2.8(a), with its decomposition on the right. The envelope at each frequency a_i , $i = 1 - 4$ is therefore the result of the Lorentzian function $\beta(f)$ and baseline:

$$\begin{aligned}
 a_1 &= \beta(f_1) + k \cdot f_1 \\
 a_2 &= \beta(f_2) + k \cdot f_2 \\
 a_3 &= \beta(f_3) + k \cdot f_3 \\
 a_4 &= \beta(f_4) + k \cdot f_4
 \end{aligned} \tag{2.13}$$

where k is the slope of the baseline and the dc term of the baseline has been ignored. Since the Lorentzian function is symmetric around f_0 , when f_c aligns with the resonance f_0 , the amplitude difference due to the Lorentzian profile $\beta(f_1) - \beta(f_2)$ and $\beta(f_3) - \beta(f_4)$ is equal to zero. The baseline induced error of the main FSK $a_1 - a_2$ comes proportional to the frequency difference Δf_1 . This holds true for the auxiliary FSK f_3 and f_4 as well. The linear baseline from the main FSK could be calibrated out if the error function is defined as

$$ERROR_{cal} = a_1 - a_2 - \frac{\Delta f_1}{\Delta f_2} (a_3 - a_4) = \beta(f_1) - \beta(f_2) - \frac{\Delta f_1}{\Delta f_2} [\beta(f_3) - \beta(f_4)] \tag{2.14}$$

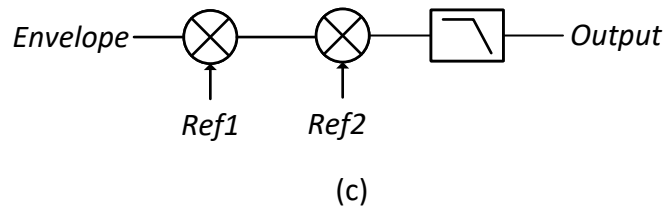
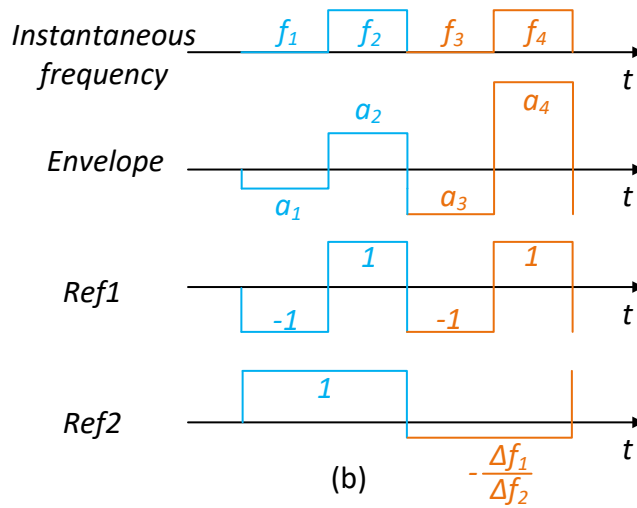
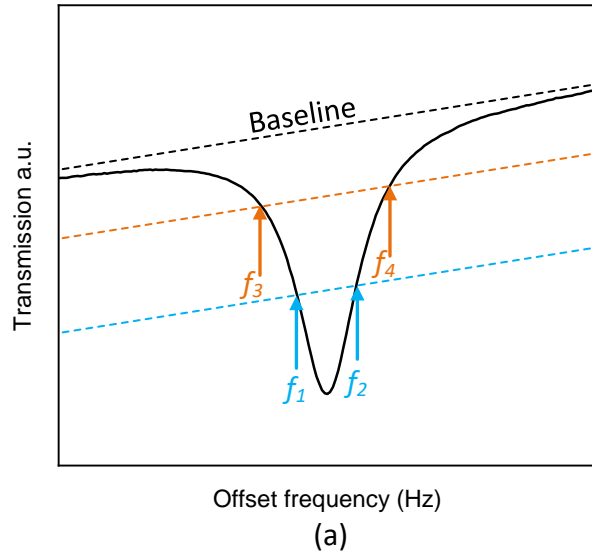


Figure 2.8: (a) Frequency tones defined by Eqn. 2.12 on the transmission profile with a tilted baseline. When f_c equal to f_0 , the error difference from main FSK and auxiliary FSK comes from the linear baseline and is proportional to $\Delta f_2 - \Delta f_1$. (b) Timing diagrams of the dual FSK scheme for instantaneous frequency, detected envelope and two reference signals Ref_1 and Ref_2 . (c) Block diagram of the dual FSK lock-in.

Obviously, the linear baseline term is canceled out from Eqn. 2.14. At the same time, a trade-off by the choice of Δf_2 between SNR and baseline cancellation is observed. With the introduction of an auxiliary FSK, the baseline is canceled as well as the output signal if Δf_2 is only slightly larger than Δf_1 . On the other hand, if Δf_2 is much larger than the Lorentzian FWHM, the assumption of a linear baseline would no longer be valid. The measured baseline which is corrupted with higher order terms would deviate from the real baseline between f_1 and f_2 . The optimal choice of f_3 and f_4 should be the frequency where the absorption reduces to a small fraction of its peak, 10% for example.

2.3 Transmitter

The transmitter is essentially a mm-Wave frequency synthesizer employed to generate a carrier frequency at the rotational resonance. The heart of a synthesizer is usually a phase-locked-loop (PLL). Because of the linear range of the dispersion curve being only a fraction of the spectral line FWHM and the small FWHM due to the high quality factor, the lock range of the molecular clock loop is limited. The frequency drift of a free running VCO at mm-Wave frequency is often much larger. The large drift is the consequence of low Q on-chip resonator tank, the power supply and temperature drift and flicker noise from the MOS transistor. The PLL allows the mm-Wave frequency to be referenced to a low frequency VCXO to ensure the initial frequency is close to the resonance.

2.3.1 PLL Introduction

Consider a general PLL block diagram in Figure. 2.9(a). To analyze the stability and noise characteristics of the PLL, it is more convenient to obtain the small signal model of the PLL. Although the PLL is inherently non-linear transitioning from unlock to lock state, the small signal approximation is valid when the loop is locked. During the locked state, the small phase perturbation of the VCO is constantly corrected by the feedback loop. As long as the PLL stays in lock, the small signal analysis is sufficiently accurate to provide useful design insights. Unlike the analysis of amplifiers, the small signal model of the PLL is performed in the phase domain, where the variable of interest is phase, instead of voltage or current. Therefore, the transfer function takes phase as its variable.

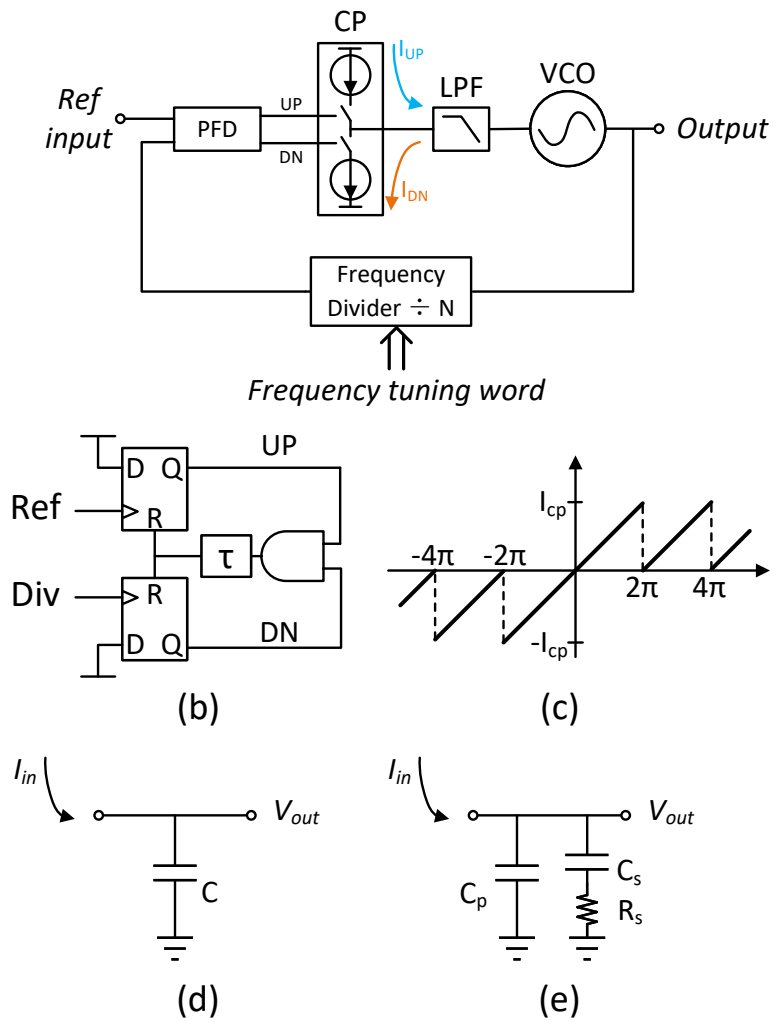


Figure 2.9: (a) PLL block diagram and (b) Circuit for phase frequency detector. (c) The transfer function of the PFD combined with a charge pump. (d) First order low pass filter, and (e) second order low pass filter with an additional zero.

2.3.1.1 PLL building blocks

In a voltage controlled oscillator, the output frequency is controlled by the voltage applied to the internal varactor. The instantaneous frequency is modified by the effective capacitance of the varactor by the voltage applied to it. Although the oscillation frequency is a non-linear function of inductance and capacitance, as well as the varactor $C - V$ curve, it is useful to define the gain from the input voltage to the frequency, denoted as K_{VCO} , in units of Hz/Volt. It should be noted that K_{VCO} is a small signal approximation of the $f - V$ characteristic of the VCO, which is dependent upon the bias voltage. Realizing that phase is the variable of concern, and that the phase change is a time integral of angular frequency change, the gain of the VCO in the phase domain is $2\pi K_{VCO}/s$, where s is the Laplace transformation of frequency. The implicit integrator in the VCO is the result of change of variable, from frequency to phase.

The phase-frequency detector is usually a pair of resettable D-flip-flops (DFFs) with a NAND gate shown in Figure. 2.9(b). The PFD is named after its ability of being sensitive to both frequency and phase. The PFD is commonly combined with a charge pump. The charge pump is a switch controlled current source, being able to source and sink current. A high performance integrator, which is required by the PLL to produce zero net phase difference, can be made of a charge pump loaded with a capacitor or capacitor/resistor network. The average output current $I_{UP} - I_{DN}$ is proportional to the frequency and phase difference between the two inputs [54] and is plotted in Figure. 2.9(c). The gain of the PFD is defined as the ratio of average output current to the phase error when the loop is locked

$$K_{PD} = \frac{I_{CP}}{2\pi} \quad (2.15)$$

The low pass filter that follows the charge pump is a capacitor/resistor network. The transfer function $F(s)$ is the impedance of the network. The simplest and lowest order PLL uses a single capacitor in the low pass filter. The capacitor and charge pump current source create a first order integrator, in which the output voltage fall in a slope of 20dB per decade in response to the input current. The PLL that employs a first order filter such as in Figure. 2.9(d) is called a second order loop, because the integrator from the

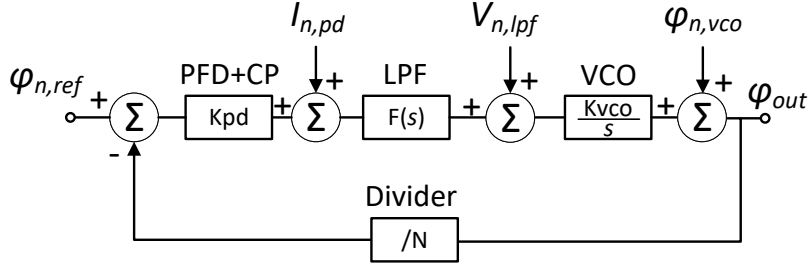


Figure 2.10: Transfer function of the PLL when it is in lock. The model also includes major noise sources, including reference phase noise, charge pump current noise, low pass filter noise and VCO phase noise, which are denoted as $\phi_{n,ref}$, $I_{n,pd}$, $V_{n,lpf}$ and $\phi_{n,vco}$, respectively.

VCO always contributes to another pole. Second order PLLs suffer from the tradeoff between locking range and loop bandwidth. To overcome such problem, second order low pass filters are widely used, as shown in Figure. 2.9(e).

The small signal model of the frequency divider is a scaling factor of $1/N$, where N is the division ratio. For a fractional divider, the scaling factor is simply the fractional frequency ratio.

2.3.1.2 PLL transfer function

The small signal model of the PLL in lock is shown in Figure. 2.10, where each block is replaced with the model derived in the previous section. In PLL designs, loop bandwidth, stability, and noise performance are among the top design goals, all of which could be analyzed with the small signal transfer function with sufficient accuracy. Loop bandwidth and stability are associated with the open loop transfer function, whereas the transfer function of each noise source to the output is evaluated individually. All circuit components generate noise in general, but the reference and the VCO are two major noise contributors. The noise is modeled as a source of random phase fluctuations at the input, denoted as $\phi_{n,REF}$ and $\phi_{n,VCO}$, for reference and VCO, respectively.

The PLL closed-loop transfer function from VCO tuning port to output can be found as

$$\frac{\phi_o}{V_{fm}} = \frac{K_{VCO}}{s + K_{PD}F(s)\frac{K_{VCO}}{N}} \quad (2.16)$$

where K_{PD} , K_{VCO} , N and $F(s)$ are phase-frequency detector gain, VCO tuning sensitivity, divider ratio and low pass filter transfer function, respectively. In a typical charge-pump based PLL, K_{PD} is proportional charge pump current:

$$K_{PD} = \frac{I_{PD}}{2\pi} \quad (2.17)$$

and $F(s)$ is a second order integrator with the following transfer function:

$$F(s) = \frac{1 + \frac{s}{\omega_z}}{s(C_1 + C_2)(1 + \frac{s}{\omega_p})} \approx \frac{1}{sC_2} \left(1 + \frac{s}{\omega_z}\right) \quad \text{if } C_2 \gg C_1 \quad (2.18)$$

where

$$\begin{aligned} \omega_z &= \frac{1}{R_1 C_2} \\ \omega_p &= \frac{C_1 + C_2}{R_1 C_1 C_2} \end{aligned} \quad (2.19)$$

The low frequency gain of $F(s)$ is infinite at dc and falls with a slope of 20 dB per decade starting from the origin and gradually approaches R_1 for frequency lower than ω_p . At very low frequency, the magnitude of Eqn. 2.16 is very low because $F(s)$ appears at the denominator. As the frequency goes higher past ω_z , the second term in the denominator becomes constant, the transfer function reduces to a single pole system and the magnitude drops at a slope of 20 dB per decade. Therefore, the transfer function overall is a band pass filter. Qualitatively, if a very low frequency perturbation is applied at the VCO input, due to the high loop gain, the feedback loop compensates the perturbation and keeps the output frequency constant. When the perturbation frequency is higher than the open loop bandwidth of the PLL, the implicit integrator in the VCO cannot keep up with the input change, resulting in a smoothed VCO output. As a consequence, the modulation frequency should be selected to lie in the pass band of the transfer function. This frequency is typically less than the bandwidth of the PLL.

2.3.2 TX specifications

In the molecular clock application, there are several key design considerations to the transmitter, namely frequency precision, phase noise and modulation capability.

2.3.2.1 Frequency resolution

The molecular rotational frequency is usually not an integer multiple of any commonly used crystal oscillators. For example, the frequency f_0 of the OCS transition line ($10 \leftarrow 9$) is approximately 121.624632 GHz. To generate this frequency, one may use a customized crystal oscillator that is a sub harmonic of f_0 together with a signal chain including multipliers and/or integer-N PLL(s). This approach has been taken by some atomic clock manufacturers. The difficulty of fractional frequency synthesis of the mm-Wave signal is then shifted to fractional synthesis of a 10 MHz signal. The 10 MHz signal is a standard reference frequency for virtually all frequency and timing equipments to synchronize and calibrate against each other. A DDS generator or a fraction-N PLL plus divider could bring an arbitrary frequency down to 10 MHz. The consequence of this approach is compromised 10 MHz performance, in terms of phase noise and spur.

The alternative approach is to use a fractional-N PLL. A fractional-N PLL is capable of generating rational frequency ratio to the reference frequency. The major upgrade over an integer-N PLL is the use of a fractional divider or counter. The frequency resolution of a fractional-N PLL is determined by the width of the accumulator in the fractional divider, as opposed to the reference frequency in an integer-N PLL. The output frequency of a fractional-N PLL can be represented as

$$f_{out} = \frac{f_{ref} \cdot (N + F)}{R} \quad (2.20)$$

where R is the reference divider, N is the integer portion of the divider ratio and F is the fraction value

$$F = \frac{FRAC}{2^K} \quad (2.21)$$

and K and $FRAC$ are the width of the accumulator and fractional coefficient, respectively. The relative resolution of a PLL is the output frequency step from 1LSB change normalized to the output frequency

$$Resolution = \frac{f_{ref}}{f_{out} \cdot 2^K} \approx \frac{1}{N \cdot 2^K} \quad (2.22)$$

For a fixed output frequency, a lower reference frequency or higher reference division increase the integer ratio N . This improves resolution, yet at the expense of worse in-

band phase noise. A wider accumulator width would also increase resolution, but it suffers from lower speed, higher power consumption and greater design complexity. For 1ppb frequency resolution, assuming a reference frequency of 100 MHz and mm-Wave frequency of 200 GHz, a minimum 19-bit fractional divider is required.

The effective frequency resolution could be improved by cascading two fractional-N PLLs. This approach essentially partitions a single wide accumulator into two lower resolution accumulators in each PLL. Note that if the second PLL is integer-N, the frequency resolution is the same as the single PLL design, because the second PLL multiplies both the output frequency and frequency step by the same ratio. It will be shown later that the effective resolution of a fractional PLL can be increased by adding a mixer in the PLL loop.

2.3.2.2 Phase noise PM-AM conversion

Phase noise in an amplitude modulated signal is generally less of a problem, because an ideal envelope detector is insensitive to carrier phase fluctuations. However, when a filter is inserted in the signal path, the phase noise in the transmitter will be converted into amplitude noise due to the frequency response of the filter [55]. This PM-AM effect becomes important when a high-Q resonator is involved. If the carrier frequency is detuned from the resonance or the resonance is asymmetric over the center frequency, the noise sidebands experience unequal attenuation. The resulting unequal sidebands are a combination of amplitude modulation and phase modulation. Phase noise in an unmodulated sine wave can be represented as

$$V(t) = V_0 \cdot \cos[\omega_0 t + \phi_n(t)] \quad (2.23)$$

where $\phi_0(t)$ is a random phase fluctuation. In the frequency domain, a signal without phase noise is a delta function at ω_0 . The random phase fluctuation adds a skirt to both sidebands of the carrier. For simplicity, the phase noise may be modeled as a collection of FM modulation of sinusoidal signals

$$\phi_n(t) = \sum \phi_p \cos(2\pi f_{offset} t) \quad (2.24)$$

where ϕ_p is the peak phase fluctuation at offset frequency f_{offset} from the carrier. The sidebands of an FM modulated signal are given by ordinary Bessel functions of the first kind $J_n(\beta)$ of order n , where β is the ratio of peak frequency deviation to carrier frequency. Assume the offset frequency of interest is f_m and small phase fluctuation, and ignoring higher order terms, $J_{\pm 1}(\beta)$ are the only remaining terms and Eqn. 2.23 is simplified as

$$V(t) \approx \cos(\omega_0 t) - \frac{\phi_p}{2} \{ \cos[(\omega - 2\pi f_m)t] - \cos[(\omega + 2\pi f_m)t] \} \quad (2.25)$$

The conversion from phase modulation to amplitude modulation through a linear filter is dependent on the filter magnitude response at each sideband [55], given by

$$V_{n,AM} = \frac{H(j\omega_u) - H(j\omega_l)}{2H(j\omega_c)} \cdot \frac{\phi_p}{2} \quad (2.26)$$

where $H(j\omega)$ is the transfer function of the filter and the subscripts l and u denote the upper and lower sideband, respectively. In the context of the gas absorption line, the filter is the Lorentzian transmission profile. The Lorentzian function could be linearized to simplify Eqn. 2.25. The slope is approximately the ratio of FWHM to the peak absorption. In the low absorption limit, $H(j\omega)$ is close to 1 and Eqn. 2.25 is rewritten as

$$V_{n,AM} \approx \frac{\alpha_p \phi_p f_m}{2FMWH} \quad (2.27)$$

2.3.2.3 Modulation capability

There are in general three locations where modulation may be injected to a PLL, all of which are illustrated in Figure. 2.11. Since a PLL is simply a frequency multiplier, in which the divided VCO output phase follows the reference. Therefore, if the reference is frequency modulated, the VCO output is also frequency modulated, with the same modulation frequency and N times larger modulation depth. There is an upper limit of the modulation frequency, because in the phase domain, the PLL is a low pass filter from the reference input to the VCO output. The loop bandwidth of the PLL sets the maximum modulation frequency that is applied to the reference. The divider is another popular device to program the PLL output frequency. In order to change the PLL output frequency, the frequency tuning word of the fractional divider should be programmed with a time-varying sequence. This corresponds to the second case in Figure. 2.11. In a binary

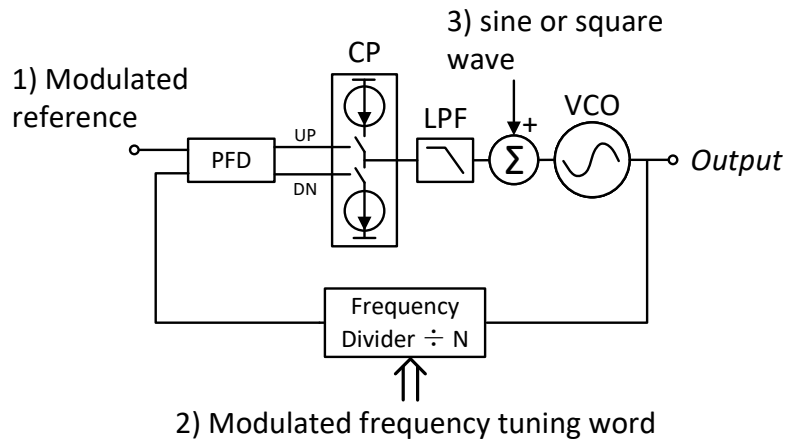


Figure 2.11: Three locations in a PLL where modulation could be applied: the modulated reference, modulated frequency tuning word and direct injection at the VCO input

Fsk modulation, the frequency tuning word is switched between two predefined values, which could be implemented very efficiently. In the situation of a sinusoidal FM, the sine wave is approximated by several discrete sampling points in one period. A smooth transition of the output frequency requires small time step of the sequence, which increases the oversampling ratio. This is essentially a $2N$ -point FSK modulation for oversampling ratio of N . Modulation may also be applied to the VCO tuning port directly, which is the third case shown in Figure. 2.11. The tuning voltage of the VCO is thus the sum of the modulation waveform and the low-pass filtered charge pump output. For such a configuration to work, the modulation frequency should be designed lower than the PLL loop bandwidth.

2.3.3 Low phase noise TX

2.3.3.1 LO generation

The phase noise of a multiplier based signal source is lower than a PLL up to a few times the PLL bandwidth, while the phase noise floor of a PLL is lower. However, when a PLL is used to generate a modulated signal, the frequency offset of interest is usually below the PLL bandwidth. A conceptual phase noise comparison between multiplier based and PLL based signal source is conducted in Figure. 2.12 (a) and (b). The passive nonlinear devices in the frequency multipliers generate little additive phase noise. An ideal frequency multiplier degrades the phase noise by $20\log(N)$, where N is the multiplication ratio. A

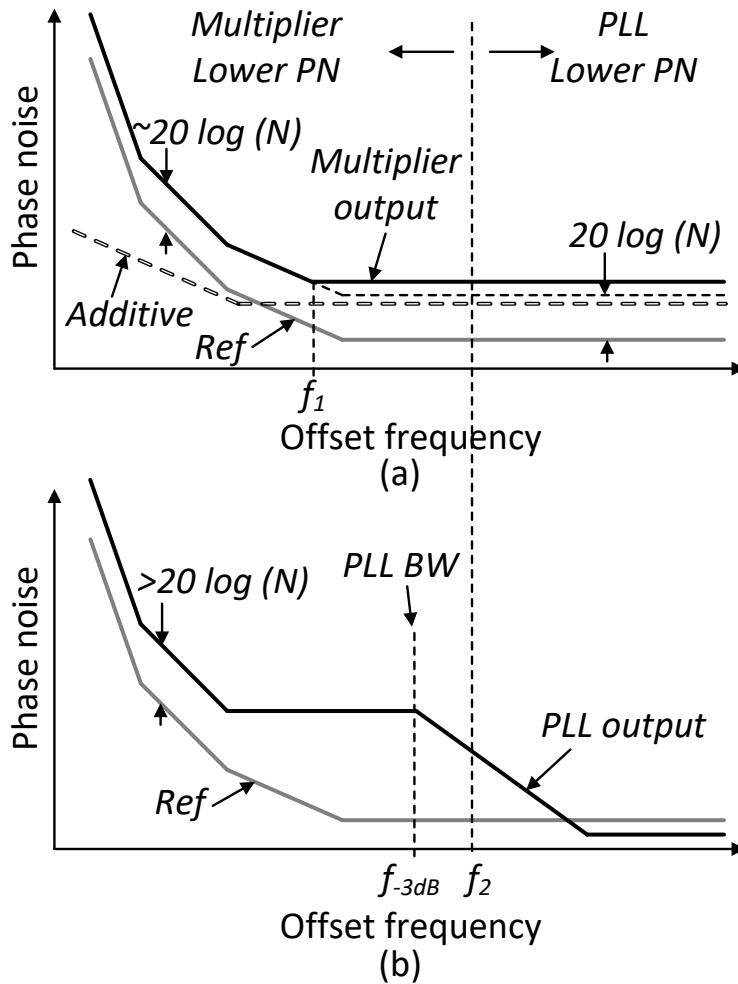


Figure 2.12: Comparison of phase noise between a multiplier based (a) and a PLL based (b) signal source.

passive frequency multiplier usually follows this relationship at a low frequency offset. The additive phase noise of a multiplier at low offset frequency is usually negligible. The slope of the additive noise at low frequency offset is $1/f$ due to the flicker phase fluctuations, whereas the phase noise slope of oscillator normally exceeds $1/f$. The hollow dashed wire in Figure. 2.12 (a), that represents the additive phase noise, stays well below the output noise in solid black trace. At higher frequency offset, thermal noise sometimes limits the phase noise performance, if the conversion loss of the multiplier is large. This is illustrated in Figure. 2.12 (a) for frequency greater than f_1 . The following example may explain how the conversion loss affects the phase noise floor. Suppose a reference with 0 dBm power is fed to a frequency doubler with 10 dB conversion loss. Further, assume that the reference

phase noise floor is thermal noise limited to is -177dBc/Hz , for equal contribution of phase and amplitude noise from the -174dBm/Hz thermal noise. The phase noise floor from an ideal doubler is -171dBc/Hz , $20\log(2)$ dB higher due to doubling the frequency. However, since the output power of the doubler is -10dBm , the thermal noise limited phase noise floor would -167dBc/Hz . It should be noted that this only happens with a very low phase noise floor source. The ideal multiplier phase noise and the thermal noise limited phase noise floor intersects at f_1 shown in Figure. 2.12 (a).

The output phase noise of a PLL is sketched in Figure. 2.12 (b). There is a plateau or hump at an offset frequency around the PLL bandwidth f_{-3dB} . At very high offset frequency, the phase noise of the VCO dominates, and the reference noise dominates at low frequency. During the transition around the PLL loop bandwidth, the noise from the reference, the phase detector, the loop filter and the VCO collectively contribute to the output phase noise. As a result, the phase noise of a PLL around the loop bandwidth is high. The frequency f_2 , which is larger than the PLL bandwidth, is where the phase noise of a multiplier and a PLL are equal. Below this frequency, the multiplier based solution is advantageous, because it is free of the noise from the phase detector and VCO. Even at very low offset frequency, where the PLL noise is generally assumed to follow the reference by $20\log(N)$, the noise of the phase detector could not be neglected for large multiplication ratio N with a low noise reference.

2.4 Detector

The receiver of the molecular clock is an envelope detector. Several common detector architectures are sketched in Figure. 2.13 (a-c). In the non-coherent or direct detection, an envelope detector at RF directly demodulates the envelope information into baseband. A homodyne detector makes use of a mixer that downconverts the AM sidebands to baseband. In the more sophisticated heterodyne detector, the carrier frequency f_c is first downconverted to an intermediate frequency f_{IF} , then followed by an envelope detector. There are some unique considerations in the molecular clock, in addition to well known differences between the three architectures. The direct detection method is the simplest

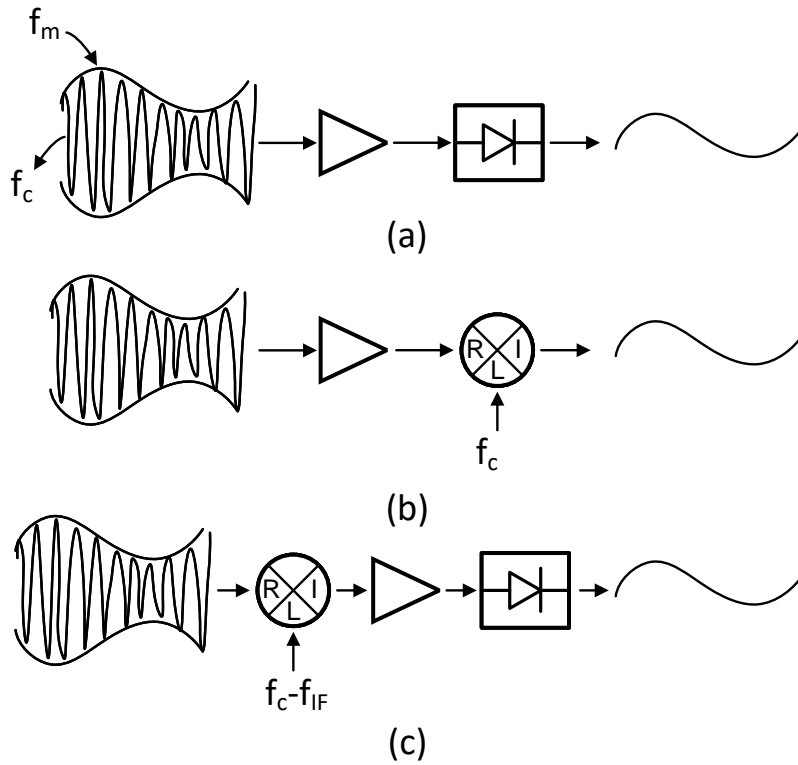


Figure 2.13: (a) A non-coherent envelope detector. (b) Homodyne envelope detector. (c) Heterodyne detector. Filters are not shown for simplicity.

among all, yet the sensitivity of the detectors would limit the SNR for small input signals. Therefore, an amplifier must be inserted before the detector to boost the signal strength, if the path loss from the Tx to Rx is large. Note that amplification at mm-Wave is difficult, depending on the received power and detector input noise, the available gain with the current technology may be insufficient to achieve good SNR. The sensitivity of heterodyne and homodyne receivers are much higher than non-coherent detectors, and an ideal homodyne receiver shows 3-dB higher sensitivity than a heterodyne system. In a homodyne system, the uncorrelated phase noise in the received signal and the LO is converted to amplitude noise, significantly degrades the SNR. For least conversion from phase noise to amplitude noise, the LO in the homodyne system must be completely coherent with the received signal. In the heterodyne detector, the phase noise of the LO is not important. It contributes to the IF phase noise, which will be rejected by the following envelope detector. By downconverting the RF to a lower intermediate frequency,

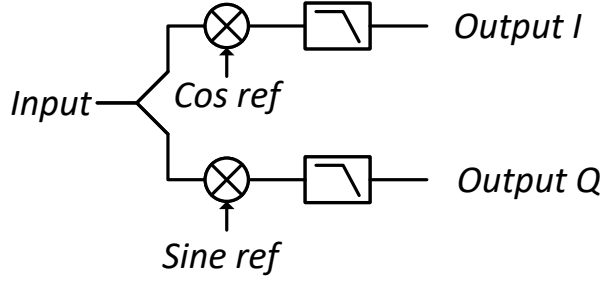


Figure 2.14: The generic block diagram of a lock-in amplifier consists of a pair of mixers followed by narrow bandwidth low-pass filters. The in-phase and quadrature outputs are further processed to obtain the magnitude and phase of the input signal.

signal amplification becomes easier and more power efficient.

2.5 Lock-in Amplifier

A lock-in amplifier is commonly used to extract signals with known frequency from a noisy environment. It generates the amplitude and phase of the signal under test with respect to the reference. The basic block diagram of a lock-in amplifier is shown in Figure. 2.14. The input signal is multiplied by the in-phase and quadrature reference. The mixer output is then low-pass filtered to average out noise and reject the second harmonic of the reference. Mathematically, a signal with random noise may be represented as

$$S(t) = A\cos(\omega t) + n(t) \quad (2.28)$$

and the reference signal with the same frequency but an unknown phase is

$$R_0(t) = A_r\cos(\omega t + \phi_0) \quad (2.29)$$

The quadrature version of the reference is simply

$$R_{90}(t) = A_r\cos(\omega t + \phi_0 + \pi/2) = -A_r\sin(\omega t + \phi_0) \quad (2.30)$$

The in-phase and quadrature output of the mixer are

$$I(t) = A \cdot A_r \frac{\cos(2\omega t + \phi_0) + \cos(\phi_0)}{2} + n(t) \cdot A_r\cos(\omega t + \phi_0) \quad (2.31)$$

$$Q(t) = -A \cdot A_r \frac{\sin(2\omega t + \phi_0) + \sin(\phi_0)}{2} - n(t) \cdot A_r\sin(\omega t + \phi_0) \quad (2.32)$$

The second harmonic is easily removed by the following low pass filter. The recovered signal magnitude is equal to the square root of the I and Q channels

$$|S_R(t)| = \sqrt{I(t)^2 + Q(t)^2} \quad (2.33)$$

with phase

$$\angle S_R(t) = -\tan^{-1} \frac{Q(t)}{I(t)} \quad (2.34)$$

The noise term $n(t)$ is modulated by the reference frequency, so that its average value

$$\mathbb{E}[n(t) \cdot \cos(\omega t + \phi_0)] = \mathbb{E}[n(t)] \cdot \mathbb{E}[\cos(\omega t + \phi_0)] \quad (2.35)$$

$$\mathbb{E}[n(t) \cdot \sin(\omega t + \phi_0)] = \mathbb{E}[n(t)] \cdot \mathbb{E}[\sin(\omega t + \phi_0)] \quad (2.36)$$

is equal to zero if noise and reference are uncorrelated. Note that the distribution of the noise is irrelevant, because virtually all kinds of noise have finite power, at least within the observation period. In practice, the expectation operation is implemented either with an analog low pass filter or a digital moving average. The noise reduction is therefore only finite, since the noise inside the filter bandwidth around the reference frequency is not removed. Effectively, the lock-in process is a high-Q band-pass filter with center frequency equal to the reference frequency. The bandwidth of the filter is proportional to the cutoff frequency of the low pass filter. Therefore, a higher signal-to-noise ratio is achieved with very low filter bandwidth or number of averaging in the digital domain. If the lock-in amplifier is used in a feedback loop or real-time observation is required, the filter bandwidth cannot be designed to be arbitrarily low. This is one of the practical limits of lock-in amplifiers.

The heart of a lock-in amplifier is the mixer or phase-sensitive detector (PSD), which can be implemented in either analog or digital domains.

2.5.1 Analog mixers

Analog mixers could be categorized into active or passive, depending on the bias current in the transistor. Figures. 2.15 (a) and (b) show the typical single-balanced active and passive mixer. The bottom transistor in the active mixer works as a transconductance stage and provides dc bias current for the LO switches. As a result, active mixers have

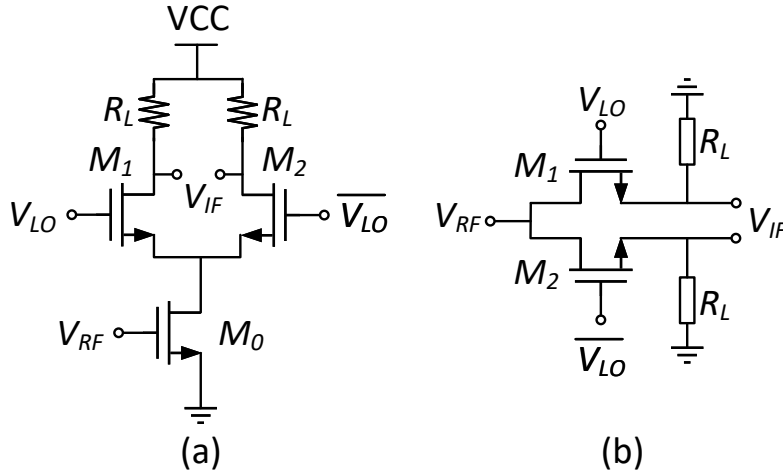


Figure 2.15: (a) A simple active mixer and (b) passive mixer.

higher conversion gain than passive counterparts, but they suffer from higher flicker noise and dc offset.

The flicker noise in a MOS transistor can be modeled empirically [56]

$$S_{I_d}(f) = \frac{KF \cdot I_d^{AF}}{fC_{ox}WL} \quad (2.37)$$

where KF and AF are process dependent parameters. In a three-transistor single-balanced mixer, the output flicker noise is dominated by the LO switches if the LO frequency is higher than the flicker corner. The noise current from the bottom transistor M_0 is chopped by the LO switches M_1 and M_2 , so that the low frequency flicker content is shifted to the LO sidebands. The exact analysis of output flicker noise from LO switches is complicated: each LO switch rotates between cutoff, saturation and triode regions during one period. Since the transistor still carries substantial current during the triode region, the current noise may be approximated by the same expression as in the saturation region. Note that such approximation essentially overestimates the noise contribution. It is also assumed that there is no current noise when the LO switch is off. Therefore, each LO switch contributes to the output voltage noise during half of a period, as is shown in Figure. 2.16, where \bar{I}_n represents the current flicker noise defined in Eqn. 2.37. The sum of the uncorrelated noise from each transistor is $\sqrt{2}$ times larger

$$\bar{V}_n = \frac{\sqrt{2}}{2} \frac{KF \cdot I_d^{AF}}{fC_{ox}WL} R_L \quad (2.38)$$

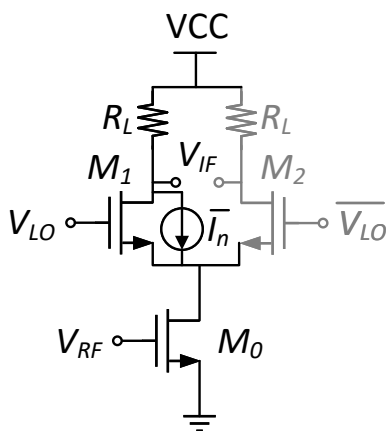


Figure 2.16: Illustration of the current noise in an active mixer when one of the LO switch M_2 is turned off.

Since the output noise increases with drain current, it is attractive to use passive mixers, which presents little $1/f$ noise with zero dc bias current. In practice, the time-varying zero-mean drain current of the transistor in the passive mixer still contributes to the output noise [57], but is substantially lower than active mixers.

The offset in the mixer comes from various sources: including transistor mismatch, LO duty cycle mismatch, LO leakage, etc. The effect of transistor mismatch is different in active and passive mixers. Several sources of mismatch could contribute to output offset: threshold voltage ΔV_{th} , transistor size $\Delta(W/L)$ and baseband load resistor ΔR_L . In the active mixer, assuming the LO swing is a fast switching square wave, the mismatch from V_{th} contributes negligible timing error. Due to the large LO swing, the transistor can be fully turned on; thus, the transistor size mismatch would have little effect. In such case, the current from the M_0 is directed to either branch, exactly at the transition edge of the LO. To evaluate the output offset, the input of the RF path is set to a dc voltage. The average current of each branch is half of I_{d0} so that output offset only comes from the resistor matching, yielding

$$V_{os,act} = \frac{1}{2} V_i \cdot gm_0 \Delta R_L \quad (2.39)$$

where the $1/2$ factor comes from the fact that each transistor is conducting for half of a period. For a single-balanced passive mixer, during each phase when one of the transistors is turned on, a voltage divider is formed with R_{on} and R_L . Here, due to the low operating

frequency in a lock-in amplifier, resistor loading is more practical than a large capacitor. The maximum offset voltage happens when R_{on} is larger and R_L is smaller

$$\begin{aligned}
V_{os,pas} &= \frac{V_i}{2} \left(\frac{R_{on} + \frac{\Delta R_{on}}{2}}{R_L - \frac{\Delta R_L}{2}} - \frac{R_{on} - \frac{\Delta R_{on}}{2}}{R_L + \frac{\Delta R_L}{2}} \right) \\
&= \frac{V_i}{2} \left(\frac{R_{on} + \frac{\Delta R_{on}}{2}}{R_L - \frac{\Delta R_L}{2}} - \frac{R_{on} - \frac{\Delta R_{on}}{2}}{R_L - \frac{\Delta R_L}{2}} + \frac{R_{on} - \frac{\Delta R_{on}}{2}}{R_L - \frac{\Delta R_L}{2}} - \frac{R_{on} - \frac{\Delta R_{on}}{2}}{R_L + \frac{\Delta R_L}{2}} \right) \\
&= \frac{V_i}{2} \left(\frac{\Delta R_{on}}{R - \frac{\Delta R_L}{2}} + (R_{on} - \frac{\Delta R_{on}}{2}) \frac{\Delta R_L}{(R_L - \frac{\Delta R_L}{2})(R_L + \frac{\Delta R_L}{2})} \right) \\
&\approx \frac{V_i}{2} \left(\frac{\Delta R_{on}/R_{on} + \Delta R_L/R_L}{R_L/R_{on}} \right) \quad \text{if } \Delta R_L \ll R_L \text{ and } \Delta R_{on} \ll R_{on} \\
&\approx \frac{V_i}{2} \left(\frac{\Delta R_{on}}{R_L} \right) \quad \text{if } R_L \text{ matching is much better}
\end{aligned} \tag{2.40}$$

The last approximation is based on the observation that R_{on} matching is dependent on transistor size W/L and threshold voltage V_{th} , which in general is less accurate than resistor matching. The output offset voltage is a scaled version of the input. The scaling factor is the absolute mismatch of R_{on} divided by the load resistor. To minimize the offset, apart from optimizing layout for device matching, reducing the on resistance R_{on} and increasing the load resistance R_L linearly reduce the dc offset. The offset of a passive mixer can be made much smaller than a passive mixer because ΔR_{on} and $1/R_L$ can be designed much smaller than ΔR_L and g_m , respectively. Literature also supports this observation [50], where the offset of a passive harmonic reject mixer reached as low as 10 μV by using a very large R_L .

LO leakage induced offset is common at RF frequency. In a RF mixer, the finite isolation due to parasitic capacitor from LO to RF mixes with LO and the mixing of identical signals causes a dc term at the output. This effect is exaggerated by device mismatches [54]. In the lock-in amplifier, where the operating frequency is in the kHz to MHz range, leakage becomes insignificant.

From the above analysis, it seems that the passive mixer is the perfect candidate for lock-in amplifier. However, several other considerations should be taken into account.

Due to its low gain, the noise and dc offset of the subsequent amplifier stage are not suppressed, which then dominates the system noise and dc performance. The need of low offset and low flicker noise amplifier makes on-chip integration difficult [49, 50]. Second, the ideal waveform of CMOS mixers is a square wave to optimize conversion gain and noise figure. In the derivation of lock-in SNR in Section. 2.5.3, it was assumed that both the reference and the signal share the same waveform. The use of square wave reference in conjunction with a sine wave DUT signal causes the noise at odd harmonics in the signal to be folded into dc. Harmonic rejection mixers should be adopted [50].

2.5.2 Digital mixers

Mixers in the analog domain can never be made perfect: they inevitably suffer from input dc offset, finite dc gain and limited harmonic rejection. The dc offset issue may be understood by revisiting Eqn. 2.35 and is reproduced here

$$\mathbb{E}[n(t) \cdot \cos(\omega t + \phi_0)] = \mathbb{E}[n(t)] \cdot \mathbb{E}[\cos(\omega t + \phi_0)]$$

The right-hand side of the equation consists of the expectations of the noise and reference. In practice, there is some dc offset from the signal under test, either from the dc bias of an amplifier or from other undesired dc offset voltage. The dc offset from the signal under test could be lumped into the noise term $n(t)$, which will be canceled out if the reference is zero mean. In analog mixers, this is only achieved with perfect device matching and LO timing. Any mismatch from the mixer and LO corrupts the output by adding a dc offset voltage, because the desired information also lies at dc. When the input signals are sampled and converted in the digital domain, numerical manipulations could be conveniently performed with low power digital circuits. For example, the reference may be generated from a lookup table, without any dc offset. With accurate accumulation in the digital domain (to the precision of the input data), the dc gain of a digital integrator is infinite in theory, as long as the accumulator is not overflowed. This can be easily satisfied in the feedback system, where the average error signal fed to the accumulator is constrained to be zero. Limited by device matching, harmonic rejection in analog mixers is typically about 30-35 dB [58]. Higher harmonic rejection is possible by means of calibration in the digital domain [59],

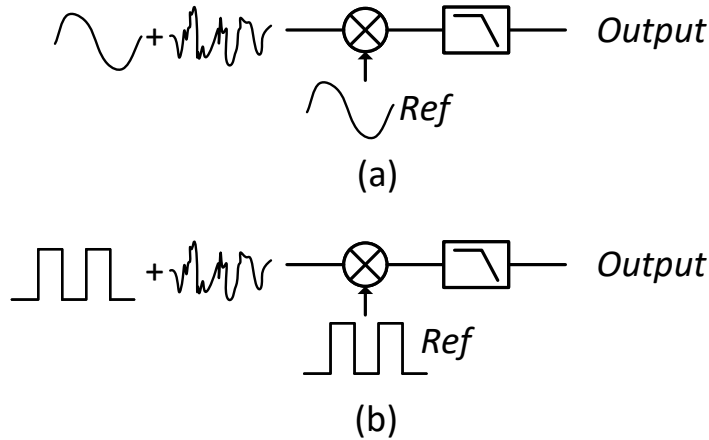


Figure 2.17: Simplified block diagram of lock-in amplifiers when the input and reference are in phase, with (a) sine wave input and (b) square wave input

but in the application of a low frequency lock-in amplifier, direct digitizing would be more efficient.

Of course, the process of data conversion from analog to digital is far from perfect, which contaminates the analog signal with the ADC quantization noise, input offset, etc. However, these issues are readily addressed, where the quantization noise can be negligible with sufficient pre ADC gain and the input offset could be easily eliminated in the digital domain.

2.5.3 SNR

One of the most important metrics of the molecular clock is SNR. Here, the SNR is calculated at the output of the lock-in amplifier. In this application, the phase shift between in the input and the reference is negligible, because the group delay at the modulation frequency is small. Thus, the quadrature path in the generic lock-in amplifier can be omitted. The input and reference signals could be either sine wave or square, as shown in Figure. 2.17(a) and (b), yet the architecture remains the same for both schemes. To compare the SNR after the lock-in amplifier, it is assumed that the reference is an ideal clean sine or square wave and the input signals have the same signal and noise power. It is further assumed, without loss of generality, that the signal and the reference are in phase. The sine wave reference has a magnitude of 1V peak, and the square wave

reference toggles between -1 and 1.

The power of a signal $f(t)$ is defined as:

$$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f^2(t) dt \quad (2.41)$$

Suppose the square wave signal toggles between $-A$ and A ; its power is just A^2 . The dc voltage at the lock in amplifier output is A , since the lock-in process reverses the $-A$ voltage into A and keeps A unchanged. For a sine wave with the same power, its amplitude should be $\sqrt{2}A$. The lock-in output is the dc component of $\sqrt{2}A \cos(\omega t)$ multiplied by $\cos(\omega t)$, and is easily calculated to be $\sqrt{2}/2A$.

The calculation of noise power is more conveniently done in the frequency domain. The output noise is computed as the input noise power spectral density (PSD) convolved with the reference PSD. For simplicity, assume that input noise is purely white. The power spectral density (PSD) of the white noise is $\overline{V_{ni}^2}$. The PSD of a signal $x(t)$ is defined as the Fourier transform of autocorrelation of $x(t)$ and is also equal to the magnitude square of $\hat{x}(f)$:

$$\overline{S_{xx}}(f) \triangleq |\hat{x}(f)|^2 \quad (2.42)$$

where $\hat{x}(f)$ is the Fourier transform of $x(t)$:

$$\hat{x}(f) \triangleq \int_{-\infty}^{\infty} e^{-i2\pi ft} x(t) dt \quad (2.43)$$

The Fourier transform of a periodic square wave consists of pulses at only odd harmonics of cycle frequency f_{sq} . The magnitude of the Fourier coefficients of a square wave with peak to peak of 2 is given by

$$|X_m| = \frac{2}{\pi|m|}, \quad m = 2n - 1, \quad n \in \mathbb{Z} \quad (2.44)$$

where m denotes the m^{th} harmonic of the fundamental frequency f_{sq} . Note that it is the two-sided spectral representation. The PSD of the square wave is plotted in Figure. 2.17(b). Hence, the PSD of noise at the lock-in output is the sum of each harmonic multiplied by the input noise PSD [54]:

$$\overline{V_{no,sq}^2} = \overline{V_{ni}^2} \sum_{n=-\infty}^{\infty} \left(\frac{2}{\pi(2n-1)} \right)^2 = \overline{V_{ni}^2} \quad (2.45)$$

Similarly, the PSD of a sine wave is $1/4$ at f and $-f$ as shown in Figure. 2.17, and the output noise PSD can be found:

$$\overline{V_{no,sine}^2} = \overline{V_{ni}^2} \cdot 2 \cdot \frac{1}{4} = \frac{\overline{V_{ni}^2}}{2} \quad (2.46)$$

With the signal power and noise PSD available, the SNR of sine and square wave are given by:

$$SNR_{sq} = \frac{A^2}{\overline{V_{no,sq}^2} \cdot B} = \frac{A^2}{\overline{V_{ni}^2} \cdot B} \quad (2.47)$$

and

$$SNR_{sine} = \frac{(\frac{\sqrt{2}}{2}A)^2}{\overline{V_{no,sine}^2} \cdot B} = \frac{A^2}{\overline{V_{ni}^2} \cdot B} \quad (2.48)$$

where B is the bandwidth of the low pass filter following the lock-in amplifier.

Chapter 3

Design of a High Power Millimeter Wave Source

3.1 Introduction

The core of the transmitter of a molecular clock is essentially a millimeter wave signal generator. There are several measures of a good signal source, including output power, dc power consumption, phase noise, etc. For a voltage controlled oscillator(VCO), tuning range is also of importance. In a molecular clock, however, the mm-Wave signal only covers a narrow modulation bandwidth around the molecular spectral line. The tuning range is used to cover the variation from process, supply voltage and temperature (PVT) to ensure the frequency lock always occurs. A wide tuning range at millimeter wave frequency is quite challenging in general, but fortunately, not necessary in this application. Since the VCO always comes with a PLL to bring the mm-Wave frequency close to the spectral line, the phase noise of the VCO is shaped by the PLL loop filter. By carefully engineering the loop bandwidth and reference frequency, the phase noise at low offset frequency could be greatly reduced. Ideally, without the noise from the phase frequency detector (PFD), the phase noise at the PLL output at an offset frequency lower than the loop bandwidth follows reference noise.

The output power and efficiency at millimeter wave frequency is especially challenging due to the limitation of the process by which the signal generator is fabricated. As a result, a good portion of the total dc power is consumed by the mm-Wave signal generation [50].

The CMOS process, massively produced in consumer electronics, has become popular in millimeter-wave applications due to its low cost. However, the high frequency performance of the CMOS device is quite limited. One of the most important specifications, maximum oscillation frequency (f_{max}), that determines the highest frequency of the device being active is quite low compared to the potentially wide millimeter wave frequency band. Oscillators do generate signals higher than the f_{max} of the device. In this case, the output signal come from the harmonics produced from the non-linearity of the device instead of the fundamental component. The harmonics are generally smaller than the fundamental in typical MOS transistors. The low f_{max} not only limits the maximum fundamental frequency generation, it also reduces the output power of fundamental oscillators.

Design techniques have been proposed to optimally extract the oscillation power out of a given transistor. It has been successfully applied to oscillator designs approaching f_{max} for decades. The component value of the embedding network is completely determined by the network parameter of the active device. To achieve high output power, the most straightforward approach is to increase the size of the active device. It will be shown later, that the inductance value scales inversely proportional to the active device size. In other words, an oscillator that generates high output power requires a very small inductance. The design technique has originally assumed a lossless embedding network composed of lossless passive elements. Not until recently has it been discovered that the loss of the passive on-chip components alters the operation condition from the lossless assumption. From the practical implementation point of view, the on-chip inductor is quite lossy with small inductance. The high loss from the inductor eventually cancels out the benefit of using a large transistor. Techniques that use a power combiner to collect the output power from multiple sources are popular when the output power requirement exceeds that of a single source. However, it is area-hungry and combiners are rarely 100% efficient.

In this work, an approach to increasing the output power of a single oscillator is proposed without causing an implementation issue of the inductor. Furthermore, two oscillators are directly combined to double the output power without combiner loss.

3.2 Theory of Stacked-FET Oscillator

In this section, a qualitative analysis of scaling in oscillator design is first provided, followed by a thorough small signal analysis of the Pi-embedding oscillator. The embedding network value of a single transistor is calculated with the small signal model in Figure. 3.4. The value of embedding components is expressed in terms of the transistor small signal parameters, in contrast to network parameters commonly used in previous works [60–62]. Finally, the proposed stacked-FET solution is analyzed and compared to a common-source stage to show its potential to alleviate the scaling difficulties.

3.2.1 Scaling of Passives

Given an arbitrary oscillator design, scaling the size of active devices by N , the impedance of passive components by $1/N$ and current by N results in a new oscillator with N times more power consumption and $10\log(N)$ dB lower phase noise [63]. Such scaling leads to N times larger capacitors and N times smaller inductors. The dashed line in Figure. 3.1 illustrates the change in output power and inductance caused by the scaling. The inductance being inversely proportional to normalized size is problematic in millimeter wave integrated circuit oscillators. On one hand, the low output power of the transistor demands operation under near optimum conditions, which limits the choice of inductance. On the other hand, on-chip inductors are generally quite lossy. The quality factor peaks at some inductance at a certain frequency. We investigate the behavior of a single turn inductor in electromagnetic simulators as an example. Figure. 3.2 plots the quality factor versus inductance where the line width and diameter of the inductor are varied. At the frequency of 180 GHz, each inductor maps to a gray circle in the 2-D $Q - L$ space. The solid line corresponds to the boundary of maximum quality factor of each inductance. The inductors in Figure. 3.2(a) and (b) share the exact same dimensions except in Figure. 3.2(b) via stacks from the top metal line to the first metal layer above the ground plane are included. We incorporate the profile of Figure. 3.2(b) into the oscillator design and plot the output power and inductance with respect to normalized power with solid line in Figure. 3.1. The transistor size is normalized to a 16- μm wide NMOS device. Only the loss from the inductor is considered. As the normalized size exceeds 6, the output power

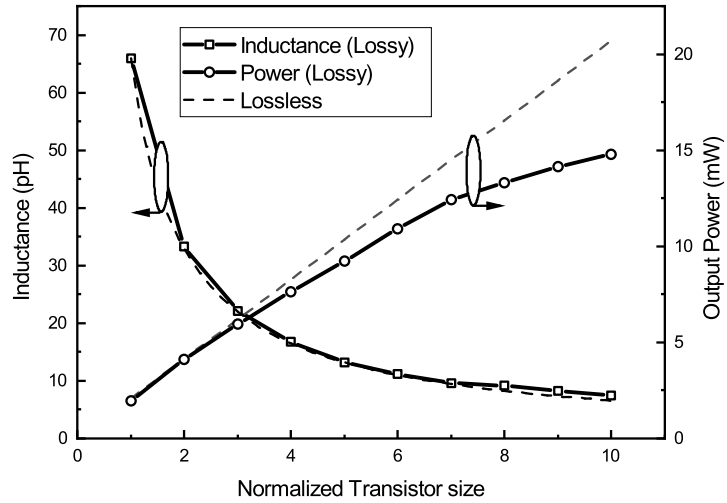


Figure 3.1: Power and inductance scaling of a Pi-embedding oscillator

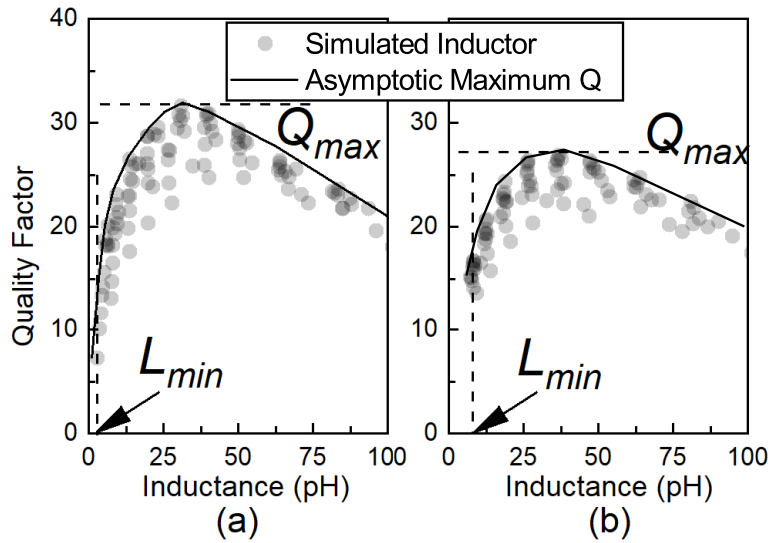


Figure 3.2: Simulated Q versus L for various inductor geometries (a) without (b) with via stack.

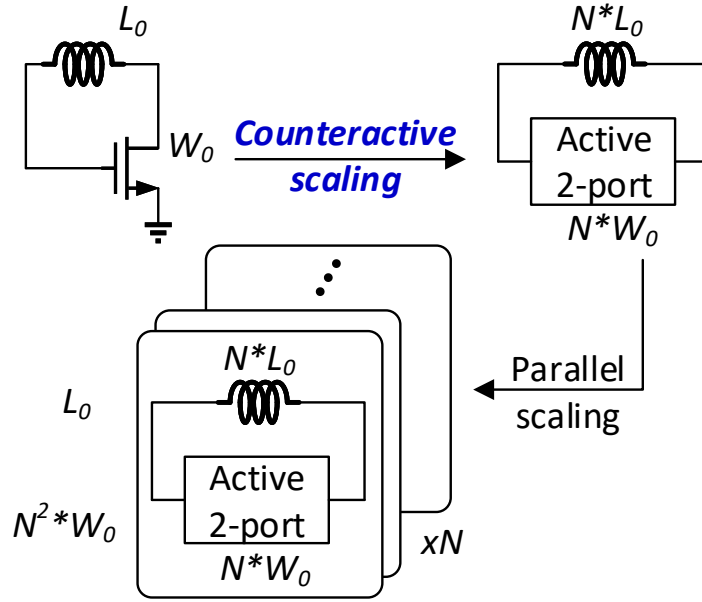


Figure 3.3: Proposed scaling consisting of parallel and counteractive scaling

significantly deviates from ideal scaling and eventually saturates due to the loss of the inductors. More importantly, at the same time, an inductance smaller than 10 pH shall be needed. In practice, the inductance of any physical implementation should also include the parasitic inductance from the vias connecting the transistor and inductor should also be included, making the core inductor even smaller and more lossy. To build such small inductor with good accuracy is another challenge at millimeter wave frequencies.

The scaling of inductance imposes a fundamental limit on the maximum width of the active device and therefore output power, due to the minimum achievable inductor. This trade-off between output power and passive value is a direct consequence of paralleling devices. Despite the fact that higher output power demands a larger total transistor width, it is not the only option to parallel these devices. Consider a scaling scheme in Figure. 3.3, in which a high power oscillator is scaled from two steps. An oscillator design denoted with size W_0 and inductance L_0 is set as reference. The width is chosen such that L_0 resides in a region in Figure. 3.2 of the highest quality factor. In the first step, a hypothetical active two-port network scales both the power and inductance by a factor of N . The two-port network consists of a total transistor width of $N \cdot W_0$, but

are arranged differently from paralleling. The detail of this scaling will be discussed in the following sections. Assuming such counteractive scaling is readily available, a second parallel scaling is applied, resulting the width of each transistor the inductance scaled by N and $1/N$, respectively. Each step scales the total transistor width by N , leading to an N^2 larger transistor. However, the scaling factors N and $1/N$ of the inductor cancels out in the end, which maintains a constant inductance L_0 for the overall oscillator. Following the proposed scaling scheme, we obtain an N^2 higher power oscillator than the reference design with the same inductor. More importantly, only one set of embedding networks is involved in the process, which significantly reduces the area.

3.2.2 Small-signal Analysis of Pi-Embedded Oscillators

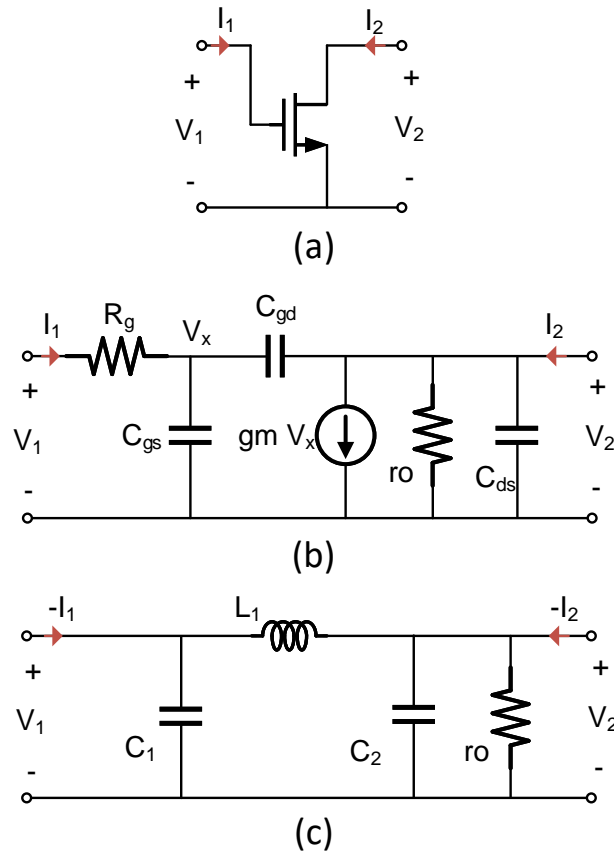


Figure 3.4: The small signal model of the transistor (a) and a generic Pi-embedding network (b)

The active device in a Pi-embedded oscillator in common-source configuration is rep-

resented by a two-port network shown in Figure. 3.4(a). Consider a small signal model of the MOS transistor in Figure. 3.4(b). A series gate resistor is included to model the parasitic resistance of the wiring. This resistor is critical to the high frequency performance of the transistor. In fact, the maximum oscillating frequency f_{max} , one of the greatest limitations to the gain of transistor arises from this resistor. As will be evident later in this section, the optimum voltage gain from Port 1 to Port 2 and thus the embedding network value depend on this resistor. The parasitic capacitance C_{sb} between source and body is ignored in the following derivation, since a PI-embedded oscillator is normally common-source configured.

Generally, the calculation of the embedding network involves two steps [60, 61, 64]. The first step is to find the optimum voltage gain A_{opt} of the two ports defined as $\frac{V_2}{V_1}$ when maximum added power is achieved. The complex gain is also denoted as $A_r + j \cdot A_i$. In the second step, the embedding network is obtained by satisfying a set of KCL equations with y-parameters and A_{opt} as their coefficients. In our case, the extra internal node V_X complicates A which now becomes

$$A = \frac{V_2}{V_X} \frac{V_X}{V_1}.$$

Without loss of generality, we instead define A^\dagger as

$$A^\dagger = \frac{V_2}{V_X}, \quad (3.1)$$

The real and imaginary parts of A^\dagger are denoted by A_r^\dagger and A_i^\dagger , respectively. While V_2 is related to V_X by (3.1), the gain from V_1 to V_X is found by writing KCL at node V_X

$$V_X = \frac{V_1}{1 + A_i^\dagger \omega R_{gs} C_{gd} + j \omega R_{gs} (C_{gs} + C_{gd}(1 - A_r^\dagger))}. \quad (3.2)$$

This is essentially a low pass response and if we assume A_i is small which shall be validated later, we obtain the cutoff frequency

$$\omega_{V_X} \approx \frac{1}{R_{gs}(C_{gs} + C_{gd}(1 - A_r^\dagger))} \quad (3.3)$$

When the operating frequency approaches ω_{V_X} , the low pass filter contributes a phase change of

$$\phi_{V_X} = \tan^{-1}\left(\frac{\omega}{\omega_{V_X}}\right). \quad (3.4)$$

Indeed (3.4) reveals that the phase shift between V_1 and V_2 is contributed by R_G . Without R_G , the optimum phase from V_1 to V_2 is always 180° .

We then seek to find the added power in terms of A^\dagger and the circuit components. Recall the definition of added power

$$P_{add} = -\frac{1}{2}\Re(V_1 I_1^* + V_2 I_2^*). \quad (3.5)$$

where the symbol "*" represents conjugate operation. The currents flowing into the two ports I_1 and I_2 are found by

$$\begin{aligned} I_1 &= \frac{V_1 - V_X}{R_{gs}} \\ I_2 &= g_m V_X + \frac{(A_r^\dagger + jA_i^\dagger)V_X}{r_o} \\ &\quad + (V_X - (A_r^\dagger + jA_i^\dagger)V_1)sC_{gd} + (A_r^\dagger + jA_i^\dagger)V_1sC_{ds}. \end{aligned} \quad (3.6)$$

The last term of I_2 in (3.6) corresponds to the current of a capacitor C_{ds} . It will be excluded in the following calculation for simplicity because it does not consume any power and can be merged into C_2 in the embedding network in Figure. 3.4(b).

The added power may be expressed in terms of V_1 , A_r^\dagger and A_i^\dagger and circuit components by substituting (3.2) and (3.6) into (3.5). Since the small signal model is linear, we can normalize the added power to $|V_X|^2$

$$\begin{aligned} \frac{P_{add}}{|V_X|^2} &= \frac{1}{2} \left[-A_r^{\dagger 2} \left(\frac{1}{r_o} + R_{gs} C_{gd}^2 \omega^2 \right) + A_r^\dagger \left(-g_m + 2R_{gs} C_{gd}^2 \omega^2 + 2R_{gs} C_{gs} C_{gd} \omega^2 \right) \right. \\ &\quad \left. - A_i^{\dagger 2} \left(\frac{1}{r_o} + R_{gs} C_{gd}^2 \omega^2 \right) - R_{gs} (C_{gs} + C_{gd})^2 \omega^2 \right]. \end{aligned} \quad (3.7)$$

Equation (3.7) is a quadratic function with variables A_r^\dagger and A_i^\dagger . The coefficients of both quadratic terms are negative and there is no $V_r^\dagger \cdot V_i^\dagger$ term. It means that the maximum of P_{add} occurs when the A_r and A_i sit at their axis of symmetry, which is given by

$$\begin{aligned} A_{r,opt}^\dagger &= \frac{r_o \left[-g_m + 2R_{gs} C_{gd} (C_{gs} + C_{gd}) \omega^2 \right]}{2 + 2R_{gs} C_{gd}^2 r_o \omega^2} \\ A_{i,opt}^\dagger &= 0. \end{aligned} \quad (3.8)$$

An important observation from (3.8) is that the phase of A_{opt}^\dagger is always 180° regardless of its magnitude. This property could simplify our derivation later. Further assuming small

R_{gs} or low operating frequency. $A_{r,opt}$ in (3.8) may be reduced to

$$\begin{aligned} A_{r,opt}^\dagger &\approx -\frac{g_m r_o}{2} \\ A_{i,opt}^\dagger &= 0. \end{aligned} \quad (3.9)$$

Equation (3.9) essentially implies that A_{opt} is reached only when the load is matched to r_o and the parasitic capacitance is completely cancelled. It is readily justified by the fact that, when R_{gs} is ignored, no power will feed back to the input of the transistor and that maximum output power is naturally satisfied with matched impedance.

Figure. 3.4(c) shows a typical two-port Pi-embedding network. In order to satisfy the oscillation condition, the sum of the current flowing into the two-port embedding network and the active device should be zero under the same excitation. As will be evident immediately, the real and imaginary parts of the KCL equation at Port1 only consists of two unknown variables C_1 and L_1 . We will solve this set of equations first and back substitute the solution to KCL equations at Port 2 to solve for C_2 and R_L . The real and imaginary part of KCL at Port1 is given by

$$\begin{aligned} &LC_1\omega^2(A_i^\dagger\omega R_{gs}C_{gd} + 1) + \\ &L\omega^2(C_{gs} + C_{gd}(1 - A_r^\dagger)) - 1 + A_r^\dagger - A_i^\dagger R_{gs}C_{gd}\omega = 0 \end{aligned} \quad (3.10)$$

and

$$\begin{aligned} &LC_1\omega^3 R_{gs}(C_{gd}(1 - A_r^\dagger + C_{gs})) \\ &- LA_i^\dagger C_{gd}\omega^2 - \omega R_{gs}(C_{gs} + C_{gd}(1 - A_r^\dagger) + A_i^\dagger) = 0, \end{aligned} \quad (3.11)$$

respectively. The set of equations is actually a linear function of the unknowns. Applying the approximated A_{opt} from (3.9) to the solution, a very compact expression for L and C_1 is obtained

$$L \approx \frac{g_m r_o}{((2 + g_m r_o)C_{gd} + 2C_{gs})\omega^2} \quad (3.12a)$$

$$C_1 \approx \frac{(2 + g_m r_o)C_{gd} + 2C_{gs}}{g_m r_o}. \quad (3.12b)$$

Similarly, C_2 and R_L may be solved from KCL at Port 2

$$R_L \approx \frac{r_o}{1 - r_o R_{gs} C_{gd}^2 \omega^2 - 4 R_{gs} \omega^2 \left(\frac{(C_{gd} + C_{gs})^2}{g_m^2 r_o} + \frac{C_{gd}(C_{gd} + C_{gs})}{g_m} \right)} \quad (3.13a)$$

$$C_2 \approx \frac{(4 + 2g_m r_o)(C_{gd} + C_{gs})}{g_m^2 r_o^2} - C_{ds} \quad (3.13b)$$

3.2.3 Interpretation of the Small-signal Analysis

Here, we summarize the result from the previous section in the following

$$C_1 \approx \frac{(2 + g_m r_o)C_{gd} + 2C_{gs}}{g_m r_o} \quad (3.14a)$$

$$C_2 \approx \frac{(4 + 2g_m r_o)(C_{gd} + C_{gs})}{g_m^2 r_o^2} - C_{ds} \quad (3.14b)$$

$$L_1 \approx \frac{g_m r_o}{(2 + g_m r_o)C_{gd} + 2C_{gs}} \cdot \frac{1}{\omega^2} \quad (3.14c)$$

$$R_L \approx \frac{r_o}{1 - r_o R_{gs} C_{gd}^2 \omega^2 - 4 R_{gs} \omega^2 \left(\frac{(C_{gd} + C_{gs})^2}{g_m^2 r_o} + \frac{C_{gd}(C_{gd} + C_{gs})}{g_m} \right)} \quad (3.14d)$$

Before we make any observation of the result, it is important to note that this solution is only valid as a small signal approximation. The approximation is attributed to the model itself, the drive level at which the large signal Y-parameter is extracted and the assumption that V_x is held constant in the normalization of (3.7). We will discuss in detail about the limitation of small signal analysis in Section. 3.2.5. Nevertheless, the linearization still proves to be useful so long as we treat the result of the analysis not as a final solution to the design, but rather as an insightful guideline towards it.

Now we take a look at the results in (3.14). Not surprisingly, the two capacitors in the embedding network are just linear combinations of some internal parasitic capacitors, with $g_m r_o$ and other numbers being the coefficient. The load resistance takes values smaller than r_o to compensate for the loss on R_{gs} . R_L converges to r_o when R_g diminishes. However, the inductor L_1 only resonates with the capacitor C_1 at the frequency of operation. This may seem counter-intuitive as the resonance frequency for a Colpitts

oscillator, for example, is determined by C_1 , C_2 , and L_1 collectively. It is in fact the result of normalization of power of (3.7) and subsequently A_{opt}^\dagger in (3.9). Thus, L_1 resonating with C_1 is completely a coincidence. The resonance frequency will depend on both C_1 and C_2 for any $\Im(A_{opt}^\dagger)$ not equal to zero. In practice, $\Im(A_{opt}^\dagger)$ takes a small value, which largely preserves the scaling property of L_1 .

Next, we examine the expression of L_1 into greater detail, which is most problematic from parallel scaling. As can be observed from (3.14c), several terms affect the value of L_1 : the C_{gd} , the intrinsic gain $g_m r_o$ and C_{gs} . Since C_{gs} is seldom a design variable, we will primarily focus on C_{gd} and $g_m r_o$ by studying their effect on L . One way to evaluate the sensitivity of C_{gd} and $g_m r_o$ to L_1 is to take the derivative of L_1 with respect to the variable. Alternatively, to simplify the calculation, we may vary the variable of interest by the same percentage while fixing others, and compute the change in L_1 . We first consider a reference common-source active device with small signal parameters: C_{gd0} , C_{gs0} , $g_{m0} r_{o0}$ and inductance L_0 is given by

$$L_0 = L(C_{gs0}, C_{gd0}, g_{m0} r_{o0})$$

Assuming a new device with *twice* the $g_m r_o$ or *half* of the C_{gd} than the reference, the inductance of this new device can be calculated from (3.14c). The normalized inductance to the reference is given by

$$\tilde{L}_{g_m r_o} = \frac{L(C_{gs0}, C_{gd0}, \mathbf{2} \cdot \mathbf{g}_{m0} \mathbf{r}_{o0})}{L_0} \quad (3.15)$$

and

$$\tilde{L}_{C_{gd}} = \frac{L(C_{gs0}, \frac{1}{2} \cdot \mathbf{C}_{gd0}, g_{m0} r_{o0})}{L_0}, \quad (3.16)$$

respectively. The normalized inductance should be interpreted in the following way. Realize that we are looking for an active device structure from which the resulting inductance scales proportionally with output power. We assume that the technique to individually tune one of the circuit parameters, e.g. the intrinsic gain, is readily available without degrading the power generation. By manipulating the $g_m r_o$ and C_{gd} , we hope to see an increase in the inductance compared to the original common source configuration. We

take the ratio of the manipulated device over the simple common source to define quantitatively the effectiveness of a certain modification. The two cases are discussed individually in the following.

3.2.3.1 The effect of the intrinsic gain

Rearrange (3.14c) by cancelling $g_m r_o$ from its numerator and denominator and set $g_m r_o$ as a variable

$$L = \frac{1}{C_{gd0} + \frac{2}{g_m r_o} (C_{gs0} + C_{gd0})} \frac{1}{\omega_0^2} \quad (3.17)$$

Clearly, (3.17) shows that the inductance L monotonically increases with the gain. Realizing that the intrinsic gain of a MOS transistor is determined to the first order by

$$g_m r_o = \frac{2}{\lambda V_{ov}} \quad (3.18)$$

where λ is the channel-length modulation coefficient and V_{ov} is the overdrive voltage of the transistor. Decreasing V_{ov} could potentially improve the gain and thus raise L . However, the benefit becomes limited once the baseline $g_m r_o$ is large. Substituting the expression (3.14c) into (3.15), we have

$$\tilde{L}_{g_m r_o} = \frac{g_{m0} r_{o0} + 2\alpha + 2}{g_{m0} r_{o0} + \alpha + 1}, \quad (3.19)$$

where α is ratio between C_{gs0} and C_{gd0}

$$\alpha = \frac{C_{gs0}}{C_{gd0}}.$$

To visualize the improvement from a high gain device, we plot $\tilde{L}_{g_m r_o}$ over $g_{m0} r_{o0}$ and various α of the reference device in Figure. 3.5. For any non-zero $g_{m0} r_{o0}$ and finite α , doubling the $g_{m0} r_{o0}$ will always yield a $\tilde{L}_{g_m r_o}$ less than 2. This relationship is quite inefficient as it suggests obtaining twice the $\tilde{L}_{g_m r_o}$, a new device would be more than doubled. As we move away from the boundary, $\tilde{L}_{g_m r_o}$ increases with α for a fixed $g_m r_o$. It quickly drops with $g_{m0} r_{o0}$ when $g_{m0} r_{o0} < 10$. Assuming a baseline $g_{m0} r_{o0}$ of 10 and α of 3 in a typical MOS transistor, doubling the intrinsic gain by reducing V_{ov} only results in an improvement of \tilde{L} of less than 37%. Furthermore, for millimeter wave designs, V_{ov}

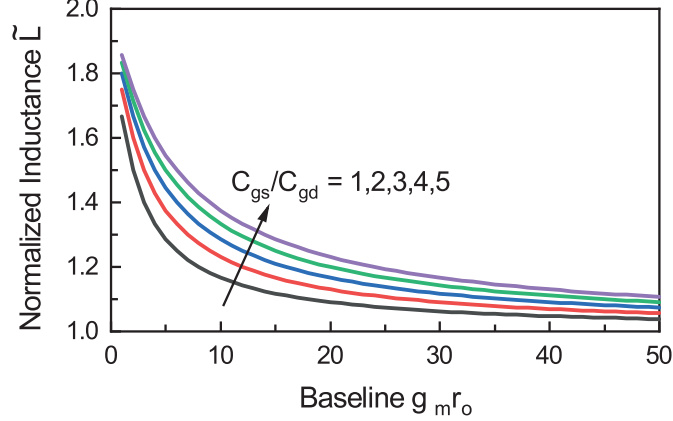


Figure 3.5: Normalized inductance \tilde{L} by doubling $g_m r_o$ (3.19) versus $g_m r_o$ for various C_{gd}

is not a free variable as it strongly affects f_{max} . The conclusion is that $g_m r_o$ alone is not an efficient tuning variable of L .

3.2.3.2 The effect of C_{gd}

Suppose there is a mechanism which reduces the C_{gd} of the reference device by half without altering other parameters. We are interested in how much the inductance changes from this. Expanding (3.16) by substituting (3.14c), we obtain

$$\tilde{L}_{C_{gd}} = \frac{2g_{m0}r_{o0} + 4\alpha + 4}{g_{m0}r_{o0} + 4\alpha + 2}, \quad (3.20)$$

$\tilde{L}_{C_{gd}}$ is plotted as a function of $g_{m0}r_{o0}$ and α in Figure. 3.6. It decreases with α and increases with $g_{m0}r_{o0}$, as opposed to the previous case. $\tilde{L}_{C_{gd}}$ reaches 2 only if $g_{m0}r_{o0}$ is infinite or α equal to zero. It is still inefficient for realistic values of $g_{m0}r_{o0}$ and α . However, since tuning C_{gd0} exhibits opposite dependency on the circuit parameters as $g_{m0}r_{o0}$, one is possible to be superior to another in some regions. For example, using the same set of baseline values with $g_{m0}r_{o0}$ of 10 and α equal to 3, $\tilde{L}_{C_{gd}}$ yields 1.5, which shows slightly higher efficiency than doubling $g_{m0}r_{o0}$.

The foregoing analysis has discussed the situation when either $g_{m0}r_{o0}$ or α is tuned individually. Both result in a larger normalized inductance to counteract the scaling problem. Yet neither is able to increase the normalized inductance linearly. If $g_m r_o$ and C_{gd} are tuned simultaneously in the same direction that helps to alleviate the scaling difficulty, a higher normalized inductance can be expected. This is exactly what a stacked-

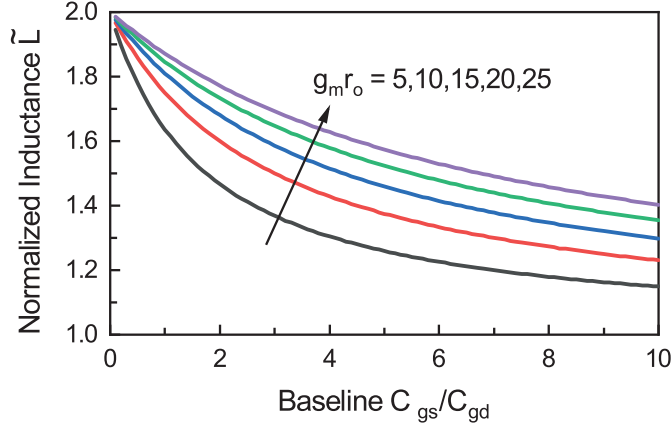


Figure 3.6: Normalized inductance \tilde{L} by halving C_{gd} versus C_{gd0} for various $g_m r_{o0}$

FET device could achieve. In the next section, we will show that the scaling difficulty is completely addressed by stacked-FET devices.

3.2.4 Stacked-FET

Before analyzing the stacked-FET configuration, we first evaluate the effective C_{gd} of single transistor common-source connection, which serves as a reference to the stack-FET technique. Again, we employ the small signal model of a common-source MOS transistor in Figure. 3.4(b) for y-parameter calculation. The reverse transfer admittance parameter y_{12} is given by

$$y_{12,CS} = -\frac{sC_{gd}}{1 + sR_g(C_{gd} + C_{gs})} \approx -sC_{gd}. \quad (3.21)$$

Notice that operating frequency is generally far less than the 3-dB frequency of the low pass filter by R_g and $C_{gd} + C_{gs}$, it is safe to approximate y_{12} as only the susceptance of C_{gd} . In other words, y_{12} is an appropriate measure of effective C_{gd} of an amplifier.

To evaluate the C_{gd} of a stacked-FET amplifier, we make the following observations. Consider a two-stacked amplifier in a normal operating condition illustrated in Figure. 3.7(a). The input voltage V_{in} , inter-stage matching network and load impedance are designed such that output power is maximized, and each transistor is free of breakdown. The first observation is that, the inter-stage matching, regardless of its physical implementation, together with the common-gate-like transistor transfers the drain current of the first stage to the output without loss. An ideal stacked-FET amplifier ensures the

current gain I_1/I_2 to be unity. Secondly, due to the equal V_{ds} swing of each transistor, V_y and Z_L must be twice V_x and Z_x , respectively. Next we replace M_2 and the inter-stage matching with an ideal current transfer element shown in Figure. 3.7(b). This current transfer component ensures the unity current gain and impedance matching property as M_2 and the inter-stage network functions. Again, ignoring the R_g in the small signal model, the y_{12} of the overall stacked FET is

$$y_{12,stack} = \frac{I_1}{V_2} = \frac{1}{2} \frac{I_1}{V_X} = \frac{1}{2} y_{12,CS} \quad (3.22)$$

where the voltage as V_X is one half of V_2 , which is enforced by the current transfer. Transistor M_1 itself is configured the same way as in the y_{12} calculation. Therefore, the y_{21} is reduced by the ratio of V_X and V_2 . This holds if more than two transistors are vertically stacked, where the y_{12} is reciprocal to the number of stacked transistors. We have performed large-signal S-parameter simulation of up to 5 stacked-FET devices, from which the effective C_{gd} is derived. Each of the transistors in the stacked-structure shares an equal size of 16 μm . The simulated result is given in Figure. 3.8 and denoted by the solid square symbol. A reciprocal function in dashed line is also plotted, taking the same value with the simulation when only one transistor is used. The simulated C_{gd} follows a similar trend of a reciprocal function, but having a smaller magnitude when more than two transistors are stacked. The behavior of the simulation is best fitted by a function of $x^{-1.5}$ with x being the number of stacked transistors. This essentially implies that for every double transistors in stack, the effective C_{gd} is reduced by 2.8 times.

With N transistors vertically stacked, the gain of the stacked-FET amplifier is N times the gain of a corresponding common-source amplifier [65]. This is because with optimal inter-stage matching and gate capacitor of the common-gate like transistors, the overall g_m is the same as the bottom common-source transistor and output resistance is multiplied by N . The C_{gd} of a stacked-FET device is less than $1/N$ times that in a common-source amplifier. We shall substitute the value $g_m r_o$ and C_{gd} of the stacked-FET into (3.14c), assuming that $g_m r_o$ and C_{gd} are scaled by N and $1/N$, respectively, although C_{gd} is in fact less than $1/N$ times the common-source device. We calculate the normalized

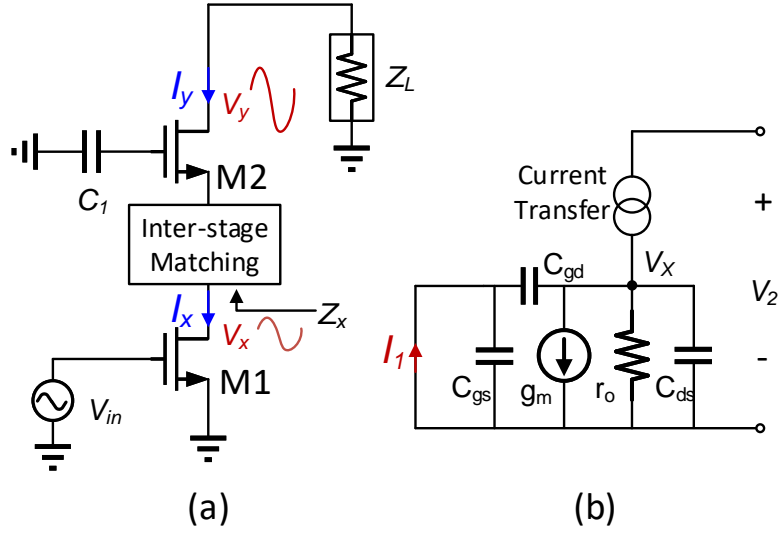


Figure 3.7: Schematic and simplified two-port representation of a two-stack amplifier

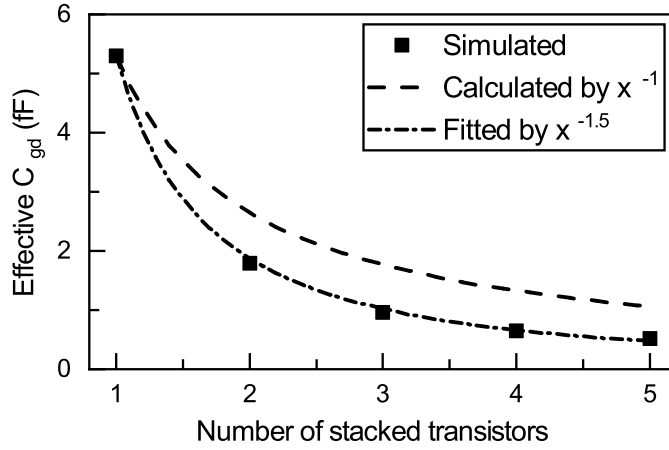


Figure 3.8: Simulated and calculated effective C_{gd} of various number of stacked transistors

inductance of a N-stack device to the common-source amplifier

$$\tilde{L}_N = \frac{L_{N_{Stack}}}{L_{CS}} = N \frac{1 + \alpha \left(1 + \frac{g_m r_o}{2} + 1\right)}{1 + \alpha \left(1 + \frac{g_m r_o}{2} + \frac{1}{N}\right)} > N, \quad (3.23)$$

The contribution from both $g_m r_o$ and C_{gd} essentially leads to a greater improvement in \tilde{L} . We notice that for any value of α and $g_m r_o$, β_N is always larger than N . Table 3.1 compares the performance of a single CS stage and a N-stacked stage. This property will enable us to scale the oscillator power to any value with a constant gate-to-drain

inductance. Suppose a reference oscillator design generates power P_R with transistor width W_R and inductance L_R . The design of an oscillator with N^2 higher output power involves a two-step scaling: (1) scale the transistor width to $N \cdot W_R$ (2) stack a total of N transistors with width of $N \cdot W_R$ vertically and design the inter-stage matching network and gate capacitance following the procedure in [66].

3.2.5 Discussion of the limitation of linear models

Linear small-signal modeling has provided useful insight about how the embedding network components are affected by the transistor parameters. It may be less attractive obtaining the exact solution to the design. Previous work [61] has discussed the limitation of linear models. We first make a short summary. The behavior of the transistor subject to large voltage swing is only captured by a large signal model where nonlinear controlled source and parasitics are included. In the context of oscillators, however, large signal circuit parameters have been applied to extract the relative accurate behavior of the transistor. During such process the drive level should be carefully determined. The normalization in (3.7) assumed that V_X is held constant over various A^\dagger . This is generally not true and especially when the quality factor of the embedding network is considered. The loss in the passive components has a much more profound effect beyond the A_{opt} : the input drive level should be decreased and in some cases the embedding component may change from a capacitor to inductor for example.

It has become clear that to obtain a set of optimum embedding network, an iterative process is necessary. Notice that the design of a stacked-FET active core is indeed independent of the embedding network. Even though the oscillator design itself involves tuning of port drive power and impedance, the active device is almost always driven to full swing. As long as this large signal condition is satisfied, the gate capacitor and inter-stage matching of the stacked-FET active devices stays optimum with small perturbation of the port excitation and termination. Thus, the design of the stacked-FET oscillator may divide into individual tasks: 1) design the gate capacitor and inter-stage matching network of the stacked-FET active device, such that maximum added power is reached without imposing reliability issues, 2) treat the stacked-FET active device in the previous

step as a two-port black box and apply the design methodology in [61] for an optimum the embedding network.

3.3 Implementation

The proposed approach is validated by a fundamental oscillator at 180 GHz in 45 nm silicon-on-insulator (SOI) process.

3.3.1 Process and Transistor

The use of silicon dioxide insulator in SOI process it beneficial to the operation of the stacked-FET amplifiers. A deep n-Well would otherwise be created for each transistor to isolate their body in a bulk CMOS process. Although stacked-FET amplifiers have been implemented in bulk CMOS process [67], SOI is favored due to the lack of body effect [66]. Additionally, a deep-N-well structure introduces two reverse biased parasitic diodes in equivalent to a series resistor and capacitor, which creates a leakage path to ground. Furthermore, the accurate measurement of these parasitic diodes may not be available at the operating frequency. The use of extrapolated data may result in inaccurate modelling.

The total width of all the transistors is determined by the required output power. The goal of the design is to demonstrate oscillation power greater than 10 dBm while maintaining high efficiency. Considering the use of a nearly lossless combiner, each oscillator core should generate at least 5 mW. To leave some margin during the implementation phase as well as in the manufacturing, we set the target power of a single oscillator as 8 mW. A two stack configuration with the use of 41.6- μm transistors meets the power requirement and the corresponding inductance falls in the high Q region. The layout basically follows the similar arrangement in [61], with four columns, each consisting of 13 fingers. Figure. 3.9(c) shows the 3-D portrait of the transistor.

3.3.2 Oscillator

A fundamental oscillator utilizing both the stack-FET and combining technique has been implemented. The schematic of one core of the proposed oscillator is sketched in Figure. 3.10 with the value of critical components labeled. The embedding network of the proposed oscillator is fundamentally similar to any Pi-embedding structures, except for

Table 3.1: Comparison between stack-FET and common-source amplifier

	Single CS	N-Stacked-FET
Gain	$g_m r_o$	$g_m r_o * N$
Gate-drain capacitor	C_{gd}	C_{gd}/N
Oscillation Power	P	$P * N$
Inductance	L	$L * N$

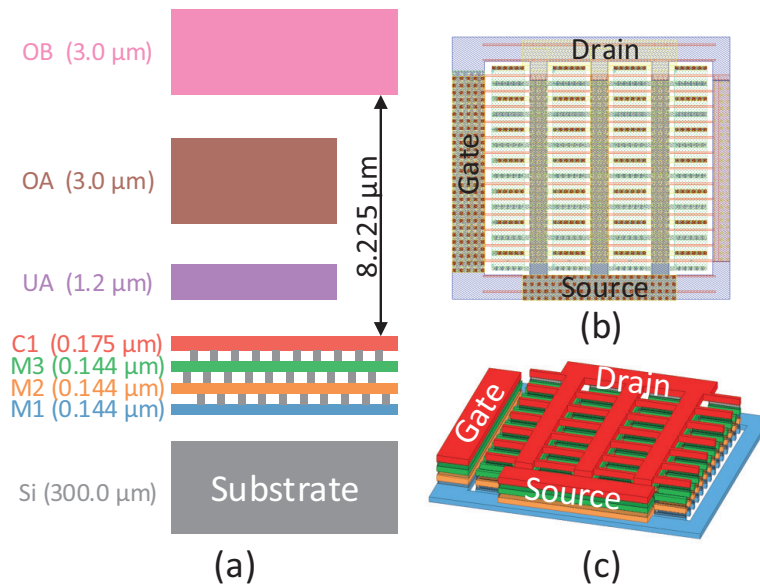


Figure 3.9: (a)Metal cross-sectional profile. (b)Layout of a 41.6μm transistor. (c)3-D portrait of the transistor.

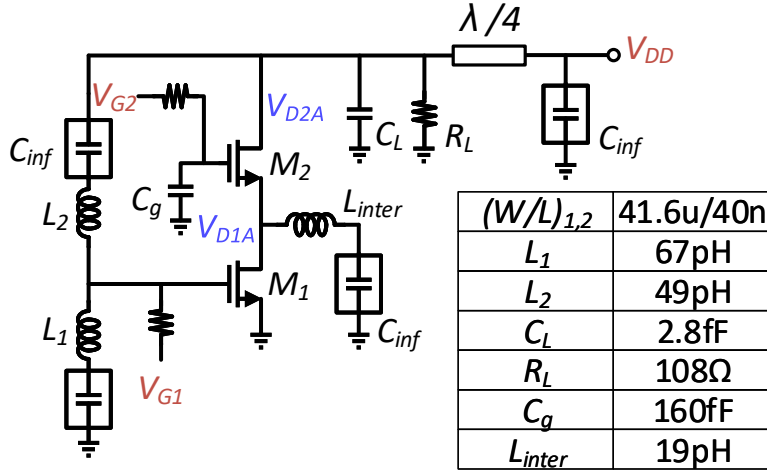


Figure 3.10: Schematic of the stacked-FET oscillator core.

some attention to the gate embedding component and the biasing scheme. In the calculation of the previous work [61], a capacitor is expected at the gate while being replaced by an inductor in Figure. 3.10. The component value in Figure. 3.10 has been carefully iterated from large-signal harmonic balance simulations. Such divergence is a collective consequence of large signal behavior and limitation on linear models. As we observe both inductors fall into the high Q region from Figure. 3.2 and thus no practical implementation difficulty has appeared.

The optimum bias for gate and drain is usually different. The gate bias voltage should be selected corresponding to the optimum current density of highest f_{max} . For this particular process, optimum current density occurs at gate bias of around 0.7V. Drain voltage bias is limited by the breakdown voltage of the process, which is 1 V for a thin oxide transistor. Oscillators with an inductor connecting gate and drain thus requires a large capacitor to block the dc path. Previous works [61,62,68] of this type seek to share the gate and drain bias without using a series capacitor. This is accomplished by choosing a proper bias voltage balancing output power and gain of the device. In this paper, the power supply of the two-stack transistor is 2V which is substantially higher than the gate bias. It would not be possible to balance the supply voltage without significantly sacrificing power or gain. In this work, the dc block capacitor is integrated with the inductor. We will discuss in detail about the implementation in Section.3.3.4.

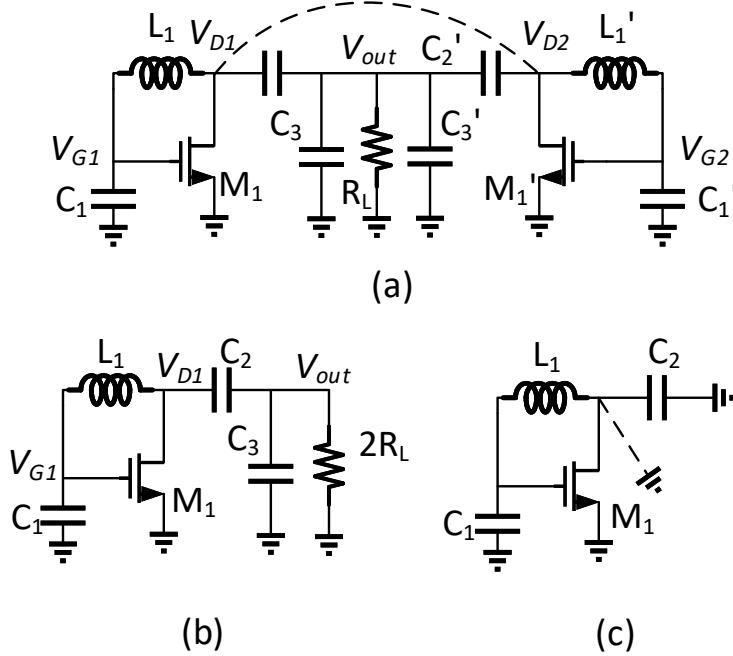


Figure 3.11: (a) 2-way combined oscillator with capacitive matching circuit, (b) common-mode half circuit (c) differential-mode half circuit

3.3.3 Combining

We have developed an efficient combining scheme which doubles the oscillator power without using any of the physical combiners. To better appreciate the concept of the proposed power combining network, Figure. 3.11(a) shows two generic Pi-embedding oscillators combined by direct paralleling. A capacitive matching network at the output transforms the optimum load impedance to 50Ω [69]. Its common-mode and differential-mode equivalent half circuit are given in Figure. 3.11(b) and (c), respectively.

Two possible modes exist in the general case. In order to ensure proper operation of the combined oscillator, only one mode of oscillation and constructive addition should be satisfied. For parallel combining, it is desired to ensure two cores operate in phase where the output current add constructively to the load. First of all, ignore the dashed line in Figure. 3.11 and focus on the common-mode operation. The nodes V_{D1} and V_{D2} share the same potential; therefore, two half circuits are split by an open circuit. The resulting half circuit is indeed a Pi-Embedded oscillator. In this mode, two oscillator cores operate

in phase. The output current of the two cores adds constructively at V_{out} and the load resistance is half of the Pi-embedding core, thus doubling the output power. Note here the output resistance of each oscillator core should be $100\ \Omega$ instead of $50\ \Omega$. Common-mode is the only mode on which the circuit operates. Differential-mode must be reliably eliminated.

In differential-mode operation, V_{D1} and V_{D2} experience out of phase voltage swing by definition. A virtual ground is thus created along the line of symmetry so that C_3 , C'_3 and R_L are deactivated. Again, ignoring the dashed lines, the remaining differential-mode half circuit Figure. 3.11(c) is also in the form of a Pi-embedded oscillator, but without a load resistor. Although the drain is loaded differently from the effective load capacitance as in Figure. 3.11(b), the absence of load resistor increases the loop gain by a factor of 2. The Barkhausen criterion could still be satisfied. Without completely eliminating the differential-mode operation, the coupled oscillator may trap into either mode during start-up.

The ambiguity of dual mode start-up may be solved simply by introducing a short between node V_{D1} and V_{D2} as highlighted with a dashed line in Figure. 3.11(a). The short has no effect in common-mode when V_{D1} and V_{D2} are in phase. In the differential mode, the drain of the transistor is forced to ground as illustrated in Figure. 3.11(c) with the dashed line. The loop gain of this mode will always be zero because by shorting drain to ground, the active transconductance g_m is deactivated. Moreover, this method is completely insensitive to the value of the embedding network. The complete schematic of the proposed oscillator is shown in Figure. 3.12.

3.3.4 Layout and Passives

A simplified layout of a generalized output combiner is sketched in Figure. 3.13(a) where the rest of the embedding network is omitted. The capacitors C_2 , C'_2 , C_3 and C'_3 are merged into series C_C and shunt C_L , respectively. Suppose C_C is implemented with interdigitized MOM capacitor, connecting the fingers that electrically connected to the drain completely eliminates the out of phase operation. Highlighted by the red thick wire in the inset of Figure. 3.13(a), due to the proximity and small size of the two MOM capacitors,

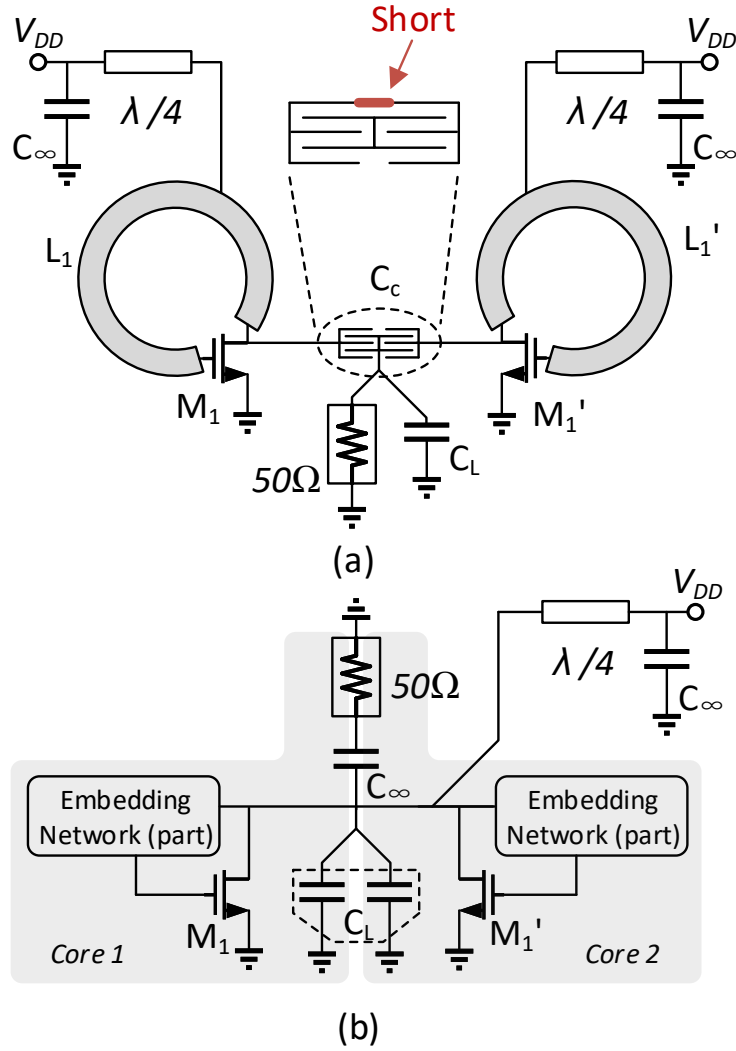


Figure 3.13: (a) layout sketch of the output combiner including capacitive matching network and (b) simplified schematic of alternative output combining.

calculated R_L is close to $100\ \Omega$ that allows for the structure in Figure. 3.13(b).

Figure. 3.9(a) shows the metal stack up profile as well as the implementation of ground and interconnects. Four lowest thin metal layers are stitched into a low impedance ground plane. The signal traces of inductors and transmission lines are constructed by OB layer. The inductor L_{PA} and L_{PB} are implemented by transmission lines with characteristic impedance of $75\ \Omega$ for optimum quality factor. C_{gA} and C_{gB} at the gate of M_{2A} and M_{2B} are parallel-plate capacitors consisting of layers M1-M3 and C1. The capacitor is designed to bury in the ground plane without any structures over the ground plane

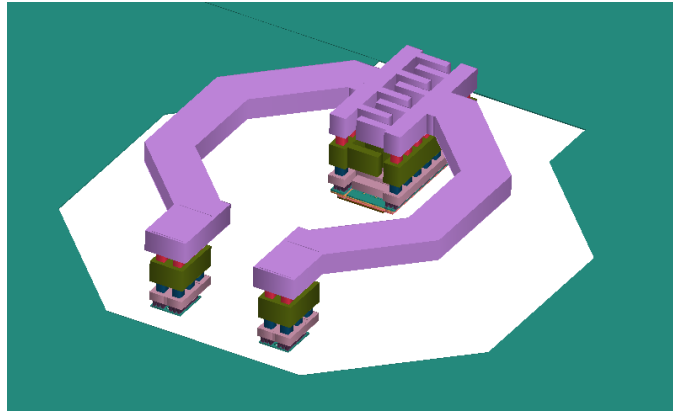
interfering with the transmission line and inductor. The output capacitor C_L is not physically implemented. Its value is very small and it is thus absorbed into the parasitics of the inductor L_{2A} and L_{2A} as well as the dc block capacitor.

The inductor L_2 and dc-block capacitor was integrated to achieve high quality factor. The 3-D rendition of the integrated inductor and capacitor as well as the ground plane is shown in Figure. 3.14(a). Placed at the center of the inductor, the capacitor breaks the inductor into two pieces. Such layout arrangement is specially chosen because the MOM capacitor made with lower level metals introduces parasitics to the transistor if two are placed close to each other. Also, to reduce the parasitic capacitance to ground, the ground plane beneath the inductor is removed. The simulated inductance and quality factor around the operation frequency are plotted in Figure. 3.14(b). The inductance and quality factor at 180 GHz is 66 pH and 16%, respectively.

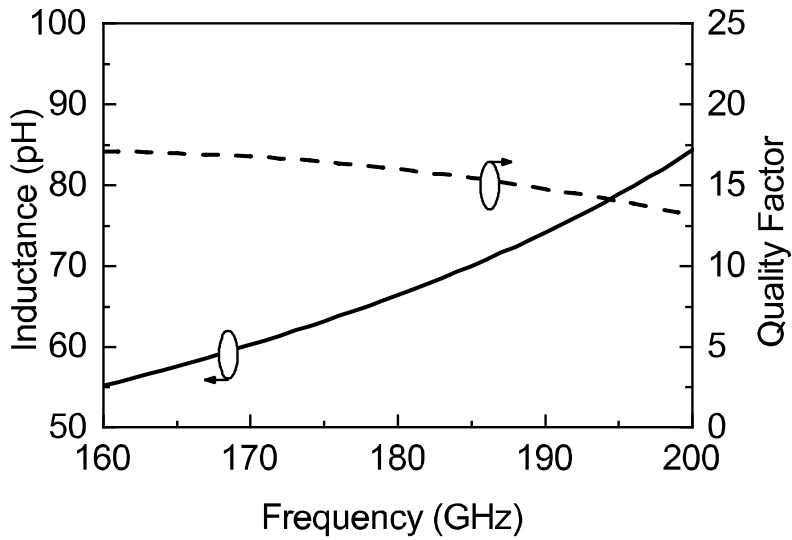
The simulated output power and dc-to-RF efficiency is 12.6 mW and 15.5%, respectively, at 180.4 GHz. Beyond the RF performance, it is necessary to check the transient waveform to keep the transistors from break down. Specifically, the top common-gate-like transistor may carry large gate-to-drain and drain-to-source voltage, causing reliability issues. The gate-to-drain and drain voltage of each transistor in one of the two identical oscillator cores are plotted in Figure. 3.15. The bottom common-source transistor biased at 1-V V_{DS} and 0.7-V V_{GS} , which sets a reference for the top transistor, is free of break down stress. Figure. 3.15(a) shows that the drain voltage of each transistor divides evenly and maintains phase alignment. From Figure. 3.15(b), we notice that the voltage swing of the top common-gate-like transistor is smaller than the common-source transistor.

3.4 Measurement of the Stacked-FET Oscillator

The measurement setup for frequency and power basically resembles that of a previous work [61] and is repeated in Figure. 3.16 for illustration. Spectrum and power are measured separately. A VDI WR-5.1 mm-head module functions as a down converter for the spectrum analyzer to extend its frequency range. The LO signal from the spectrum analyzer is tripled and fed to the WR-5.1 module, which is multiplied internally and then



(a)



(b)

Figure 3.14: (a) 3-D profile and (b) simulated performance of the proposed integrated inductor

mixed with the oscillator output. The IF signal returns to the spectrum analyzer and is displayed after frequency adjusting. The power measurement is conducted with a VDI Erickson PM5 power meter. The output power of the oscillator is calibrated with the loss of the RF probe and waveguide.

3.4.1 Calibration

The one port S11 measurement of the probe [70] is adopted to calibrate its loss. The calibration setup is shown in Figure. 3.17. First, the TRx mm-head module is calibrated using mechanical calibration standards corresponding to Step 1 in Figure. 3.17(a). In the

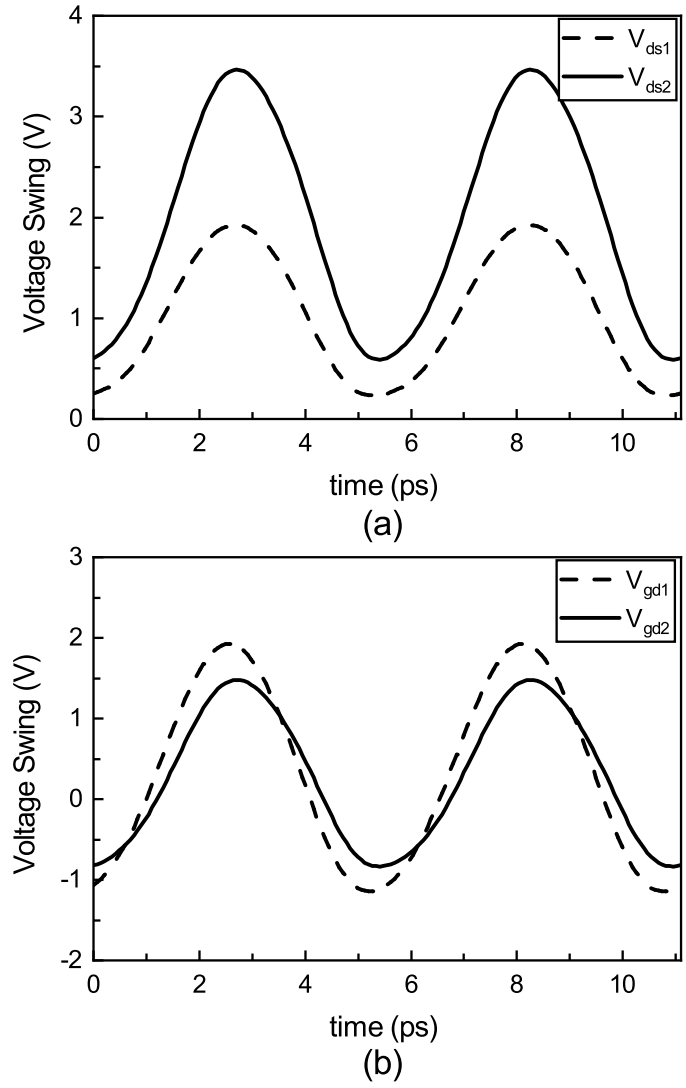


Figure 3.15: Simulated voltage waveform of drain and gate-to-drain swing

next step, the probe and s-bend waveguide are connected to the TRx mm-head module shown in Figure. 3.17(b). The measured S11 when the probe is shorted with gold substrate indicates the round trip loss through the probe and the waveguide. The measured round-trip loss is 11.2 dB, yielding an insertion loss of 5.6 dB including the s-bend and RF probe.

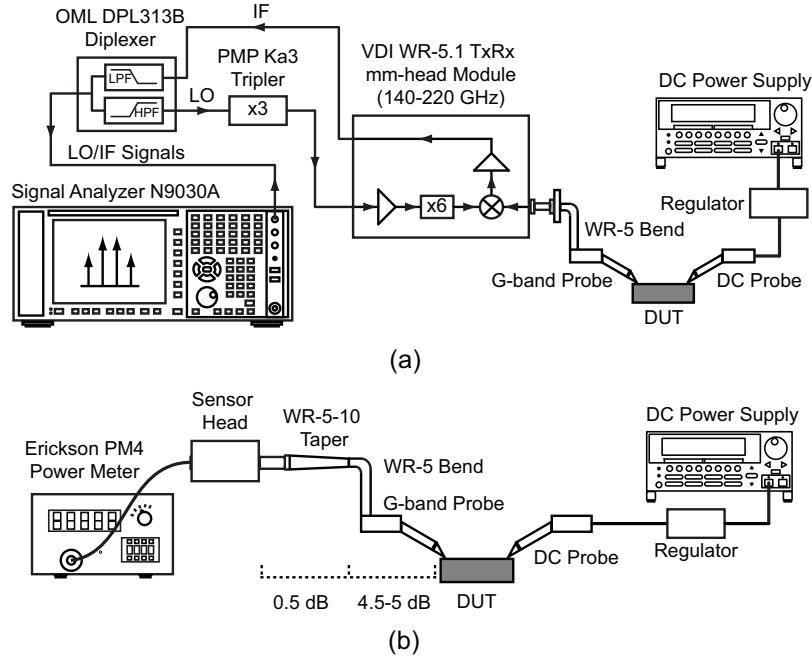


Figure 3.16: Measurement setup for (a) frequency and (b) power.

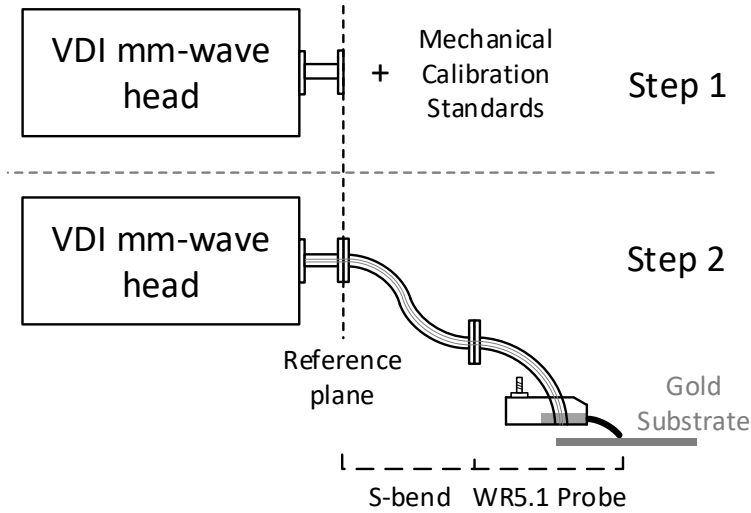


Figure 3.17: Calibration procedure of the total loss from s-bend and RF probe.

3.4.2 Measured Power and Efficiency

The measured power and efficiency are plotted in Figures. 3.16 (a) and (b). An optimum output power is achieved at gate bias of 0.75 V to 0.8 V depending on the supply voltage. The dc-to-RF efficiency stays constant over gate bias from 0.55 V to about 0.75 V. The measured output power is about 3mW less than the simulation with dc-to-RF efficiency

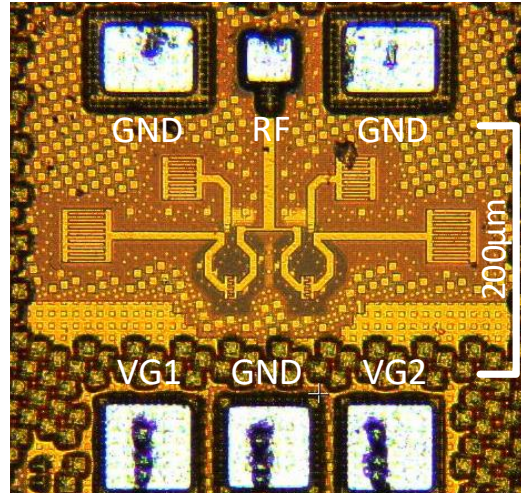
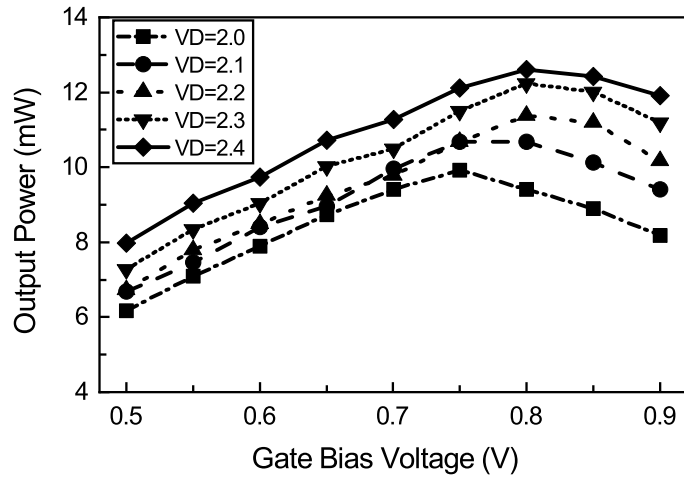
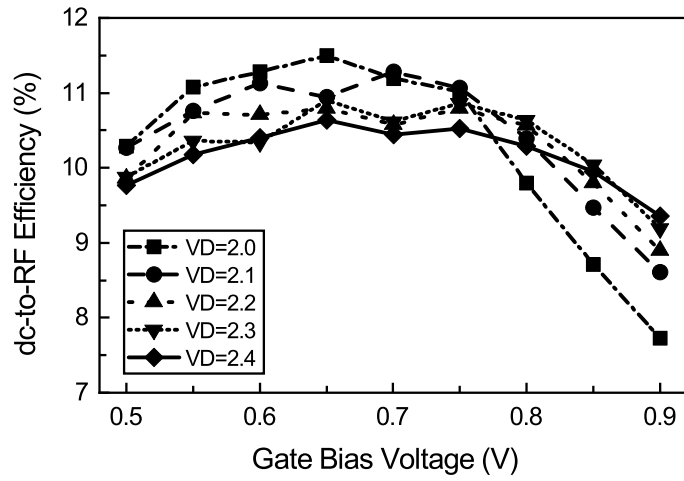


Figure 3.18: Die photography of the fabricated chip.

drops from 15.5% to 11%. The output power increases and matches to the simulation when the supply voltage is raised to 2.4 V. The difference in the output power and efficiency may partly be attributed to the dummy filling into the design that decreases the quality factor of inductors and transmission lines.



(a)



(b)

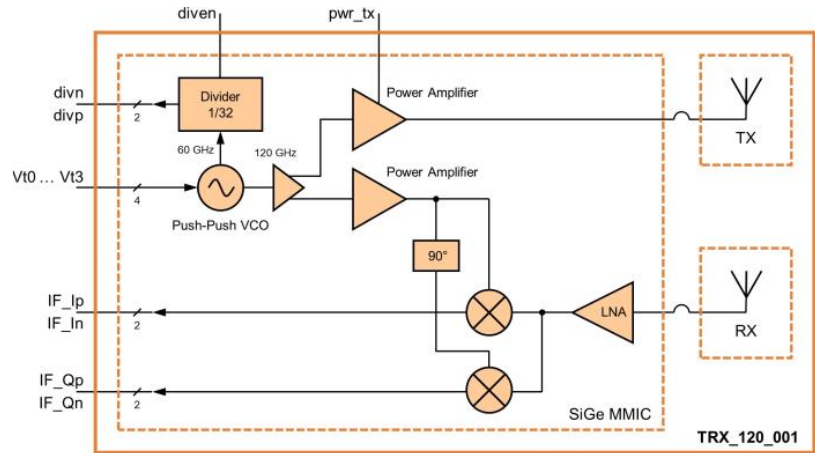
Figure 3.19: Measured output power and dc-to-RF efficiency of the proposed oscillator.

Chapter 4

Molecular Clock Implementation and Experiment

The system design was based on a commercially available millimeter wave transceiver chip: TRX_120_001 from silicon radar. The choice of the chip must correspond with the gas molecule, which is OCS. The advantage of OCS gas is that there are many equally spaced (approximately 12 GHz) spectral lines from below ~ 30 GHz all the way to above ~ 1 THz. On the other hand, the absorption strength of each spectral line increases with frequency for frequency below 500 GHz. An ideal transceiver would have a high center frequency, while covering at least one of the spectral lines with sufficient bandwidth to overcome process variations. There were and are not many millimeter wave transceivers available on the market. The TRX_120_001 was eventually decided due to operating frequency and integration of the antenna. It was the highest frequency transceiver at the time of design that covers one of the OCS spectral line (about 121.6 GHz).

The block diagram of the transceiver is reproduced in Fig. 4.1(a) [71]. It consists of a push-push millimeter wave oscillator with fundamental frequency of about 60 GHz. Frequency tuning of the VCO is achieved with a 4-bit varactor bank. The varactors may be individually tuned with a different tuning sensitivity of each bit, or connected together for maximum tuning range. The second harmonic from the VCO is extracted and divided into two paths: one for the transmitter and the other for the receiver. Each 120 GHz signal path is followed by a power amplifier. The power amplifier in the transmitter path



(a)



(b)

Figure 4.1: (a) Block diagram of the millimeter wave transceiver TRX_120_001 from Silicon radar. (b) The chip photograph [71].

directly drives an on-chip patch antenna array. In the receiver path, the antenna couples the signal from free space, followed by a low noise amplifier (LNA). The output of the LNA is split and fed into a pair of I-Q downconverter mixers, driven by the in-phase and quadrature versions of the 120 GHz signal from the VCO. The downconverted intermediate frequency is directly routed to the output. One of the useful features of this product is that no high speed PCB routing is needed outside the chip. The highest frequency signal coming out of the chip is a divided-by-32 output of the VCO fundamental tone. The 120 GHz signal from the transceiver is internally coupled to an antenna array fabricated on an interposer layer. On one hand, the high integration simplifies the PCB design. On the other hand, it becomes almost impossible to modify any mm-Wave components to improve the performance. For example, the output power of the internal PA is only about -3 dBm and the antenna half-power beam width is as wide as 70 degrees [71]. Later it will be shown that, the system SNR is limited by the link budget and the high amplification

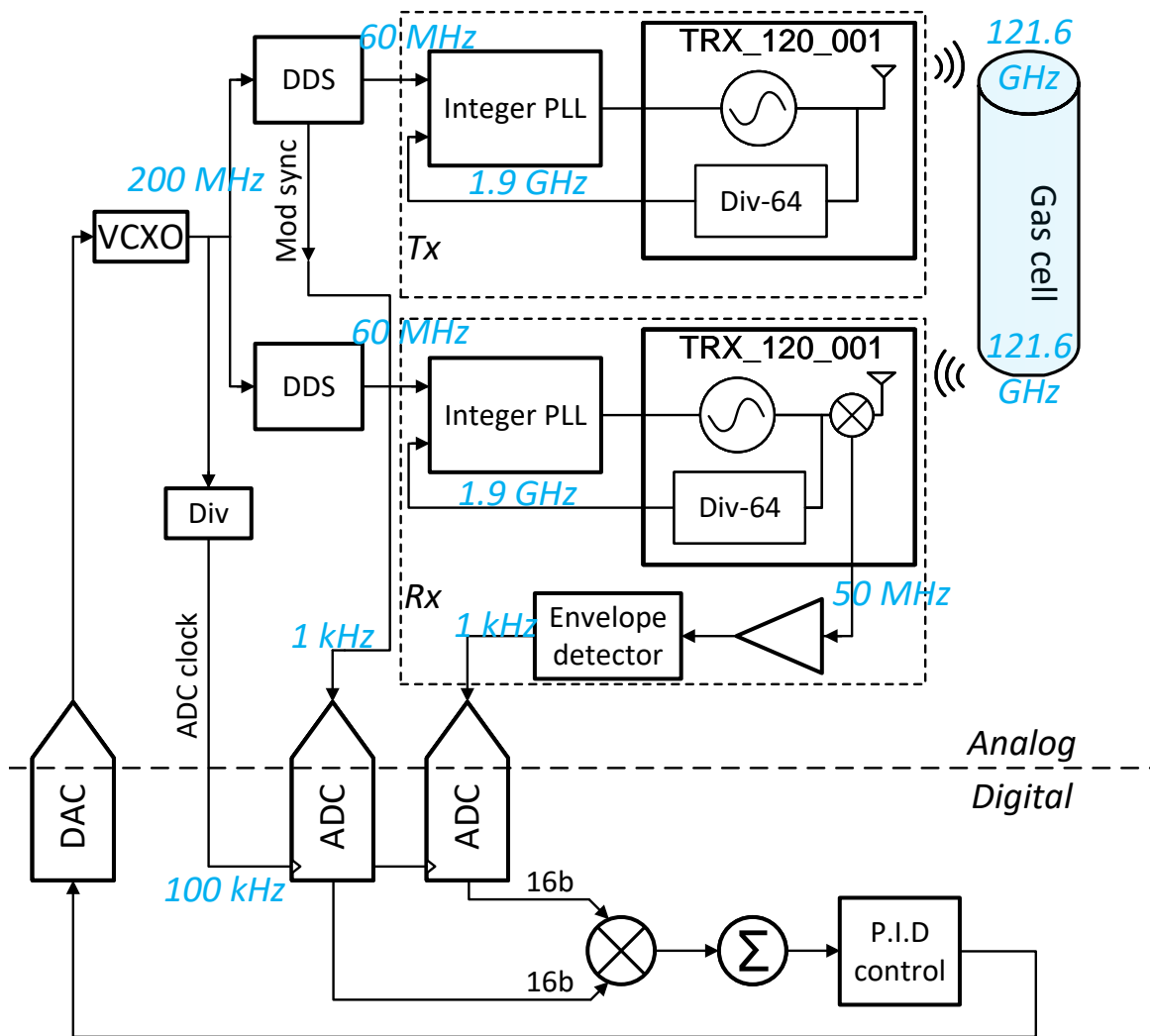


Figure 4.2: The simplified block diagram of the molecular clock.

is required in the receiver.

4.1 Implementation of a 120 GHz molecular clock system

4.1.1 Top-level block diagram

The molecular clock system built around the 120 GHz transceiver is described in Figure. 4.2. The two-port measurement is carried out by a transmitter and receiver pair, facing each other at each side of the gas cell. As described in the previous chapter, a free

running VCO cannot be locked to the molecular spectral line, due to the large frequency drift of the mm-Wave VCO and the narrow linear range of the dispersion curve. The VCO must be locked to a more stable VCXO, usually through a phase-locked loop. In this early work, the VCO locks to a low frequency VCXO in two steps. A DDS generator synthesizes a fractional frequency of the about 60 MHz as a reference to an integer PLL that locks to the transceiver's 1.9 GHz external LO output. The heterodyne receiver shares the LO section of the transmitter, with additional IF amplification and envelope detection circuits. The Tx and Rx share the same PCB design, because the integer PLL and power supply could be reused. The IF amplifier and detection circuit are unpopulated on the transmitter board, since they are not used. The envelope that contains the frequency offset information, as well as the modulation reference, are simultaneously sampled into digital domains. The lock-in operations and subsequently the PID control are performed numerically. The output of the PID controller drives a digital-to-analog converter, which then feeds to the tuning port of the VCXO.

Fractional frequency synthesis could be achieved by a fractional PLL alone, but the introduction of a DDS enables higher frequency resolution and easier modulation. The 32-bit frequency tuning word of AD9913 provides much higher frequency resolution than the PLL used in this design, although less than 26-bit resolution is sufficient. The frequency of the PLL is usually programmed through SPI serial bus, which is flexible and easy to use for one-time PLL setting. However, it is difficult to continuously program the fractional divider in the PLL with good timing. Microcontrollers are usually the top choice for programming the SPI bus. The problem of the microcontroller is that the clock is not synchronized with the molecular system clock, upon which the PLL frequency update depends. This results in a jittery modulation frequency, which is obviously undesired.

The DDS is a simpler and more precision solution in the prototype design. The DDS is usually equipped with an internal frequency modulator, and the modulation frequency is directly derived from the DDS system clock. Frequency modulation could be implemented without continuous programming the SPI bus. Further, since the system frequency of the DDS is referenced to the molecular spectral line, when the clock locks, the modulation

frequency is stable. The major disadvantage of the DDS generator is high power consumption. The core of the DDS is a digital-to-analog converter (DAC) that runs at least twice the output frequency. The high clock speed of the DAC is one of the major reasons of the high power consumption.

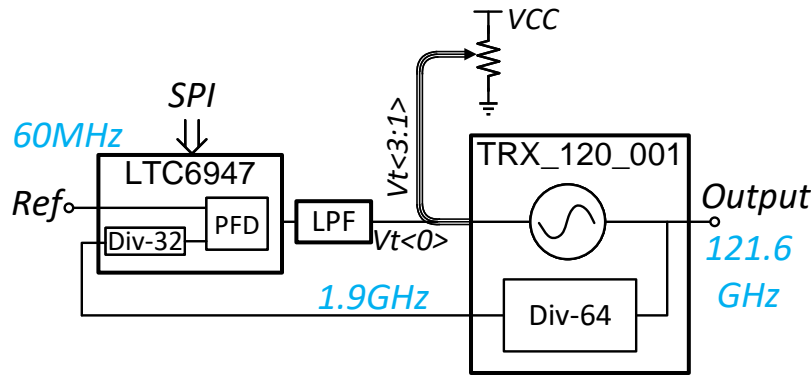
It should be noted that, for low power applications, a customized PLL integrated with a frequency modulator should be a more efficient solution. The tuning word length of the fractional PLL could be designed to be sufficiently wide to meet the resolution requirement. It is not uncommon to incorporate a 32-bit or wider fractional denominator in modern commercial PLL designs [72, 73]. A customized PLL design also simplifies the frequency modulation, which could be accomplished by accessing the internal parallel frequency tuning word register from a look-up table. None of these difficulties associated with commercial PLL chips are fundamentally challenging, and the choice of a DDS instead of a PLL as the fractional frequency synthesis is only a matter of convenience for demonstration.

4.1.2 The LO generation

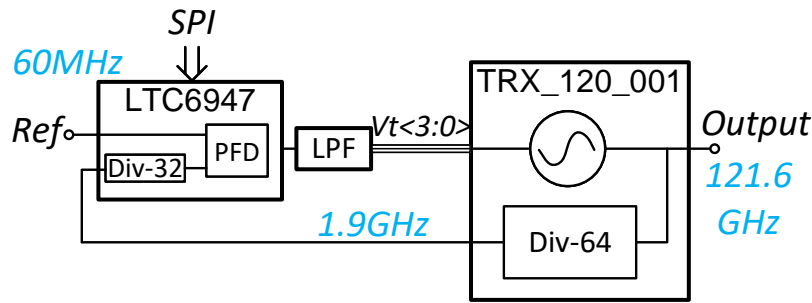
4.1.2.1 PLL

The LO section of the transceiver consists of a PLL chip LTC6947 from Analog Devices that operates in the integer mode. The divider ratio of the PLL is programmed to the minimum allowed value, which is 32. From the noise perspective, a low divider ratio helps to reduce the total phase noise. Even with a noiseless reference and loop filter, the PLL in-band phase noise is approximately equal to the PFD phase noise multiplied by the divider ratio. For a large divider ratio, it is the PFD noise that ultimately limits the PLL performance. The lower limit of the divider ratio depends on the PFD and divider design. Since the PFD is made of logic gates and flip-flops, the speed of the logic gates determines the upper limit of the speed. However, the divider design of this chip poses another constraint: only a minimum divider ratio of 32 is allowed. With a fully customized PLL design, the minimum divider ratio would only be limited by the maximum PFD frequency.

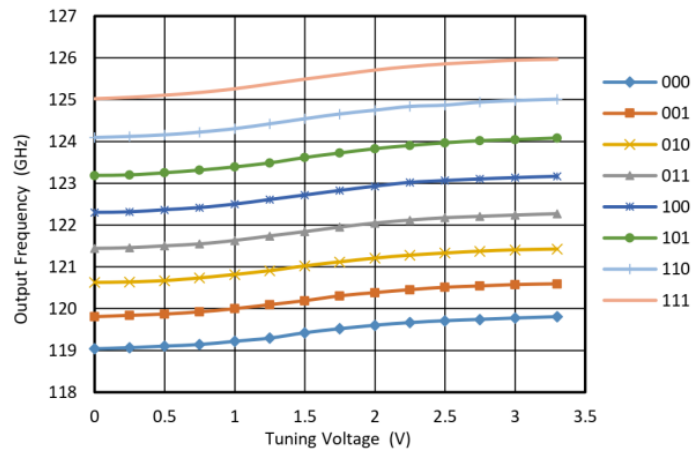
Due to the unique design of the VCO tuning, the VCO may be controlled by a single



(a)



(b)



(c)

Figure 4.3: The block diagram of the transmitter and LO generation for the receiver. The PLL LTC6947 is programmed in the integer mode that multiplies the reference by 32. (a) An initial design where the PLL only controls the LSB of the VCO tuning port, while the higher bits are connected to a potentiometer. (b) All 4 bits from the VCO tuning port are connected together and controlled by the PLL. (3) The VCO tuning characteristic of TRX_120_001 reproduced from the datasheet [71]. The frequency labels in the figure are only approximations.

bit or multiple bits of the 4-bit tuning port. The four frequency tuning bits $Vt0$, $Vt1$, $Vt2$ and $Vt3$ of the VCO frequency are binary weighted. Each bit is associated with a varactor in the VCO tank, where the tuning ranges for higher order bits are larger. Suppose the upper three bits $Vt < 3 : 1 >$ are controlled digitally, while the $Vt0$ is continuously tuned, the tuning characteristic of the VCO is plotted in Figure. 4.3 (c). If none of the eight tuning curve sets the desired center frequency, $Vt < 3 : 1 >$ may be connected together and applied with a voltage between 0 and 3.3V. As a result, the tuning curve would be shifted to any location in between the curve from 000 and 111. To reduce the PLL phase noise, a low tuning sensitivity of the VCO is usually preferred. The concept of low sensitivity VCO tuning is illustrated in Figure. 4.3 (a), where the PLL controls $Vt0$, with $Vt < 3 : 1 >$ controlled collectively by a potentiometer. The potentiometer adjustment is performed before the PLL operation. Ideally, the potentiometer is adjusted such that, the VCO output equals to the desired frequency, when $Vt0$ is applied with a mid-supply voltage. Then, the PLL stabilizes the VCO through continuous tracking of $Vt0$. While this architecture works quite good initially, after a long period of time, when the center frequency of the VCO drifts more than half of the LSB tuning range, the PLL fails to lock. In an alternative approach shown in Figure. 4.3 (b), $Vt < 3 : 0 >$ are wired together, forming a much wider tuning range VCO. The PLL stays in lock as long as the drift does not exceed half of the total tuning range, which never happened during the measurement. Phase noise was not degraded by increasing the tuning sensitivity, because the limitation for this setup is the noise from the reference.

4.1.2.2 DDS

The output frequency of the DDS is approximately 60 MHz, which is 1/32 of the 1.9 GHz LO output from the transceiver. Since a DDS is a DAC based signal generator, Nyquist sampling theorem demands that the sampling frequency is at least twice the analog signal frequency. Usually, a DDS can generate an analog frequency up to 40% of the system clock. A DDS generator running at least 150 MHz should be able to generate the 60 MHz output. The 200 MHz VCXO is selected as the clock source for DDS, because DDS phase noise performance is degraded when the analog output frequency approaches the

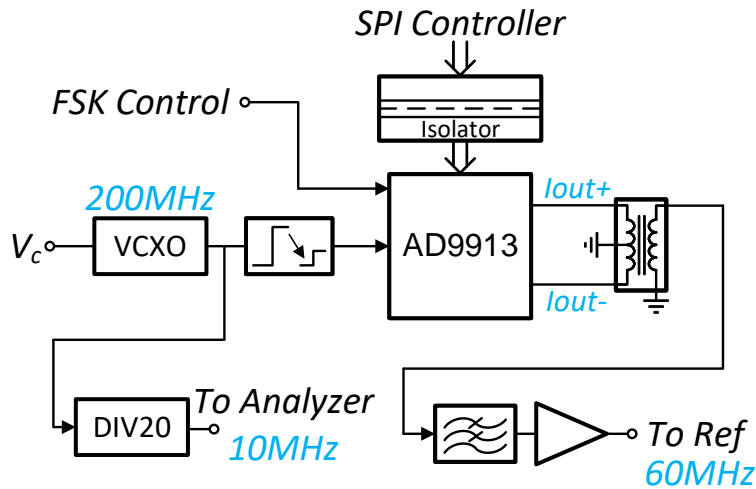


Figure 4.4: Block diagram for fractional frequency synthesis using AD9913.

Nyquist limit. The AD9913 was eventually selected, because 250 MHz maximum clock frequency matches the VCXO frequency, and the chip is also equipped with a versatile FSK modulation capability.

Some external circuits are needed to ensure proper operation of the DDS chip. A digital galvanic isolator is placed between the SPI controller and the DDS, to reduce the noise coupling from the microcontroller to the sensitive analog circuit. Between the VCXO and the AD9913 is a level shifter that converts a 3.3V signal into 1.8V. A resistive divider could also possibly work, but the CMOS output driver in the VCXO achieves optimal performance with a capacitive load. Here, a level translator LMK00804B from TI is used for its good noise performance. The VCXO frequency is divided down to 10 MHz by AD9515, a programmable divider from Analog Devices, to the frequency stability analyzer. The DAC of the AD9913 is a differential open-drain output, which requires a proper termination. A balun converts the differential output to single-ended, where the center-tap at the primary is connected to ground. In this way, the dc current of the DAC output flows to ground through the center-tap. The output of the balun must be low-pass filtered first before being amplified. Amplifier intermodulation products from images around the clock frequency at the unfiltered DAC output may fall into the frequency band of interest. The process of intermodulation is illustrated in Figure. 4.5. Suppose a

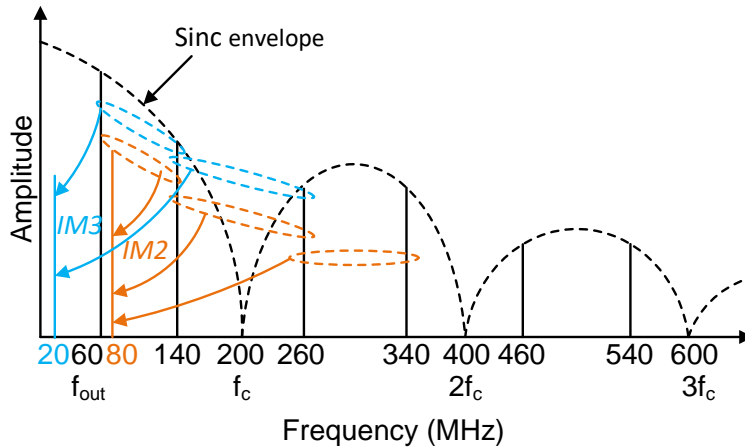


Figure 4.5: Conceptual illustration of intermodulation generation from the unfiltered DDS output.

60 MHz analog output from a DDS clocked at 200 MHz. The unfiltered DDS spectrum consists of tones at $n \cdot f_c \pm f_{out}$, with n equal to positive integers. The amplitude of these tones follows the sinc function. Due to the nonlinearity of the amplifier, intermodulation products of these image frequencies will be generated. In this example, 80 MHz spur is the result of second order IMD of 140 MHz and 260 MHz, 260 MHz and 340 MHz, etc., and the 20 MHz spur arises from third order IMD of 60 MHz and 140 MHz, 140 MHz and 260 MHz, etc. The proximity of the IM components to the signal of interest makes it difficult to be removed by filtering. This intermodulation tone degrades the clock performance by introducing a periodic jitter. Therefore, sufficient filtering at the DDS output is necessary before any nonlinear circuits.

4.1.3 The Receiver

The heterodyne receiver is adopted instead of the direct envelope detection at RF frequency for the following two reasons. The theoretical free space path loss at 120 GHz for 20 mm distance is approximately 60 dB. The loss in practice is higher, because there is reflection at each air-glass interface. Therefore, the received power being comparable to the sensitivity of the diode receiver significantly impacts SNR. To boost the signal, amplification may be performed either at RF or IF. In terms of gain and power consumption,

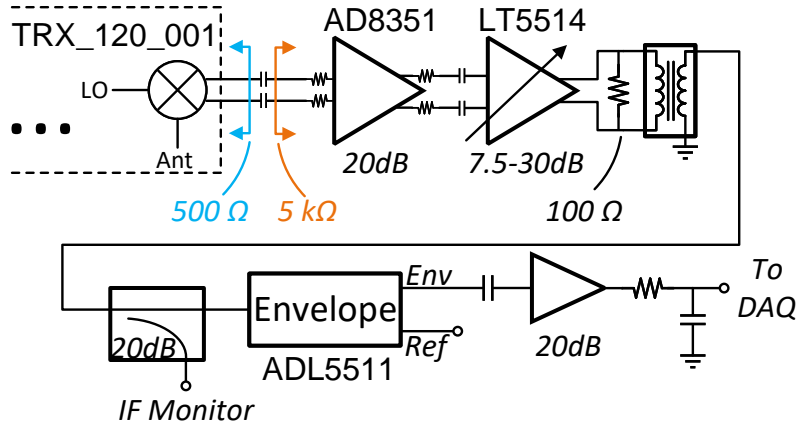


Figure 4.6: Receiver IF amplification and envelope detection.

however, amplification at IF is more efficient. Here, an IF frequency of 50 MHz is chosen as a trade-off between noise and bandwidth. The block diagram of the receiver including IF amplification and detection is given in Figure. 4.6.

The IF output from the TRX_120_001 is not 50- Ω terminated. The 500 Ω output impedance requires an amplifier with high input impedance, to avoid loading of the mixer. Op amps are usually preferred, due to the high input impedance and resistor adjustable gain. On the other hand, to buffer and amplify a differential signal, a differential amplifier should be used, otherwise, half of the signal is lost. Standard commercial high-speed fully-differential amplifiers usually require four matched resistors to set the gain, and the input impedance of the circuit is directly dependent on the resistance. Without a fully-differential amplifier, the differential output could be amplified by two single-ended amplifiers. Here, a dedicated IF amplifier was used to simplify the design, with a fixed input impedance of 5 k Ω and a single gain setting resistor. The unmatched interface between the mixer and the amplifier potentially gives rise to ringing or oscillation due to the amplifier parasitic. The unwanted ringing results from the inductance of the wirebond inside the chip in resonance with the parasitic capacitor from the PCB trace and internal transistor [74]. Without proper damping of the resonance, unwanted tones could present in the IF output spectrum. Depending on the PCB layout, these tones may well reside in the IF bandwidth, thereby corrupting the desired signal. Small series resistors in the

range of 10 to 20 Ω at the input and output of the amplifier would sufficiently suppress the oscillation.

The first stage amplification is followed by a variable gain amplifier (VGA) LT5514. Although the gain of AD8351 is adjustable by a single resistor, it is almost fixed at maximum in practice, because the gain from AD8351 alone is insufficient. The gain of the VGA is set by 4 digital control bits, ranging from 7.5 to 30 dB. The minimum gain step is 1.5 dB. A balun converts the differential output of the LT5514 to single-ended. The output stage of LT5514 is a gm-cell, therefore, a 100 Ω resistor is placed in parallel with the balun primary, for impedance matching. A 20 dB coupler follows the balun, where the coupled port is used as IF monitoring. This function is found to be useful to adjust the VGA gain for optimum input level into the envelope detector. Automatic level control was considered but abandoned, because the gain does not change drastically over time for a fixed setup. Initial adjustment of the gain is sufficient. ADL5511 is a versatile detector, which incorporates an envelope detector as well as an RMS detector. Although an RMS detector could demodulate slow amplitude variations, general purpose AM detection makes use of envelope detectors. The output of ADL5511 is pseudo-differential: the envelope is indicated by the voltage difference between Env and Ref pins. For dc envelope measurements, a differential amplifier extracts the envelope by subtracting V_{Ref} from V_{Env} . This is a necessary step in measuring dc envelope, because V_{Ref} tracks V_{Env} over temperature when there is no input [75]. In the molecular clock, all the information needed to control the VCXO is the ac component in the envelope, while the dc content will be completely removed by the subsequent lock-in operation. Thus, the envelope pin is ac coupled to an op amp with 20 dB gain. The output of the op amp is followed by a low pass filter before entering the ADC. The low pass filter prevents aliasing, if the ADC in the data-acquisition (DAQ) module is not equipped with one. The DAQ module that was used in this experiment does not have a anti-alias filter. The analog bandwidth of the ADC is several times higher than the maximum sampling rate. This configuration enables some sub-sampling experiments that may be useful in other applications. However, in this experiment, the unfiltered out-of-band noise folded into the Nyquist frequency degrades

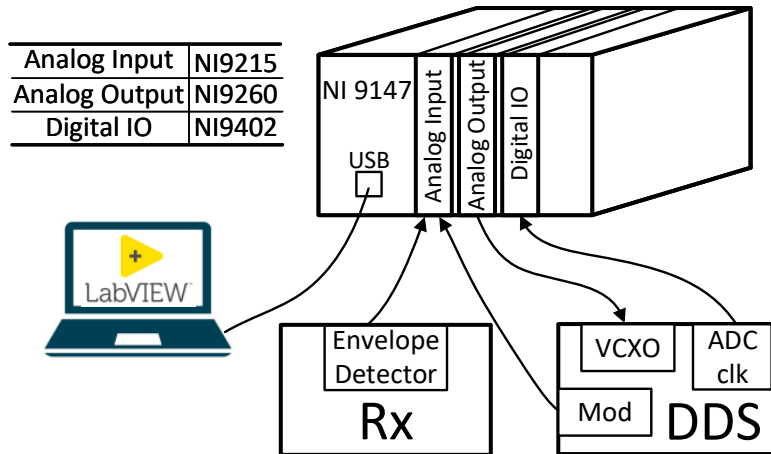


Figure 4.7: The setup of DAQ system with the circuit boards and LabVIEW. The transmitter and gas cell are not shown.

the SNR for about 20 dB.

4.1.4 The Lock-in

The lock-in operation is performed digitally. The data conversion between analog and digital, as well as the numerical computations, are implemented with the hardware and software from National Instruments, shown in Figure. 4.7. The hardware is based on the compact DAQ module NI9147 together with several IO modules. The NI9147 is a USB expansion chassis for NI C-series modules. The computer running a LabVIEW program communicates with the NI9148 through the USB interface. Three IO modules are installed to complete the setup: one analog input, one analog output and one digital IO. The model of each module is listed in the table in Figure. 4.7. The 24-bit analog output drives the voltage tuning pin of the VCXO on the DDS board. The four-channel 16-bit input module takes two inputs: one from the envelope detector and the other from the modulation generator, which are simultaneously sampled. The sample clock of the analog input module comes from the divided VCXO at 100 kHz, which is synchronized with the modulation frequency.

Without coherence between the analog signal and the sample clock, the lock-in output will be periodically modulated even if the amplitude of the analog signal is constant. In

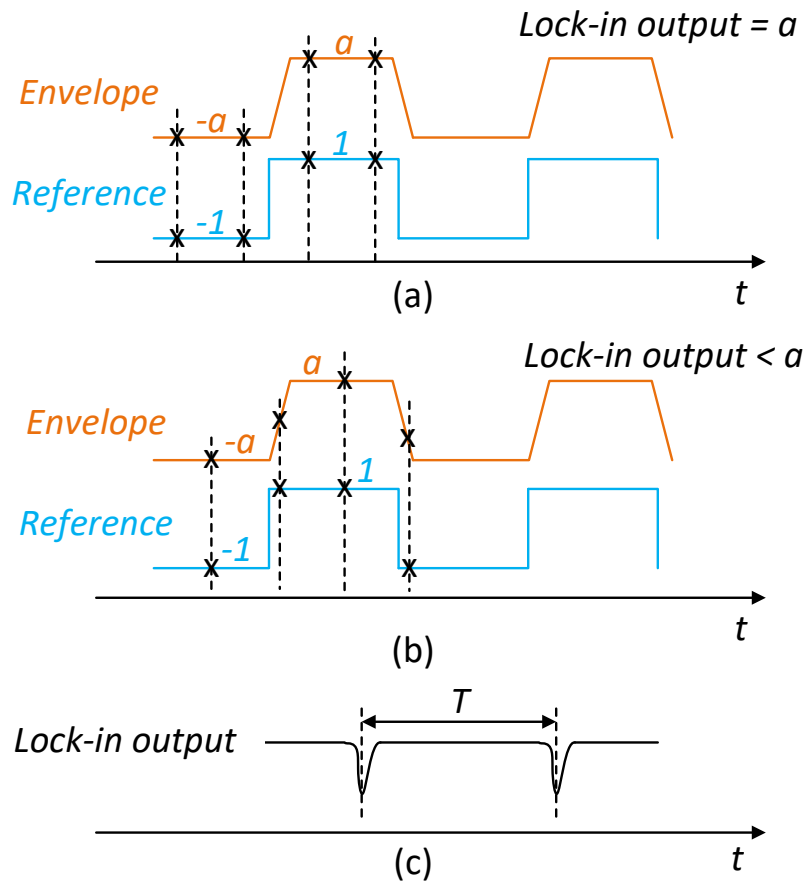


Figure 4.8: The analog waveform of the envelope and the reference, each sampled by an ADC with a sample rate four times the analog fundamental frequency. (a) ADC timing where square wave transitions are avoided. (b) ADC timing where two samples are located at the rising and falling edge of the envelope. (c) The resulting lock-in output for an unsynchronized ADC clock over time, when the analog input is constant.

Figure. 4.8 (a) and (b), signal and reference waveforms with different timing with respect to the ADC sample clock are plotted. The envelope waveform is modeled as a square wave with finite rising and falling time, due to the finite bandwidth of either the envelope detector or the ADC anti-aliasing filter. The reference is a perfect square wave, whose samples are adjusted to either $+1$ or -1 in the digital domain. Consider an ADC with a sample rate four times the analog bandwidth. Four samples will be collected in one period. The cross marks in Figure. 4.8 (a) and (b) denote the sampled value. In Figure. 4.8 (a), the sampled envelope and reference take values of $\pm a$ and ± 1 , respectively. Therefore, the lock-in output, calculated as the dot product of the sampled envelope and reference normalized to the number of sampling points, is equal to a . For an ADC sample clock unsynchronized to the analog signal, the sampling phase shifts at a different time. In Figure. 4.8 (b), two samples are collected at the rising and falling edge of the envelope signal. This results in a lock-in value that is smaller than a . If the lock-in value is plotted versus time with a constant input envelope, a periodic dip could be observed, whose period T is related to the frequency difference between the analog signal and the ADC sample clock.

The feedback control of the system is implemented in LabVIEW. The top-level architecture is a producer-consumer model, shown in Figure. 4.9 (a). After initialization of the data converters, the ADCs sample the envelope and the reference continuously, which is called the producer. The sampled data are pushed into a queue every Δt . An individual consumer loop, which takes the last queue element and performs lock-in and integration operations, runs in parallel with the producer loop. The numerical calculation updates faster than the digitizing loop, ensuring an empty queue for most of the time. The separation of slow acquisition and fast update loop eliminates the stalling of the loop due to the slower process.

The details of the dual FSK lock-in and integration loop are given in Figure. 4.9 (b). If dual FSK is not used, the *AuxFSK* input could be deactivated by tying to a high level without changing the code. All the inputs in this diagram are digitized and all the calculations are done in the digital domain. The two reference signals are processed first

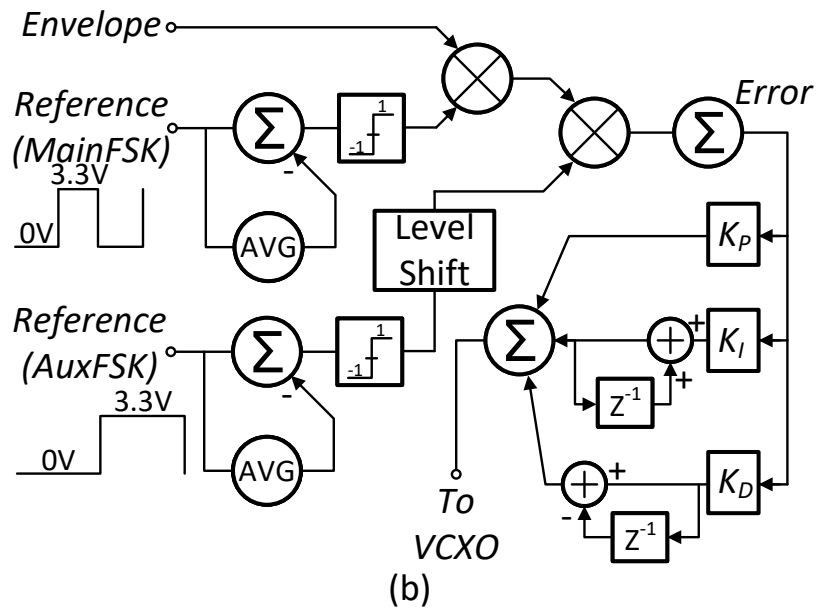
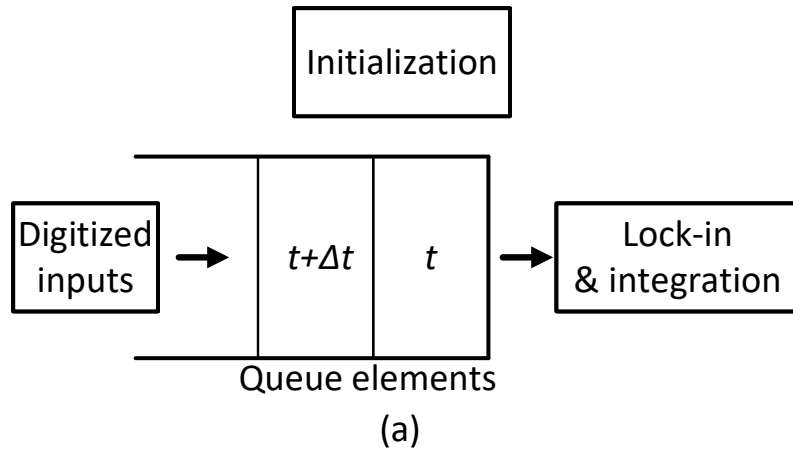


Figure 4.9: (a) The top-level architecture for the LabVIEW program, which is a classical producer-consumer model. (b) Detailed block diagram for the lock-in and integration operation.

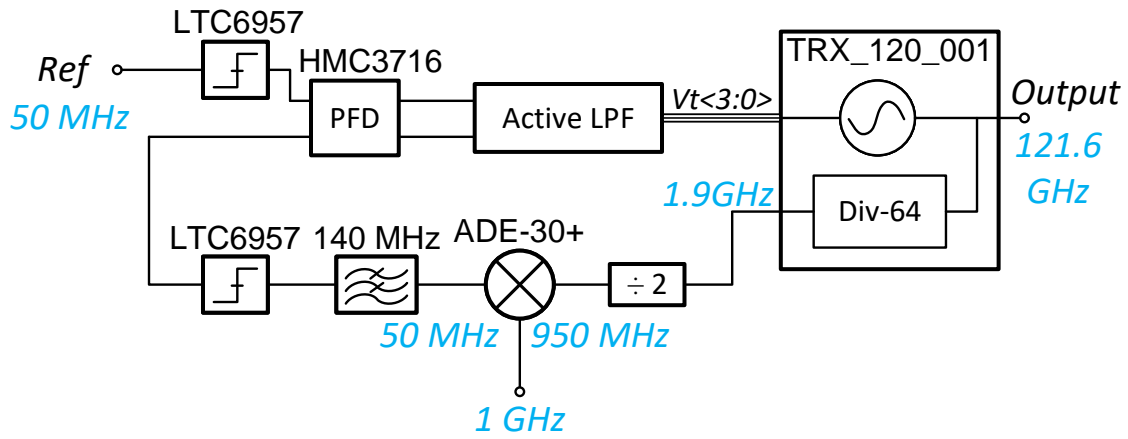
before multiplying with the envelope. An ideal square wave reference is noiseless, takes values of ± 1 only, and has no dc offset. In this setup, the reference signals are generated by the modulation generator in the DDS. Those 3.3V CMOS logic waveforms swing from 0V to 3.3V, with a dc offset of 1.65V. In order to be used as a lock-in reference, the dc offset of both the main FSK and the auxiliary FSK signals must be removed. Each of the zero mean reference signals is then followed by a comparator, which outputs only +1 and -1. The comparator eliminates any noise in the high and low states. For the auxiliary path, the lock in reference should be scaled and level shifted to be equal to the *Ref2* in Figure. 2.8 (b), which is denoted as a Level Shift block in Figure. 4.9 (b). The error signal is subsequently fed to a standard PID controller. The output of the PID controller drives the voltage tuning port of the VCXO, closing the feedback loop.

It should be noted that, it is not necessary to sample the main and auxiliary FSK references, if the digital computations are implemented on a application-specific integrated circuit (ASIC). The reference signals from the DDS may be used to trigger the logic gates directly to modify the sign of the acquired data. Therefore, the processing associated with the sampled reference to remove the dc and noise can be removed. Also, significant power and area saving are possible by eliminating the two analog-to-digital converters.

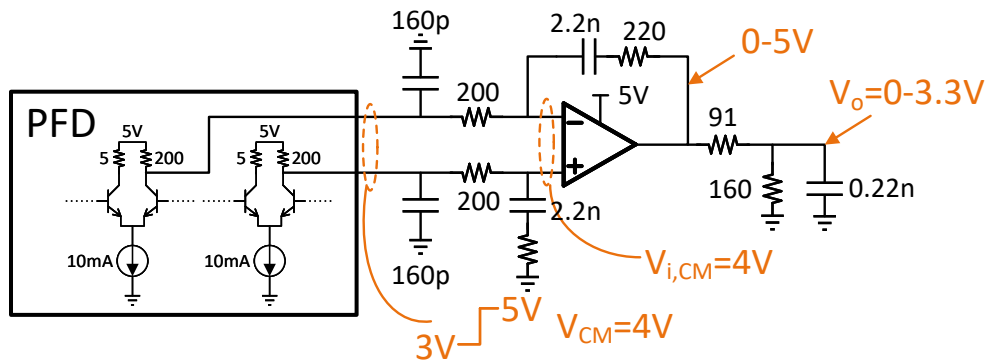
4.2 A Low Phase Noise Transmitter

4.2.1 The Transmitter

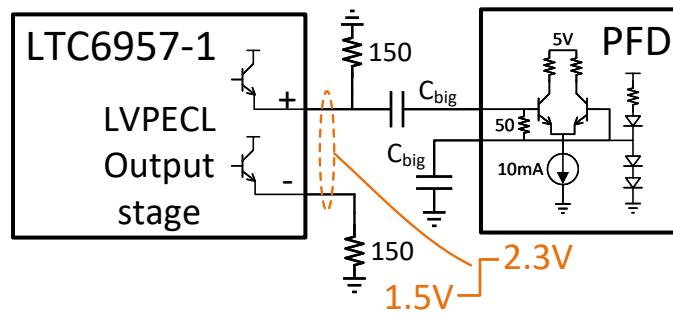
The phase noise of the improved system is reduced by introducing a mixer in the PLL loop. The schematic of the new system is shown in Figure. 4.10 (a). The mixer downconverts the LO signal from the mm-Wave transceiver by a clean 1 GHz signal. Essentially, the mixer and the 1 GHz LO replace the divider in a conventional PLL. From the PLL noise transfer function, the noise contribution from the reference and the PFD is the input referred noise of each component multiplied by the divider ratio. For a fixed reference frequency, reducing the divider ratio improves the PLL output phase noise. Compared to the block diagram in Figure. 4.3 (a), the divider ratio reduced from 32 to 2. The ratio of 2 comes from the divider after the 1.9 GHz output from the mm-Wave transceiver.



(a)



(b)



(c)

Figure 4.10: (a) System block diagram of LO generation of the improved transceiver. (b) The design considerations of the active low pass filter. (c) The interface between the comparator LTC6957-1 and the PFD HMC3716.

The main purpose of this divider is to provide some buffering between the mm-Wave LO output and the mixer input, and to match with the 1 GHz LO frequency. In the mixer path, the output of the mixer must be low-pass filtered, because the nonlinearity in the mixer produces numerous intermodulation components of the LO and RF signals. To further improve the phase noise, a dedicated low phase noise PFD HMC3716 was used. Optimal phase noise of this PFD was achieved with a square wave input, so both inputs to the PFD were converted to the sine wave by low phase noise comparators. Finally, the output of the PFD does not include a charge pump, which then requires an active filter and integrator.

The design of the active filter is straight forward. The filter parameters in Figure. 4.10 (b) were designed using the PLL design tool provided by Analog Devices. This set of parameter results in a loop bandwidth of about 500 kHz. Two design considerations are associated with the active filter, one at the op amp input and one at the output. The PFD outlined in the PFD block in Figure. 4.10 (b) is not equipped with a standard logic output. If the differential pair is driven by large inputs, the 10 mA tail current would be conducted by either branch of the differential pair. Combined with a 200 Ω load resistor, the output swing of the PFD is between 3V and 5V, which translates to a common mode output voltage of 4V. Since there is little current flowing through the 200 Ω resistor, the 4V common mode voltage is applied to the input of the op amp. Therefore, the op amp must be able to handle 4V input common mode range without degradation of the performance. Considering the op amp supply voltage of 5V, the high input common mode voltage puts one constraint on the op amp that can be used. While it is possible to increase the power supply of the op amp, this would require redesigning the power management of the TRX as well as the power supply PCB. Another concern is raised by the voltage swing at the op amp output. The tuning voltage of the VCO ranges from 0V to 3.3V, which suggests the op amp to support rail-to-rail output. On the other hand, the 5V supply voltage of the op amp enables a rail-to-rail output op amp to generate voltages up to 5V during start up or transient. Driving the gate tuning voltage over its limit could potentially cause breakdown of internal transistors in the transceiver. In Figure. 4.10 (b), a resistive divider

is placed at the op amp output, attenuating the maximum output voltage to about 3.3V.

The LTC6957 is a family of sine wave to square wave converters. Four models exist, with different output driver circuit. The one used in this design, LTC6957-1, has an LVPECL output, which is essentially an emitter follower. The simplified output stage is sketched in Figure. 4.10 (c). From the datasheet, the typical voltage swing of the single-ended LVPECL output is from 1.5V to 2.3V, or 800mV peak to peak. The RF input of the PFD consists of a 50- Ω terminated and self-biased differential pair. The recommended and absolute maximum input power is 5 dBm and 13 dBm [76], respectively. Equivalently, it corresponds to an rms input voltage of 0.4 V and 1 V. In an early version of the design, the PFD is driven by the LTC6957-1 differentially through ac-coupling. The differential output swing is 1.6 Vpp or 0.8 Vrms for the square wave. Although the input swing to the PFD is less than the absolute maximum, two PFD chips were damaged during the measurement. In a modification later, one arm of the LTC6957-1 output was terminated, as well as one of the PFD input, as shown in Figure. 4.10 (c). This reduces the input voltage swing by a half, and the PFD was never damaged again.

4.2.2 The LO Generation and the Reference

A dedicated 1 GHz source, shown in Figure. 4.11 (a), was designed to serve as the local oscillator for the mixers in the transmitter and receiver. The 1 GHz signal is generated from a $\times 10$ multiplier chain, consisting of one doubler and one $\times 5$ frequency multiplier. The 100 MHz VCXO output is first doubled, filtered and amplified before a $\times 5$ frequency multiplier. The order of the frequency multipliers is not particular important here, because all the spurious are eventually cleaned out by the filters. Moreover, since the 1 GHz source serves as the LO for the mixer in the PLL, as long as the spurious is not an image of the 1 GHz LO, all other spurs are filtered either by the low pass filter that follows the mixer, or the loop filter of the PLL. The 100 MHz VCXO first drives a doubler. The output of the doubler consists of multiple harmonics of 100 MHz. It was amplified by a 10 dB amplifier and then filtered before going into another 10 dB amplifier. The reason the filter was sandwiched between two amplifiers is that the output impedance of the multiplier is unknown, and that the filter performance could be degraded with a high

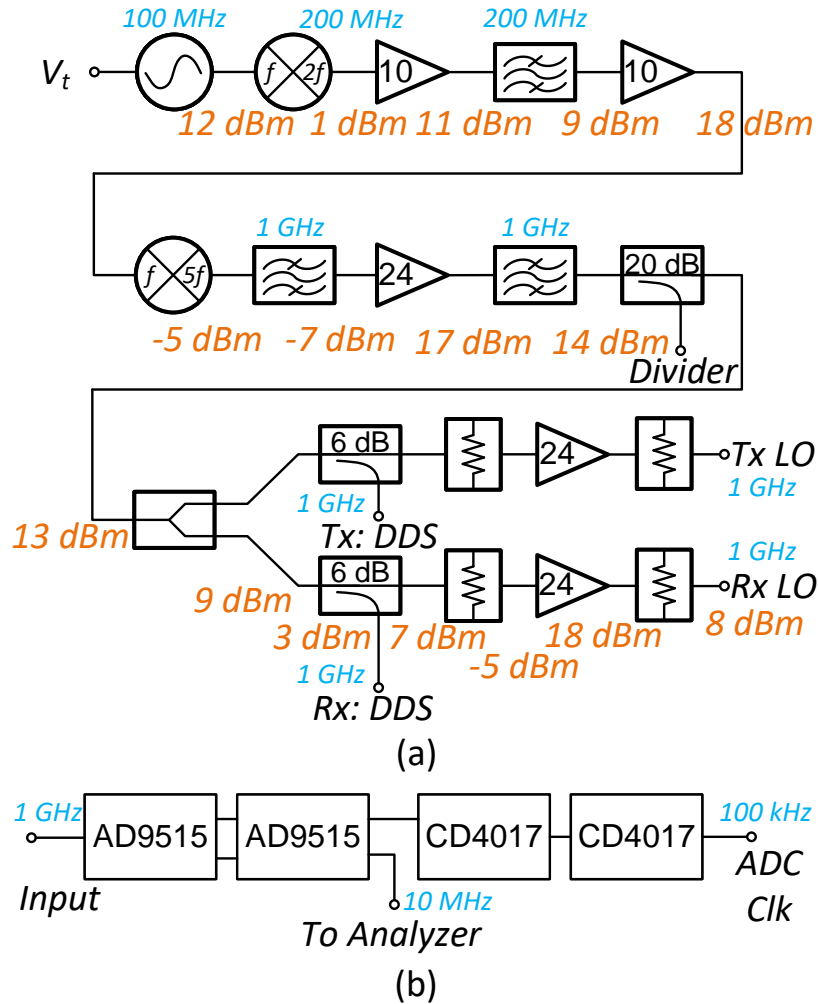


Figure 4.11: (a) A multiplier chain from 100 MHz VCXO to 1 GHz. The 100 MHz input is first multiplied by 2 and then by 5. Amplifiers are inserted between multiplication stages to boost the signal up, and bandpass filters remove the spurious tones. (b) The divider chain from 1 GHz to obtain the 10 MHz signal for stability measurement and the 100 kHz ADC sampling clock.

Description	Model	Manufacturer
Frequency doubler	RK-3+	Mini-Circuits
Frequency multiplier $\times 5$	RMK-5-13+	Mini-Circuits
Amplifier 10 dB	GVA-81+	Mini-Circuits
Bandpass filter 200MHz	RBP204+	Mini-Circuits
Amplifier 24 dB	ADL5545	Analog Devices
Bandpass filter 1 GHz	CBP-1023A+	Mini-Circuits
Coupler 20 dB	BDCA-10-25+	Mini-Circuits
Power splitter	GP2S+	Mini-Circuits
Coupler 6 dB	BDCA-6-16+	Mini-Circuits
High speed divider	AD9515	Analog Devices
Logic counter	CD74HC4017	Texas Instruments

Table 4.1: A list of devices used in the 1 GHz frequency multiplier chain.

VSWR termination. The second amplifier brings the power to a decent level to driving the $\times 5$ frequency multiplier. There are two filters following the second multiplier, hoping to provide better filtering than a single filter. Notice that the input to the first filter is not well-matched, and therefore it may not be as efficient as the second one. Due to the high conversion loss of the multiplier, an amplifier with 24-dB is used to boost up the signal level. A portion of the filtered amplifier output feed to the divider to generate a 10 MHz signal for stability measurement and a 100 kHz sample clock to the ADC. The main path of the coupler is divided, one for the Tx LO and the other for the Rx.

The signal path for Tx and Rx following the divider is designed for high isolation. To understand how the leakage between the two LO paths causes an issue, the mixers in the transceiver PCBs have to be included. Consider a simplified block diagram in Figure. 4.12, where two LO paths from a simple power divider are connected to two transceiver boards. An ideal mixer with RF frequency f_{RF} and LO frequency f_{LO} generates an intermediate frequency of $f_{RF} \pm f_{LO}$. The output of a physical mixer consists of numerous intermodulation tones of $m * f_{RF} + n * f_{LO}$, where m and n are integer numbers.

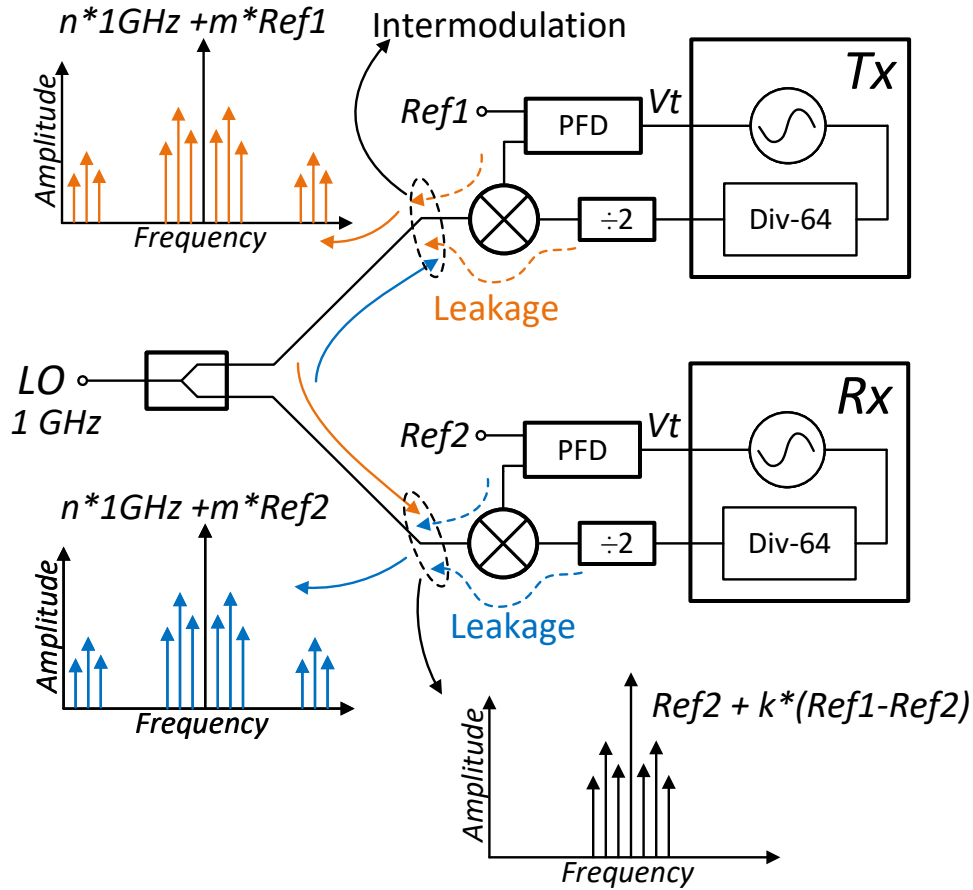
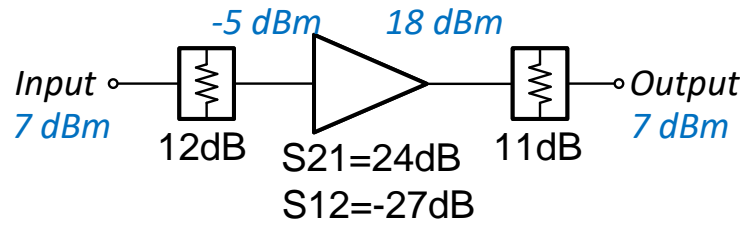


Figure 4.12: Simplified block diagram of the low phase noise transceiver and the illustration of LO leakage induced intermodulation.

These intermodulation tones leak to the LO port due to the finite isolation between IF and LO ports. The leakage of the transmitter and receiver are sketched in orange and blue in Figure. 4.12, respectively. Since the LOs from the transmitter and receiver are derived from a power divider, the Tx and Rx leakage may travel through the power divider to the other side. For example, the Tx mixer leakage contents would go through the splitter, attenuated by the splitter output isolation, and reach the LO port of the Rx mixer. The LO of the Rx mixer is thus contaminated with the intermodulation spurs from the Tx, which may be represented as $n \cdot 1\text{ GHz} + m \cdot \text{Ref1}$. The LO will be further mixed with the divided-by-2 signal from the mm-Wave transceiver. Intermodulation of the mixer produces beat tones of Ref1 and Ref2 around the desired frequency Ref2 . Consequently, these spurious frequency tones show up at the mixer IF output, as well as



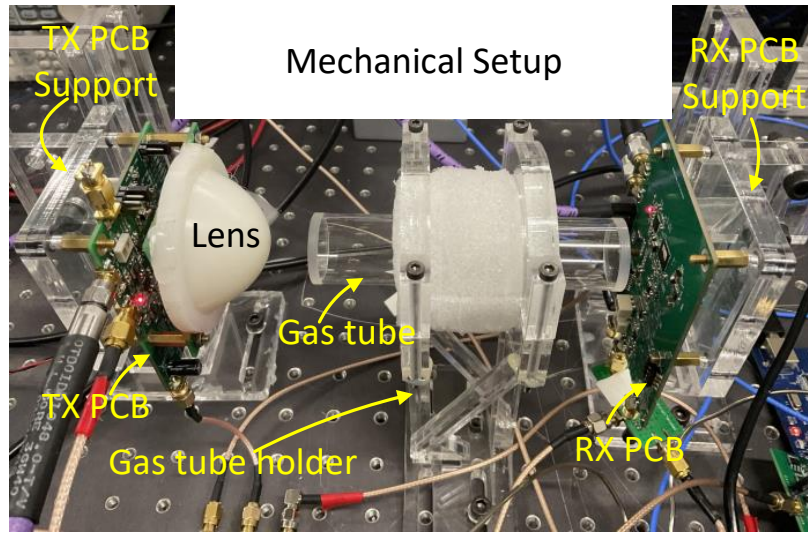
Total forward gain = 0dB

Total reverse gain = $-27 - 12 - 11 = -50\text{dB}$

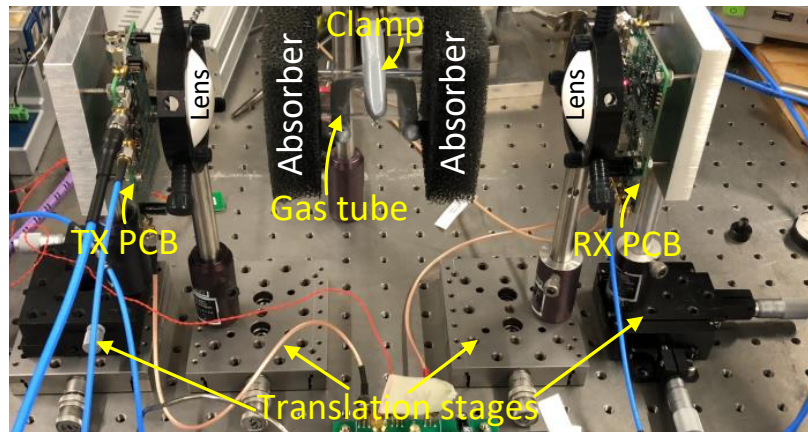
Figure 4.13: LO buffer for high reverse isolation.

the output of the mm-Wave transceiver.

The problem arises from the intermodulation and poor L-I isolation of the mixer, which however are the inherent limitations of a mixer. Fortunately, the problem could be solved if the isolation between the Tx and Rx LO is improved. The additional isolation is provided with a combination of amplifiers and attenuators, which is redrawn in Figure. 4.13. The required output power for a level-7 mixer is 7 dBm. The input power to the buffer is also 7 dBm, which translates to a required forward gain of 0 dBm. The input attenuator brings the power to the amplifier down to -5 dBm, which is about the level of the 1-dB input compression point. The resulting 18-dBm output power is then reduced to 7dBm by another 11 dB attenuator. Although the forward gain is no different from a wire, this configuration provides significant attenuation for the reflected wave. The reverse isolation comes from three parts: the two passive attenuators which attenuates the reverse signal the same way as the forward wave, and the reverse isolation of the amplifier. The combined reverse isolation is about 50 dB. It should be noted that, the major disadvantage of combining the amplifier and attenuator to increase isolation is the high power consumption. The implementation in this work only serves as a proof-of-concept design. A more elegant way to design a buffer amplifier with high isolation is using a common-gate or common-base amplifier. Isolation up to 150 dB was achieved by double cascading a common-base amplifier [77].



(a)



(b)

Figure 4.14: Two iterations of the mechanical setup: (a) an early attempt: System 1 and (b) a modified version: System 2.

4.2.3 The DDS

4.3 Mechanical Setup

Two iterations of the mechanical setup were carried out: System 1 and System 2. The first version of the setup makes use of acrylic boards to hold the gas tube and the transmitter and receiver PCB, shown in Figure. 4.14 (a). In a later version, metal pieces replaced the laser cut acrylic materials to provide a sturdy support for the gas tube and the PCBs, shown in Figure. 4.14 (b).

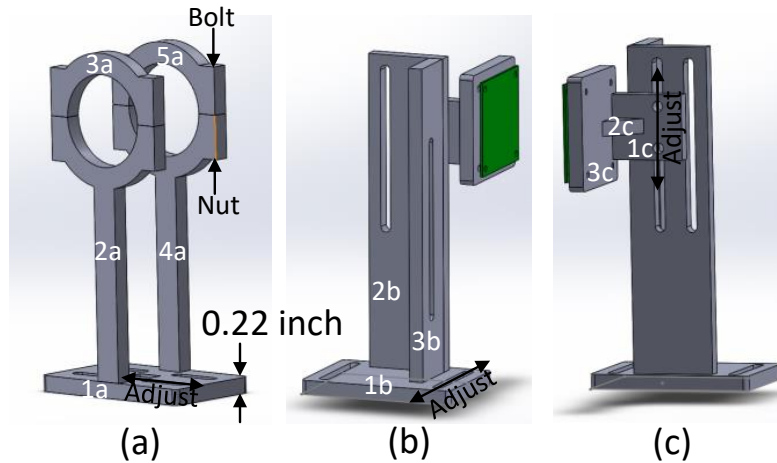


Figure 4.15: 3-D view of the acrylic fixtures: (a) gas tube holder, (b) front view of the transceiver PCB holder and (c) back view of the transceiver PCB holder.

The gas tube is made of quartz with 25-mm in diameter and 100-mm long. The tube is filled with carbonyl sulfide at a pressure of 75mTorr. It was manufactured from Precision Glassblowing Inc. Safety and cost were the major concerns regarding the design of the gas container. The prebuilt gas tube eliminates the need to handle the hazardous gas in the lab, and the sealed glass tube remain leak-tight so long as it is not broken. However, it should be noted that the design of a hermetically sealed gas package is usually an important topic in atomic clocks. A compact atomic clock calls for an integrated physics package with the electronics. These topics could be studied in the future work.

A 3-D view of acrylic fixtures for holding the gas tube and the transceiver PCBs are shown in Figure. 4.15 (a), (b) and (c). Each piece is fabricated from the 0.22-inch thick acrylic board, and multiple pieces are glued together. In the gas tube holder, piece 2a and 4a are glued to the base 1a, which is mounted on the optical breadboard. Piece 3a and 5a mate with 2a and 4a, respectively, clamped by long bolts and nuts at the sides. There are slots on the base piece 1a, which allows adjustment in one direction on the optical table. The PCB holder are made of 6 pieces, excluding the PCB itself. Pieces 2b and 3b are fixed onto the base 1b, which together form a post for the piece that the PCB is attached to. The PCB, labeled as green, is mounted on 3c through copper spacers, which are not shown in the graph. The piece 3c is attached to 1c by a triangular stiffener 2c, which moves vertically through the slots on 2b and holes on 1c. The base piece 1b

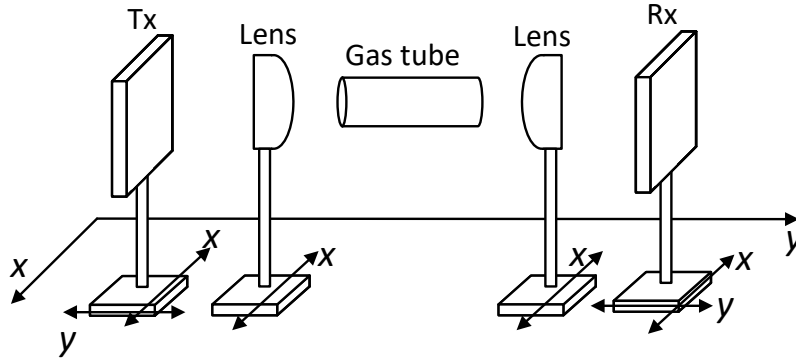


Figure 4.16: Illustration of the improved setup, where the transceiver and the lenses are mounted on translation stages through metal fixtures. The transmitter and receiver move in both x and y directions, and the lenses are adjustable in the x direction.

also provides adjustment in the horizontal direction. While tuning slots were added to the design, experimental results show that fine-tuning the position is difficult. Also, the acrylic components are not sufficiently stiff to support the PCBs. Therefore, it suggests an updated version to use metal fixtures and translation stages for sturdy support and precision control.

The improved physical setup in Figure. 4.14 (b) is redrawn in Figure. 4.16, to illustrate the adjustment of the transceiver and lenses. The gas tube is fixed in the middle, in which the chemical clamp that holds the tube is omitted. The height of the gas tube is tuned so that it is aligned with the transceiver and the lenses. All the adjustment occurs in the x-y plane. Both the lenses could be adjusted in the x-direction, while the transceivers move in both x- and y-direction. The x-direction adjustment of the transceiver aligns the off-axis antenna array in Figure. 4.1 (b) to the focal spot of the lens. The spacing of the transceiver to the lens in the y-direction is determined by maximizing the received power. The tuning range of the y-direction translation stage must be larger than half of the hole pitch on the optical bread board, for continuous adjustment over a wide range.

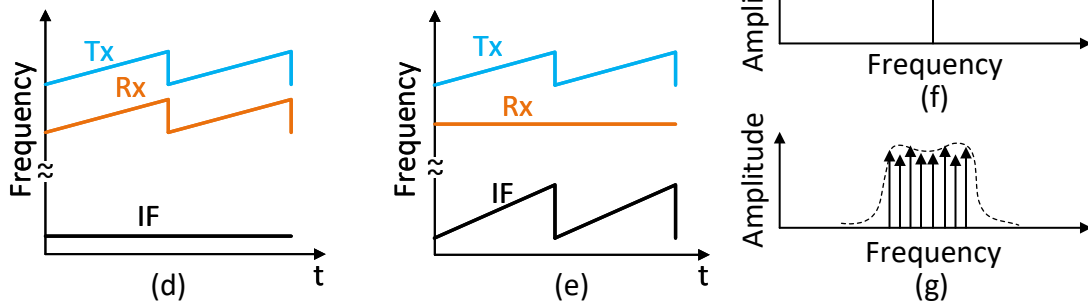
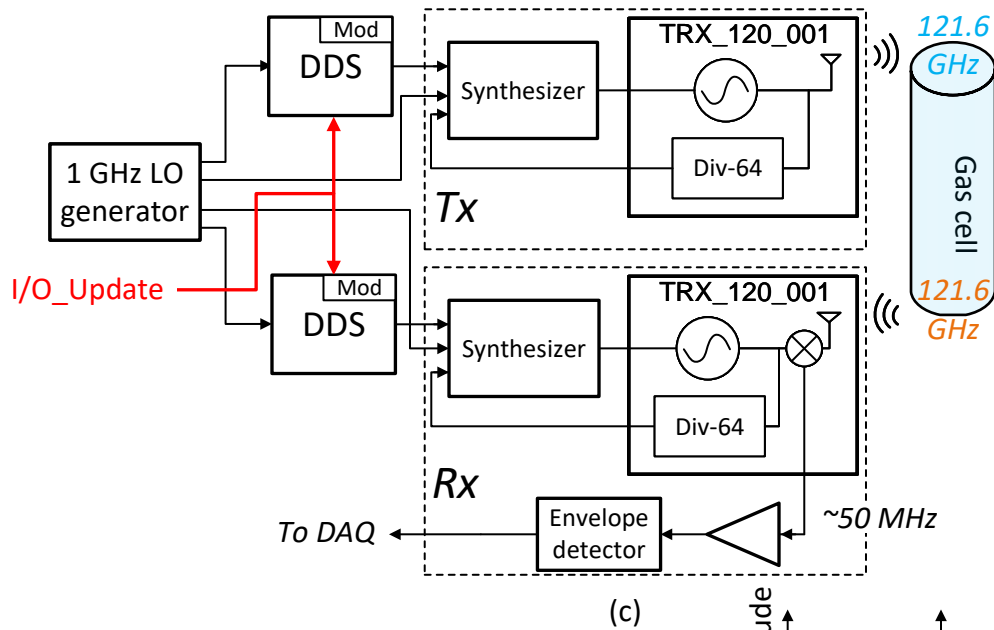
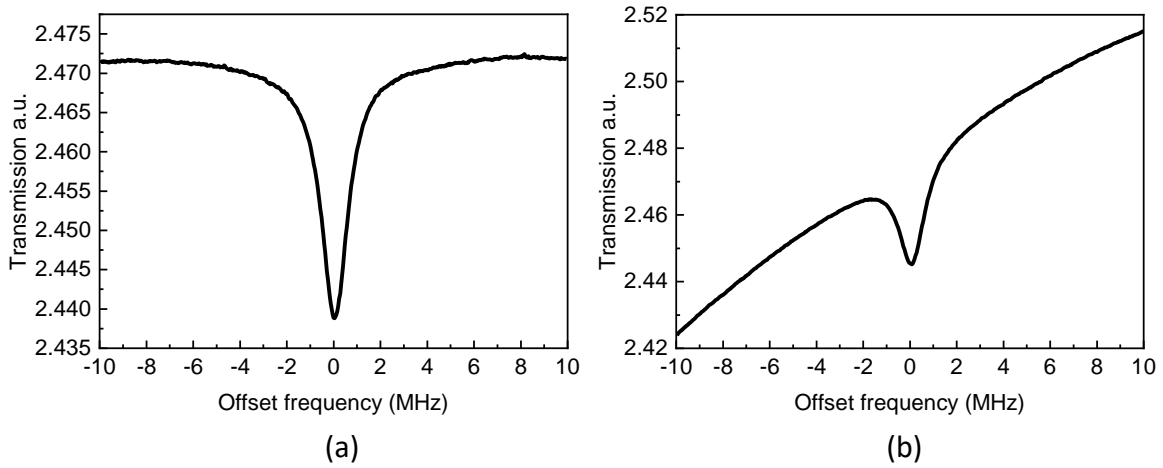


Figure 4.17: Measured gas absorption profile at about 121.6 GHz with (a) synchronized Tx and Rx LO and (b) fixed Rx LO frequency. (c) The test bench for measuring the gas absorption with System 2. (d) Tx and Rx frequency chirps when they are aligned results in a fixed IF frequency. (e) Time varying IF frequency with a Rx frequency. (f) and (g) are the associated IF spectrum of (d) and (e), respectively.

4.4 Measurements

4.4.1 Measured Absorption Profile

The measured gas absorption profiles are shown in Figure. 4.17 (a) and (b), with the hardware based on the System 2. The profile is measured by sweeping the transmitter frequency while measuring the received power. The DDS generator in the Tx performs the frequency sweep. The measured absorption profile in Figure. 4.17 (a) is based on the synchronized Tx and Rx LO, whereas in Figure. 4.17 (b), the receiver LO is constant. Raw measured data is shown in both plots, without baseline cancellation. In either cases, the center frequency of the IF equals to 50 MHz.

The setup for measuring the gas transmission is shown in Figure. 4.17 (c). The feedback loop that stabilizes the VCXO is removed, and the modulation scheme in the Tx is changed from FM or FSK to a sawtooth chirp. The Tx chirp is generated from the internal chirp source in the AD9910 DDS through frequency multiplication by the synthesizer. The transmitted chirp at 121.6 GHz corresponds to the blue curve in Figure. 4.17 (d) and (e). The frequency transitions of the Tx and Rx chirp are synchronized by issuing a common I/O_Update to both DDS generators, highlighted in red in Figure. 4.17 (c). For the receiver LO, two options are available: one is an offset LO frequency to the RF that follows a fixed IF frequency; the other is a constant LO frequency with a modulated IF. The LO and IF frequency in the receiver are illustrated in Figure. 4.17 (d) and (e), in orange and black, respectively. Notice that the modulation depth in these plots are exaggerated only for illustration purposes.

The spectrum of a modulated IF in Figure. 4.17 (f) occupies wider bandwidth than a single tone IF, shown in Figure. 4.17 (g). Since the frequency response of the IF circuits, including the amplifier and the detector, could not be perfectly flat, they contribute to the baseline of the measured gas absorption profile. In the heterodyne system, the absolute bandwidth at RF equals to the IF bandwidth, yet the fractional bandwidth at IF is significantly higher. Assuming the amount of gain variation per fractional bandwidth at different frequencies is comparable, the smaller the fractional bandwidth, the flatter the baseline. The FMCW chirp from the transmitter takes up 20 MHz bandwidth, which

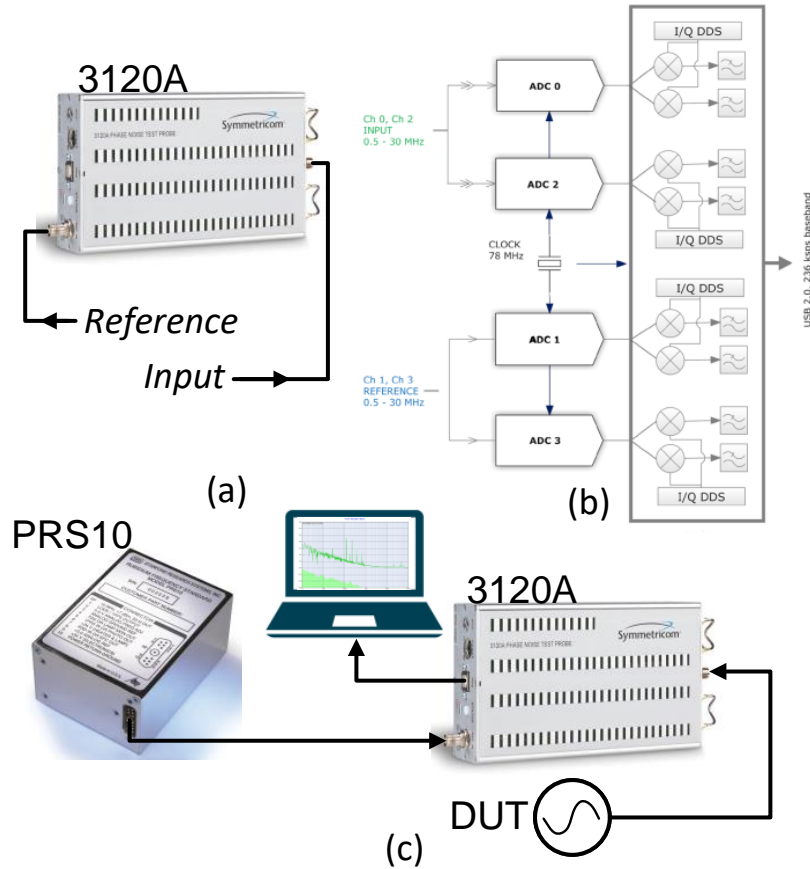


Figure 4.18: (a) Microsemi 3120A phase noise probe. (b) Architecture of the 3120A, where each input channel is power-split before sampled by two identical analog-to-digital converters [78]. (c) A typical setup for measuring phase noise and stability. A Rubidium atomic clock PRS10 from thinkSRS serves as the reference to the 3120A.

translates to a fractional bandwidth of 0.16‰ at RF and 40% at IF. As a result, the baseline in Figure. 4.17(b) is considerably steeper.

4.4.2 Phase Noise

The phase noise and stability measurement in this work are based on the 3120A phase noise probe, shown in Figure. 4.18(a), from Microsemi, formerly Symmetricom. The 3120A is a direct digitizing cross-spectrum analyzer, which converts the clock inputs to the digital domain and perform the downconversion and cross-correlation numerically. Compared to the popular phase detector based phase noise analyzer, where an internal or external synthesizer generates a copy of DUT in quadrature, the digital implementation eliminates the need for a high quality microwave synthesizer. As a result, a single reference

could measure the performance of a DUT with arbitrary frequency. The disadvantage is that, the input frequency in 3120A must satisfy the Nyquist criterion, which is half of the ADC sampling frequency. For the 78 MHz sampling clock used in the 3120A, the maximum frequency of either the reference or the DUT is limited to 30 MHz. One of the most important features in a modern phase noise analyzer is the cross-correlation technique. Essentially, the noise of two independent paths are uncorrelated, for example, the ADC0 and ADC2 in the input channel in Figure. 4.18 (b). The cross correlation analyzer rejects the uncorrelated noise and recovers the similarity in the two paths, which is the signal under test. By averaging multiple samples, the noise floor of a cross-correlation analyzer could be lower than any individual analyzer paths. The details of the cross-correlation technique were explained in [79].

The basic phase noise and stability analyzing setup is given in Figure. 4.18 (c). The 3120A analyzer itself only compares the difference of two inputs, which then requires a reference, against which the DUT is compared. The reference should be selected such that the performance is much better than the DUT. In the diagram, a Rubidium atomic clock PRS10 from thinkSRS serves as the reference to the phase noise analyzer. The exceptional stability of the atomic clock is sufficiently accurate for measuring Allan Deviation of the molecular clock. In terms of phase noise measurement, the reference phase noise must be at least several dB lower than the DUT. Since the reference frequency and the DUT frequency could be different due to the unique architecture of the 3120A, the phase noise floor at the DUT frequency is the maximum of the instrument and the reference

$$PN_{floor,DUT} = \max \left[PN_{floor,3120A}, PN_{floor,REF} + 20 \log \left(\frac{f_{DUT}}{f_{REF}} \right) \right] \quad (4.1)$$

The choice of the reference could affect the accuracy of the measured phase noise floor. Suppose the thermal phase noise of a 10 MHz DUT is below -165 dBc/Hz and a 5 MHz and 25 MHz reference with -160 dBc/Hz phase noise floor are available. Further, assume that the phase noise of the instrument is negligible. As a result, the phase noise floor at 10 MHz from the 5 MHz and 25 MHz reference are -154 dBc/Hz and -168 dBc/Hz, respectively. Apparently, only the higher frequency reference measures the phase noise floor accurately. Therefore, care must be taken for low phase noise floor measurements.

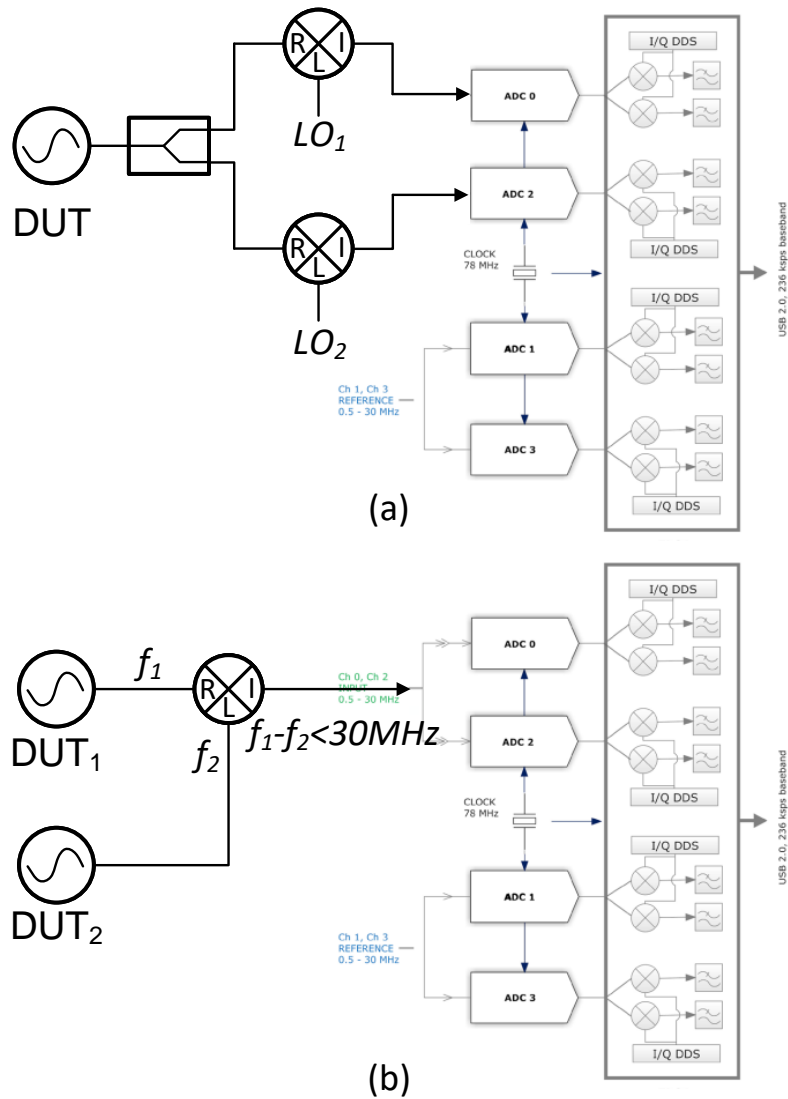


Figure 4.19: (a) Frequency extension of the 3120A using two uncorrelated LOs, leveraging the cross-correlation technique of the instrument. (b) Two identical DUTs with one acting as the RF and the other as the LO.

The maximum DUT frequency is below 30 MHz, due to the internal limitation of the instrument. It is possible to apply external RF downconverters to reduce the input frequency to the instrument, provided that the conversion process increases negligible additive phase noise. Owing to the accessibility to the internal ADCs of the DUT channel, cross-correlation could be used to cancel out the phase noise of the local oscillator in the RF downconverters. The block diagram for measuring phase noise above the capability of the 3120A is shown in Figure. 4.19 (a). Notice that the phase noise of the two local oscillators should be close to each other and no worse than the DUT by 20 dB, which is achieved by averaging over 10000 samples.

In a special case, where two DUTs are available and both are frequency tunable, a modified frequency extension setup could be adopted, as shown in Figure. 4.19 (b). Notice that the phase noise of the IF frequency is the contribution by the uncorrelated noise from two DUTs. If the phase noise of the two DUTs are completely uncorrelated and similar in magnitude, the measured IF phase noise is 3 dB higher than uncorrelated phase noise from each DUT. In the limiting case, if the uncorrelated phase noise of one DUT is significantly higher than the other, the measured IF phase noise is equal to the DUT with worse phase noise. In any case, the measured IF phase noise is less than 3 dB higher than the DUT with worse phase noise. On the other hand, the correlated phase noise from the two DUTs will be cancelled at the IF port, assuming negligible noise from the mixer. The noise which is shared by both DUTs is considered as correlated.

Recall that the mm-Wave transceiver is integrated with a local oscillator and mixer, the setup in Figure. 4.19 (b) could be used to measure the uncorrelated phase noise of the transceivers at 121.6 GHz. The setup to measure the mm-Wave phase noise is drawn in Figure. 4.20. The simplified block diagrams of System 1 and 2 are shown to identify the correlated and uncorrelated noise source. In System 1, correlated noise comes from the VCXO and in System 2, the noise from both the VCXO and the multiplier chain are correlated. The uncorrelated noise sources for System 1 and 2 are similar. The dominant uncorrelated in-band noise includes that of the DDS, the PFD and the additive noise from the divider inside the mm-Wave transceiver. The mixers used in System 2 contribute little

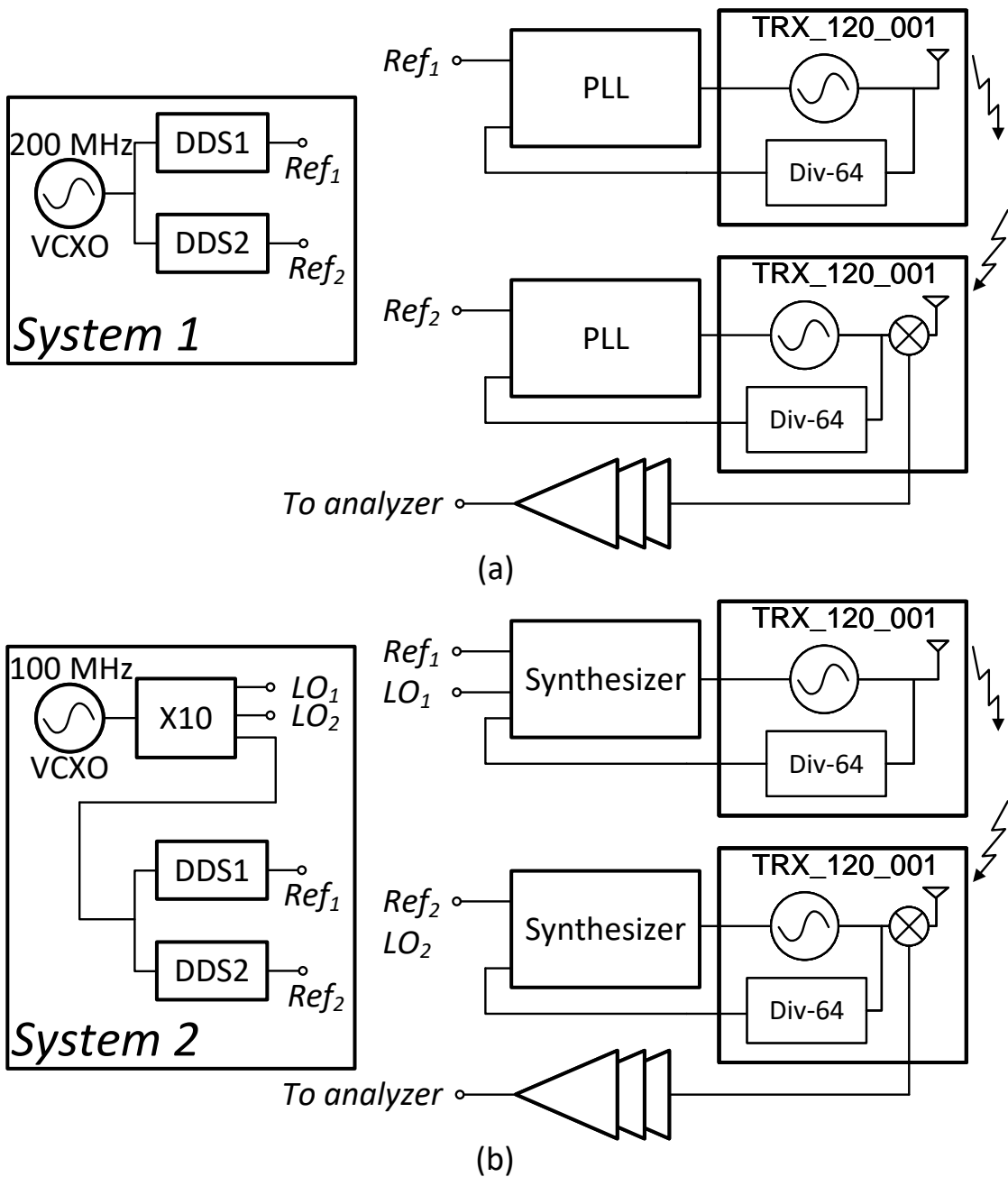


Figure 4.20: Phase noise measurement of the 120 GHz signal. (a) The setup of System 1 where the frequency synthesizer is a simple PLL. (b) The setup of System 2, which differs from System 1 by the low phase noise synthesizer.

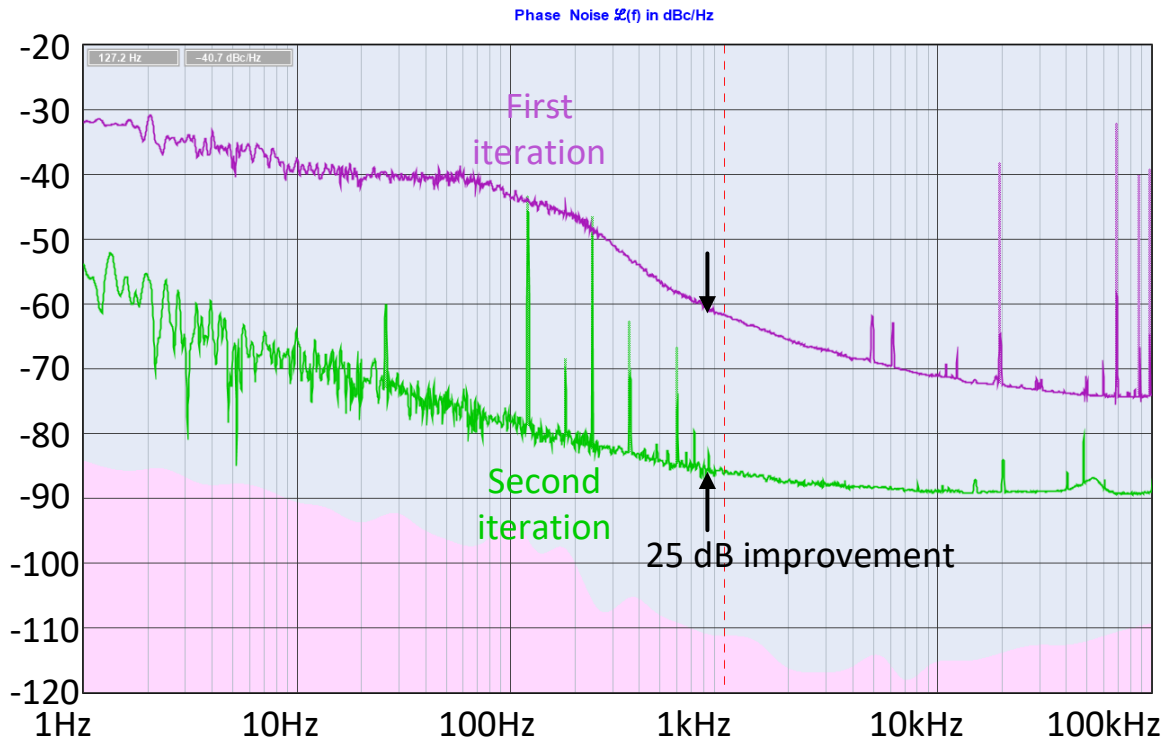


Figure 4.21: Measured phase noise of the 120 GHz transceiver. Actual phase noise is 3-dB lower for equal contribution of the oscillator in the Tx and the Rx.

phase noise and are neglected. The uncorrelated out-of-band noise is contributed by the VCO phase noise. Therefore, the setup in Figure. 4.20 measures the additive phase noise of the 121.6 GHz synthesizer, which equals to the absolute phase noise with a noiseless VCXO and/or multiplier chain.

The oscillators in both the transmitter and receiver are synthesized by the same approach. Thus, the approximation that each DUT contributes the equal phase noise is valid. The measured IF phase noise of the two systems are plotted in Figure. 4.21. At 1 kHz offset, where the modulation frequency resides, the phase noise of System 2 improves by 25 dB over System 1. The 60-Hz spur and its harmonics in the phase noise plot were due to the external amplifier.

4.4.3 Stability

The stability measurements in Figures. 4.22 (a), (b) and (c) were conducted with System 2 and the setup in Figure. 4.14 (b). The inconsistent results conducted with Figure. 4.14

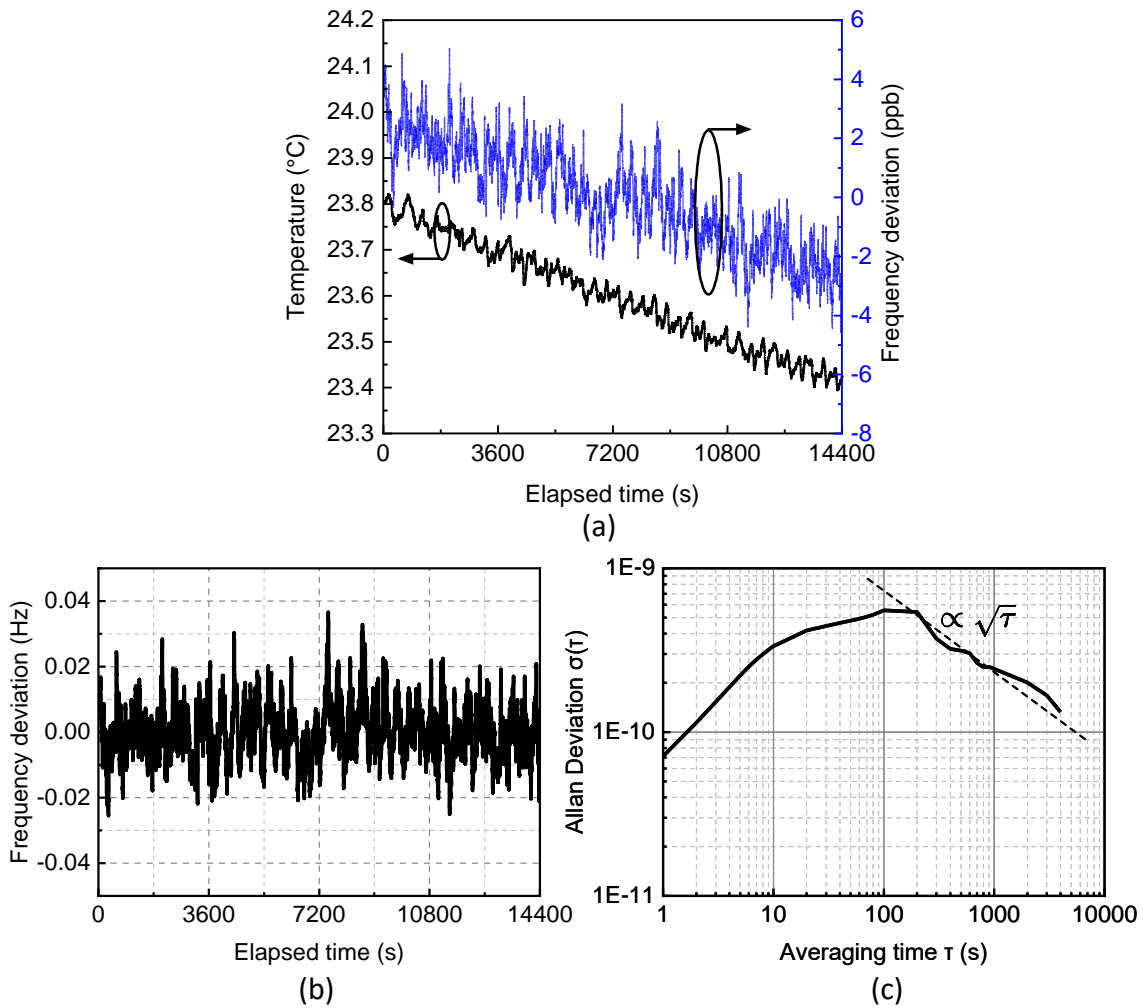


Figure 4.22: (a) Raw measured frequency and temperature data in 4 hours. (b) Measured frequency deviation from 10 MHz with temperature drift cancelled. (c) Measured Allan deviation with temperature drift cancelled.

(a) were discarded due to the lack of precision control of the transceiver to the lenses. It will be shown later that the locked frequency is a strong function of the position due to the standing wave problem. The raw measured frequency data in Figure. 4.22 (a) shows a linear correlation to the temperature. The frequency drift canceled data is plotted in Figure. 4.22 (b), where the peak-to-peak deviation is reduced to about 6 ppb. The Allan deviation of the drift-canceled frequency data is plotted in Figure. 4.22 (c). For integration time below about 10 s, which corresponds to the loop bandwidth of about 0.1 Hz, the Allan deviation increases with respect to the integration time, due to the free running VCXO. With an integration time greater than 10 s, the random frequency fluctuations are averaged out, resulting in a decreased slope. Between about 200 s and 4000 s, the ADEV decreases with a slope of $\sqrt{\tau}$, indicating a zero-mean frequency noise. The ADEV at 1000 s is 2.3×10^{-10} . In this system, the ADEV at 1000 s is limited by the SNR, due to the path loss between the transceivers, as well as the reflection at the air-glass interfaces. With a dedicated design of the gas cell and the interface between the gas and the electronics, the ADEV between 200 s and 4000 s is expected to be lower.

To study the temperature dependence in greater details, the setup in Figure. 4.23 (a) is configured. The gas cell is enclosed in a styrofoam box, which provides good thermal insulation to the ambient. A heater is placed inside the styrofoam box, which is made of two current sources mounted on a copper sheet. The schematic of the heater is shown in Figure. 4.23 (c). The voltage applied to the heater is chosen to dissipate approximately 10-W power. This results in a temperature range of 15°C. Higher heater power would increase the temperature range, but potentially introduces a stronger thermal diffusion, where the gas temperature in the tube may be less uniform. Power cycling was applied to the heater, to see whether the temperature was the only factor that affects the frequency. During each cycle, the heater turns on for $0 < t < 500$, $1000 < t < 2000$ and $3000 < t < 5000$, and turns off for the rest of the time. The temperature inside the foam box was monitored with a precision sensor LMT70 from Texas Instruments. The measured temperature and the frequency drift are shown in Figure. 4.23 (d), where the state of the heater is also illustrated. Clearly, the temperature drift is inversely proportional to the temperature,

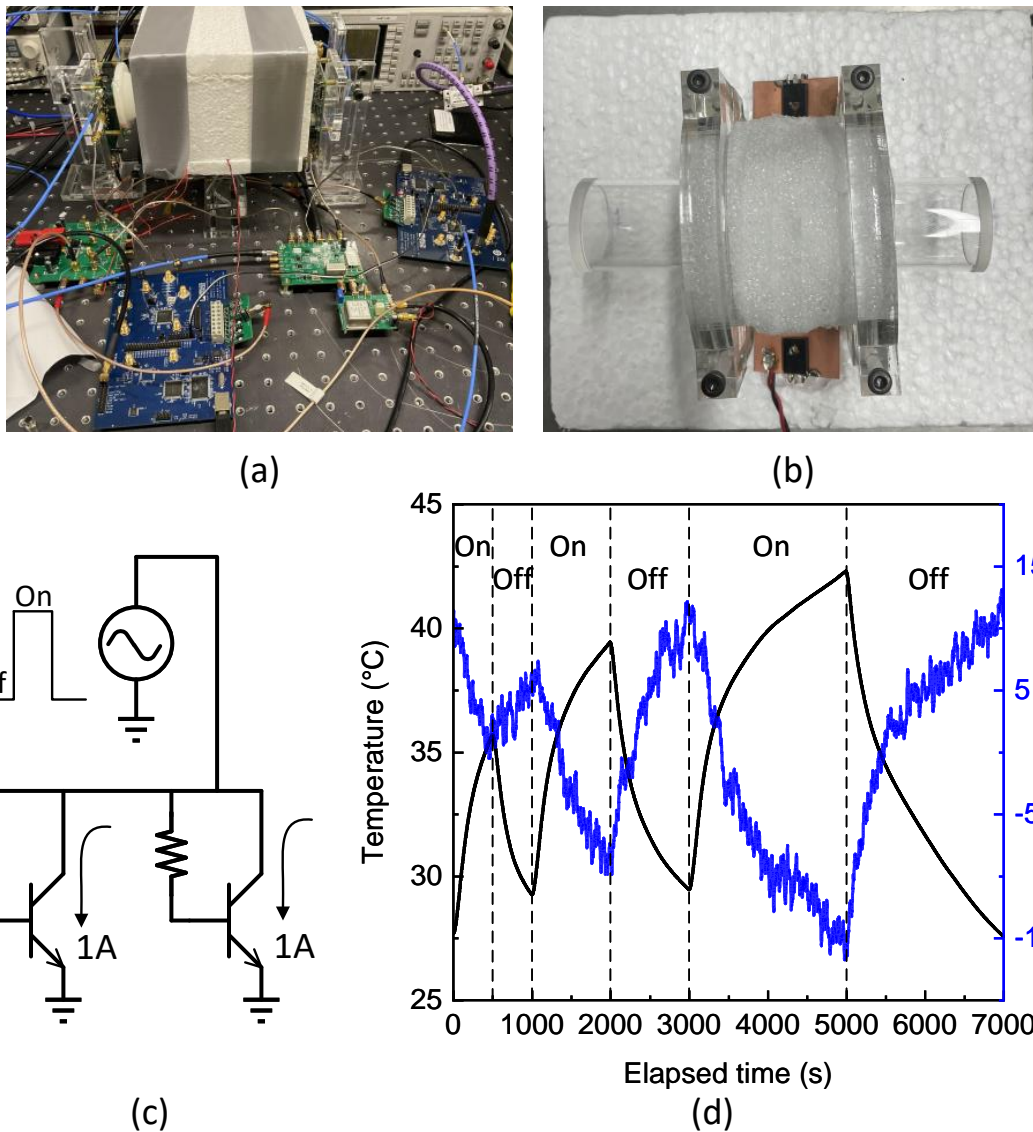


Figure 4.23: (a) The bench setup for measuring temperature dependence, where the gas tube is enclosed in a styrofoam box. (b) The inside of the styrofoam box where the heater is placed beneath the glass tube. (c) The heater circuit. (d) The temperature inside the styrofoam box and frequency drift in one period of the power cycle of 7000s. The measurement was conducted for 15 hours with multiple power cycles.

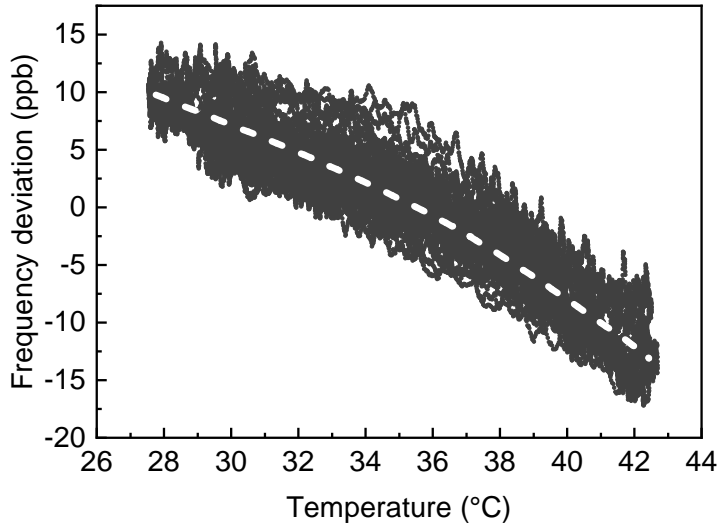
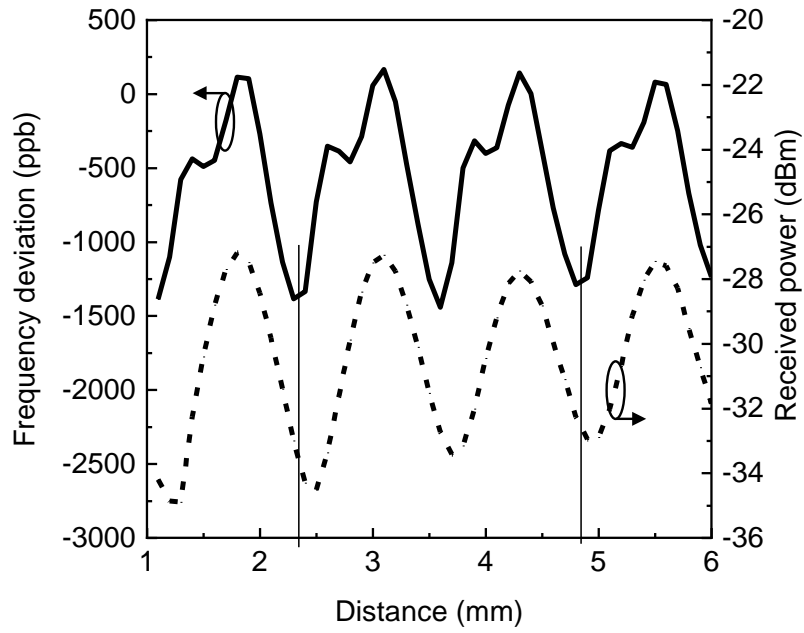


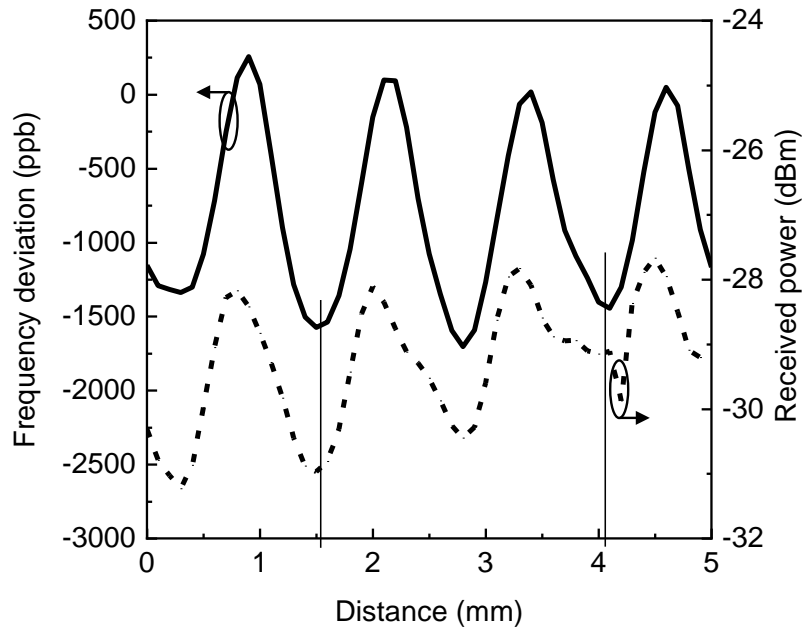
Figure 4.24: Scatter plot of the measured temperature drift with respect to temperature. Each dot is the measured frequency and temperature at each second. The dashed line indicates the most probable frequency-temperature relationship.

as opposed to the result in Figure. 4.22 (a). To better view the temperature dependence, a scatter plot of the frequency versus temperature is constructed in Figure. 4.24, where the measurement took a total of 15 hours with multiple periods in Figure. 4.23 (d). The measured frequency and temperature at each second are shown as a dot in Figure. 4.24. The frequency drifts approximately 30 ppb in the 15°C temperature range. A quadratic relation between the frequency and the temperature is observed, as indicated by the white dashed line. It is also observed that, at each temperature point, the frequency drift spans as large as 10 ppb, which suggests that temperature may not be the only factor that affects the frequency.

Due to the mismatch in the dielectric constant, reflection occurs at the air-quartz interface, which forms a standing wave with the incident wave. To study the effect of standing wave on the molecular clock frequency, the setup in Figure. 4.16 was used. The x-direction adjustment compensates the off-axis antenna on the transceiver chip, shown in Figure. 4.1 (b). Since if the transceiver and the glass tube are perfectly paralleled and aligned, the standing wave only happens in the y-direction. In the following measurements, the x-direction positions of all linear stages are fixed. The system is configured in



(a)



(b)

Figure 4.25: (a) Measured clock frequency and received IF power as a function of the distance between the Tx and the lens. (b) Measured clock frequency and received IF power as a function of the distance between the Rx and the lens.

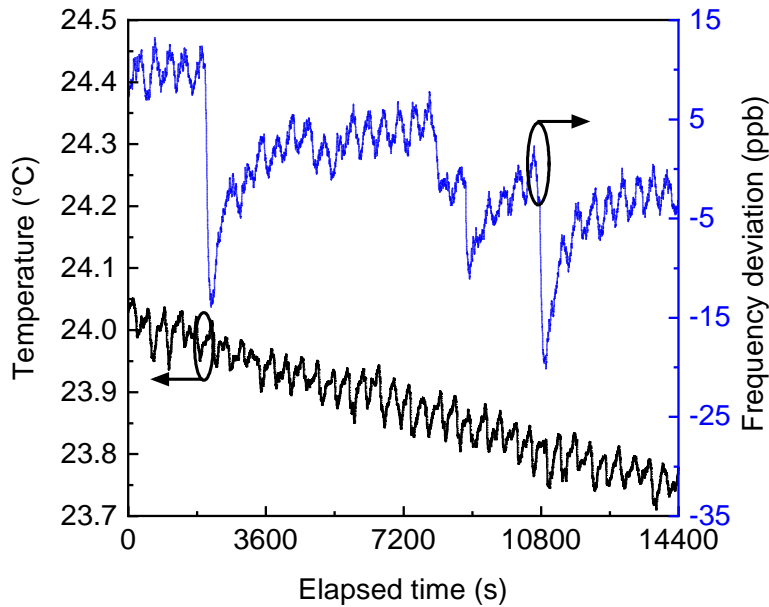


Figure 4.26: A measurement of the frequency and temperature of the molecular clock, where the positions of the transceivers were adjusted arbitrarily.

a closed-loop molecular clock, where the mm-Wave frequency is locked to the molecular spectral line. Moving the transmitter or receiver along the y -direction, while fixing the position of all other components, the received IF power and the instantaneous frequency are measured and plotted in Figures. 4.25 (a) and (b). The received power as a function of distance is a sign of the standing wave pattern. The distance between the minimums or maximums is approximately 1.25 mm, corresponding to half of the free space wavelength of at 121.6 GHz. The locked frequency is also a function of distance of both the transmitter and the receiver, where the peak frequency deviation in one period is approximately 1500 ppb. In addition, the period of the frequency change equals to the period of the received power. It should be noted that the phase of the frequency deviation and received power is arbitrary. The relationship changes after reinstalling the transceiver or the glass tube. The strong dependence of the frequency on the standing wave significantly impacts the stability of the clock. When the position of the transceiver is at the middle of standing wave peak and valley, the frequency sensitivity to the distance is high. Compared to Figure. 4.22 (a), where the transceivers are moved to a position least sensitive to distance,

Figure. 4.26 shows another measurement where the transceivers were adjusted arbitrarily. In addition to a higher temperature sensitivity than in Figure. 4.22 (a), there are sudden frequency jumps which do not follow the temperature curve. Consider that the only difference in the two setups in Figure. 4.22 (a) and Figure. 4.26 was the transceiver position, the sudden frequency jumps may be attributed to the frequency sensitivity to the standing wave. The exact mechanism that the standing wave changes the molecular clock frequency is unclear at the moment, which is left to the future experiments for validation. One hypothesis is that, since the transmitter is frequency modulated, the frequency of the incident and the reflected wave that shares the same propagation constant, are different. At any time, the absorbed EM energy equals to the absorption at two different frequencies. Since the instantaneous frequency difference of reflected and the incident wave depends on the time delay, which is proportional to the distance, the absorption of the reflected wave is expected to be a function of distance. Of course, future experimental validation is required to support this hypothesis.

Chapter 5

Summary

In this work, the theory and design of a molecular clock was presented. The design methodology of a high power millimeter wave source was also introduced. The design considerations of the molecular clock, including the architecture, the baseline issue, the transmitter phase noise and the lock-in amplifier, etc., were discussed. A molecular clock based on the $10 \leftarrow 9$ transition of carbonyl sulfide was implemented based on a 121 GHz mm-Wave transceiver. A low phase noise transmitter was designed to reduce the PM-AM noise. The lock-in process and subsequent integration were implemented in the digital domain. The factors of the clock frequency stability were discussed. To overcome the design difficulty of the high power millimeter-wave signal source in CMOS, a stack-FET technique was integrated to the embedding network method, providing an extra degree of freedom to achieve optimum inductor quality factor. In-phase power combining was also adopted in the design to double the output power without using explicit power combiners.

5.1 Future Work

The molecular clock frequency drift could be explored in more detail in the future work. In particular, a deeper understanding of the standing wave on the frequency drift is missing. It is also expected to characterize the frequency dependency on the temperature and measure the frequency stability with temperature drift canceled. In the oscillator design, the next challenge would be the tuning range, which plays an important role in the molecular clock to cover the process and temperature variations.

REFERENCES

- [1] H. Sonderegger, “History of portable sundials,” *The Compendium, Journal of the North American Sundial Society*, pp. 19–35, Mar. 2020.
- [2] W. A. Marrison, “The evolution of the quartz crystal clock,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 510–588, 1948. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1948.tb01343.x>
- [3] M. Headrick, “Origin and evolution of the anchor clock escapement,” *IEEE Control Systems Magazine*, vol. 22, no. 2, pp. 41–52, Apr. 2002.
- [4] W. S. Eichelberger, “Clocks—ancient and modern,” *Science*, vol. 25, no. 638, pp. 441–452, Mar. 1907.
- [5] A. R. Willms, P. M. Kitanov, and W. F. Langford, “Huygens’ clocks revisited,” *Royal Society Open Science*, vol. 4, Sep. 2017.
- [6] W. Milham, *Time and Timekeepers*. Macmillan, 1945.
- [7] F. J. Britten, *The Watch & Clock Makers’ Handbook, Dictionary, and Guide*. Skyhorse, 2011.
- [8] M. Lombardi, “The evolution of time measurement, part 2 - quartz clocks,” *IEEE Instrumentation and Measurement Magazine*, no. 14, 10-01 2011.
- [9] “Ultra-stable oscillators HSO14 and HSO13 for ground based applications.” [Online]. Available: <https://www.rakon.com/news/ultra-stable-oscillators-hso14-and-hso13-for-ground-based-applications>
- [10] J. Levine, “The history of time and frequency from antiquity to the present day,” *European Physical Journal H*, vol. 41, no. 1, Apr. 2016.
- [11] T. Heavner, T. Parker, J. Shirley, and S. Jefferts, “NIST F1 and F2,” 04 2009. [Online]. Available: <https://tf.nist.gov/general/pdf/2500.pdf>
- [12] E. Donley, T. Heavner, M. Tataw, F. Levi, and S. Jefferts, “Progress towards the second-generation atomic fountain clock at NIST,” in *Proceedings of the 2004 IEEE International Frequency Control Symposium and Exposition, 2004.*, 2004, pp. 82–86.
- [13] K. Beloy, M. I. Bodine, T. Bothwell, S. M. Brewer, S. L. Bromley, J.-S. Chen, J.-D. Deschênes, S. A. Diddams, R. J. Fasano, T. M. Fortier, Y. S. Hassan, D. B. Hume, D. Kedar, C. J. Kennedy, I. Khader, A. Koepke, D. R. Leibbrandt, H. Leopardi, A. D. Ludlow, W. F. McGrew, W. R. Milner, N. R. Newbury, D. Nicolodi, E. Oelker, T. E. Parker, J. M. Robinson, S. Romisch, S. A. Schäffer, J. A. Sherman, L. C. Sinclair, L. Sonderhouse, W. C. Swann, J. Yao, J. Ye, and X. Zhang, “Frequency ratio measurements at 18-digit accuracy using an optical clock network,” *Nature*, vol. 591, no. 7851, pp. 564–569, Mar 2021. [Online]. Available: <https://doi.org/10.1038/s41586-021-03253-4>

- [14] F. Major, *The Quantum Beat: Principles and Applications of Atomic Clocks*. Springer New York, 2007. [Online]. Available: https://books.google.com/books?id=tmdr6Wx_2PYC
- [15] E. Pedrozo, S. Colombo, C. Shu, A. Adiyatullin, Z. Li, E. Mendez, B. Braverman, A. Kawasaki, D. Akamatsu, Y. Xiao, and V. Vuletic, “Entanglement on an optical atomic-clock transition,” *Nature*, vol. 588, pp. 414–418, 12 2020.
- [16] S. Kolkowitz, I. Pikovski, N. Langellier, M. D. Lukin, R. L. Walsworth, and J. Ye, “Gravitational wave detection with optical lattice atomic clocks,” *Phys. Rev. D*, vol. 94, p. 124043, Dec 2016. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevD.94.124043>
- [17] C. Thornton and J. Border, *Radiometric Tracking Techniques for Deep Space Navigation*. John Wiley Sons, 01 2005.
- [18] H. Schuh and D. Behrend, “VLBI: A fascinating technique for geodesy and astrometry,” *Journal of Geodynamics*, vol. 61, pp. 68–80, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0264370712001159>
- [19] K. Kebkal, O. Kebkal, I. Glushko, V. Kebkal, L. Sebastiao, A. Pascoal, J. Gomes, J. Ribeiro, S. H., and M. Ribeiro, “Underwater acoustic modems with integrated atomic clocks for one-way travel-time underwater vehicle positioning,” in *Proc. of the 4th Underwater Acoustics Conference and Exhibition (UACE)*, Sep. 2017.
- [20] L. Essen and J. V. L. Parry, “An Atomic Standard of Frequency and Time Interval: A Cæsium Resonator,” , vol. 176, no. 4476, pp. 280–282, Aug. 1955.
- [21] “SI brochure: The international system of units (SI).” [Online]. Available: <https://www.bipm.org/en/publications/si-brochure>
- [22] E. F. Arias, “Report of the international association of geodesy 2011-2013.” [Online]. Available: https://iag.dgfi.tum.de/fileadmin/IAG-docs/Travaux2013/08_BIPM.pdf
- [23] “Record number of frequency standards contribute to international atomic time.” [Online]. Available: <https://www.bipm.org/en/-/2021-12-21-record-tai>
- [24] H. E. Ives and G. R. Stilwell, “An experimental study of the rate of a moving atomic clock,” *J. Opt. Soc. Am.*, vol. 28, no. 7, pp. 215–226, Jul 1938. [Online]. Available: <http://opg.optica.org/abstract.cfm?URI=josa-28-7-215>
- [25] J. C. Hafele and R. E. Keating, “Around-the-world atomic clocks: Observed relativistic time gains,” *Science*, vol. 177, no. 4044, pp. 168–170, 1972. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.177.4044.168>
- [26] S. Herrmann, F. Finke, M. Lülf, O. Kichakova, D. Puetzfeld, D. Knickmann, M. List, B. Rievers, G. Giorgi, C. Günther, H. Dittus, R. Prieto-Cerdeira, F. Dilssner, F. Gonzalez, E. Schönemann, J. Ventura-Traveset, and C. Lämmerzahl,

- “Test of the gravitational redshift with Galileo satellites in an eccentric orbit,” *Phys. Rev. Lett.*, vol. 121, p. 231102, Dec 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.121.231102>
- [27] T. Bothwell, C. J. Kennedy, A. Aeppli, D. Kedar, J. M. Robinson, E. Oelker, A. Staron, and J. Ye, “Resolving the gravitational redshift within a millimeter atomic sample,” 2021, arxiv:2109.12238[physics.atom-ph].
- [28] K. Akiyama, K. Bouman, and D. Woody, “First M87 event horizon telescope results. I. the shadow of the supermassive black hole,” *Astrophysical Journal Letters*, vol. 875, 04 2019.
- [29] Event Horizon Telescope Collaboration, T. Event Horizon Telescope Collaboration, K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A. Baczko, D. Ball, M. Baloković, J. Barrett, D. Bintley, L. Blackburn, W. Boland, K. Bouman, G. Bower, M. Bremer, C. Brinkerink, R. Brissenden, S. Britzen, A. Broderick, D. Brogiere, T. Bronzwaer, D. Byun, J. Carlstrom, A. Chael, C. Chan, S. Chatterjee, K. Chatterjee, M. Chen, Y. Chen, I. Cho, P. Christian, J. Conway, J. Cordes, G. Crew, Y. Cui, J. Davelaar, M. De Laurentis, R. Deane, J. Dempsey, G. Desvignes, J. Dexter, S. Doeleman, R. Eatough, H. Falcke, V. Fish, E. Fomalont, R. Fraga-Encinas, P. Friberg, and C. Gammie, “First M87 event horizon telescope results. II. array and instrumentation,” *Astrophysical Journal Letters*, vol. 875, no. 1, Apr. 2019, publisher Copyright: © 2019. The American Astronomical Society..
- [30] D. Dondurur, “Marine seismic data acquisition,” in *Acquisition and Processing of Marine Seismic Data*. Elsevier, 2018, ch. 2, pp. 37–169. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128114902000025>
- [31] C. Cai, D. A. Wiens, W. Shen, and M. Eimer, “Water input into the mariana subduction zone estimated from ocean-bottom seismic data,” *Nature*, vol. 563, no. 7731, pp. 389–392, Nov. 2018. [Online]. Available: <https://doi.org/10.1038/s41586-018-0655-4>
- [32] S. Shimizu, M. Katagiri, Y. Watarai, Y. Ueda, P. Sack, and F. Gonzalez, “Semipermanent ocean-bottom seismic node: Toward practical reservoir monitoring,” *The Leading Edge*, vol. 38, pp. 716–719, Sep. 2019.
- [33] A. T. Gardner and J. A. Collins, “Advancements in high-performance timing for long term underwater experiments: A comparison of chip scale atomic clocks to traditional microprocessor-compensated crystal oscillators,” in *2012 Oceans*, Oct. 2012, pp. 1–8.
- [34] R. Eustice, H. Singh, and L. Whitcomb, “Synchronous-clock, one-way-travel-time acoustic navigation for underwater vehicles,” *Journal of Field Robotics*, vol. 28, pp. 121 – 136, Jan. 2011.

- [35] R. M. Eustice, L. L. Whitcomb, H. Singh, and M. Grund, “Experimental results in synchronous-clock One-Way-Travel-Time acoustic navigation for autonomous underwater vehicles,” in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 4257–4264.
- [36] M. Lombardi, T. Heavner, and S. Jefferts, “NIST primary frequency standards and the realization of the SI second,” *NCSLI Measure: The Journal of Measurement Science*, vol. 2, pp. 74–89, 12 2007. [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=50602
- [37] Consultative Committee for Time and Frequency (CCTF), “Report of the 21st meeting (8-9 June 2017) to the international committee for weights and measures.” [Online]. Available: <https://www.bipm.org/documents/20126/30132574/CCTF21.pdf/525bb9fb-23db-377f-03d5-947b58731c2d>
- [38] “Primary frequency standard 5071a.” [Online]. Available: <https://www.microsemi.com/product-directory/cesium-frequency-references/4115-5071a-cesium-primary-frequency-standard#resources>
- [39] “Chip scale atomic clock SA.45s.” [Online]. Available: <https://ww1.microchip.com/downloads/en/DeviceDoc/00002982.pdf>
- [40] P. Cash, W. Krzewick, P. Machado, K. R. Overstreet, M. Silveira, M. Stanczyk, D. Taylor, and X. Zhang, “Microsemi chip scale atomic clock (CSAC) technical status, applications, and future plans,” in *2018 European Frequency and Time Forum (EFTF)*, 2018, pp. 65–71.
- [41] J. Camparo, “The rubidium atomic clock and basic research,” *Physics Today - PHYS TODAY*, vol. 60, p. 12, 12 2007. [Online]. Available: <https://physicstoday.scitation.org/doi/10.1063/1.2812121>
- [42] “PRS10 rubidium frequency standard.” [Online]. Available: <https://www.thinksrs.com/downloads/pdfs/catalog/PRS10c.pdf>
- [43] H. Zhang, H. Herdian, A. T. Narayanan, A. Shirane, M. Suzuki, K. Harasaka, K. Adachi, S. Yanagimachi, and K. Okada, “29.4 Ultra-low-power atomic clock for satellite constellation with 2.2×10^{-12} long-term allan deviation using cesium coherent population trapping,” in *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2019, pp. 462–464.
- [44] W. V. Smith, J. L. G. De Quevedo, R. L. Carter, and W. S. Bennett, “Frequency stabilization of microwave oscillators by spectrum lines,” *Journal of Applied Physics*, vol. 18, no. 12, pp. 1112–1115, 1947. [Online]. Available: <https://doi.org/10.1063/1.1697592>
- [45] D. Sullivan, “Time and frequency measurement at NIST: the first 100 years,” in *Proceedings of the 2001 IEEE International Frequency Control Symposium and PDA Exhibition (Cat. No.01CH37218)*, 2001, pp. 4–17.

- [46] D. J. Wineland, D. A. Howe, M. B. Mohler, and H. W. Hellwig, "Special purpose ammonia frequency standard a feasibility study," *IEEE Transactions on Instrumentation and Measurement*, vol. 28, no. 2, pp. 122–132, 1979.
- [47] R. Besson, "A new 'electrodeless' resonator design," in *31st Annual Symposium on Frequency Control*. IEEE, 1977, pp. 147–152.
- [48] "OXCO 8600, 10 times better than any other OCXO." [Online]. Available: <http://www.sungwhatech.com/product/pdf/case/8600.pdf>
- [49] C. Wang, X. Yi, J. Mawdsley, M. Kim, Z. Hu, Y. Zhang, B. Perkins, and R. Han, "Chip-scale molecular clock," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 914–926, 2019.
- [50] C. Wang, X. Yi, M. Kim, Q. B. Yang, and R. Han, "A terahertz molecular clock on CMOS using high-harmonic-order interrogation of rotational transition for medium-/long-term stability enhancement," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 2, pp. 566–580, 2021.
- [51] M. Kim, C. Wang, Z. Hu, and R. Han, "Chip-scale terahertz carbonyl sulfide clock: An overview and recent studies on long-term frequency stability of OCS transitions," *IEEE Transactions on Terahertz Science and Technology*, vol. 9, no. 4, pp. 349–363, 2019.
- [52] "The HITRAN database." [Online]. Available: <https://hitran.org/>
- [53] C. Wang, X. Yi, J. Mawdsley, M. Kim, Z. Wang, and R. Han, "An on-chip fully electronic molecular clock based on sub-terahertz rotational spectroscopy," *Nature Electronics*, vol. 1, pp. 421–427, Jul. 2018.
- [54] B. Razavi, *RF Microelectronics*. Pearson Education, 2011. [Online]. Available: <https://books.google.com/books?id=zTnD1RgHbbkC>
- [55] M. Danielson, "AM to PM conversion of linear filters," 2018, arXiv: 1805.07224[eess.SP].
- [56] K. Hung, P. Ko, C. Hu, and Y. Cheng, "A physics-based MOSFET noise model for circuit simulators," *IEEE Transactions on Electron Devices*, vol. 37, no. 5, pp. 1323–1333, 1990.
- [57] W. Redman-White and D. Leenaerts, "1/f noise in passive CMOS mixers for low and zero IF integrated receivers," in *Proceedings of the 27th European Solid-State Circuits Conference*, 2001, pp. 41–44.
- [58] B. Razavi, "The harmonic-rejection mixer [a circuit for all seasons]," *IEEE Solid-State Circuits Magazine*, vol. 10, no. 4, pp. 10–14, Nov. 2018.

- [59] N. A. Moseley, Z. Ru, E. A. M. Klumperink, and B. Nauta, "A 400-to-900 MHz receiver with dual-domain harmonic rejection exploiting adaptive interference cancellation," in *2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, Feb. 2009, pp. 232–233,233a.
- [60] O. Momeni and E. Afshari, "High power terahertz and millimeter-wave oscillator design: A systematic approach," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 3, pp. 583–597, March 2011.
- [61] H. Wang, J. Chen, J. T. S. Do, H. Rashtian, and X. Liu, "High-efficiency millimeter-wave single-ended and differential fundamental oscillators in CMOS," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 8, pp. 2151–2163, Aug 2018.
- [62] R. Han and E. Afshari, "A CMOS high-power broadband 260-GHz radiator array for spectroscopy," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3090–3104, Dec 2013.
- [63] A. Imani and H. Hashemi, "Frequency and power scaling in mm-wave colpitts oscillators," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 5, pp. 1338–1347, 2018.
- [64] M. Vehovec, L. Houselander, and R. Spence, "On oscillator design for maximum power," *IEEE Transactions on Circuit Theory*, vol. 15, no. 3, pp. 281–283, Sep 1968.
- [65] A. Agah, J. A. Jayamon, P. M. Asbeck, L. E. Larson, and J. F. Buckwalter, "Multi-drive stacked-FET power amplifiers at 90 GHz in 45 nm SOI CMOS," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 5, pp. 1148–1157, 2014.
- [66] H. Dabag, B. Hanafi, F. Golcuk, A. Agah, J. F. Buckwalter, and P. M. Asbeck, "Analysis and design of stacked-FET millimeter-wave power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 4, pp. 1543–1556, 2013.
- [67] A. Chakrabarti and H. Krishnaswamy, "High-power high-efficiency class-E-like stacked mmwave PAs in SOI and bulk CMOS: Theory and implementation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 8, pp. 1686–1704, 2014.
- [68] Y. M. Tousi, O. Momeni, and E. Afshari, "A novel CMOS high-power terahertz VCO based on coupled oscillators: Theory and implementation," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 12, pp. 3032–3042, Dec 2012.
- [69] H. Wang, D. Kuzmenko, B. Yu, Y. Ye, Q. J. Gu, H. Rashtian, and X. Liu, "A compact 213-GHz CMOS fundamental oscillator with 0.56 mW output power and 3.9% efficiency using a capacitive transformer," in *2017 IEEE MTT-S International Microwave Symposium (IMS)*, June 2017, pp. 1711–1714.
- [70] R. Campbell, M. Andrews, L. Samoska, and A. Fung, "Membrane tip probes for on-wafer measurements in the 220 to 325 GHz band," in *18th Int. SYmp. Space Terahertz Technology*, 01 2007.

- [71] “120-GHz highly integrated IQ transceiver with antennas in package in silicon germanium technology.” [Online]. Available: https://siliconradar.com/datasheets/Datasheet_TRX_120_001_V1.4.pdf
- [72] “LMX2582 high performance, wideband pllatinum™ RF synthesizer with integrated VCO.” [Online]. Available: https://www.ti.com/lit/ds/symlink/lmx2582.pdf?ts=1645339106986&ref_url=https%253A%252F%252Fwww.ti.com%252Fstore%252Fti%252Fen%252Fp%252Fproduct%252F%253Fp%253DLMX2582RHAT
- [73] “ADF4155 integer-N/fractional-N PLL synthesizer.” [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/adf4155.pdf>
- [74] “AD8351 low distortion differential RF/IF amplifier.” [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/ad8351.pdf>
- [75] “DC to 6 GHz envelope and TruPwr RMS detector.” [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/ADL5511.pdf>
- [76] “HMC3716LP4E HBT digital phase frequency detector 10 - 1300MHz.” [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/HMC3716.pdf>
- [77] C. Nelson, F. Walls, M. Sicarrdi, and A. De Marchi, “A new 5 and 10 MHz high isolation distribution amplifier,” in *Proceedings of IEEE 48th Annual Symposium on Frequency Control*, Jun. 1994, pp. 567–571.
- [78] “3120A high performance low phase noise test probe.” [Online]. Available: <https://www.microsemi.com/product-directory/phase-noise-and-allen-deviation-testers/4131-3120a>
- [79] W. Walls, “Cross-correlation phase noise measurements,” in *Proceedings of the 1992 IEEE Frequency Control Symposium*, May 1992, pp. 257–261.

Jingjun Chen
June 2022
Electrical and Computer Engineering

A Millimeter-wave Molecular Clock in Silicon

Abstract

In this work, a molecular clock based on the $10 \leftarrow 9$ transition of the carbonyl sulfide gas is implemented. Techniques to remove the linear baseline in the absorption profile and to reduce the transmitter phase noise to prevent PM-AM conversion are discussed, as well as sources that affect the frequency stability. A high-power and high-efficiency millimeter-wave oscillator has been designed in CMOS, to address the signal generation difficulty in millimeter-wave transmitters.