Electrical and Computer Engineering

Ph. D. Program Dissertation Proposal

Three-Dimensional Microelectronics Integration:
Design, Analysis and Characterization

Zeynep Dilli
Advisor: Dr. Neil Goldsman

September 20, 2005
Contents

1 Motivation 5

2 A Study of Performance Improvement by 3-D Integration 8
  2.1 Introduction ................................................................. 8
  2.2 System Design ............................................................ 8
    2.2.1 Ring Oscillator Operating Frequency .............................. 8
    2.2.2 Internal and External Ring Oscillators ............................ 10
  2.3 Measurement Results .................................................... 12
  2.4 3-D Chip Design ......................................................... 14
  2.5 Heat Generation and Dissipation ....................................... 15
    2.5.1 Modeling and Simulation .......................................... 15
    2.5.2 Measurement ....................................................... 17
  2.6 Further Research ....................................................... 18

3 An Example Smart Dust System 19
  3.1 Introduction .............................................................. 19
  3.2 Overview of the Proposed System ..................................... 19
  3.3 Process Information ..................................................... 20
  3.4 Photodiodes: Design Issues ............................................ 21
    3.4.1 Photocurrent Calculation ........................................ 21
    3.4.2 Photodiode Design ................................................. 24
  3.5 Circuit Simulations ..................................................... 28
  3.6 Layout ................................................................. 31
    3.6.1 Tier 1: The Local Oscillator ..................................... 31
    3.6.2 Tier 2: The Capacitor .............................................. 32
    3.6.3 Tier 3: Photodiodes, Input and Output ........................... 33
  3.7 Further Research ....................................................... 34

4 On-chip Inductor and Transformer Structures 35
  4.1 Introduction .............................................................. 35
  4.2 Basics of On-Chip Inductors and Transformers ....................... 35
    4.2.1 The Concept of Inductance ...................................... 35
4.2.2 Physical Design .................................................. 36
4.2.3 Modeling Overview .............................................. 37
4.2.4 Inductor and Transformer Design Parameters and Rules-of-Thumb .......... 39

4.3 Designs and Measurements ........................................ 42
  4.3.1 Data Analysis and De-embedding .............................. 42
  4.3.2 Preliminary Results: Inductors ................................. 44
  4.3.3 Preliminary Results: Transformers ............................. 48
  4.3.4 Preliminary Results: Photoelectric Effect ..................... 51

4.4 Exploiting the Self-Resonance ..................................... 52

4.5 Further Research .................................................... 52
1 Motivation

Microelectronics integration can be at any of several layers. Just like system design, the integration hierarchy is often modular, and follows the general lines of devices integrated into gates, gates integrated into modules, modules integrated into chips (“integrated circuits”), chips integrated into board-level circuits, and boards put together on a main board to create a system.

Considerations of speed, compactness and robustness have created the trend for tighter integration at every level. At the chip-to-board level, an intermediate step is the design of multi-chip modules (MCMs). A multi-chip module is composed of several chips packaged together [1, 2]. The individual dies in a MCM are mounted on and are connected through a single substrate, designed using different technologies: creating metal conduction patterns over ceramics or plastic laminates or deposited metals over a dielectric substrate.

While this method is an improvement over each semiconductor die being individually packaged and soldered on a PCB to form the system, it still requires electrical signals to go off-chip while moving between components. In addition, if the packaging method requires bonding (either wirebonding or flip-chip bonding), each signal output needs to go through a bonding pad on the chip.

The output of any circuit that has to drive the large load capacitance of an off-chip connection needs to go through a buffer, creating its own problems. The method of buffering that causes the least delay requires a certain number of buffers, increasing in size. The number of buffers and the exponential increase in size between stages is determined by the ratio of the load capacitance to the input capacitance of the first buffer stage [9]. These buffers can get quite large; hundreds or even thousands of microns are possible. This in turn leads to quite a bit of chip real estate being given over simply to the chip’s interaction with the next stage in the system. Moreover, the large MOSFETs, drawing large currents, are power drains and unwanted heating sources.

It is desirable to have a method of connecting different dies without having to go through bonding pads and off-chip traces while conserving chip space. Chip stacking, or three-dimensional (3-D) integration, emerged as a possible answer [15, 16].

3-D integration potentially plays a central role in the development of a new technology that is currently drawing considerable interest. Commonly called “smart dust systems”, these are conceived to be a network of many very small circuits deployed over a certain geographical area and in communication with the other members of the system. Each circuit in a smart dust system is ideally miniaturized, self-powered, and self-contained. These requirements imply that mixed-signal systems
on a single unit are desirable. However, single-chip integration of mixed signal systems have several well-documented problems; noise-coupling between digital and analog parts of the circuit is the most prevalent one [7, 8, 10]. It is possible that separate physical substrates will allow analog and digital circuits to be integrated together without the attendant performance degradations.

This also provides motivation for investigating devices and structures typically used in RF systems in the context of 3-D integration. Especially for the communication circuits that would be included in smart dust units, on-chip RF components such as inductors, antennas and transformers should be investigated. There has recently been considerable work done on the design and modeling of on-chip inductors [33, 34, 45]. There has also been some interest in quasi-three-dimensional inductor and transformer structures: These structures use more than one metal layer, however are still essentially planar since they are built using conventional CMOS integration. However, true 3-D integration would allow the inductor designer to investigate newer degrees of freedom.

The advantages of 3-D integration can be summarized as follows:

1. Net system size reduction;
2. Increased ratio of active silicon substrate area to of the chip footprint;
3. Delay reduction, making faster clock speeds and higher signal bandwidths possible, through
   • Shorter interconnects,
   • Less parasitic impedance and load impedance;
4. Potential noise reduction in the transmissions between different parts of a system;
5. Potential substrate noise reduction, allowing mixed-signal integration;
6. The possibility to integrate components which require substrates of different materials—for example an optical system with an electrical system;
7. Higher degree of geometric freedom, especially for the design of geometry-dependent structures like inductors and transformers;

On the other hand, the main disadvantages of this approach are

1. Potential increase in heat-dissipation problems;
2. Increased (geometric, computational, routing) design complexity.

Several 3-D stacking techniques have been proposed [17, 18, 19, 21, 22]. One approach is to route the signals on the planar dies to the sidewalls of the stack and route inter-die interconnections on
the sidewalls, or on a specifically designed cap chip at the very top. Another approach is to use through-
viats, whether through the handle- and over-oxide of SOI stack layers or through the silicon of the dies
themselves. At the Laboratory of Physical Sciences, a new 3-D integration technique is currently under
development. This proposed dissertation work plans to use, help develop and characterize this process.

The remainder of this proposal is structured as follows: For each section, highlights of previous
work done is presented, followed by proposed work and methodology. We start by describing our work
on characterizing different methods of intra-chip connections. The next section involves the design of a
complete, self-contained system designed to be integrated three-dimensionally. The final section covers
work about on-chip inductors, transformers, and self-resonant LC structures.
2 A Study of Performance Improvement by 3-D Integration

2.1 Introduction

We have designed a system to demonstrate the extra load and subsequent speed loss caused by having to go through bonding pads. We have measured the operation speeds of circuits integrated through off-chip bonds vs. on-chip bonds. We have done the preliminary design of a chip that will be integrated with 3-D vias on which to repeat the same measurement. We are planning to characterize and model the extra noise and error-susceptibility caused by off-chip vs. on-chip and 3-D interchip connections. We also intend to look into modeling and investigating the heating problem in three-dimensionally integrated systems.

2.2 System Design

2.2.1 Ring Oscillator Operating Frequency

The central structures in our test system are ring oscillators—a simple ring of an odd number of inverters feeding back into itself.

A ring oscillator is essentially a positive-feedback system. Once the inverters are powered, random noise in the input of one stage may cause the output this stage to flip, which leads the output of the next stage to flip, and so on; if the number of inverters in the chain is odd, all inverters keep changing state as long as the power is kept up. The oscillation frequency of the system, with the output taken from between any two stages, depends on the propagation delays:

\[
\frac{1}{N(t_{PLH} + t_{PHL})},
\]

where \(N\) is the number of inverters \(t_{PLH}\) and \(t_{PHL}\) are the high-to-low and low-to-high propagation delays respectively, measured between the 50% point of the input and the 50% point of the output. A five-stage ring oscillator is shown in Figure 2.2.1.

To approximate these delays we use the digital model for an inverter [9] shown in Figure 2. The calculation treats the inverter switching process as the charging of a capacitor through the large signal equivalent switching resistance. For an inverter loaded with (driving) another inverter, when the output is going low-to-high (PLH), a load capacitor of \(C_{inn} + C_{inp} + C_{outn} + C_{outp}\) (the net capacitance from the output to the AC ground) is being charged up to \(VDD\) through \(R_p\). The same capacitor is discharged through \(R_n\) for the high-to-low transition (PHL).
In this model, the input and output capacitances are related to the transistors' oxide capacitances: $C_{in} = 1.5C_{ox}$ and $C_{out} = C_{ox}$. The switching resistances are the average resistances of the relevant MOSFETs between the off and the on stages of the full swing: $R_{n,p} = V_{DD}/I_{Dn,p}$.

Therefore the low-to-high and high-to-low propagation delays are given by

$$t_{PLH} = 0.7R_p(2.5C_{oxn} + 2.5C_{oxp}), \quad t_{PHL} = 0.7R_n(2.5C_{oxn} + 2.5C_{oxp}),$$

(2)

and used in Eqn. 1 to find the operating frequency.

If there is an extra capacitive load between two of the stages, as shown in Figure 3, this will change the propagation delay of the preceding stage:

$$t_{PLH,\text{load}} = 0.7R_p(2.5C_{oxn} + 2.5C_{oxp} + C_{load}).$$

(3)
Similarly for $t_{PHL,load}$. In this case, the total propagation delay for the N-stage ring oscillator becomes 

\[(N - 1)(t_{PHL} + t_{PLH}) + (t_{PHL,load} + t_{PLH,load}).\]

\[\text{Figure 3: A five-stage ring oscillator with a capacitive load between two of the stages.}\]

\[\text{2.2.2 Internal and External Ring Oscillators}\]

In our preliminary work, we have designed integrated circuits with two versions of ring oscillators on a single die. Two versions of the chips were made: one using the AMI 1.6 \mu m (ABN) technology and the other the AMI 0.6 \mu m (C5N) technology. On each chip, among some other structures, were internal and external ring oscillators.

The internal ring oscillator on each chip was designed to have no extra load between its stages. Utilizing the basic construction for, e.g., clock signal generation on many digital circuits, we had a 19-stage internal ring oscillator on the ABN chip and a 31-stage oscillator on the C5N chip (shown on Figure 4).

The output of each internal oscillator circuit went through an extra inverter, serving as a buffer, to a counter. In the C5N chip, this is a 6-bit counter comprised of D-flip-flops (which divides the output frequency by 64). The counter was designed to serve as an on-chip testing structure, so that the internal oscillator operation would not be too fast for our test equipment. The layout of the oscillator and counter is shown on Figure 5.

It can be noted that the buffer inverter acts as an extra capacitive load inserted between two stages, with $C_{load} = C_{inn} + C_{imp} = 1.5C_{oxn} + 1.5C_{exp}$. 
Figure 4: The 31-stage ring oscillator comprised of minimum-size transistors in the 0.6 \( \mu \text{m} \) technology chip. The output is taken from between the stages on the lower and upper right hand corners and goes through an additional inverter, which serves as a buffer (not shown).

Figure 5: The layout for the 31-stage ring oscillator and 6-bit counter/divide-by-64 frequency divider.

The external ring oscillators were laid out as individual inverters with their input and outputs connected to separate bonding pads. We used standard bonding pads with ESD (electro-static discharge) protection, provided by MOSIS for these technologies [14]. The C5N chip had eleven inverters laid out for this purpose (three are shown in Figure 6). When these inverters are linked externally to form the ring oscillator, between each stage there is the extra load of two bonding pads, bonding wires to package, and the out-of-the-chip connection elements. This was intended to simulate situations where two chips are integrated in a system and a signal has to go off chip to get from one part of the system to the other.
2.3 Measurement Results

Comparing the performances of the 11-stage external ring oscillator with that of the 31-stage internal ring oscillator, we can conclude that the pads, pins and breadboard connections caused nearly 800 times reduction in the operation frequency.

We first formed the 11-stage external ring oscillator on a breadboard and took measurements using a digital oscilloscope. Figure 8 shows the output waveform, with a frequency of 398.3 kHz.

The output of the divide-by-64 counter was then measured using the same scope. The output waveform, at 1.76 MHz, is shown in Figure 9.
This implies that the internal oscillator is running at $1.76 \times 64 = 112$ MHz.

For a comparison between the two structures, we look at what would happen if both had only 3 stages each. From Eqn. 1, ignoring the errors introduced by the buffer stage for the internal oscillator and the probe load for the external oscillator, we can find the equivalent frequencies. We can also calculate the single-stage delays for either structure. The results are given in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Internal Oscillator</th>
<th>External Oscillator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Freq.</td>
<td>112 MHz (31-stage)</td>
<td>398 KHz (11-stage)</td>
</tr>
<tr>
<td>3-stage equivalence</td>
<td>1.16 GHz</td>
<td>1.46 MHz</td>
</tr>
<tr>
<td>One-stage Delay</td>
<td>287 ps</td>
<td>228 ns</td>
</tr>
</tbody>
</table>

Table 1: The performance comparison of internal and external ring oscillators.

Thus the speed improvement when the signal stays on-chip is $228 \text{ns} / 287 \text{ps} = 794$-fold.

This effect is the direct result of the extra capacitive load on the oscillator inverters caused by the bonding pads (which the layout program Cadence extracts to be around 0.24 pF each, not including the ESD protection diode capacitances), bonding wires and the breadboard. To estimate this capacitance, we used the large-size inverter on the 1.6 μm ABN chip, inserting varying additional load capacitances between its output and the ground and measuring the rise, fall and delay times.
Extrapolating from the data points thus obtained, we can deduce an extra load capacitance of about 15 pF on the inverters of the external ring oscillator.

2.4 3-D Chip Design

We have developed a variation of the ring oscillator chip to be integrated three-dimensionally. The aim of this design is to provide comparison data for this alternative method of inter-chip communication: Vertical vias between dies as opposed to off-chip-on-chip connections. The layout of the chip, which was fabricated by the MOSIS fabrication facility, is shown in Figure 10.

![Figure 10: Chip designed for 3-D integration.](image)

The heart of this design is another 31-stage ring oscillator, broken into six groups of five or six inverters. Each of these groups is connected to an “input pad” and an “output pad”, squares of metal that are placed to be the connection points for vertical vias between the chips. The chip is designed with post processing in mind, during which one copy would be turned upside down (substrate-side up) and aligned with the bottom copy. Vertical vias would then go through the silicon substrates of both layers, connecting the metal pads. In this way, the signal would go up and down between dies six times. Figure 11 shows a closeup of this structure.

This design requires signal communication between two chips through vertical vias. We would be able to determine the extra delay this connection would cause by measuring the oscillation frequency of the ring oscillator. The vertical via landing pads have a to-substrate capacitance of the order of the input capacitance of a minimum-size inverter in this technology. Thus it is expected that this form of inter-chip communication will introduce much smaller delays than going through bonding pads.
2.5 Heat Generation and Dissipation

2.5.1 Modeling and Simulation

Our research group has previously worked on the modeling of heat generation and transmission on semiconductor chip systems [3, 4, 5]. This section outlines the methodology developed for our in-house simulation program.

The algorithm is based on the creation of a thermal network between the devices on a chip. By modeling heat sources, storage and transmission as thermal current sources, thermal capacitors and resistors, a thermal network is created and solved by using a set of KCL equations. This network is represented in Figure 12 (reproduced from [3]).

The values of the thermal capacitor and resistors are obtained from layout geometry and process parameters. The heat current sources are calculated as part of the self-consistent solution of the device-level equations and the full-chip heating network.

Our solver includes quantum-effects at the device level, represented by the Schroedinger Equation. Other equations are the drift-diffusion equations for electrical modeling (the Poisson Equation and electron and hole continuity equations respectively), and the lattice heat flow and population equations, which last is used for calculating the electron density. These equations are given below.

\[ E_i \Psi_i(y) = -\frac{\hbar^2}{2m^*} \frac{d^2 \Psi_i(y)}{dy^2} - q\phi(x,y)\Psi_i(y), \quad (4) \]

\[ \nabla^2 \phi = \frac{-q}{\epsilon}(p - n + D), \quad (5) \]

\[ \frac{\partial n}{\partial t} = \nabla \cdot (-n \ast \mu_n \nabla \phi + \mu_n V_{th} \nabla n) + GR_n, \quad (6) \]

\[ \frac{\partial p}{\partial t} = \nabla \cdot (p \ast \mu_p \nabla \phi + \mu_p V_{th} \nabla p) + GR_p, \quad (7) \]
\[
C \frac{\partial T}{\partial t} = \nabla \cdot (\kappa \nabla T) + H,
\]

\[
n = \frac{m^* k * T}{\pi \hbar^2} \sum_i [\lvert \Psi_i \rvert^2 \ln(1 + \exp(\frac{(E_F - E_i)}{kT}))].
\]

In these equations, \(E_i\) and \(E_F\) are the sub-band and Fermi energy; \(\Psi\) is the wave function; \(\phi, D, GR_n\) and \(GR_p\), \(n\) and \(p\), the electrostatic potential, doping concentration, electron and hole net generation-recombination and concentrations; \(C\) and \(\kappa\), the heat capacity and the thermal diffusion constant; \(H\) and \(T\), the generated heat and the lattice temperature, respectively. The heat generation mechanism is Joule heating: \(H = -J \nabla \phi\), where \(J\) is the net current density.

The need to solve these equations self-consistently stems from the temperature dependence of the parameters \(V_{th}\), the intrinsic carrier concentration and the built-in potential, the carrier mobilities, saturation velocity and bandgap, and the thermal diffusion constant itself.

The algorithm can be outlined as follows. Once the thermal resistance and capacitances are determined for each device, an initial temperature is set. Equations 4-9 are solved for a representative device and the heat generation is calculated based on their result. Extending this result across the chip, the heat generation of each node in the thermal KCL network, representing a single device, is estimated depending on the probability that a device in a given region of the chip will be on or off. The KCL network solution sets new temperature levels across the chip, which in turn modify the temperature.

Figure 12: A chip-wide thermal network model.
dependent parameters within Equations 4-9, which are solved anew. The simulation continues until the resulting temperature profile and device characteristics converge.

Our research group has applied this model to planar integrated circuits. Recently, we have started expanding the methodology for use in three-dimensionally integrated chips. In Section 2.6, we outline our plans about merging this line of research with our work 3-D system characterization.

2.5.2 Measurement

The heat-dependent semiconductor device parameters described in Section 2.5.1 implies that the basic operation characteristics of semiconductor devices such as pn-junction diodes and MOSFETs are altered with device temperature. Indeed, this effect has been included in the device models for circuit simulators like SPICE and SPECTRE.

For a first-order look at this effect, consider the ideal diode current equation:

\[ J(V_A) = J_o(\exp\left(\frac{V_A}{V_{th}}\right) - 1) , \quad (10) \]

where \( J \) is the diode current density, \( J_o \) is the diode saturation current, and \( V_A \) and \( V_{th} \) are the applied and thermal voltages respectively. \( V_{th} = kT \) has a direct, linear temperature dependence. The saturation current depends on the intrinsic carrier concentration and thus on the bandgap, and on the electron and hole diffusion constants and thus mobility: \( J_o = qn_i^2(D_p/N_D L_p + D_n/N_A L_n) \) for a long-base diode [6]. These parameters all have different temperature dependencies. Therefore the current for a given applied voltage \( V_{A0} \) will have a complex nonlinear dependence on temperature:

\[ J(T) = qn_i^2(T)\left(\frac{D_p(T)}{N_D L_p(T)} + \frac{D_n(T)}{N_A L_n(T)}\right)(\exp(\frac{V_{A0}}{kT}) - 1). \quad (11) \]

This in turn implies that after some calibration, the I-V characteristics of given devices on a chip can be used as device temperature indicators. Possible calibration methods include measurements taken at a set temperature of some heat source or sink much larger than the chip or IR imaging of the chip surface while in operation.

Some chip designs were considered for this purpose. These chips have two main features: A heating mechanism, either in the form of poly resistors or switching transistor arrays, and an array of pn-junction diodes to obtain a temperature map of the silicon surface, the resolution being determined by the diode array size.

Since MOSFET parameters also depend on temperature, the operation characteristics of transistor circuits can also be used as temperature indicators. An example is the operation frequency of a
ring oscillator; this measurement is interesting since the oscillating circuit is a contributing heat source as well as the meter. With measurements on such a chip, we will be able to experimentally verify our device-to-chip-scale heat generation and distribution modeling.

2.6 Further Research

We are currently revising the design of the 3-D-integrated ring oscillator chip. Our plan is to measure the oscillation frequency to deduce the load characteristics of the three-dimensional vias.

In addition to the speed improvement, we intend to investigate how noise characteristics are altered when two subsystems on silicon are integrated three-dimensionally. Substrate noise coupling has been identified as a obstacle in especially mixed-signal integration on a single substrate [7, 8, 10, 11, 12]. Following established research methods on substrate noise characterization, measurement and modeling, we will design digital and analog circuits intended to be integrated three dimensionally and investigate the noise and performance degradation in circuit operation after stacking. In particular, we propose to use network-analyzer measurements on copies of different pairs of structures integrated on a planar chip and three-dimensionally. Some of the structures we are considering vary in complexity from a single switching inverter in the vicinity of an LNA to a phase-locked-loop designed with both analog and digital building blocks. Starting with the works cited above, we will do further literature searches to finalize a modeling method for the theoretical analysis of this phenomenon.

Finally, to investigate the heat generation and dissipation properties of 3-D integrated systems, we will start by the layout and fabrication of planar chips, specifically designed with our existing simulation program in mind and with integrated on-chip temperature sensors as described in Section 2.5.2. This will allow us to obtain the experimental data to compare with our simulations.

In the meantime we will finalize the expansion of the planar self-consistent mixed-mode full-chip heat equation solver to simulate 3-D-integrated systems. There has already been some unpublished work in progress towards this end. In particular, we need to adapt the thermal network generation process to fit with the integration method developed at the Laboratory of Physical Sciences. Then we will design a 3-D version of our planar heat generation/measurement chip, which can be characterized both by measurement and simulation. Such a measurement system, using the structures on chip as temperature sensors, would have a significant advantage in characterizing the heat generation and distribution properties in 3-D integrated systems, as traditional methods such as IR imaging are not likely to be applicable for stacked chips. We expect this work to lead to the development of design practices to alleviate the heating problem in chip stacks.
3 An Example Smart Dust System

3.1 Introduction

Part of the work on developing a self-contained sensor/computing network ("smart dust systems") involves gaining a better understanding of the challenges inherent in creating self-powered sensor units. There are several approaches which can be taken to solve this problem. Two of the main possibilities depend on harvesting ambient power from the environment, either from RF electromagnetic waves or light. The design to be described here makes use of the latter in a 3-D framework.

The Lincoln Laboratories at MIT (MIT-LL) have developed a fully-depleted silicon-on-insulator process. Recently, they have taken steps to adopt this process to 3-D chip stacking by means of dense inter-tier vias. In late 2004, MIT-LL put out a call for proposals for an open run in this process. Submissions for circuits to be implemented in a three-tier architecture within this process were solicited. We submitted a proposal for the design and implementation of a self-powered local oscillator. This project was accepted and assigned a square layout area, 250 micrometers on one side. The design has been submitted, and MIT-LL has informed us that mask production commenced in early June.

In the following sections, we will first provide an outline of the system, and highlight the challenges involved in the design, including details of the process features that are relevant. A detailed section on the design will be followed by some simulation results and discussion.

3.2 Overview of the Proposed System

Our system concept for this design uses the three tiers available in the process. Each tier is visualized as fulfilling a certain general function in a specific way:

1. The bottom-most tier, Tier 1 as labeled by the process facility, houses the functional electronics. In the case of this design, our functional block is a local oscillator, comprised of three inverters connected in a positive feedback loop followed by a two-stage output buffer.
2. The middle tier, Tier 2, is set aside for some storage function. In our system, we have placed a capacitor in this tier. In a computing or sensor system, one can visualize data storage elements placed on this tier.
3. The top tier, Tier 3, is considered for sensor placement. In our case, the energy harvesting necessary for the operation of the system is placed on this tier in the form of photodiode arrays. It is possible to imagine data sensors fabricated on the top level of a similar system.
Figure 13 provides a schematic visualization of this system concept. Figure 14 shows the particular circuit being built in this case.

![Figure 13: Three tiers in a conceptual 3-D system design: The sensor, storage and electronics levels.](image)

3.3 Process Information

The process MIT-LL uses in this run, labeled “3DL1”, is a 0.18 µm, fully depleted silicon-on-insulator (FDSOI) process. This is a three-metal, single-poly process [21]. Three pairs of dopants are provided: CBN and CBP are the p-type and n-type body threshold adjustment implants, both at $5 \times 10^{17}\, cm^{-3}$.

PSD and NSD are the degenerately doped p-type and n-type source/drain implants. Finally, CAPP and CAPN are p-type and n-type island implants provided to allow for low-temperature and low-voltage-coefficient capacitors. These implants are doped at $5 \times 10^{18}\, cm^{-3}$ and $1 \times 10^9\, cm^{-3}$ respectively. The undoped silicon is p-type and has a dopant density of about $10^{14}\, cm^{-3}$.

A single tier in the process consists of 50-nm thick silicon islands built on a 400-nm thick layer of buried oxide (BOX), lying over a supportive silicon substrate. The gate oxide is 4.2 nm thick.

The full 3-D chips are assembled after all three tiers are independently fabricated [22]. Tier 1, the bottom tier, retains its silicon substrate and is kept device-side up. Tier 2 is flipped over, aligned
and bonded to Tier 1. The silicon substrate is then removed from Tier 2. 3-D vias are etched through the Tier 2 oxides and the topmost oxide layer in Tier 1. Tungsten is deposited to create the tier-to-tier connections. After Tier 3 is also flipped over, aligned and bonded to Tier 2 and its silicon substrate removed, the 3-D via fabrication process is repeated. Finally, bond pads are etched down to metal 1 on Tier 3. Figure 15, adapted from [22], displays the final 3-D structure.

From the viewpoint of our particular application, it is important to note that since the top tier is inverted, the metal or polysilicon layers will not block outside light from passing through the bottom oxide—net 600 nm thick cap oxide plus BOX, in this case—to reach the semiconductor islands which will house the photodiodes.

3.4 Photodiodes: Design Issues

3.4.1 Photocurrent Calculation

When photons of sufficient energy fall on the depletion region of a pn-junction, they may be absorbed to create electron-hole pairs. The built-in electric field in the junction then separates these carriers and sweeps them away, the electrons drifting towards the n-side and the holes towards the p-side. Once these carriers hit the bulk regions, if they can diffuse without recombination to the device edges, they contribute to the current outside the device, creating a photocurrent.
There are several factors that determine how much photocurrent will be obtained from a photodiode with a given amount of incident optical power falling on the junction [23]. These factors are brought together in the definition of responsivity, $R$:

$$i_p = R \times P_{inc},$$  \hspace{1cm} (12)$$

where $i_p$ is the photocurrent in amperes, $P_{inc}$ is the incident optical power on the photosensitive region in watts, and $R$ is thus given in units of A/W.

To calculate the incident power, we need to know the incident intensity and the light-sensitive area: $P_{inc} = I_{ph} \times A$. As a rough guideline, the sunlight intensity on a bright day is about 1000
W/m²=1×10⁻⁹ W/µm² [24], and a GaInP laser operating at around 670 nm can put out a power of the order of 5 mW, which when focused on 1 µm² yields an intensity of 5×10⁻³ W/µm² [23].

The responsivity is given by

\[ R = \eta \frac{\lambda}{1.24}, \]  

(13)

where \( \eta \), quantum efficiency, is defined as the ratio of the number of electron-hole pairs that are created to the number of incident photons on the photosensitive area. With that definition in mind, we can show how to arrive at Eqn. 13 by relating the photocurrent \( i_p \) to the electron flux, \( \phi_e \) and through that, to the photon flux \( \phi_p \):

\[ i_p = q\phi_e = q\eta\phi_p, \]  

(14)

\[ \phi_p = \frac{P_{inc}}{h\nu}, \]  

(15)

\[ \Rightarrow i_p = \eta \frac{q}{h\nu} P_{inc}. \]  

(16)

\[ \frac{q}{h\nu} = \frac{q\lambda}{hc} = \frac{\lambda}{1.24}, \]  

(17)

\[ \Rightarrow i_p = \eta \frac{\lambda}{1.24} P_{inc}. \]  

(18)

Hence the definition in Eqn. 13.

The quantum efficiency is the factor that depends most on the structural design of the photosensitive device. It is given as a combination of three factors:

\[ \eta = (1 - R)\xi(1 - \exp(-\alpha d)) \]  

(19)

Let us examine these factors one by one.

- \( R \) is the optical power reflectance from the surface. In our device there will be two surfaces that will cause some of the incident power to be reflected: Air-silicon dioxide and silicon dioxide-silicon. For a simple calculation, we will consider normal incidence and ignore multiple reflections in the oxide layer. The reflectance from the surfaces will then be

\[ R_1 = \left(\frac{n_{air} - n_{SiO_2}}{n_{air} + n_{SiO_2}}\right)^2 = 0.04, \]  

(20)

\[ R_2 = \left(\frac{n_{SiO_2} - n_{Si}}{n_{SiO_2} + n_{Si}}\right)^2 = 0.16. \]  

(21)

Thus, 96% of the incident optical power will reach the semiconductor surface, and 84% of this will enter the semiconductor—that is, about 80% of the incident optical power.
• $\xi$ is the fraction of created photocarriers that reach the outer circuit without recombination. With a value between 0 and 1, it is mostly governed by the material quality and surface recombination. In silicon photodetectors, if the material has been treated by antireflection coatings, it is possible to get quantum efficiencies close to 1, due to the high quality of the silicon crystal. Let us assume that in our case $\xi$ is 0.9, and thus

$$ (1 - R)\xi \approx 0.75 . \quad (22) $$

• The expression $\left( 1 - \exp(-\alpha d) \right)$ indicates how much of the photon flux streaming through the photosensitive material will be absorbed. Here $d$ is the total thickness of the material that the photons pass through, and $\alpha$, the absorption coefficient, is a material property. For silicon, $\alpha$ is about $3.5 \times 10^{-4} \text{ nm}^{-1}$ at 633 nm (red light). It increases with the photon energy and is around $10^{-3} \text{ nm}^{-1}$ at UV frequencies.

In our design, this factor turns out to be the bottleneck. Considering top-illumination, the available silicon depth is only 50 nm. This yields $\left( 1 - \exp(-\alpha d) \right) = 0.017$ for red light.

Following this discussion, we can obtain a rule-of-thumb about how much photocurrent we can generate in our system per micron-square of photosensitive area, assuming red light (0.633 $\mu$m) and an intensity of 1000 W/m²:

$$ i_p = \eta \times \frac{\lambda}{1.24} \times 10^{-9} \text{ W/\mu m}^2 = \eta \times 0.51 \times 10^{-9} $$

$$ = (0.75 \times 0.017) \times 0.51 \times 10^{-9} = 0.013 \times 0.51 \times 10^{-9} , $$

$$ \Rightarrow \quad i_p = 6.63 \text{ pA/\mu m}^2 \quad (24) \quad (25) $$

This current can be increased by using a higher intensity light source like a laser and picking an optimal wavelength—simply increasing the wavelength to exploit the $\lambda/1.24$ component will not work due to decreasing absorption at lower photon energies.

### 3.4.2 Photodiode Design

The useful area for photocurrent generation in our case will be the top-view cross-section area of the diodes’ depletion regions. This area in turn depends on the layout and depletion region width. For our diodes, we can choose pairs of doping implants from the list of available implants given in Section 3.3.

The option that will yield the widest depletion region and thus the largest photosensitive area is using the n-type threshold adjust implant (CBP) and undoped silicon. Using the relation

$$ W_d = \left[ \frac{2\epsilon_S}{q} V_b \frac{N_A + N_D}{N_A N_D} \right]^\frac{1}{2} \quad (26) $$

24
where \( V_{bi} = V_{thermal} \cdot \ln \left( \frac{N_A N_D}{n_i^2} \right) \) is the built-in potential of the junction, this will yield a depletion region width of about \( 1.5 \, \mu m \). If we design the diode layout such that the junction is \( 10 \, \mu m \) wide (across), this yields an area of \( 15 \, \mu m^2 \) and, from Eqn. 25, about 99.5 pA per diode.

However, using the very lightly doped “substrate” material (p-type silicon, doping around \( 10^{14} \, cm^{-3} \)) is not recommended by the fabrication facility [25]. Among the possible problems is the concern that such “intrinsic” regions will be vulnerable to possible surface accumulation or inversion if a poly or metal layer over the region becomes charged. Since we need to cover our metallurgical junctions with polysilicon in order to provide silicide protection due to the specific fabrication process steps, this is a real concern for us. Thus we chose not to rely solely on this design type.

Figures 16 through 18 display this diode: Only the implant regions and contacts, implants plus poly, and the full layout shown, respectively. The junction is between the CBP and “intrinsic” regions, i.e. there are two junctions, one to either side of the central strip, in Figure 16. The heavy-n-doping NSD implant is for ohmic contact to the n-side, and PSD for the p-side. Polysilicon covers all the intrinsic region and the junctions and serves the dual roles of silicide protection and implant alignment during fabrication. 26 of these structures are combined to obtain 52 diodes of this type.

Figure 16: Implants for the pin-type diode.

Figure 17: Implants plus poly for the pin-type diode.
The next best option, in terms of wide-depletion-region diodes, is using the two threshold adjust implants, CBN and CBP, as the p-side and n-side respectively. This results in a depletion region 0.0684 µm wide. To easily stack many of these diodes into an array, we have chosen an annular design, where the pn-junction is a square 2 microns long on one side; thus the photosensitive area per diode is approximately 0.55 µm² and we can obtain 3.63 pA per diode.

Figures 19 through 21 display the layout of this diode: Similar to the case for the lateral diode, only the implant regions and contacts shown, implants plus poly shown, and the full layout shown, respectively. The junction is between the CBP and CBN regions, i.e. as a ring around the CBP square in the center, which is two microns per side. 2062 of these diodes are stacked in the final layout.

The area assigned to us for this fabrication run is, as mentioned, 250 µm by 250 µm. Within this area we also need some space for bonding pads to probe the circuit. Making use of all the space available, we have managed to lay out a system with 2062 annular diodes (of the CBN/CBP type) and 52 lateral diodes (of the “intrinsic”/CBP type).

With this total number of diodes and under the rather stringent illumination conditions assumed in Section 3.4.1, we expect to obtain about 12.6 nA of photocurrent. The question is whether this current will be sufficient to run a local oscillator of the design given in Figure 14.
Figure 19: Implants for the CBN/CBP diode.

Figure 20: Implants plus poly for the CBN/CBP diode.

Figure 21: The full layout for the CBN/CBP diode.
3.5 Circuit Simulations

Assuming a 10 nA photocurrent and using a 30 pF capacitor as the middle tier energy storage device, we simulated the circuit using the MOSFET models provided by the fabrication facility. Since our version of Spectre was not recent enough to be able to simulate with the original BSIMSOI level models, we used bulk FET models modified to reflect the SOI process by Dr. Wyatt [25].

The overall circuit operation is given in Figure 22. The three plots in the figure are, from top to bottom, the voltage across the capacitor, the signal supplied to a 15 fF load 1 by the output buffer and the signal between the oscillator stages. As can be seen from the figure, the photocurrent starts charging the capacitor, which raises the rail voltage of the inverters. At a certain point, the gain of the three-stage positive-feedback loop gets high enough for oscillation to start. As the circuit oscillates, the inverters are drawing current and discharging the capacitor, which eventually leads to the capacitor voltage reaching an equilibrium point. For this input current and transistors this level is 227 mV, at which rail voltage the transistors are operating in their linear regions [26].

Zooming in on the simulation results during full oscillation reveals that the circuit is operating at 1.29 MHz. This is shown in Figure 23. Figure 24 focuses on where oscillation is just beginning.

Finally, if we assume that we can obtain a higher photocurrent 2 for our simulation, results are given in Figure 25, where the photocurrent is 40 nA and the resulting output signal has both higher amplitude and frequency—about 4 MHz.

---

1 This is the approximate expected capacitance between the output pad and the silicon support substrate.
2 Or supply the circuit with an external voltage source. In the layout, we have made provision for this case. This would forward-bias the photodiodes, however if we keep the bias level below the diode threshold, the net current drawn is of the order of milliamperes.
Figure 22: System operation with $i_p = 10$ nA and storage capacitor 30 pF.

Figure 23: Zoom in on the system operation with $i_p = 10$ nA and storage capacitor 30 pF.
Figure 24: Zoom in to Figure 23 where oscillation is beginning.

Figure 25: System operation with $i_p = 40$ nA and storage capacitor 30 pF.
3.6 Layout

3.6.1 Tier 1: The Local Oscillator

Figure 26: Local oscillator (Tier 1—bottom tier) layout. The structures on the left and right are the 3-D vias coming from the tiers above, for GND and VDD rails respectively.

Figure 27: A zoom of the local oscillator layout, its output buffers and the 3-D via carrying the output signal up.
3.6.2 Tier 2: The Capacitor

Figure 28: The capacitor tier (Tier 2—middle tier) layout. The capacitor top plate is a poly square, 67.9 µm on one side. The bottom plate is produced using the CAPN n-type implant. The expected capacitance of this structure is 30 pF and the Cadence-extracted capacitance is 29 pF.
3.6.3 Tier 3: Photodiodes, Input and Output

Figure 29: The diode tier (Tier 3—top tier) layout. The bondpads have overglass cuts that will allow us to access the Metal 1 layer of the tier with probes or bondwires. Most of the layer is taken by an array of 2062 annular CBN/CBP diodes, and there are also 52 lateral “pin” diodes near the top of the layout. The full layout is 250 µm on one side.
3.7 Further Research

We plan to create a variant of this system using the 3-D integration technique being developed at LPS. Instead of using SOI devices, this technique integrates full silicon dies thinned to around 20 microns. This will be an advantage in photocurrent generation, as it allows for greater active photosensor depth, going around the bottleneck explained in Section 3.4.1.

The first step in demonstrating the viability of this technique is to create the whole system using planar technology in a traditional CMOS process. Having fabricated this system, we will build on our experience in designing for a 3-D chip geometry, based on the work described in Section 2, to create either two or three distinct chips to be stacked with the LPS process.

An alternate method of powering such systems is the use of rectifying antennas, which make use of power transfer through electromagnetic waves. This method has been investigated for microwave power transmission, recently and notably in the design of RFID tag-type circuits [27, 28, 29, 30]. We have done some preliminary investigation on the design of a rectifying antenna system on printed circuit board level. Using planar and dipole antennas and an RF source operating at 462 MHz (the FRS band) and Schottky rectifying diodes chosen for their speedy response and low voltage drop, we have obtained DC voltage levels up to two volts. We intend to link this work with our work on on-chip passive structures, to be described in Section 4, and study the viability of using antennas built on a semiconductor substrate for this purpose.

We expect to expand this work into the codification of a design methodology for self-powering systems. For a rectenna-based circuit, this will involve analyzing, independently and together, requirements for the system antenna, whether a transformer will be required and the specifications of a fitting transformer, the rectifier diode impedance and switching characteristics and load circuit characteristics. For a photodiode-based circuit, the elements to investigate are photodiode characteristics and power generation ability, requirements for a charge storage antenna and for a power regulation circuit, and load circuit characteristics.

Finally, the fabrication and post-fabrication processes being investigated by groups connected to this project include the use of different dielectric materials. It is possible to shape this research project as an application for a high-k dielectric for use in the charge storage capacitor for the photodiode-based system, or to study the use of a low-k dielectric in the layout of an on-chip antenna for an RF structure with a higher resonant frequency.
4 On-chip Inductor and Transformer Structures

4.1 Introduction

Radio frequency circuits are not exempt from the drive for integration and miniaturization. One specific challenge they have offered is their need for passive elements, like capacitors, inductors and transformers, that are space-consuming and require modeling in a semiconductor system [31].

Research on such passive structures in microelectronics has intensified since the latter half of 1990s [33, 34]. Different modeling approaches for these devices are briefly reviewed in Section 4.2.3. Partly as a result of this modeling work, and partly following many empirical studies, rules of thumb and trade-offs have been described about the design of such structures, outlined in Section 4.2.4. Starting out from these rules, we plan to look into the advantages and disadvantages of inductors and transformers designed using our 3-D integration technique.

It has also been proposed that such devices, with their natural LC resonance, can be used as passive-LC tanks in circuits such as VCOs [45]. During experiments in our previous work, we observed that the resonance point of these devices can not only be affected by the design, but also shifted during operation by using the properties of the semiconductor substrate to change the charge concentration underneath. Some of these preliminary results will be presented in Section 4.3.4. An electrically or optically tunable LC tank would both be an interesting and potentially useful device to investigate. We plan to combine our group’s previous research into the modeling of charge concentration in semiconductor devices with our attempts on this question.

4.2 Basics of On-Chip Inductors and Transformers

4.2.1 The Concept of Inductance

In circuit-theory terms, the inductance of an inductor links the potential “induced” across the device to the time rate of change of the current flowing across it:

\[ V(t) = L \frac{dI(t)}{dt} \]  \hspace{1cm} (27)

This is related to the physical concept of inductance between two current loops: A measure of the electromotive force induced in one loop is proportional to the rate of change of the total flux linkage set up over that loop by the current in the other loop [41, 46]. Thus \( L_{12} \) is the inductance between loops 1 and 2:

\[ L_{12} = \frac{\Phi_{12}}{I_2}, \]  \hspace{1cm} (28)
where $\Phi_{12}$ is the magnetic flux in loop 1 caused by the current in loop 2. It is found by integrating the flux density over $S$, the surface defined for flux linkage:

$$\Phi(\vec{r}) = \oint_S \vec{B}(\vec{r}) \cdot d\vec{s}.$$  

(29)

This definition gives the mutual inductance between two current loops. This can be generalized to calculate the self-inductance of a current or a current loop, which is based on the flux linkage on that loop arising from the magnetic field of the current itself [53]:

$$L_{\text{self}} = \frac{1}{I} \int_S \vec{B}(\vec{r}) \cdot d\vec{s}.$$  

(30)

Part of the flux induced by a current flowing through a conductor will be within the conductor. This part of the flux linkage, leading to internal self inductance, is hard to calculate with the approach of Eqn. 30 since defining $S$ is usually not intuitive. For this part of the problem, we can use the definition of magnetic energy stored in a system:

$$\frac{1}{2}LI^2 = \int_V \frac{\mu}{2} \mathcal{H}^2 dV \Rightarrow L = \frac{\mu}{I^2} \int_V \mathcal{H}^2 dV.$$  

(31)

Another approach, when considering a certain conductor shape, is to calculate the internal impedance of the conductor by treating the electric field on the surface as an applied field and finding the resulting internal current profile, then using the definition $Z(\omega) := \mathcal{E}/I = R(\omega) + jL(\omega)$.

The external self inductance of a loop of current is then calculated from the remaining part of the surface integral in Eqn. 30, where $S$ is now taken as surfaces external to the conductor.

Finally, although all our definitions so far have referred to current loops, it has been shown that a representative inductance between current segments can be calculated using the concept of partial inductance [41].

4.2.2 Physical Design

On-chip inductors and transformers are commonly built by using the metal layers available in the semiconductor process. Like the components used in microwave electronics and built on a dielectric substrate, the prevalent geometric design is a spiral. Sometimes octagonal or other polygonal spirals are used, however the most common style is a square spiral, as shown in Figure 30.

This design style allows several degrees of freedom for a planar inductor:

- Number of turns
- Total length of inductor traces
Figure 30: Top and side views of typical square spiral on-chip inductors.

- Length of the first or last segment
- Width of inductor traces (W)
- Spacing between turns (S) (Setting three of the parameters listed up to this point uniquely determines the other two.)
- Metal layer used (This sets both the oxide thickness to substrate and the thickness of the metal strip.)
- Substrate doping

An additional degree of geometrical freedom is obtained by using several metal layers for a stacked inductor geometry [32, 33, 37].

For an on-chip transformer design, it is possible to visualize two spirals on different metal layers laid out on top of each other; three spirals, two in series, laid out on top of each other; spirals placed around each other or interwound spirals [45, 50, 51]. All of the above degrees of freedom apply.

### 4.2.3 Modeling Overview

On-chip spiral inductors are nonideal inductors. They exhibit non-zero resistance and are capacitively coupled to the substrate they rest on and to other metal structures nearby. Part of the inductance they exhibit arises from inductive coupling between their turns and segments; however these segments are also capacitively coupled to each other. Finally, the semiconductor substrate is lossy, so eddy currents are induced, and some energy is lost due to substrate resistance; there is also substrate capacitance.
An on-chip inductor can be modeled for circuit simulations by a lumped-element model. Several such models have been proposed over the years, increasing in complexity. A very basic lumped-element model for an on-chip spiral inductor merely includes the serial inductance and resistance, the capacitance between the windings adding up to a capacitance parallel to this serial branch, and the oxide capacitance in series with the substrate resistance presented as a shunt element [33, 35]. A more complicated model includes substrate capacitance as well [36, 37]. An approach to representing the frequency-dependence of the series inductance and resistance is to put another reactive branch in parallel with the series resistance [39].

Another approach is treating the individual segments of a spiral inductor and the coupling between them. The basis of this approach is the lumped-element modeling of a single microstrip and coupled microstrips on a lossy substrate [40] as illustrated in Figure 31.

![Figure 31: Lumped-element models for a single microstrip and coupled microstrips on a lossy substrate.](image)

The substrate resistance and capacitance element of the coupling between two strips can be used in a lumped-element model for a spiral inductor [37, 43], in which the effects of these elements are collected together as another parallel branch to the main inductor/resistor series branch.
For these modeling methods, parameters are extracted by measuring the manufactured inductor’s two-port parameters and a parameter fitting procedure. Different techniques exist for modeling from the other direction, that is, starting from the device geometry and electrical parameters to arrive at the circuit model. By using the microstrip lumped element models in conjunction with partial element equivalent circuit methods [41] and using a complex image method to include the lossy substrate effects [40], the series inductance and resistance of a spiral inductor can be modeled [10, 42]. An alternative way of modeling the frequency-dependence of the conductor skin and proximity effects and of substrate losses is to use transformers or mutual inductance in the circuit model [44]. The metal-to-metal capacitive components of the models have been studied by assuming a certain voltage drop pattern across the segments, depending on the geometry and current flow direction in the inductor [37].

Finally, a method combining many of the features described above to create a segment-by-segment model of an on-chip inductor has been proposed in [45]. The elements of these models can be calculated using classic methods developed by R. Grover in the 1940s, or the work of H. Wheeler or H. Hasegawa, M. Furukawa and H. Yanai in the 1970s [46, 47, 48].

Modeling on-chip transformers is a similarly active research area [49, 50].

A rather different modeling approach uses full-wave simulations, through programs such as HFSS or numerical analysis techniques such as alternate-direction-implicit finite-difference-time-domain (ADI-FDTD) programming. While this method is the most time consuming, it can yield the most accurate results.

4.2.4 Inductor and Transformer Design Parameters and Rules-of-Thumb

In the literature about on-chip RF passive devices, the inductance of such a device is usually defined as the imaginary part of its impedance. The common method is to take two-port scattering (S-) parameter measurements of the fabricated device with a vector network analyzer, convert these to impedance and admittance parameters [52], and extract the inductance and the quality factor [35, 45]:

\[
L = \frac{\text{Im}\{1/Y_{11}\}}{\omega},
\]

\[
Q = \frac{\text{Im}\{1/Y_{11}\}}{\text{Re}\{1/Y_{11}\}}.
\]

Possible on-chip inductor geometries have been extensively investigated in the literature [32, 33, 37, 38, 45]. As the on-chip “inductor” is in fact a device with series and shunt reactive impedance and admittances, the inductance of such a typical structure defined by Eqn. 32 exhibits the kind of
behaviour depicted in Figure 32: From low to high frequency, we have inductive, self-resonance, and capacitive regions.

![Figure 32: The “inductance” of a typical on-chip spiral inductor vs. frequency, as defined by Eqn. 32.](image)

This response has several features important to circuit designers: Mainly, the low-frequency inductance and the self-resonant frequency (as a measure of the frequency range in which the inductor is usable as an inductor). Related to this are other design concerns such as the device area and the parasitic coupling between the inductor and nearby structures, as well as the inductor sensitivity to outside stimuli such as ambient RF signals. The quality factor is also significant in several aspects of RF circuit design, depending on the application: filter bandwidths, oscillator frequency stability and noise sensitivity, and loss in matching networks are all affected by $Q$.

We illustrate how the inductor characteristics change with the model parameters described in Section 4.2.3. Figure 33 shows how the inductance curve shifts with each changing parameter in a simple model with a series inductance ($L$), series resistance ($R$), oxide capacitance ($C$) and substrate resistance ($R_{\text{sub}}$). In a nutshell, the series inductance in the model determines the low-frequency inductance, the resonance peak and the self-resonant frequency (top left corner). The series resistance (of the metal) and the shunt resistance (of the substrate) both cause a softer resonance switch when increased (top right and bottom right corners). The shunt capacitance, when increased, reduces the self-resonant frequency and the resonance peak (bottom right corner). It is worth noting that the ratio of the self-resonant frequencies for the curve pairs is determined by the ratio $f_{sr2}/f_{sr1} = \sqrt{L_1C_1/L_2C_2}$. 

40
Combining insights obtained from the various modeling methods and basics of inductance concepts, it is possible to come up with rules of thumb for inductance design. Some samples for these are given here.

- Current segments in opposite directions exhibit negative mutual inductance, while those in the same direction exhibit positive mutual inductance. Therefore the designer should attempt to put parallel current segments as close to and anti-parallel segments as far from each other as possible. In a planar Manhattan geometry, that makes a square the most desirable design.
- Longer conductors yield higher inductance but also higher serial resistance, which lowers Q. The same effect makes wider tracks preferable, as the serial resistance decreases; however see below.
- Increased oxide capacitance and increased capacitance between metal segments both reduce self-resonant frequency and peak Q. Therefore planar inductors built on higher metal layers are preferable, as they yield higher Q. Further, there is a trade-off with respect to the track width mentioned in the previous rule.

Figure 33: The model changes reflected by the inductance curve changes.
• Small separation between metal tracks increases inductance due to higher mutual coupling; however, high frequency capacitive coupling between segments reduces Q.

There is a sizeable body of literature investigating different configurations for both inductors and transformers [32, 33, 34, 36, 37, 38, 45, 50, 51]. For our focus on 3-D integration, we have concentrated on different geometrical shapes for these devices. Soem preliminary results are presented in Sections 4.3.2 and 4.3.3.

4.3 Designs and Measurements

As part of our preliminary work, we have had several chips featuring different inductive structures fabricated. This section describes our measurement methodology, some of the structures and the measurement results.

4.3.1 Data Analysis and De-embedding

As described in Section 4.2.4, we take S-parameter measurements using a vector network analyzer. To be able to do so using direct RF probing, we lay our structures out placed within probe pad structures designed for GSG (ground-signal-ground) type probes that are available to us. A typical inductor structure within such a pad setup is shown in Figure 34.

![Figure 34: An on-chip inductor within a probe-pad structure designed for two-port S-parameter measurements.](image)

The probe size constraints mean that the parasitic impedance of these structures is considerable, and some de-embedding procedure must be followed. Therefore we also have “open” and “thru”
pad structures (Figure 35) to be used in de-embedding. We have also written data analysis software to relatively automatize the de-embedding and inductance/Q extraction processes.

The de-embedding process works by converting the scattering parameters of measured devices to impedance and admittance parameters as appropriate to extract the effect of the measuring pads [52, 54, 55]. A schematic illustration of the process is given in Figure 36.
4.3.2 Preliminary Results: Inductors

We have designed several chips and fabricated them through the MOSIS production facility. The probe pad designs have allowed us to take two-port measurements with a vector network analyzer. We will present a sample of our results here.

- First, the effects of de-embedding on the inductance plots is displayed in Fig. 37. As the shunt capacitance of the probe pads are subtracted, the self-resonance frequency shifts higher; it is also possible to observe the extra serial inductance of the pad connections and pads being removed.

![Figure 37: Results of the de-embedding process.](image)

- To illustrate one of the rules of thumb presented in Section 4.2.4, the measured inductance and Q-factors of three inductors designed with the same geometrical layout but on three different metal layers (M1, M2 and M3) are presented in Fig. 38. The trace geometries of all three are the same, yielding very close low-frequency inductances and serial resistances. They are built on the same type of substrate, causing similar substrate resistances as well, although the eddy currents induced necessarily differ somewhat. The most significant difference between the three inductors is the capacitance to the substrate, which is clearly reflected in the $f_{sr}$ shift to higher frequencies for the higher metal layers. As expected, the M3 inductor exhibits the highest quality factor.
- Figure 39 shows the layout of the planar inductor built on the third metal layer alongside the layout of a stacked inductor built using three metal layers. A side-view of this inductor and the
current flow direction are also given in the figure. The total metal length is the same for both; however the planar inductor takes up 58072 µm² of silicon space for six turns whereas the stacked inductor takes 22500 µm² for nine turns. In addition to the greater number of turns, the stacked inductor has the design advantage of clumping more segments with parallel current directions together; however, this causes increased segment-to-segment capacitance. The net result, higher inductance with lower $f_{sr}$, is presented in Figure 40.
Figure 40: Inductance and quality factor plots planar and stacked inductors.

- Another possible way of exploiting multiple metal layers is attempting to emulate the macro wound-around-a-core coil inductor geometry using metal lines in a semiconductor process. A schematic of the basic geometry obtained thus is the first example shown in Figure 41. Variants of this, attempting to increase single “turn” areas and bring positive mutual-inductance segments closer together while keeping negative segments farther apart, are the second and third examples in this Figure, coil2 and coil3. Compared to the latter two, coil1 exhibits a very low inductance, and therefore Figure 42 shows the layout for variants coil2 and coil3 only. Figure 43 displays the inductance and Q-factor for these two inductors, layout area 47800 $\mu\text{m}^2$, as compared to the planar inductor built on metal 3 with an area of 58072 $\mu\text{m}^2$. While the serial inductances of these coils are about one-third of that of the planar inductor, and the metal trace lengths being the same yields similar serial resistances and therefore a lower Q, it is interesting to point out that this low L translates to a higher self-resonant frequency. Coil2 in particular seems to exhibit a comparable parasitic capacitance.
Figure 41: Schematic representations of the coil-type inductors: “Coil1,” “coil2” and “coil3”.

Figure 42: Layouts of the coil inductor variants coil2 and coil3.

Figure 43: Inductance and quality factor plots for planar and coiled inductors.
4.3.3 Preliminary Results: Transformers

Impedance conversions, feedback path couplers, and voltage/current gain are a few possible functions for on-chip transformers in RF circuits, whose possible geometries have also been investigated in literature [38, 50, 51]. We have designed and fabricated some transformers with representative geometries for our preliminary research.

Figure 44 displays the schematic representations for four different transformer designs: Two-metal stacked (“M2:M3”), three-metal stacked (“M2:M3M1”), interwound and spiral-within-spiral transformers, top left, right, bottom left and right respectively. Figure 45 shows the layout for the M2:M3 and the spiral-within-spiral transformers.

![Schematic representations of transformers](image)

Figure 44: Different transformers: M2:M3, M2:M3M1, interwound, spiral-within-spiral.

We examine the reflection coefficients (S11), transmission coefficients (S12), “turn ratios” calculated by assuming a load of 50 Ω on the secondary side and using the expression $\sqrt{Z_{11}/50}$, and voltage gain calculated by starting from the impedance matrix for a transformer, replacing $V_2$ and $I_2$ by \((N_2/N_1)V_1\) and \((N_1/N_2)I_1\) respectively, and solving the resulting system of equations for $N := N_1/N_2$.
Figures 46 and 47 present these results for the M2:M3 and M2:M1M3 transformers. Figures 48 and 49 present the results for the interwound and spiral-within-spiral transformers. The most efficient device for coupling is the M2:M3 transformer, followed by the two planar transformers and the M2:M1M3 transformer, which latter naturally exhibits the highest reflection. The coupling efficiency of the two planar transformers exhibit more frequency dependency than the stacked transformer geometries. The “turn ratio” is observed to be highest for this last device, however, as could be expected from comparing the designs. Therefore, if a high voltage gain is preferred over efficient coupling, this would be the design of choice.

Figure 46: Reflection and transmission coefficients, for the M2:M3 (blue) and M2:M1M3 (red) transformers.
Figure 47: “Turn ratio” (left) and voltage gain (right) for the M2:M3 and M2:M1M3 transformers.

Figure 48: Reflection and transmission coefficients for the interwound (green) and spiral-within-spiral (black) transformers.

Figure 49: “Turn ratio” (left) and voltage gain (right) for the interwound and spiral-within-spiral transformers.
4.3.4 Preliminary Results: Photoelectric Effect

One interesting and promising observation during our experiments was how the inductive structures on a semiconductor substrate reacted to ambient light. When we turn on the microscope light while taking a VNA measurement, the devices all undergo a characteristic property shift, illustrated in Figure 50.

Figure 50: The measured inductances of an metal 3 on-chip spiral inductor on p-substrate and on an n-well. Solid lines: External light off. Dashed lines: Light on. This is data prior to de-embedding.

We can interpret the effects of light using the rules of thumb outlined in Section 4.2.4: Zooming on the low-frequency response shows a very slight (\(\sim 0.1\) nH) increase in serial inductance. This shift is not all that is responsible for the \(f_{sr}\) drop. Rather, the shunt capacitance of the devices are increased enough to shift the self-resonant frequency down by 150 MHz for the device on p-substrate and 200 MHz for the device on n-well (see Figure 51). This increase in capacitance should have decreased the peak impedance. However, the sharper drop to the resonance from the peak indicates an effective decrease in the series resistance, which also increases the peak impedance (see Figure 52).

After de-embedding is done to negate the effects of the shunt impedance of the probe pads, the \(f_{sr}\) shifts are about 570 MHz for the p-substrate device and 1.14 GHz for the n-well device.
Figure 51: Zoom near the \( f_{sr} \) point of Figure 50.

Figure 52: Zoom near the peak impedance point of Figure 50.

4.4 Exploiting the Self-Resonance

In many RF circuits, especially communication circuits like LNAs and signal-processing circuits like mixers, a self-resonant LC tank is used as a tuned load for ensuring a certain frequency response \([31, 45]\). Just by themselves, LC filters have been long in use and recently adapted to IC fabrication \([56]\).

There has been some investigation on purposefully integrating an inductor-structure with a capacitor-structure in series or shunt configurations. But it should be noted that the on-chip planar inductor is by itself a self-resonant structure. There has been made studies of coupling between nearby inductors on the same substrate \([10, 39]\), and it has been suggested that this coupling, if intentionally designed, might be utilized to tune the \( L \) and \( Q \) characteristics of on-chip inductors \([45]\).

The results presented in Section 4.3.4 point out another possibility, that of changing the characteristics of the semiconductor substrate as a modulation method. It is possible to conceive uses for such a tunable structure in the design of ideally tunable circuits such as filters.

4.5 Further Research

As good-quality reactive components are important as integrated elements in both general high-frequency analog circuit design and in applications significant to 3-D integration, such as mixed-signal circuit stacks and power harvesting structures, there are a number of research tracks we plan to follow.

- **3-D inductors in chip stacks** In Section 4.3.2 we have overviewed several “three-dimensional” inductor designs, which are in fact simply inductors that use the different metal layers available in planar technologies. Our 3-D stacking process gives us the new degree of geometrical freedom
to pursue truly three-dimensional structures. Moreover, the process capabilities at LPS include the deposition of a low-k dielectric as a post-processing step [20]. This is promising for inductors and transformers with low parasitic capacitance and therefore higher resonant frequencies and quality factors. We intend to work on modeling such structures with a commercial program such as HFSS and create mask designs for fabrication at LPS.

- **Different stacked inductor and transformer geometries** Our plan is to use one of the lumped-element models available in literature to model the different inductor and transformer geometries that we have data for. We hope to use the insight so gained in the design of structures as described in the previous item. We are also currently exploring the inductor structures that comprise our transformer designs in order to gain more insights into the transformers’ behaviours.

- **Tunable Self-Resonant Structures** We can characterize the specific changes in substrate conductivity with incident light by taking more precise measurements, possibly using a laser beam with a known intensity. Using this and the lumped-element circuit models as described above as a starting point, a methodology for controlled tuning of the self-resonant structures can emerge. We also intend to investigate methods of electrically controlling substrate conductivity. The starting point of this investigation is using our semiconductor modeling code, as described in Section 2.5.1, to simulate changing carrier concentrations in different regions on semiconductor devices according to applied bias. An application for such a tunable circuit, for instance a mixer, should be developed simultaneously.
References


[25] Dr. Craig Keast and Dr. Peter Wyatt, personal communication.


