Progress and Challenges in the Direct Monolithic Integration of III-V Devices and Si CMOS on Silicon Substrates


1Raytheon Integrated Defense Systems, Andover, Massachusetts, USA; 2IQE Inc., Bethlehem, Pennsylvania, USA; 3Teledyne Scientific Company, Thousand Oaks, California, USA; 4Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; 5Raytheon Systems Limited, Glenrothes, Fife, United Kingdom; 6SOITEC, Bernin, France; 7CEA-LETI, MINATEC, Grenoble, France

Abstract

We present results on the direct monolithic integration of III-V devices and Si CMOS on a silicon substrate. Through optimization of device fabrication and material growth processes III-V devices with electrical performance comparable to devices grown on native III-V substrates were grown directly in windows adjacent to CMOS transistors on silicon template wafers or SOLES (Silicon on Lattices Engineered Substrates). While the results presented here are for InP HBTs, our direct heterogeneously integration approach is equally applicable to other III-V electronic (FETs, HEMTs) and opto-electronic (photodiodes, VSCLS) devices and opens the door to a new class of highly integrated, high performance, mixed signal circuits.

Index Terms — CMOS integrated circuits, Heterojunction bipolar transistors, Indium Phosphide, Monolithic integrated circuits, Silicon

I. INTRODUCTION

The future of integrated circuits will include the integration of high performance III-V electronic and/or opto-electronic devices with standard Si CMOS. While traditional hybrid approaches, such as wire bonded or flip chip multi-chip assemblies (Fig 1, left), may provide short term solutions, the variability and losses of the interconnects and the limitation in the placement of III-V devices relative to CMOS transistors will limit the performance and utility of these approaches. Recently, investigators have successfully demonstrated heterogeneous integration of InP HBTs and Silicon CMOS using variations on wafer bonding techniques [1,2] where the III-V epitaxial layers or completed devices are bonded to the surface of a completed Si CMOS wafer. A more attractive approach is the direct integration of CMOS and III-V devices on a common silicon substrate (Fig 1, right). In this way circuit performance can be optimized by the strategic placement of III-V devices adjacent to CMOS transistors and cells. While the direct growth and fabrication of III-V devices on silicon substrates has been pursued for over 30 years [3], recent advances in strain and lattice engineered materials and epitaxial growth techniques have enabled the direct growth of high quality III-V device layers on silicon substrates.

In this work we present the challenges and recent progress on the direct heterogeneous integration of InP HBTs and Si CMOS on a silicon substrate. As a demonstration vehicle we designed and fabricated a high speed, low power dissipation differential amplifier which serves as the basic building block for high performance mixed signal circuits such as ADCs and DACs.

Fig. 1. Traditional hybrid assembly (left) and direct monolithic integration of III-V and CMOS on SOLES substrate (right).

II. RESULTS AND DISCUSSION

Our direct integration approach is based on a unique “engineered” silicon substrate which is similar to a standard SOI wafer. The SOLES (Silicon-on-Lattice Engineered Substrate), invented at MIT [4,5] and manufactured by SOITEC using their Smart-Cut™ Process [6, 7], contains a buried III-V template layer that enables the direct growth of high quality III-V epitaxial material in windows directly on the
silicon substrate (Figure 2). At present the buried III-V template layer is Ge, although the substrate fabrication process is compatible with GaAs or InP template layers as well. SOLES have been successfully scaled to 200mm diameter wafers and are compatible with and can be readily inserted into a standard silicon CMOS foundry.

**Fig. 2. Schematic cross section of SOLES wafers showing placement of III-V device in windows.**

While our main efforts have focused on the fabrication of InP HBTs on SOLES, effectively creating a high performance InP BiCMOS process (similar to the SiGe BiCMOS process), the approach is equally applicable to other III-V electronic (FETs, HEMTs) and opto-electronic (photodiodes, VSCLS) devices. In fact the process flow is similar to a SiGe BiCMOS process flow: 1) Si CMOS device fabrication; 2) HBT epitaxial growth and device fabrication; 3) multilayer interconnect fabrication. In our approach, after the completion of CMOS device fabrication, windows are lithography defined and etched into the SOLES wafer to reveal the III-V template layer. Since the III-V growth windows are defined as part of the CMOS fabrication process, the III-V epitaxial material can be grown selectively and arbitrarily across the substrate as required for the particular circuit or applications.

Figure 3 shows an example of a SEM image of a completed InP HBT in close proximity to a Si CMOS transistor prior to heterogeneous interconnect formation. To facilitate the interconnecting of the III-V devices and CMOS transistors, the thickness of the III-V epitaxial layers and depth of the windows are optimized such that the III-V devices and CMOS transistors are planar. Figure 4 shows an example of a daisy chain test structure interconnecting InP HBTs and Si CMOS. With this truly planar approach, interconnect lengths (III-V – CMOS separation) as small 2.5 um have been demonstrated.

One of the biggest challenges of this approach is the growth of high quality III-V epitaxial material in windows on the Ge template layer. (Note: all of the III-V epitaxial material reported in this work is grown by MBE.) For the InP HBT, we first grow a GaAs nucleation layer, whose growth conditions are optimized to minimize the formation of anti-phase domains (APDs). GaAs is chosen as it is nearly lattice matched to Ge. Then a metamorphic buffer layer is grown followed by the InP device layers. Optimization of the windows etch and epitaxial growth processes are key to achieving high quality device layers. Figure 5 (left) shows an example of III-V growth in windows for unoptimized windows etch and epitaxial growth processes. Note the surface roughness, poor edge definition and formation of nanowire material due to nucleation and growth of III-V material on impurities at the windows edge. High quality III-V epitaxy, well defined windows edges and repeatable (wafer to wafer), uniform growth across a 100mm diameter wafer in windows as small as 15um x 15um are readily obtained with an optimized process (Figures 5 right, 6).

A detailed report on the growth of high quality InP HBT epitaxial material in windows on SOLES has been previously published [8]. With optimized growth conditions low dislocation density (<10^7) material with good surface morphology (surface roughness < 1nm as measured by AFM) and well defined X-ray spectra are easily achieved.

**Fig. 3. SEM image of a completed InP HBT in close proximity to a Si CMOS transistor prior to heterogeneous interconnect formation.**

**Fig. 4. SEM image of a heterogeneous interconnect daisy chain test prior to final interconnect metallization. InP HBT – Si CMOS interconnect spacing is < 2.5um.**

**Fig. 5. SEM Image of InP HBT device epitaxy material grown in windows on SOLES for unoptimized process (left. Note nanowire growth) and optimized process (right. Note: well defined windows down to 15um x 15um windows dimensions)**
The electrical performance of InP HBTs fabricated on SOLES is comparable to HBTs grown directly on native InP substrates [9]. Figures 7 and 8 show the Gummel characteristics and small signal parameters of a 0.5 x 5 \( \mu \text{m}^2 \) emitter HBT grown in a 15 x 15 \( \mu \text{m}^2 \) window on a SOLES substrate. Gain (beta), \( f_t \), and \( f_{\text{max}} \) of 40, > 200GHz and > 200GHz, respectively are achieved.

Using the InP HBT described above and standard CMOS a differential amplifier test vehicle was designed and fabricated. Figure 9 shows an optical image of a completed differential amplifier circuit. In addition to the core differential amplifier, the circuit contains a bias circuit and all HBT output buffer. The role of the output buffer is to attenuate the output of the core differential amplifier to facilitate the characterization of the differential amplifier.

Because of our truly monolithically integrated, planar approach we were able to include multiple design variants within a reticle on a wafer, effectively creating a design optimization design of experiments (DOE) within the reticle. Each design variant is step and repeated across the 100mm SOLES wafer. The planar approach also facilitates automated on-wafer probing for circuit characterization and the collection of circuit performance and uniformity data for the different design variants.

The following test results are for one of these design variants which utilizes a 3-2x5\( \mu \text{m}^2 \) HBT in each diff amp branch (6-total) with 6-finger (2um gate length, 19.2um wide) PMOS devices for the amplifier loads. For all the measurements that are shown, the differential amplifier core was biased at a \( V_{\text{dc}} = 6 \text{V} \) and \( I_{\text{dc}} = 10 \text{mA} \) (\( P_{\text{diss}} = 60 \text{mW} \)).
Separate DC supply inputs are provided for the amplifier core and output buffer circuits to ensure an accurate measurement of the dissipated power of the core.

4-port S-parameter measurements were made to determine the low frequency amplifier gain and unity-gain bandwidth. Measurements were made from 1MHz-20 GHz using on-wafer differential GSGSG probes. A probe tip calibration was performed using a GGB Industries calibration substrate.

Measurements from 1-50MHz were used to extract the low frequency gain of the differential amplifier. The low frequency voltage gain of the differential amplifier core was determined by measuring the gain of the chain of the differential amplifier with output buffer and correcting for the attenuation of the amplifier such that $A_v,\text{diff amp} = S_{21,\text{chain}}-S_{21,\text{buffer}}$. The output buffer amplifier has a low frequency attenuation of $\sim 25$dB – a value that agreed well with simulations.

Figure 10 shows the low-frequency gain of the core differential amplifier. A peak low frequency gain of 454V/V was measured. At lower frequencies, the gain is observed to decrease slightly. We believe this is due to device self-heating (increased output conductance of HBT).

Fig. 10. Measured $S_{21}$ of amplifier chain and output buffer test circuits at low frequencies. Low frequency gain $A_{v,DA}$ is given by $A_{v,DA}=S_{21,\text{chain}}-S_{21,\text{O/B}}$. Core diff amp utilized 3-2x5um$^2$ HBT in each diff amp branch with 6-finger 2um PMOS devices. $I_{ss} = 10mA$, $V_{ss} = -6V$.

The high frequency gain measurements are extracted using a similar scalar approach for determining the core amplifier characteristics. Figure 11 shows the corrected high frequency characteristics for the same amplifier as shown in Figure 10. A deviation in the slope of the roll-off was observed at the higher end of the frequency band. The cause of this discrepancy has not been determined. To determine the unity gain cut-off frequency of the amplifier, this portion of the frequency response was not utilized. Instead, the unity gain frequency was extrapolated from the intercept of data taken from 1-15 GHz. A unity-gain frequency of $>25$ GHz was extracted from the measurement shown in Figure 11. From the DC-gain measurement, the DC-gain*unity gain bandwidth product is measured to be $1.1 \times 10^4$ V/V GHz.

Slew-rate measurements were made on the same amplifier shown in Figures 10 and 11. For the slew rate measurement, a 250 MHz input signal was provided from a signal generator. A differential input signal was generated using a 180° balun. Both outputs from the amplifier were provided to a high-speed Agilent sampling oscilloscope and the differential amplifier output was determined using the mathematical functions of the oscilloscope. Figure 12 shows the measured output waveform of the amplifier when driven to saturation. A peak output swing of 420mV was measured from the output buffer stage. Correcting for the measured attenuation of the output buffer (25.5dB from S-parameters) the corresponding voltage swing of the amplifier core is 9V peak to peak (4.5V single ended).

Similar results were achieved for other differential amplifier design variants and for different wafers highlighting the manufacturability of our approach.
III. SUMMARY

In this work we presented results on the direct monolithic integration of InP HBTs with Si CMOS on a silicon substrate. Our direct growth approach yields InP HBTs with similar RF performance to HBTs fabricated on InP substrates. Our truly planar approach allows tight device placement (InP HBTs - Si CMOS transistors separation as small as 2.5um) and the use of standard wafer level multilayer interconnects. While the results presented here are for InP HBTs directly integrated onto the silicon substrate, the approach is equally applicable to other III-V electronic (FETs, HEMTs) and opto-electronic (photodiodes, VSCLS) devices and opens the door to a new class of highly integrated, high performance, mixed signal circuits.

ACKNOWLEDGEMENT

This work is supported in part by the DARPA COSMOS Program (Contract Number N00014-07-C-0629). The authors would like to thank Mark Rosker (DARPA), Harry Dietrich (ONR) and Karl Hobart (NRL).

REFERENCES


[7] Smart-Cut™ is a registered trademark of Soitec.
