WAFER-LEVEL FAN-OUT FOR HIGH-PERFORMANCE, LOW-COST PACKAGING OF MONOLITHIC RF MEMS/CMOS

Rameen Hadizadeh, Anssi Laitinen and David Molinero, Ph.D.
WiSpry, Inc.
Irvine, CA, USA
rameen.hadizadeh@wispry.com

Nelson Pereira and Márcio Pinheiro
Amkor Technology, Inc.
Porto, Portugal

ABSTRACT
Navigating the trade-offs between performance, size, cost and reliability can be a challenge when considering integrated circuit (IC) packaging and the end-application. The integration of micro-electromechanical systems (MEMS), either monolithic or heterogeneous, introduces yet another level of complexity that has only recently been a major focus of multi-device packaging [1]. Wafer-level fan-out (WLFO) technology can enable improvements in several areas, primarily the reduction in size of parasitic interconnects and, proportionally, a drastic decrease in overall form-factor as compared to more ubiquitous chip-scale packaging solutions. Widespread adoption of WLFO packaging [2] has driven implementation costs to a level competitive with traditional fan-in wafer-level packaging.

This study quantifies the benefits of WLFO versus flip-chip land grid array (FCLGA) packaging for a radio frequency (RF) MEMS digital tunable capacitor array integrated with 180nm CMOS technology. RF performance hinges critically upon the ability of the package to transfer signals with minimal impedance, necessitating shortened redistribution layer (RDL) paths and reduced, or removed, solder interconnects. Flip-chip packaging requires a multi-layered substrate and an intermediate solder interconnect while chip-first WLFO packaging makes use of direct Cu RDL bond to die pads and a single-level of routing to the ball grid array. Die size in both cases is 1.8x2.2x0.3mm, which enables a direct comparison of form-factor between the two package types. Manufacturability is addressed in this study; a primary challenge of monolithically integrated MEMS/CMOS is the ability to survive typical IC packaging processes wherein thermal, mechanical and electrical overstress may occur.

Daisy chain packaged parts were subjected to board-level mechanical shock. Functional packaged parts were subjected to component-level reliability stresses. RF characterization of functional packaged parts was conducted on printed circuit boards (PCB); the primary figure of merit being self-resonance frequency (SRF) as a result of overall parasitic losses.

Key words: Fan-out, wafer-level packaging, RF MEMS

INTRODUCTION
Wireless devices benefit from tunable RF MEMS components as they enable high performance over a broad range of carriers and frequency bands in the radio spectrum. RF front-end (RFFE) architecture complexity is reduced when designed specifically for tunable components, resulting in a savings of power, space, RF losses, cost and time-to-market [3]. Therefore the ability to package RF MEMS components competitive with incumbent technologies becomes vital to market acceptance. Maintaining device and package reliability while significantly decreasing size and cost will enable MEMS to become a pervasive solution for global connectivity.

The FCLGA package is robust in terms of device protection and is a well-exercised process at many outsourced suppliers of assembly and test (OSATs). Package footprint is dictated primarily by the complexity of interconnects, and can be comparable to die size if the pin-out density is low. Flip-chip is synonymous with low parasitics due to the absence of bond wires, however parasitics can be further decreased with wafer-level packaging technologies that eliminate substrate interposers altogether. WLFO packaging is an attractive solution as it enables reduced parasitic interconnects as well as a robust, over-molded form factor. Traditional fan-in wafer-level chip-scale packaging (WLCSP) is comparable in interconnect complexity to WLFO but does not provide die protection nor the ability to integrate multiple dies in a single package.

FABRICATION
Monolithic MEMS/CMOS Foundry Process
The RF MEMS digital tunable capacitor array (DTC) is integrated into the back-end-of-line (BEOL) of a 0.18um HV-CMOS process on 200mm wafers. Among the final BEOL process steps is hermetic wafer-level encapsulation of the MEMS devices, which are suspended metal-oxide beam structures capable of electrostatic actuation (Figure 1 and Figure 2).

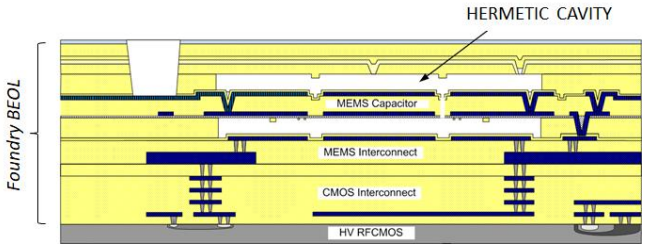


Figure 1. Cross-section schematic of MEMS/CMOS structure

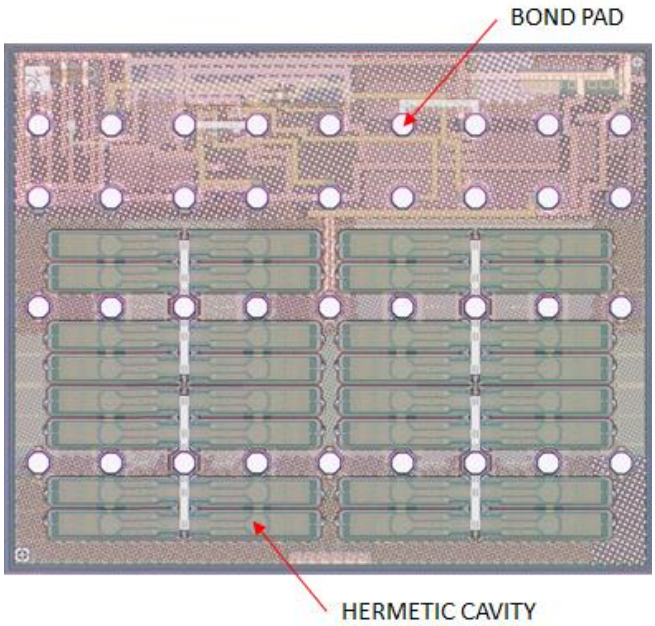


Figure 2. Optical image of the DTC integrated with CMOS

Upon completion of the foundry process the MEMS and CMOS are monolithically integrated within a die area of 1.8mm x 2.2mm, enabling the wafers to proceed through a typical single-die IC packaging solution.

FCLGA Packaging Process

Surface micro-machined MEMS structures present various challenges (e.g. thermal and mechanical overstress) when implementing traditional IC packaging processes. There are a number of concerns including, but not limited to, cavity lid fracture, plastic deformation or fracture of MEMS beams and triboelectric charging that can significantly decrease performance yield of final packaged parts. The MEMS structure has been thoroughly characterized to ensure its survival within the typical FCLGA packaging process parameters.

After polymer re-passivation, wafer bond pads are plated with a thick Cu UBM and Sn-Ag flip-chip bump. Singulated dies are mounted to a four-layer organic laminate substrate. Substrate panels are subsequently epoxy compression over-molded; the over-molding stage also serves as the molded underfill (MUF) process. The resultant pressure from the MUF process is below the known fracture pressure of the MEMS cavity lid, which has been shown to withstand 2000 psi. Special care must be taken to ensure excessive electrostatic charging/discharging events do not occur, for example the omission of ultrasonic rinsing baths and direct DC bias plasma processing.

The DTC in FCLGA format has been qualified to moisture sensitivity level 3 (MSL3) JEDEC specifications, with an upper limit operating temperature of 85°C. Board-level drop testing (JESD22-B111, condition B) produced no package-PCB interconnect failures after 30 drops.

WLFO Packaging Process

The WLFO and FCLGA packaging processes begin with the same MEMS/CMOS device wafers. Commonality in foundry output is ideal for comparison of final product metrics across varying packaging technologies. Chip-first, face-down WLFO (Figure 3) was chosen for this study primarily for the direct Cu RDL to bond pad interconnect with the intention of eliminating flip-chip solder parasitics. Wafer preparation consisted of mechanical backgrind and step-cut dicing. Singulated dies were reconstituted into an over-molded 300mm wafer, which continued through a standard four-mask RDL process (polymer, Cu RDL, polymer, Cu UBM) with pre-formed Sn-Ag-Cu (SAC) solder ball drop.

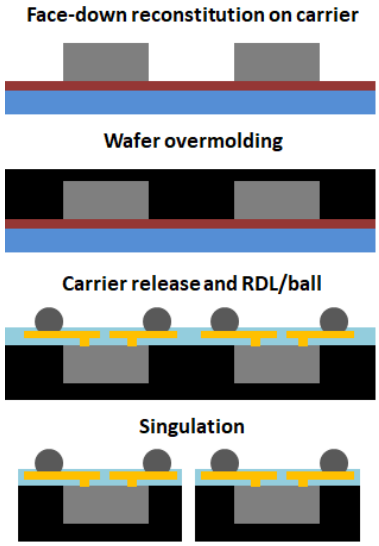


Figure 3. Chip-first, face-down WLFO process

The RDL layout was designed as a multi-project wafer (MPW) configuration in order to maximize RF and package characterization results with minimal process hardware. Two variations of ball-grid array (BGA) placement were designed into the MPW reticle field with the intention of studying the electro-mechanical effects of populating and depopulating interconnects directly above the MEMS cavity area (Figure 4).

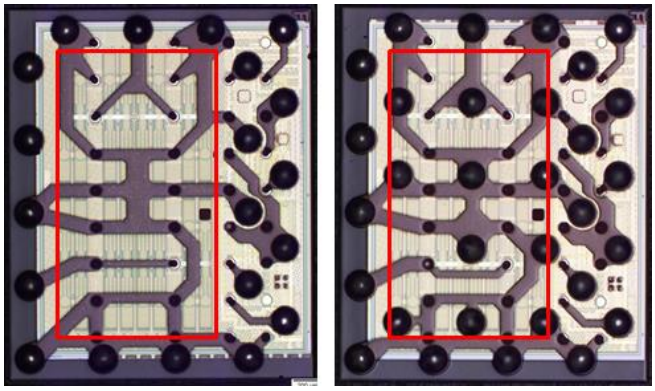


Figure 4. "Sparse" (left) and "dense" (right) BGA variations, MEMS area boxed in red

Package Comparison

The DTC in WLFO package format resulted in an XY area of 5.519mm^2, which is slightly smaller than the 5.738mm^2 consumed by the FCLGA package (Figure 5, bottom). Major size reduction is evident in the package height; the WLFO package, excluding solder ball standoff, is approximately 56% thinner than the FCLGA package (Figure 5, top). WLFO enables very close proximity of dies within a multi-die package thus further XY area savings could be realized. WLFO offers a packaging cost savings of approximately 48% over FCLGA (excludes test cost).

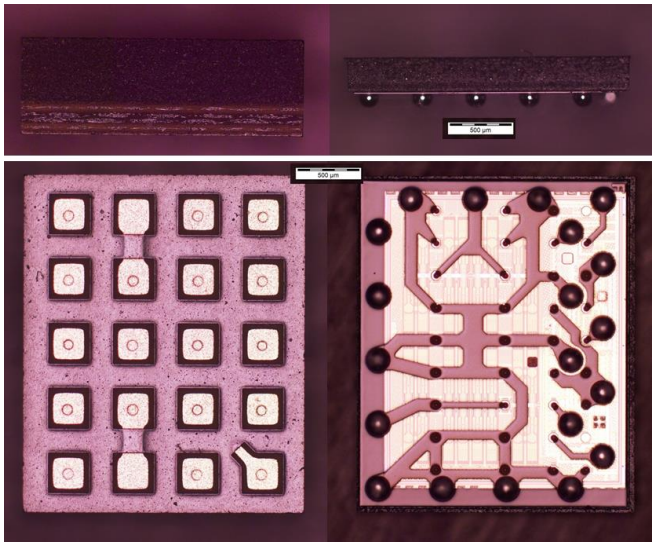


Figure 5. FCLGA (left) versus WLFO (right)

Die-to-PCB interconnects between the two package types can be expressed as a network of parasitic components, namely capacitance and inductance. For the FCLGA package, the flip-chip solder bump (subscript FC in Figure 6) and four layers of substrate routing (subscripts S,1-S,4 in Figure 6) contribute to a decrease in SRF of the DTC. For the WLFO package, these same interconnect parasitics are replaced by a more direct path (subscript RDL in Figure 6) and thus an increase in SRF of the DTC can be realized as discussed in a later section.

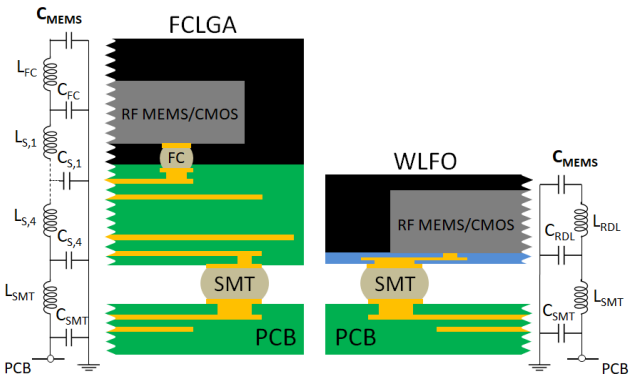


Figure 6. Comparison of parasitic inductance and capacitance for each package type

RF CHARACTERIZATION

Analytical Model

Package-level redistribution layers and interconnects result in parasitic losses that limit device RF performance. Effects are especially severe for tunable components, be it MEMS or Silicon-on-Insulator (SOI) technology. Device operating frequency range is affected by the inductance associated with packaging redistribution levels and solder interconnects. A simplified model derived from Figure 6 can be used to quantify the effects of package metal inductance wherein a simple inductance-capacitance (LC) lumped model is valid. Capacitance is related to the DTC while the inductance is related to the packaging interconnects. The input impedance of the series LC is calculated as:

$$Z_{IN} = j(\omega L - \frac{1}{\omega C}) \quad (1)$$

where $j = \sqrt{-1}$, ω is the angular frequency, L and C are the model equivalent inductance and capacitance values, respectively. As capacitance is the expected performance metric for the DTC, Equation (1) should be modified as:

$$Z_{IN} = -j\frac{1}{\omega C}(1 - \omega^2 LC) \quad (2)$$

or as:

$$Z_{IN} = Z_C \cdot \left(1 - \left(\frac{f}{SRF}\right)^2\right) \quad (3)$$

where Z_C is the expected DTC capacitive impedance and SRF is defined as:

$$\text{SRF} = \frac{1}{2\pi\sqrt{LC}} \quad (4)$$

SRF defines the frequency limit above which the DTC no longer behaves as a capacitor, but rather as an inductor, and is thus a critical parameter for defining the device operating frequency range. Capacitance of the DTC is expected to increase non-linearly as the frequency of operation nears the SRF limit. During operation, capacitance is ideally

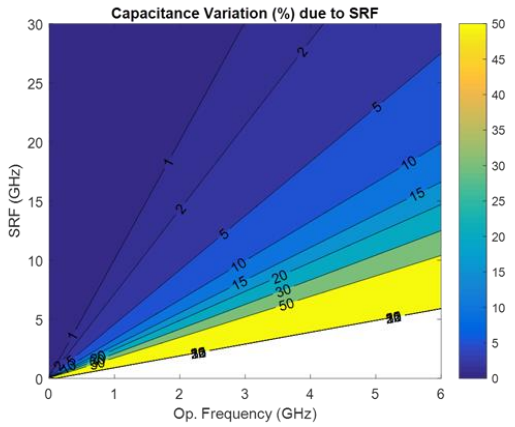
independent of frequency and thus a higher SRF value is desirable for tuning versatility. From Equation (3), we can evaluate this capacitance variation as:

$$C = \text{imag}\left(\frac{1}{Z_{IN}}\right)\Big/_{\omega} \quad (5)$$

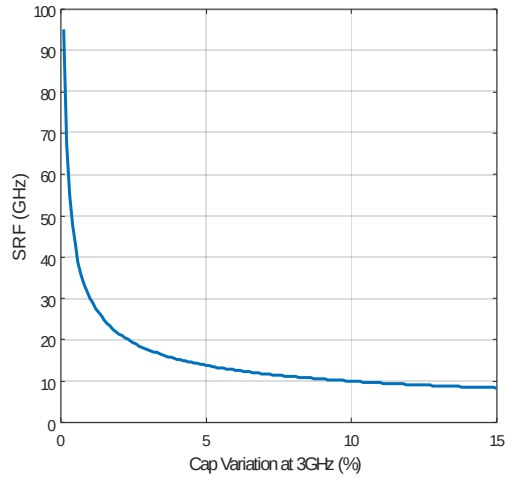
Finally, the measured device capacitance as influenced by proximity to SRF can be written as:

$$C = \frac{C_0}{1-\left(\frac{f}{SRF}\right)^2} \quad (6)$$

where C_0 is the nominal DTC capacitance value and C is the extracted or measured capacitance. Figure 7a shows how the normalized capacitance (C/C_0) behaves as a function of SRF and operating frequency. Figure 7b shows the same behavior for a fixed operating frequency of 3GHz. These plots define what SRF is required to achieve a desired capacitance variation tolerance. For example, an SRF of 13GHz would result in a 5% maximum variation of capacitance at 3GHz operating frequency.



(a)

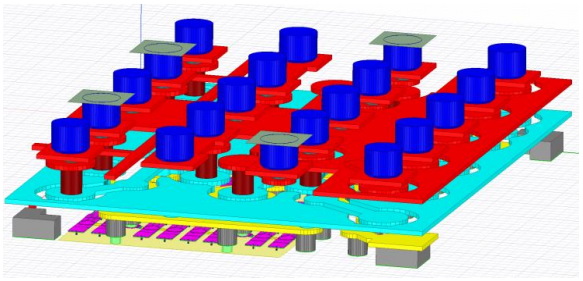


(b)

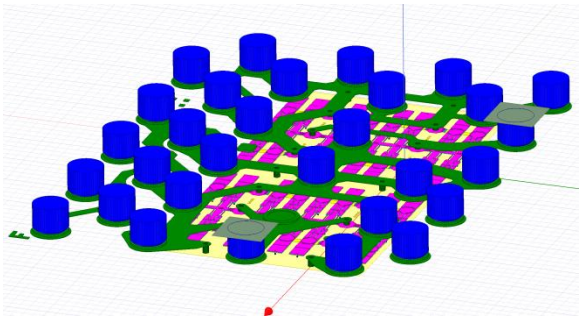
Figure 7. (a) Normalized capacitance variation as a function of SRF and operating voltage. (b) SRF as a function of normalized capacitance variation for an operating frequency of 3GHz

Finite Element Model

ANSYS high-frequency structure simulator (HFSS) models were constructed for the FCLGA and WLFO package types using a single RF port configuration. Simulations yielded the RF scattering (S) parameters. In Figure 8 the topside of the model is adjacent to the PCB.



(a)



(b)

Figure 8. ANSYS HFSS models for (a) FCLGA and (b) WLFO package types

Device capacitance can be extracted from the simulated S-parameters using Equation (5). Figure 9 shows the extracted maximum capacitance (C_{max}) as a result of actuating all eight DTC beams in the C4 bank (Figure 10), which is defined as a worst-case scenario for SRF evaluation.

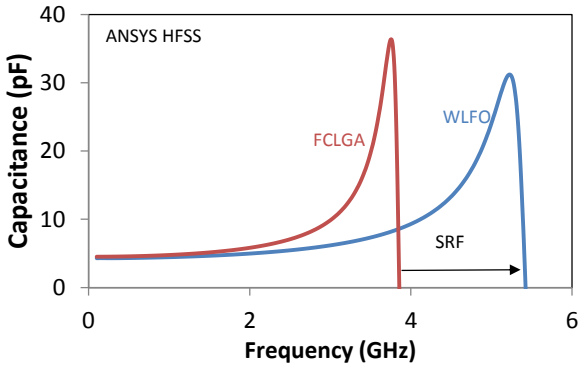


Figure 9. FCLGA (red) and WLFO (blue) maximum capacitance extracted from ANSYS HFSS Simulations

The DTC in FCLGA format was designed to have a useful operating frequency range up to 3GHz, which can be extended by the design and application of WLFO packaging. Simulated results (Table 1) show an extended SRF of 1.6GHz, or approximately 42%, for the WLFO package.

Table 1. Simulated capacitance, SRF and inductance values

	Cmax (pF) << SRF	SRF (GHz)	L (nH)
WLFO	4.29	5.4	0.20
FCLGA	4.51	3.8	0.40

It is important to note that all RF device technologies incur parasitic losses regardless of packaging approach. Degradation of SRF is therefore expected and accounted for within product design and development.

EXPERIMENTAL RESULTS

Component-Level Reliability (CLR)

Functional WLFO packaged parts (dense BGA variant) were subjected to various temperature and moisture conditions as outlined in Table 2. Pass or fail was determined by the same socketed test metrics used to previously qualify the FCLGA packaged parts; focus was primarily on MEMS performance as well as die-to-package interconnect continuity verification.

Table 2. CLR stress plan for WLFO parts

Stress / Test	Method	Conditions	Requirement	Sample Size / Variant
High Temperature Storage Life (HTSL)	JESD22-A103	Service Condition A 125°C (-0/+10°C)	1000 hrs	30
Low Temperature Storage Life (LTSL)	JESD22-A119	Service Condition A -40°C (-10/+0°C)	1000 hrs	30
Preconditioning (PC-MSL1)	JESD22-A113 J-STD-020	Bake 125°C, 24 hours MSL1 soak = 85°C/85%RH, 168 hr 3X reflow, 260°C peak per JEDEC	MSL1	60
Highly Accelerated Stress Test (u-HAST)	JESD22-A110	130°C, 85% RH, 33.3 psia, unbiased	96 hrs	30 (from MSL1 group)
Temperature Cycling (TC-CL)	JESD22-A104	Service Condition N -40°C to +85°C, 15 minute soak (mode 4)	1000 cycles	30 (from MSL1 group)

The component-level reliability stress plan was completed without any failures. MSL1 is certainly more desirable than MSL3 as it enables an indefinite shelf life for packaged parts that have been removed from their moisture barrier bag [4].

Board-Level Reliability (BLR)

Dummy silicon daisy chain versions of each functional variant were included in the WLFO MPW to address the board-level reliability ramifications of such large area depopulations in the BGA (Figure 4). Side A and Side B of a JEDEC-compliant BLR PCB were designed to accommodate the dense and sparse BGA daisy chain packages, respectively. The entire BGA of each daisy chain was treated as a single resistive network and was monitored in-situ for high resistance and/or discontinuity during stress. Drop testing (JESD22-B111, condition B) produced no package-PCB interconnect failures for either BGA variant after 30 drops.

RF Measurements

Empirical S-parameter data was gathered in order to validate the finite element model simulation. RF measurements were taken using an Agilent N3383A programmable network analyzer (PNA) with an available frequency range of 0.1MHz to 9GHz. Packaged parts were mounted on a five-layer, high-performance Rogers RO4350B PCB. Figure 10 shows a CAD model of the evaluation board as well as the DTC array and bank assignments C1 through C4. Although each device has four separately addressable capacitor banks, measurements focused on a single bank as there is little difference among them.

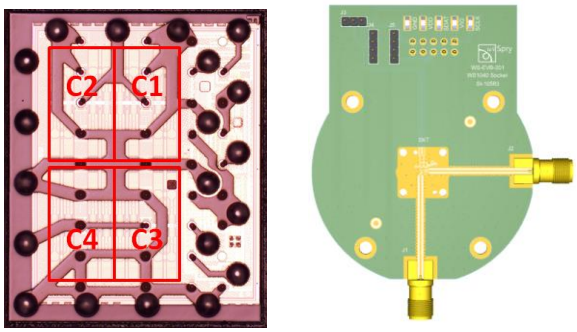


Figure 10. Capacitance bank assignments (left) and WLFO RF test PCB (right)

The PNA was calibrated following a short-open-load-thru (SOLT) methodology up to the board connectors. Further de-embedding was required to bring the RF reference plane up to the package-PCB solder interconnects [5]. Figure 11 shows agreement in extracted C_{max} values for both simulated and empirical results.

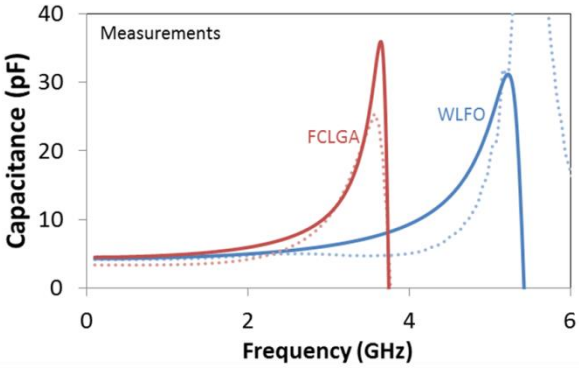


Figure 11. Measured (dashed line) versus simulated (solid line) C_{max} for FCLGA and WLFO packaging

An SRF increase of more than 2GHz is observed for the WLFO package, with minimal capacitance variation below 4GHz. Further RF characterization (e.g. harmonics, IIP3 and large signal) was completed to ensure that signal integrity remains unaltered by the WLFO package.

CONCLUSION

WLFO packaging produces several benefits compared to traditional FCLGA packaging of monolithic RF MEMS / CMOS. Improvements were realized in packaging cost and reliability, package size and RF performance.

1) Major reductions in interconnect complexity resulted in a notable increase in SRF and a decrease in capacitance variation at higher operating frequencies.
2) Packaging cost, sans test, decreased by approximately 48% thanks primarily to the reduction in complexity and assembly consumables.
3) Functional devices passed MSL1 reliability specifications, an improvement from the previously qualified MSL3 specifications. BGA depopulation above the MEMS area had no critical effect on board-level mechanical shock.
4) Package z-height, sans BGA, was reduced by approximately 56%, enabling a form factor that more closely resembles the incumbent handset packaging technology.

WLFO packaging enables multi-die (heterogeneous) integration in which further consolidation of the RF front-end architecture is possible with the aid of chip-package co-design.

ACKNOWLEDGEMENTS

The authors would like to acknowledge and thank the WiSpry engineering and test teams for their valuable package characterization efforts. WiSpry would like to thank Amkor Technology Europe, Portugal for their support and efficacy in prototyping this technology.

REFERENCES

[1] A. Pizzagalli, Yole Development, "MEMS & Sensors Packaging: Wafer-Level-Packaging Technology and Market Trends", Proceedings of SEMICON West, July 2016.

[2] J. Azemar, Yole Development, "Fan-Out: Technologies and Market Trends 2017", Proceedings of SEMICON Taiwan, September 2017.

[3] Hilbert, Jeffrey L. 2016. *Tunable RF Components and Circuits: Applications in Mobile Handsets.*

[4] IPC/JEDEC, "Joint Industry Standard: Handling, Packing, Shipping and Use of Moisture, Reflow, and Process Sensitive Devices", J-STD-033D, April 2018.

[5] D. Molinero, S. Cunningham and A. Morris, "Capacitance Extraction De-Embedding Method to Address Board Variability", Asia-Pacific Microwave Conference, November 2018.