Between 2D and 3D: WLFO Packaging Technologies and Applications

Minghao Shen
Altera (now part of Intel)

June 9th, 2016
Outline

- The 2.n D

- WLFO technologies
  - Process and architect options
  - Feature comparison

- WLFO’s application for Si – Interposer – Substrate co-architecture
  - Introduction
  - <10um pitch bumping

- Summary
Chip-to-Package Interconnect Technology

Low-end application

1957
Wire Bonding (WB)

Mid-range application

1964
Flip Chip Bonding (TAB)

1966
Tape Automatic Bonding (TAB)

High-end application

Mixed technologies

Cost-effective
SSF & cost driven

Heterogeneous integration

Performance driven

Heterogeneous integration

3D

2.nD

2D
Multi-die Integration Key Factors

Cost
- Cost adder compared to monolithic
- Figure of merit cost

Size
- How many dies can be accommodated?
- Max body size with reasonable warpage.

Scaling
- Can routing be easily scaled down?
- Scaling limit?
- Cost for scaling?

Connection / routing density
- 50~300 /mm inter-die channels, depending on application
- Area of routing

Substrate / interposer

Quality, supply chain, schedule...
- FLI, BLR, auto...
- Multi-options?
- OSAT turn-key?
- Technology qual available?
<table>
<thead>
<tr>
<th></th>
<th>CoWoS</th>
<th>WLFO</th>
<th>EMIB</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSV</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Interposer size</td>
<td>Covers all dies, reticle limit, or high cost reticle stitching</td>
<td>Covers all dies, reticle limit, or high cost reticle stitching</td>
<td>Only at the die interface. No area limitation</td>
</tr>
<tr>
<td>Connection / routing density</td>
<td>4 layers of full area routing, &lt;4um pitch</td>
<td>&gt;=4 layers of full area routing. ~4 - 20um pitch</td>
<td>Only inter-die routing. Others go through looser PKG routing</td>
</tr>
<tr>
<td>Substrate layer count</td>
<td>Same or add 2 layers wrt. monolithic</td>
<td>Partially absorbed in interposer. Possible to reduce substrate 2 layers without PTH or no substrate at all.</td>
<td>Add min. 2 layers wrt. monolithic.</td>
</tr>
</tbody>
</table>
WLFO Schematics

Substrate-less

On substrate

WL mold

Die 1

Die N

FO routing

Substrate
# WLFO Technology Overview

<table>
<thead>
<tr>
<th>Description</th>
<th>Die First (eWLB like)</th>
<th>Die Last</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Si based</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. 2~3 layers of stacked RLD.</td>
<td>1. upto 4 layers of &lt;=2/2um routing</td>
<td>1. upto 3 layers of 2~3um minimal line width RDL</td>
</tr>
<tr>
<td>2. 2~3um minimal line width RDL</td>
<td>2. Si interposer w/o TSV, sacrificial Si removed</td>
<td>2. Die to interposer join by ubump.</td>
</tr>
<tr>
<td></td>
<td>3. Die to interposer join by ubump.</td>
<td></td>
</tr>
<tr>
<td><strong>Example</strong></td>
<td>InFO, eWLB, Deca*</td>
<td>SLIM, SLIT</td>
</tr>
<tr>
<td><strong>Pro.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Simpler process, no bumping involved.</td>
<td>1. High routing density, can support HBM and XCVR</td>
<td>1. Die last --&gt; known good die and known good routing</td>
</tr>
<tr>
<td>2. No Si interposer</td>
<td>2. Die last --&gt; known good die and known good interposer</td>
<td>2. No Si interposer</td>
</tr>
<tr>
<td></td>
<td>3. High routing process yield (natural FEOL Cu damascene process)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. No technical limitation on routing layer count</td>
<td></td>
</tr>
<tr>
<td><strong>Con.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Low routing density</td>
<td>1. Si interposer --&gt; cost adder</td>
<td>1. ubump --&gt; Cost adder</td>
</tr>
<tr>
<td>2. Die committed prior to low yield stacked RDL process</td>
<td>2. ubump --&gt; Cost adder</td>
<td>2. Low routing density</td>
</tr>
<tr>
<td>3. Limited routing layer count</td>
<td></td>
<td>3. Limited routing layer count</td>
</tr>
</tbody>
</table>
Pick the right 2.n Package

- No one-solution fits all 2.n package technology
  - Size, cost, performance, and other factors...

- Small body size, low connection density applications favor die-first WLFO, due to its simplicity, and lower cost

- Super large body size applications are more suitable for EMIB technology due to less limitation on the die box size.

- In between...
  - TSV based interposer solution and die-last TSV-less interposer solutions
  - With the advance of fine pitch and multi-stack of Cu RDL, die-first WLFO may take some advantages away from die-last WLFO
  - Deca panel-like WLFO...
  - And more…
The 2.n D

WLFO technologies
- Process and architect options
- Feature comparison

WLFO’s application for Si – Interposer – Substrate co-architecture
- Introduction
- <10um pitch bumping

Summary
From Interface Co-Design to Co-Architecture

Design 1

Si

Package

Design 2

The Design
From Interface Co-Design to Co-Architecture

Design 1
Si

Package

Design 2

The Design
From Interface Co-Design to Co-Architecture

Design 1

Si

Package

Design 2

The Design
From Interface Co-Design to Co-Architecture
From Component Co-Design to System Co-Architecture

Baby 2.XD

Baby WLFO/PLFO
The Si ↔ Interposer Co-architecture

- Interposer L/S <= Si top metal and Substrate L/S
- Interposer area > Si area
- Interposer absorbs some layers from Si and substrate to reduce total layer count

**Si FEOL**

- Lower level metal routing L/S <<1/1um
- "FO" routing on Si to match to RDL

- >20um pitch

**WL-RDL**

**Si FEOL**

- 2~4um pitch Si Die top metals

**Interposer / FO**

- Top metal absorbed into interposer with comparable L/S

- ~2-4um pitch
The **REAL** Si ↔ Interposer Co-architecture

- Ubump bonded interposer structure is **NOT REAL** Si ↔ interposer co-design
- Si top metal still needs to FO to provide large pitch ubump pads
  - Large routing area taken by ubumps.
  - “FO” design is needed to redistribute the routing given the constrains.

- Bonding pitch ≅ routing pitch **IS REAL** Si ↔ interposer co-design
First Step: Direct Cu-Cu diffusive bonding technology

- **Direct Cu-Cu bond**
  - Cu bump to Cu bump diffusive bonding under low pressure and relatively low temperature.
  - Small bump pitch of 6um, comparable to die top metal <=4um pitch.
  - Reduce the unnecessary FO from die to interposer.
  - Move top metals to interposer
  - Reduce total layer count and cost

© 2016 Altera—Public
Second Step: Direct Cu-Cu diffusive bond TSV-less WLFO

Remove TSV for further cost reduction
Third Step: Bumpless TSV-less WLFO

- **TSV-less high density interposer**
- **Direct die trace to die trace bonding < 4um pitch**
  - No Die to interposer FO at all → easily exchange die top metals with interposer metals.
  - Remove bumping process → lower cost
  - Low temperature, low pressure, high throughput process → lower cost
- **Reduced layer low cost coreless substrate**
Summary: Co-architecture

- Co-development is a more advanced co-design
- Co-architecture is a more advanced co-development
- Different interconnect levels use different type metal circuitries
- Different type of metal circuitries can be substituted
- Thicker metal is more efficient for power supply
- Thinner metal is more efficient for IO
- Reduced form factor reduces discontinuity
- Reduced discontinuity reduces overhead on interconnect circuitry

CO-ARCHITECTURE = LOW COST
Thank You