Next Generation Packaging Technologies for cognitive computing devices

Yasumitsu Orii, PhD.
Senior Technical Staff Member
Senior Manager of Science & Technology, IBM Research Tokyo
Outline

• Computer System for the Era of Big Data

• What is “Cognitive Computer”?

• Requirement for Cognitive Computing Devices
  – Interposer Technologies
  – Packaging Technologies
Outline

- Computer System for the Era of Big Data
- What is “Cognitive Computer”?
- Requirement for Cognitive Computing Devices
  - Interposer Technologies
  - Packaging Technologies
Info-Plosion

The oceans of data to be produced in the world. The size of data generated per Day in 2010 is “1ZB”, which is equal to 3.7 trillion years News Paper Data.

M.Kitsuregawa, J.IEICE, Vol.94, No.8, 2011
Data volume is increasing!

Source: IDC Digital Universe

2012: 3 zettabytes
2020: 40 zettabytes
Internet of Things
Big data to be accelerated by IoT

Ratio of the data generated by sensors

2005 11% 2020 42%

Source: IDC Digital Universe
Capacity of communication Infrastructure exceeds its limit in 2020
Outline

- **Computer System for the Era of Big Data**

- **What is “Cognitive Computer” ?**

- **Requirement for Cognitive Computing Devices**
  - *Interposer Technologies*
  - *Packaging Technologies*
Cognitive Systems: A new era of computing

3rd Generation
Cognitive Systems Era

Programmable Systems Era

Tabulating Systems Era

Computer Intelligence

Time

Yasumitsu Orii, IBM Research - Tokyo/Slide 10
IBM Watson Project

- 15TB Memory
- 2,880 Processor core
- 200KW

- Ken Jennings, who won a record 74 consecutive "Jeopardy!" games in 2004-05
- Brad Rutter, who is the biggest all-time money winner.

It is expected to apply the technologies for Watson to many applications.
- System to backup more rapid and accurate medical diagnosis.
- System to check potential coactions between medicines.
- System to refer past authority by lawyers and benches.
- System for hypothesis scenario and legal compliance in finance sector.

Yasumitsu Orii, IBM Research - Tokyo/Slide 11
Computers and the Brain are Different and Complementary

3.55 GHz, sequential, linear  10 Hz, parallel, high fanout

200kW  20W

12,000 L  1.2 L

~5,000 connections/cm²  10⁶ neurons/cm²

Separates memory  Integrates memory
IBM Creating New Cognitive Technologies

Power, space, memory, and communication scale with synapses

Crossbar architecture
Brain-inspired Chip

~$10^{10}$ synapses/cm$^2$  ~$10^6$ neurons/cm$^2$

~$10^4$ neurons/cortical column

~$5\times10^3$ long range axons @ 1 Hz

~$5\times10^3$ long range axons @ 1 Hz

~$10^4$ neurons/cortical column

~$10^6$ intersections/cm$^2$ @ 100 nm pitch

5x$10^8$ transistors/cm$^2$

Layered cortical circuits with $\sim10^6$ neurons/cm$^2$

Multi-Gbit/sec Digital comms

Electronic Brain

Outline

• Computer System for the Era of Big Data

• What is “Cognitive Computer”? 

• Requirement for Cognitive Computing Devices
  – Interposer Technologies
  – Packaging Technologies
Human’s brain

- Human’s brain: $10^{11}$ (100Billion) neurons
  $10^{15}$ (1000Trillion) synapses

Eye: 1B neurons
10T synapses

J. DiCarlo et al., Neuron, 2012
Brain size of mammals

Mouse’s brain:
- 70M neurons
- 700B synapses

Mouse’s vision (est.):
- 0.7M neurons (~1M)
- 7B synapses (~10B)

Human’s brain:
- 100 Billion neurons
- 1,000 Trillion synapses

Human’s vision:
- 1B neurons
- 10T synapses

S. Herculano-Houzel 2009
Package Outline on Implementation

**Assumption:** Mouse’s eye (synapse=10B, # of neurons=1M)

**Number of I/O estimation**
- Total signal I/O of neurons: 1M x 2
- Total I/O with power & ground: 4M
- Layers of RBM: 16 (= number of stacked chip)
- Total I/O with power & ground of a stack (4 dies): 1M x 2 x 2 / 16 = **250,000**

**Die size estimation**
- Total drivers of a stack (4 dies): 125,000
- # of driver on a die: 125,000 / 4 = 31,250
- Area of driver (45nm): 25 x 50 = 1,250um²
- Area of die: 0.00125 x 31,250 um² = 39.0625mm² -> **6.25mm²**

**Package outline**
- 4 die stacked with 7mmSQ chips.
- Area arrayed 125,000 signal I/Os and same counts of power and ground per chip in 14μm pitch.
- 16 die stacks on interposer.
- Chained synapse to synapse connection between die stacks.
Wiring Study of Interposer

- Interposer technology innovation required
  - Most likely less than 0.5μm line and spacing
  - Land-less capability for via connection between layers
  - Appropriate pad surface finishing for 10μm interconnection
Outline

- Computer System for the Era of Big Data
- What is “Cognitive Computer”?
- Requirement for Cognitive Computing Devices
  - Interposer Technologies
  - Packaging Technologies
10micron pitch interconnection

- Bump/Pad structure
  - SnAg: 3μm
  - Ni: 3μm / Au: 0.1μm
  - Post: 3μm (Cu, Ni)
  - Al wiring
  - Solder bump: 5μm
  - Ni/Au pad: 6μm

- Wiring on chip/substrate
  - 4x4 10 μm-pitch cluster bumps
  - 30x30 100 μm-pitch cluster array

Substrate: 10mmx10mm

Table:

<table>
<thead>
<tr>
<th></th>
<th>As depo.</th>
<th>After hydrogen radical treatment + wetback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu-post</td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
</tr>
<tr>
<td>Ni-post</td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
</tr>
</tbody>
</table>

Chip 7mmx7mm

10μm pitch SnAg bump

10μm pitch Ni/Au pad

Yasumitsu Orii, IBM Research - Tokyo/Slide 21
10micron pitch interconnection
- Bonding with contact-retry alignment method
- Bonding condition: 300 deg. C, 10N

Cu-post: Good continuity, IMC growth
Ni-post: Void generated due to consumed solder
IMS (Injection Molded Solder) technology

- Molten Solder Injection Method -
  - Very simple process
  - Green process
    - No flux, No formic acid
    - Minimum solder waste
  - Pure solder injection
    - Higher solder volume
    - Extendible to very fine pitch
    - Solder alloy flexibility
  - One pass for multiple solder sizes

- Solder Paste Printing Method -
  - Solder Paste (Solder powder + Flux)
  - Reflow + Cleaning

Reference
Wafer IMS technology: Demonstration with 8’ wafer

Demonstrated Wafer IMS bumping with 8 inch wafer

Wafer IMS demonstration with various opening size in 8 inch wafer!!!

50um

solder bump

50um diameter

75um

solder bump

75um diameter
Experimental Results of Substrate IMS Technology

• **A: Mask IMS (132μm pitch)**

*Mask IMS 53.3μm over SR.*

<table>
<thead>
<tr>
<th></th>
<th>Mask IMS bump height over solder resist</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>54.3</td>
</tr>
<tr>
<td>B</td>
<td>55</td>
</tr>
<tr>
<td>C</td>
<td>55.5</td>
</tr>
<tr>
<td>D</td>
<td>51</td>
</tr>
<tr>
<td>E</td>
<td>52.9</td>
</tr>
<tr>
<td>F</td>
<td>52.7</td>
</tr>
<tr>
<td>G</td>
<td>53.2</td>
</tr>
<tr>
<td>H</td>
<td>51.4</td>
</tr>
<tr>
<td>I</td>
<td>54.4</td>
</tr>
<tr>
<td>MAX</td>
<td>55.5</td>
</tr>
<tr>
<td>MIN</td>
<td>51</td>
</tr>
<tr>
<td>AVE</td>
<td>53.38</td>
</tr>
</tbody>
</table>

IMS bump height distribution tighter than with paste printing

• **B: Mask-less IMS (132μm pitch)**

No residue on solder resist

~ ideal for low cost (very fine pitch) ~

<table>
<thead>
<tr>
<th></th>
<th>Solder height above solder resist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Std</td>
</tr>
<tr>
<td>17.9 μm</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Solder volume and bump height are depended on solder resist opening design
Terminal devices enable cognitive processing. Workload is balanced between data center and terminals.

Data center responsible for advanced cognitive processing using accumulated valuable information.

Terminal devices responsible for low-end cognitive processing and delivery to data center, to avoid information explosion.
Thank you for your attention!