Outline

- Introduction
- Smartphone Requirements
- WideI0
- 3D Smart Partitioning
- Conclusion
Smartphones ➔ Environment aware Computers

GPRS 56-114 Kbits

1000 times increased peak rate in 10 years

LTE >150 Mbps

13 Mhz

From voice only to the power of a computer

>1.5GHz 8000+DMIPS

Communication

Entertainment & Internet

Augmented Reality

Digital Convergence

Image: Raygun Studio
What do we need in a Smartphone?

- PC-like computing platform
- High Performance Graphics Processing
- MODEM for ultra fast network connections
- Sensors
- Interfaces
- ...

Differentiator is **Performance**
### Thermal & Power Constraint

<table>
<thead>
<tr>
<th>Terminal</th>
<th>Feature Phone</th>
<th>Smart Phone</th>
<th>Smart Phone ++</th>
<th>Small Tablet</th>
<th>Large Tablet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display Size</td>
<td>-</td>
<td>4”</td>
<td>5”</td>
<td>7”</td>
<td>10”</td>
</tr>
<tr>
<td>Total Max Power</td>
<td>2W</td>
<td>4W</td>
<td>5W</td>
<td>8W</td>
<td>15W</td>
</tr>
</tbody>
</table>

Power efficiency = Performance

![Diagram showing power consumption and temperature for Laptop, Tablet, and Smartphone]
The FD-SOI advantage

**Faster**
- +35% max freq.
- 2.5GHz

**Cooler**
- -25% at 1.85GHz
- -50% at 800MHz
- 5K DMIPS@0.65V

**Simpler**
- -20% for same performance
- -20% for video recording
- -20% for video playback

**Computing**
- More headroom for protocol stack improvements

**Graphics**
- +20% max freq.

**Multimedia**
- +20% max freq.

**Modem**
- -12% LTE data call

Next step performance with same GPU

Next step performance

More performance available to implement new standards on same architecture
DRAM Bandwidth Requirements
## Memory Options and BW

<table>
<thead>
<tr>
<th></th>
<th>LPDDR2 PoP</th>
<th>LPDDR3 PoP/Discrete</th>
<th>WideIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW (Gbyte/s)</td>
<td>8.5(1)</td>
<td>12.8 (1)</td>
<td>12.8</td>
</tr>
<tr>
<td>possible BW evolution</td>
<td>-</td>
<td>17 (2)</td>
<td>17 (3)</td>
</tr>
<tr>
<td>(Gbyte/s)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max package density</td>
<td>4x4</td>
<td>4x4</td>
<td>1x4 (4)</td>
</tr>
<tr>
<td>(Gbit)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power efficiency</td>
<td>78</td>
<td>67</td>
<td>42</td>
</tr>
<tr>
<td>(mW/Gbyte/s)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samples availability</td>
<td>OK</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>volume maturity</td>
<td>2011</td>
<td>2012</td>
<td>2013</td>
</tr>
</tbody>
</table>

(1) 32b dual channel configuration assumed
(2) LPDDR3E: clock from 800 to 1066MHz. Standardization at JEDEC in progress
(3) WideIO clock frequency from 200MHz to 266Mhz: already specified at JEDEC
(4) 4x4Gbit supported with JEDEC 4 die memory stack
WIOMING – WideIO Application Processor

- ST-Ericsson has successfully tested its WIOMING 3D application processor
- WIOMING which stands for WideIO Memory Interface Next Generation was developed in cooperation with STMicroelectronics and CEA-Leti and provides a major breakthrough for performance increase in low power mobile devices.

- More than 200Mhz at Vmin
- Less than 4 mW/Gbit/s

<table>
<thead>
<tr>
<th>WideIO controller and WideIO memory frequency (MHz)</th>
<th>333</th>
<th>313</th>
<th>294</th>
<th>278</th>
<th>263</th>
<th>250</th>
<th>238</th>
<th>227</th>
<th>217</th>
<th>208</th>
<th>200</th>
<th>192</th>
<th>188</th>
<th>179</th>
<th>172</th>
<th>167</th>
</tr>
</thead>
<tbody>
<tr>
<td>V max.</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
</tr>
<tr>
<td>V nom.</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>FAIL</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
</tr>
<tr>
<td>V min.</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
<td>PASS</td>
</tr>
</tbody>
</table>

(1) 13N MBIST test pattern at ambiant temperature
WIOMING Technology

- WIOMING is based on the WideIO SDRAM JEDEC memory standard released in Jan 2012
- Target is to offer same bandwidth as a quad-channel 32-bit LPDDR2 interface but with half the power consumption
- Technology is based on a large bus interface (512 bits) operating at low frequency (200 MHz) in Single Data Rate (SDR) mode
- WideIO DRAM die is stacked on top of the mobile processor in the same package to reduce interconnect capacitance
- Face to Back stacking with Through Silicon Vias (TSVs) in the mobile SoC flip-chip die
- TSVs: Ø 10 μm, AR 8, Pitch 40 μm
- Cu Pillars: Ø20 μm, Height 20 μm, Pitch 40 μm
But why did WideIO not take off yet?

WideIO in Smartphone?

- Complex business model
- Less BW required
- Cost
- LPDDR3/4 available
- Thermal
Wide-IO Business Model

Ownership of SOC and memory sourcing

The consignment model is considered as sole solution to be acceptable by all three parties involved

Consignment Cage

Handset Manufacturer

Purchase Order → Billing → Final Goods → KGD DRAM → Chip set Manufacturer

Purchase Order → Billing → KGD DRAM → Memory Supplier

Yield compensation to be negotiated

Has to cope with delayed payment (after assy + test)
## Memory Options - Evolution

<table>
<thead>
<tr>
<th></th>
<th>LPDDR2</th>
<th>LPDDR3</th>
<th>WideIO</th>
<th>LPDDR4</th>
<th>WideIO2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW (Gbyte/s)</td>
<td>8.5(1)</td>
<td>12.8 (1)</td>
<td>12.8</td>
<td>34(1)</td>
<td>51.2</td>
</tr>
<tr>
<td>possible BW evolution</td>
<td>-</td>
<td>17 (2)</td>
<td>17 (3)</td>
<td>Not yet defined</td>
<td>Not yet defined</td>
</tr>
<tr>
<td>(Gbyte/s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max package density</td>
<td>4x4</td>
<td>4x4</td>
<td>1x4</td>
<td>Not yet defined</td>
<td>Not yet defined</td>
</tr>
<tr>
<td>(Gbit)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>power efficiency</td>
<td>78</td>
<td>67</td>
<td>42</td>
<td>Not yet defined</td>
<td>Not yet defined</td>
</tr>
<tr>
<td>(mW/Gbyte/s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>volume maturity</td>
<td>2011</td>
<td>2012</td>
<td>2013</td>
<td>Not yet defined</td>
<td>Not yet defined</td>
</tr>
<tr>
<td>relative memory cost</td>
<td>1</td>
<td>~1.1</td>
<td>~1.2</td>
<td>Not yet defined</td>
<td>Not yet defined</td>
</tr>
<tr>
<td>for equivalent density</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1) 32b dual channel configuration assumed
(2) LPDDR3E: clock from 800 to 1066MHz. Standardization at JEDEC in progress
(3) WideIO clock frequency from 200MHz to 266Mhz: already specified at JEDEC
(4) Estimates based on memory supplier survey (memory cost only)
Do we need all that bandwidth?

Example: Video Encode Directed Cache

Direct fetch from DDR
- Read data rate: 1.25 GB/s
- Short bursts (32B)
- 50% efficiency -> 2.5GB/s BW
- Latency: 2.5KB in 2μs

Fetch thru 2 x 96 lines buffer
- Read data rate: 0.25 GB/s
- Long bursts (4KB)
- 90% efficiency -> 0.28GB/s BW
- Latency: 60KB in 250μs
THERMAL - WIDE IO VS POP

**Observation:** POP thermal performance better than WideIO

- TSV requires silicon die to be reduced to 50-70um, which results in poor lateral heat distribution
- Thermally tightly coupled WideIO DRAM heats up much faster than in POP
- WideIO DRAM performance reduced at $T_j > 85^\circ C$ due to increased refresh cycle requirements
  
  $\Rightarrow$ Max performance peak limited by WideIO structure

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Time to reach memory limits (s)</th>
<th>Time to reach SOC limit (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>85°C</td>
<td>95°C</td>
</tr>
<tr>
<td>PoP</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>WideIO</td>
<td>0.08</td>
<td>0.21</td>
</tr>
</tbody>
</table>

- 4” smartphone mechanics, typical chipset
- Application of 10W SOC perf peak starting from 2W SOC steady state
Research for new materials and methods to reduce heat resistance

- Phase change materials (PCM)
  - Peak power “thermal buffering” close to heat source by implementing small pockets of PCM directly into the silicon or package

- Graphite based coatings on silicon or substrate
  - Reduction of lateral thermal resistance with graphene foils and coatings

- Polymers and carbon based nano tubes (CNT)
  - New thermal interface materials which can significantly reduce thermal resistance between die, package and casing

- Active cooling and heat extraction
  - Mechanical, piezo and MEMS based fans, liquid cooling and heat piping on system level
3D Test Challenges

- New TSV and stacking defect types to be covered
- Limited test access to individual die pins
- Additional test steps to be introduced

**TSV fault models**

**Yield loss**

- Pre-bond or wafer test
- Back grinding (‘TSV exposed’)
- Backside processing
- Thin wafer handling
- Die bonding
- Packaging (molding)

**Increasing Cost**

Assembly process

Post-bond or package test

vertical probing on micro-bumps ‘unexposed’ or ‘exposed’ TSV
But why did WideIO not take off yet?

- Business model complex compared to established POP solution
- LPDDR3E will reach same BW as WideIO in same production time frame at much lower cost - TSV, wafer backside processing and fine pitch Cu pillar assembly add significantly to product cost
- LPDDR4 will enable higher BW than WideIO at similar power levels and lower cost
- Memory BW requirements for given GPU and CPU performances may be lower than initially expected through improved memory hierarchy architectures and system cache strategies
- Thermal performance of WideIO not on par with external LPDDR solutions
Was WideIO a bad dream?

Clearly No!
Many products will benefit from derived technology
- Memory footprint reduction through 3D TSV stacking – on the market
- WideIO did drive TSV and backside processing technology to production maturity
- Many useful applications outside smartphone such as FPGA, server market, gaming consoles with WideIO like 2.5D and 3D solutions

But for the smartphone?
- WideIO technology not yet on radar – LPDDR3 and LPDDR4 will take this spot
- Increased power density with 3D stacking limits thermal performance
Any Future for 3D in Smartphones?

3D Image: IBM/3M
Cost per Transistor - Evolution
3D Smart Partitioning – High Potential

**Lower Cost** (at the condition of high TSV and assembly yield)
- IP design in best suited process (analog/high voltage/high perf digital)
- Reduction of “high cost” die area in < 20nm process

**Modularity and TTM**
- Mixed-signal IP reuse for different flavors of digital performances and high speed cmos process node

**Power**
- Overall Leakage power can be lowered by removing non critical circuitry from 20nm and below process nodes. Dynamic power to be monitored due to higher voltage and RC.

**Formfactor reduction**
- Analog/mixed-signal and logic functions in single package
- Reduced package thickness without POP (DRAM in MCP with eMMC)

**Manufacturing capacity**
- A smaller digital die size will help in alleviating capacity issues seen in advanced cmos process nodes

**What is missing?**
- Thermal performance to be improved by thermal aware design of silicon, package and smartphone mechanics
- DFT and test bricks to be developed for pin count reduction
- 3D pathfinding and design tools have to mature
ST-Ericsson 3D Logic on AMS Prototype

**Backside Cu-Pillars**
- Diameter 70 μm
- Height: 80 μm
- Pitch 180 μm
- Number ~100

**F2F Cu µPillars**
- Diameter 25 μm
- Pitch 60 μm
- Height 30-40 μm
- Number 480

**RDL**
- Width 20 μm
- Pitch 40 μm
- Height 7 μm

**TSV middle AR 8:1**
- Diameter 10 μm
- Pitch 40 μm
- Height 80 μm
- Number ~100

**VFBGA Package**
- Ball Matrix 9x9
- Size 4x4mm
- Ball Pitch 0.4mm
Conclusion

- 3D TSV and wafer backside process technology is ready for mass production
- WideIO technology not yet competitive fit for main stream Smartphones – LPDDR3 and LPDDR4 will take this spot
- 3D smart partitioning options could allow higher performances, lower power, smaller formfactors and faster time to market cycles – and due to increased cost/transistor on future silicon technology nodes all that at lower cost!
- Increased power density with 3D does not help thermal performance ➔ design and technology need to take care of that
Thank you

Contributors:
Stephane Lecomte, Gianni Qualizza