Modeling and Analysis of the System Bus on the SoC Platform

Eun Ju Choi , Young Sin Cho , Kyoung Rok Cho

Communication Circuits and System Design Lab.
Chungbuk National University
12, Gaeshin-dong, Cheongju-city, Rep. of Korea

Contents

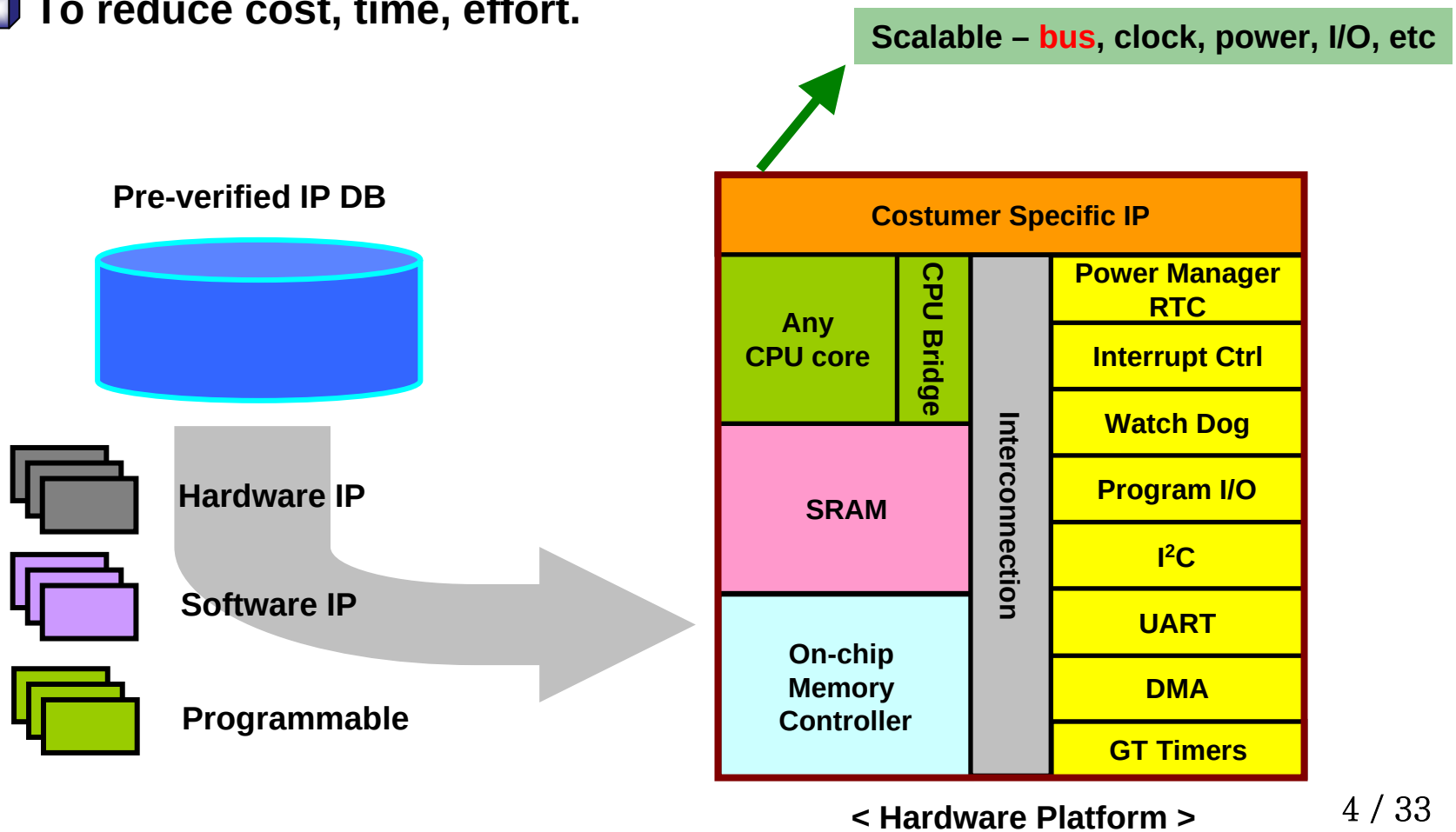
- SoC platform & shared-bus
- Design issue
- Proposed latency model
- Simulation & result
- Conclusions

Contents

- SoC platform & shared-bus
- Design issue
- Proposed latency model
- Simulation & result
- Conclusions

Platform based SoC design

- Design methodology, Verification environment, ...etc
- IP reuse
- To reduce cost, time, effort.



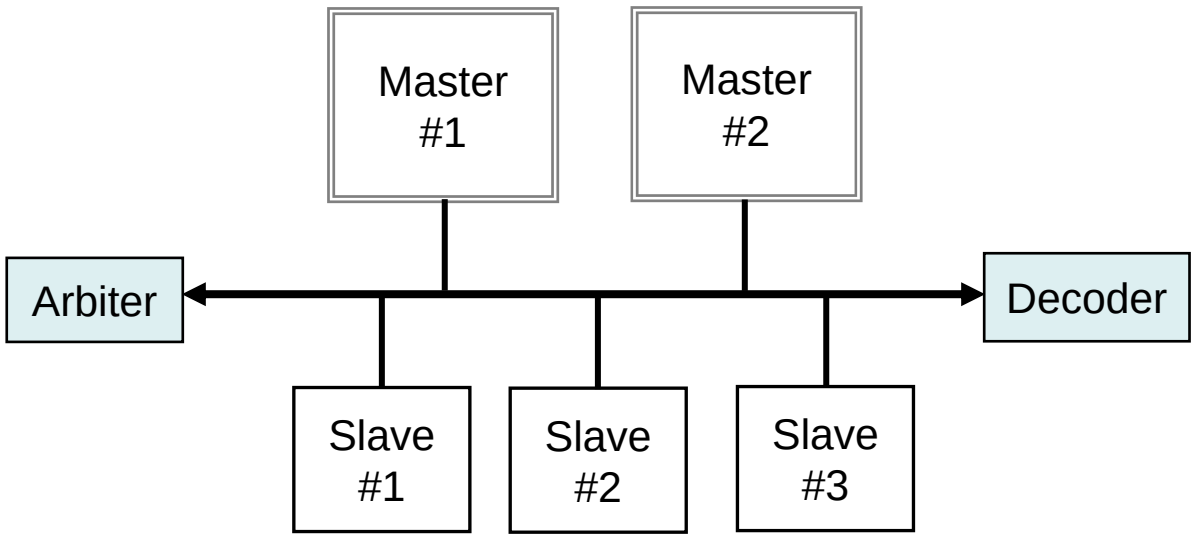
< Hardware Platform >

Shared bus for interconnection

- Simple architecture
- Totally reusable
- Lower speed than resident cores
- Performance depends on an arbitration
- Efficient solution in the current design flows

Single-layer bus

- Number of IPs on a bus
- Only one master grabs a ownership at a time



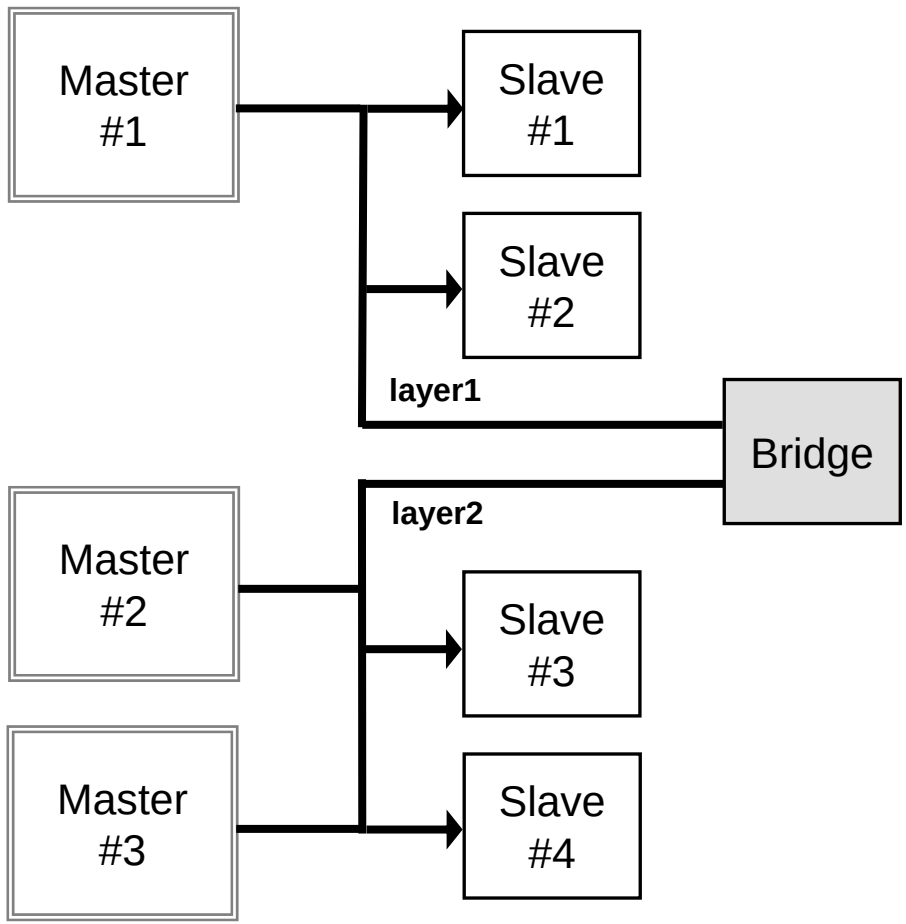
Multi-layer bus

Multi-path between master and slave

- Each layer can be simple
- Increase bandwidth

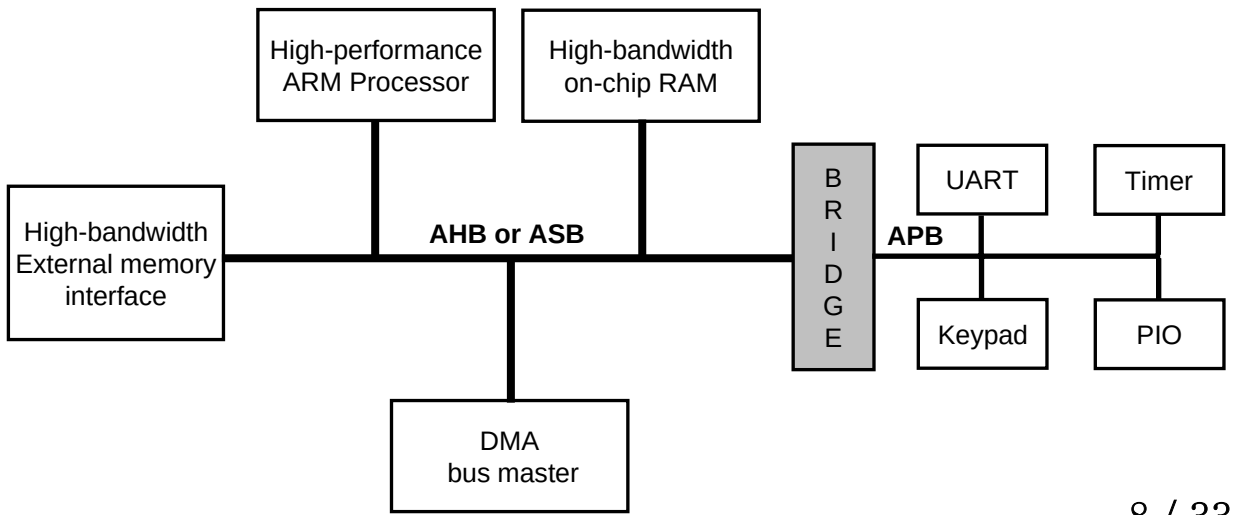
Weak point

- Hardware resource
- Power
- Design complexity



AMBA: popular standard for SoC

- **Open standard, on-chip bus specification by ARM**
- **AHB, ASB, APB, AXI**
- **Support multi-layer architecture**
- **Advanced High-performance Bus**
  - Pipelined operation
  - Non-tristate implementation
  - Multiple bus masters
  - Burst transfers
  - Split transactions



Contents

- SoC platform & shared-bus
- Design issue
- Proposed latency model
- Simulation & result
- Conclusions

Design issue

- **How can you estimate a throughput from the present shared-bus before actual design?**
  - Number of masters
  - Number of layers
  - Transfer properties

Contents

- SoC platform & shared-bus
- Design issue
- **Proposed latency model**
- Simulation & result
- Conclusions

IS(Ideal-Slave) latency model

- A slave has no latency to response to a master.
- L_{Bus} – Latency of shared-bus
- $L_{Complex\_Bus}$ – Latency of shared-bus including multiple master
- L_{Single_Layer} – Latency of single-layer bus
- L_{Multi_Layer} – Latency of multi-layer bus

Parameter	Description
N_M	Number of masters
N_L	Number of layers
N_D	Number of data
S	Single transfer ratio
B	Burst size
U	Usage of bus
A	Active bridge ratio

Modeling (1/8)

Latency for shared bus

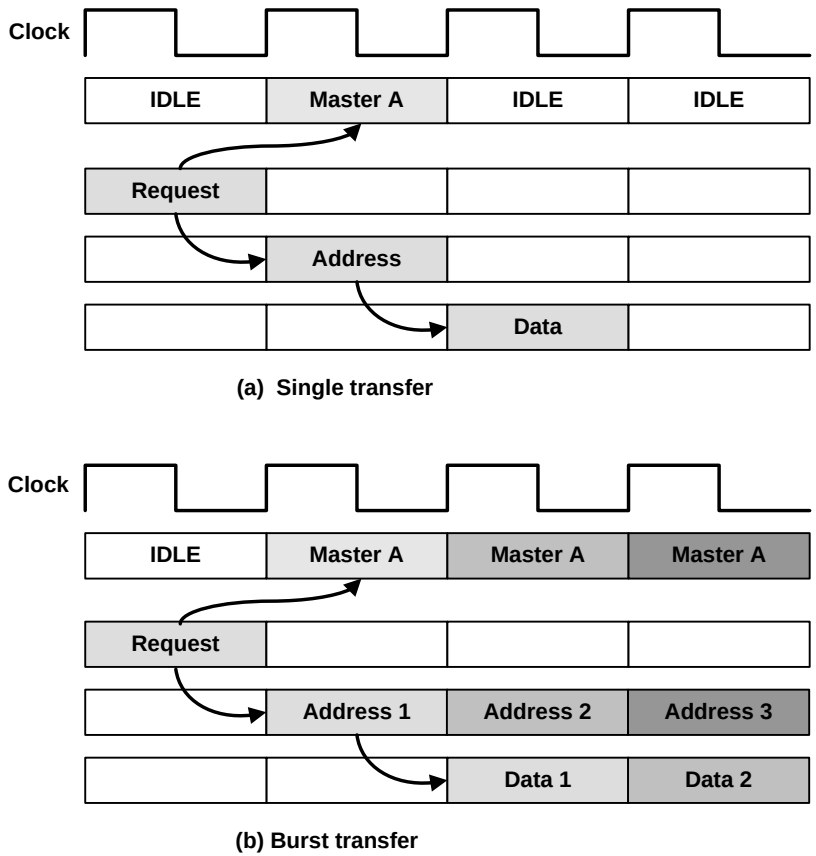
$$L_{Bus} = 1 + N_D \quad (1)$$

,where N_D is number of data and
'1' indicate the request cycle
getting approval from a bus arbiter

Consider an effect of transfer mode and pipelined architecture

$$L_{Bus} = 3 \cdot N_D \cdot S + \left\{ Ceiling\left(\frac{N_D \cdot (1-S)}{B} \right) + N_D \cdot (1-S) \right\} \quad (2)$$

,where S($0 \le S \le 1$) is a ratio of single transfer
and B is a burst data size



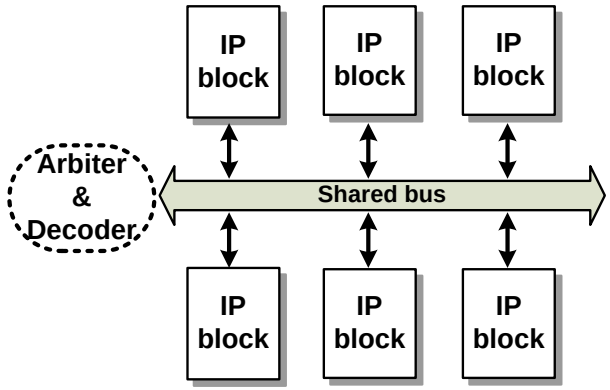
<An example of two transfer type in shared bus>

Modeling (2 / 8)

Latency for single-layer shared bus

$$L_{\text{Single\_Layer}} = N_{\text{M}} \cdot L_{\text{Bus}} \quad \textbf{(3)}$$

,where N_M is number of masters



<The general single-layer structure>

- All master Ips are connected to the single layer bus and are controlled by an arbiter
- This one master IP latency occupies the shared bus

Modeling (3/8)

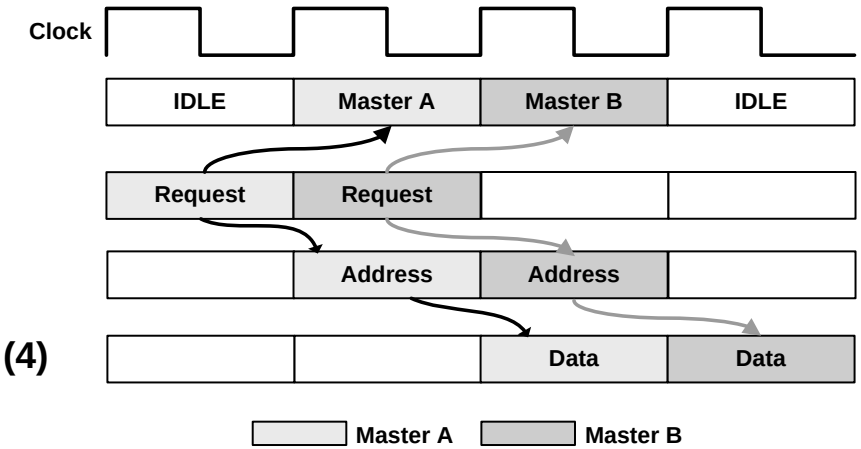
Latency for single-layer shared bus

$$L_{\text{Single\_Layer}} = N_M \cdot L_{\text{Bus}}$$

Pipeline effect by several masters

$$L_{\text{Bus\_Complex}} = (3 - 2 \cdot U) \cdot N_D \cdot S + \left\{ Ceiling\left(\frac{N_D \cdot (1 - S)}{B} \right) + N_D \cdot (1 - S) \right\} \quad (4)$$

,where U(0≤U≤1) is usage of bus which is a probability of continuing single transfer

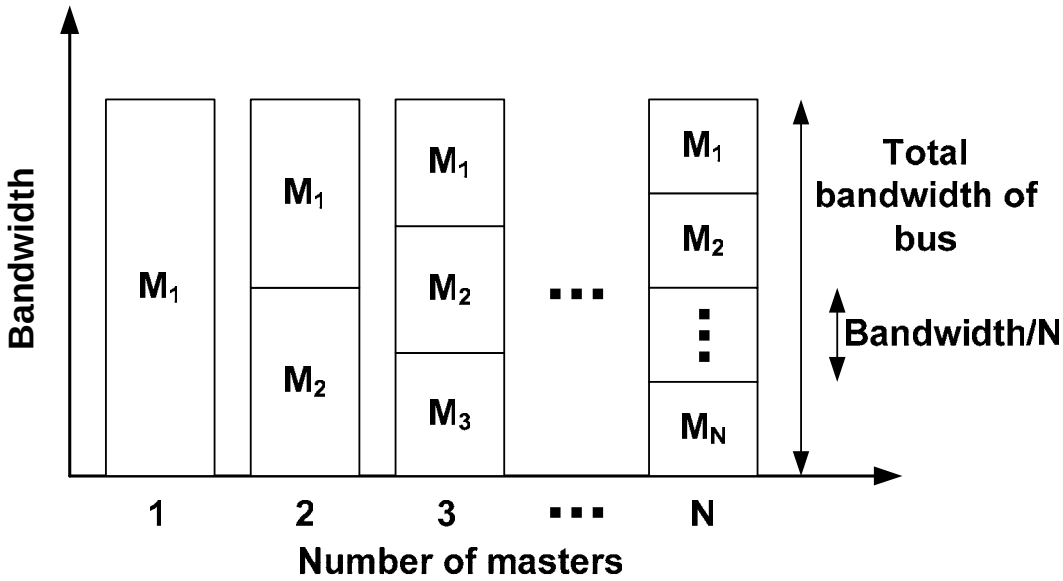


<An example that shows two master transfer the data continuously>

- If two or more master Ips are connected to the bus, address and data cycle access the bus simultaneously.
- the effect of the pipeline architecture depends on the bus usage

Modeling (4/8)

A partition of bandwidth according to number of masters



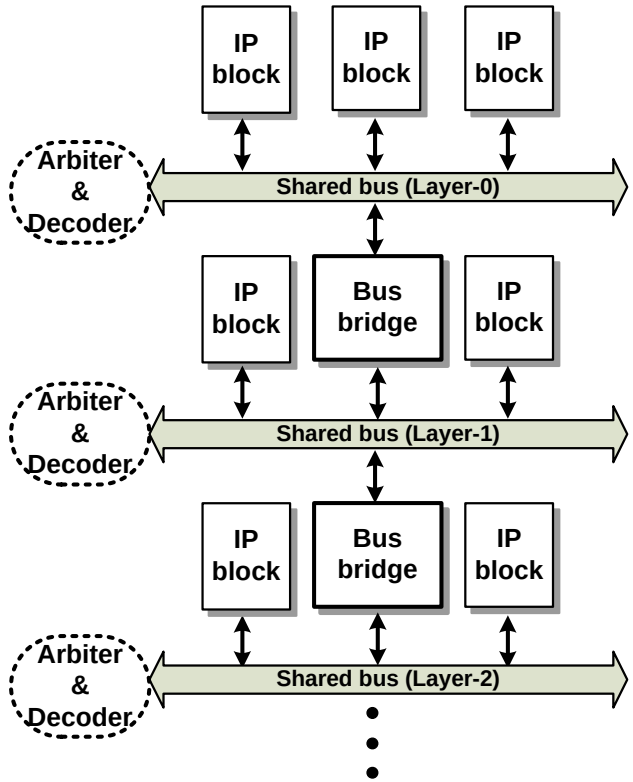
- Iincrease in bus usage means increase the probability of the continuing data processing.
- Total bandwidth is equal to total bandwidth of each master IP.

Modeling (4/8)

Latency for multi-layer shared bus

$$L_{\text{Multi\_Layer}} = \frac{N_M}{N_L} \cdot L_{\text{Bus\_Complex}} \cdot (1 - A) + \alpha \cdot A \quad \textbf{(5)}$$

,where A(0≤U≤1) is a probability making a data path through a bridge module. Bridge factor, α, is latency overhead caused by bridge module. N_L is number of layers.

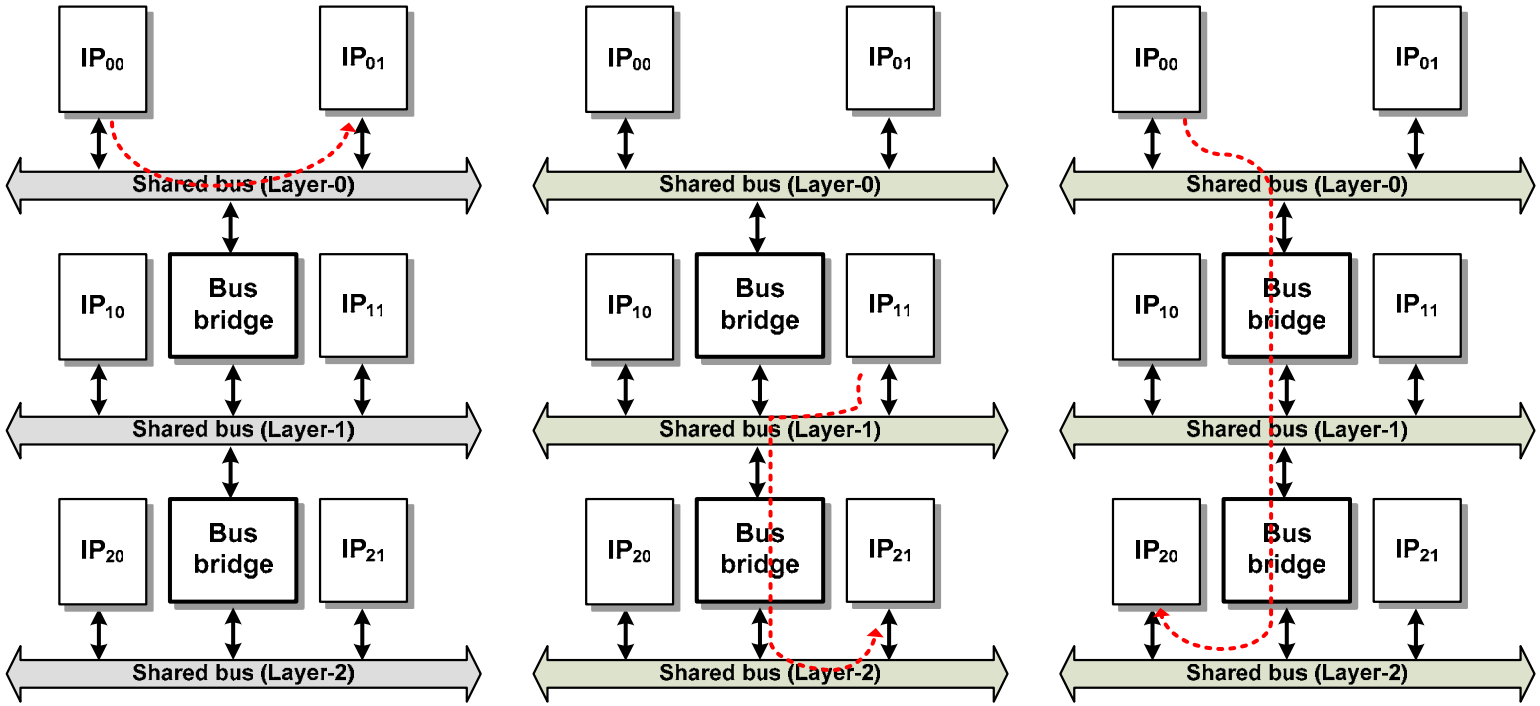


<The multi-layer structure with bridge module>

Modeling (5/7)

Latency for multi-layer shared bus

-The latency is increased due to bridge modules.
-If two layers are connected through a bridge module, one IP should be a master of both layers.
-It cannot offer entire bandwidth of two layers.



<The configuration of data path with 3-layer bus>

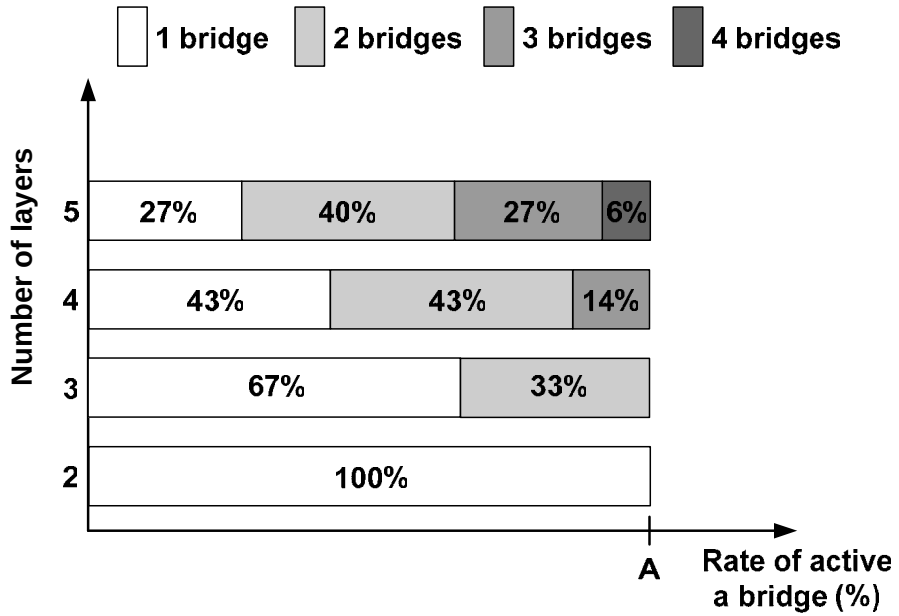
Modeling (6 / 7)

Latency for multi-layer shared bus

Data paths which use same number of bridge modules

$$\alpha = \sum_{i=1}^{N_L-1} \left(\frac{C_i^{N_L-1}}{\sum_{j=1}^{N_L-1} C_j^{N_L-1}} \cdot \frac{N_M}{N_L - i} \cdot L_{\text{Bus\_Complex}} \right) \quad (6)$$

Total number of data paths using bridges which can may appeared on multi-layer bus



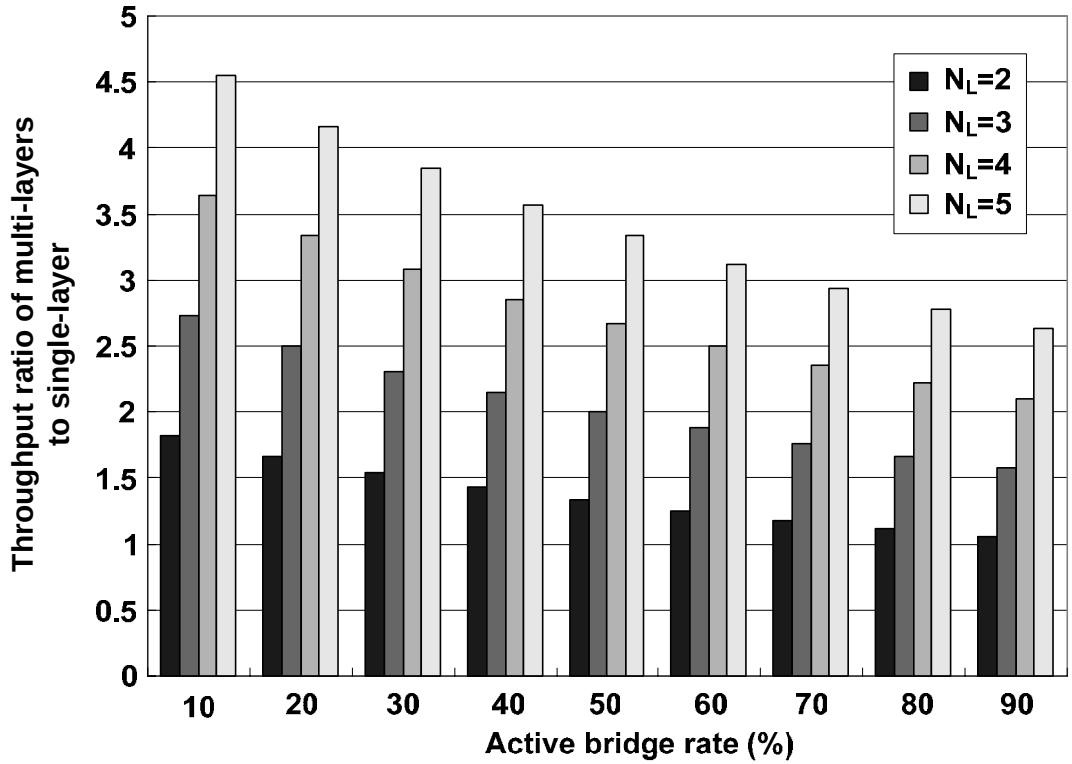
<The distribution of probability A by combination of data path>

-Bridge factor is the latency overhead by the using bridge.

-Bridge factor depends on the number of data path.

Modeling (7/7)

Throughtput ratio of multi-layers to single-layer



- the throughput is inversely proportional to A and proprtional to number of layer

Contents

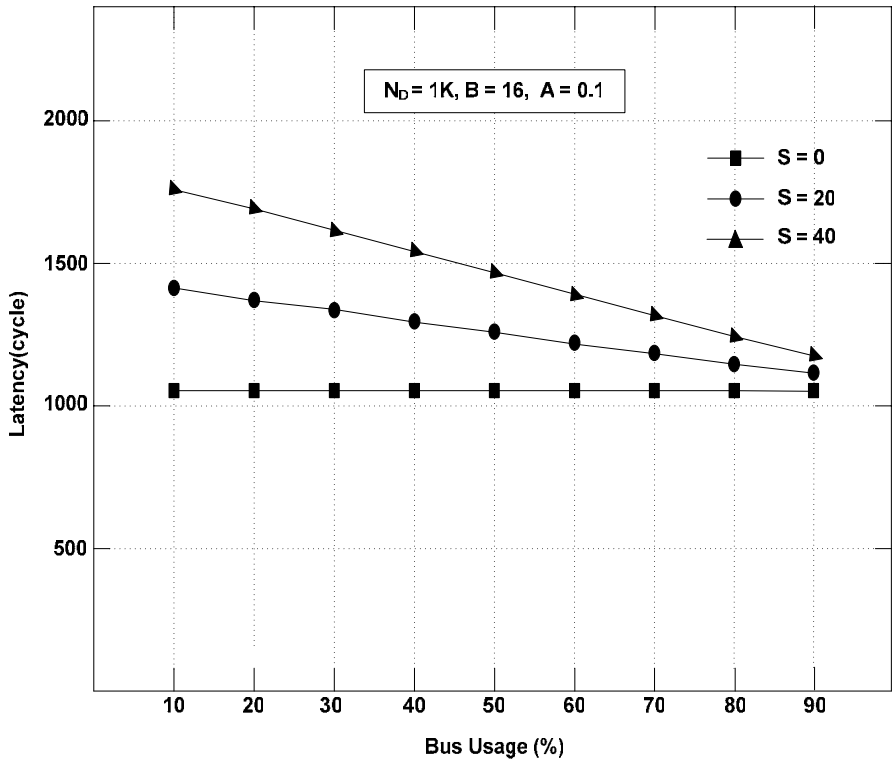
- SoC platform & shared-bus
- Design issue
- Proposed latency model
- **Simulation & result**
- Conclusions

Simulation & Result (1/9)

Result of latency model for shared bus (N_D = 1000)

-Case1) If same U (bus usage), S ↑, latency ↑ .
-Case2) If same S (single transfer rate), U ↑, latency ↓.

If the system which has high U, it doesn't have to much consider about S.



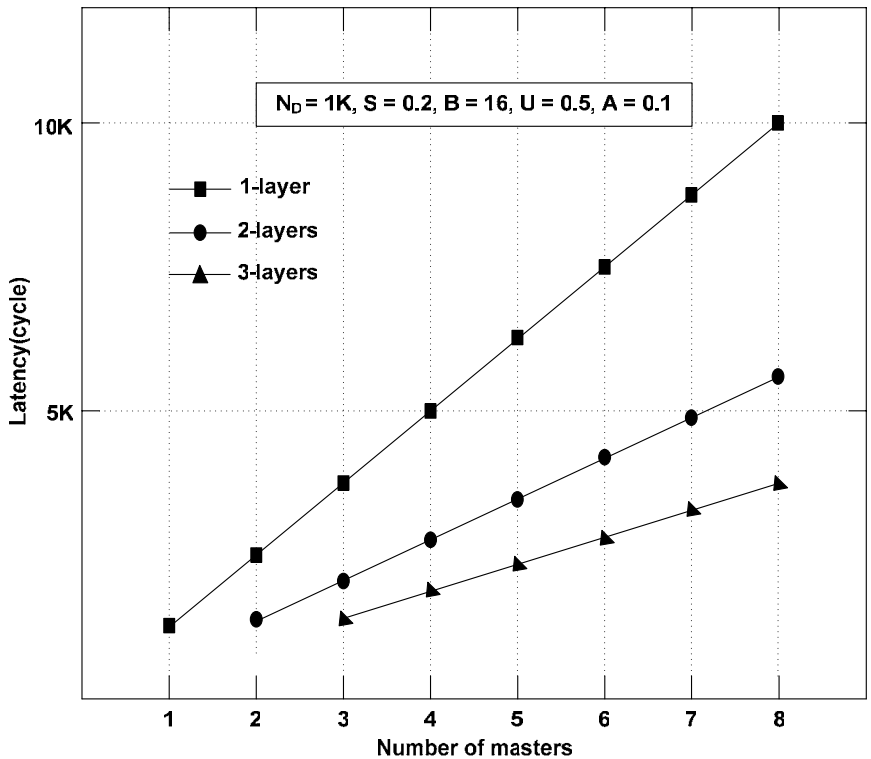
<The variation of latency according to increase of bus usage and single transfer>

Simulation & Result (2/9)

Result of latency model for shared bus

-The latency is reduced when compare multi-layers with single-layer.
(2-layers 45%↓, 3-layers 63%↓)

-The condition (S, B, U, A) depends on characteristic of SoC.



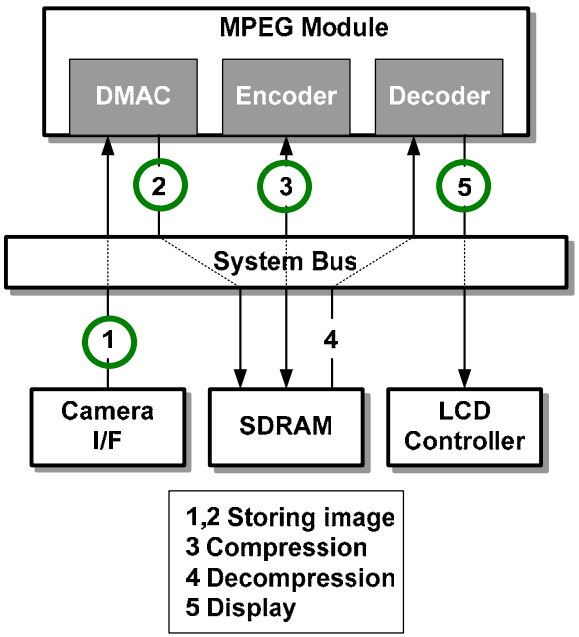
<The latency difference of each shared bus by parameter number of master IPs>

Simulation & Result (3/9)

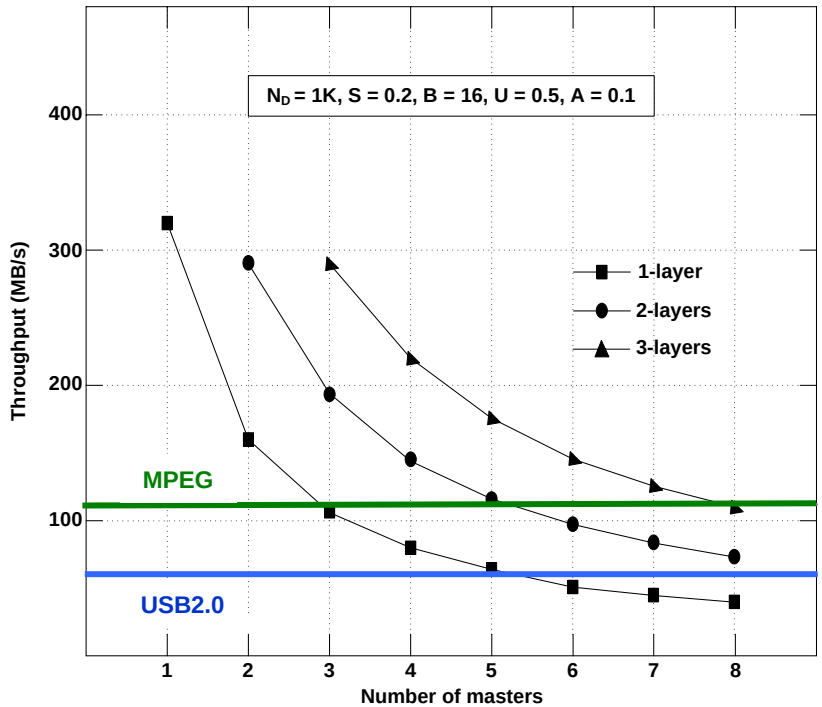
Result of latency model for multi-layer bus

Condition: VGA(640x480), 30frame/s
Job requirement: 27.65[MB/s] = 640X480X3[Byte]X30[frame]
Total requirement: 110.6[MB/s] = 27.65[MB/s]X4

Maximum throughput of USB2.0 is 480[Mb/s] (= 60[MB/s])



<The simple example of image processing by MPEG>

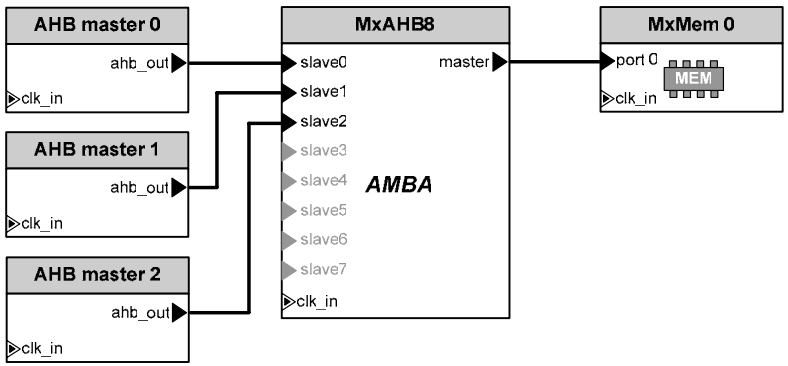


<The expected throughput of each shared bus according to increase number of masters>

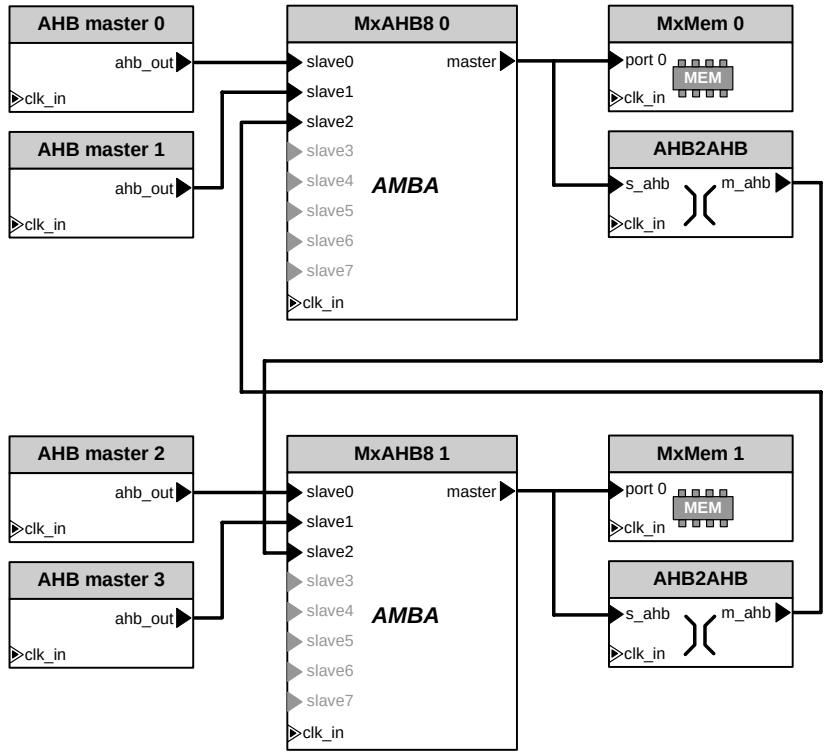
Simulation & Result (4/9)

We use MaxSim for a comparison of simulation results

- **Modeling & simulation tools for SoC designs**
- **Cycle-accurate models**



(a) single-layer architecture



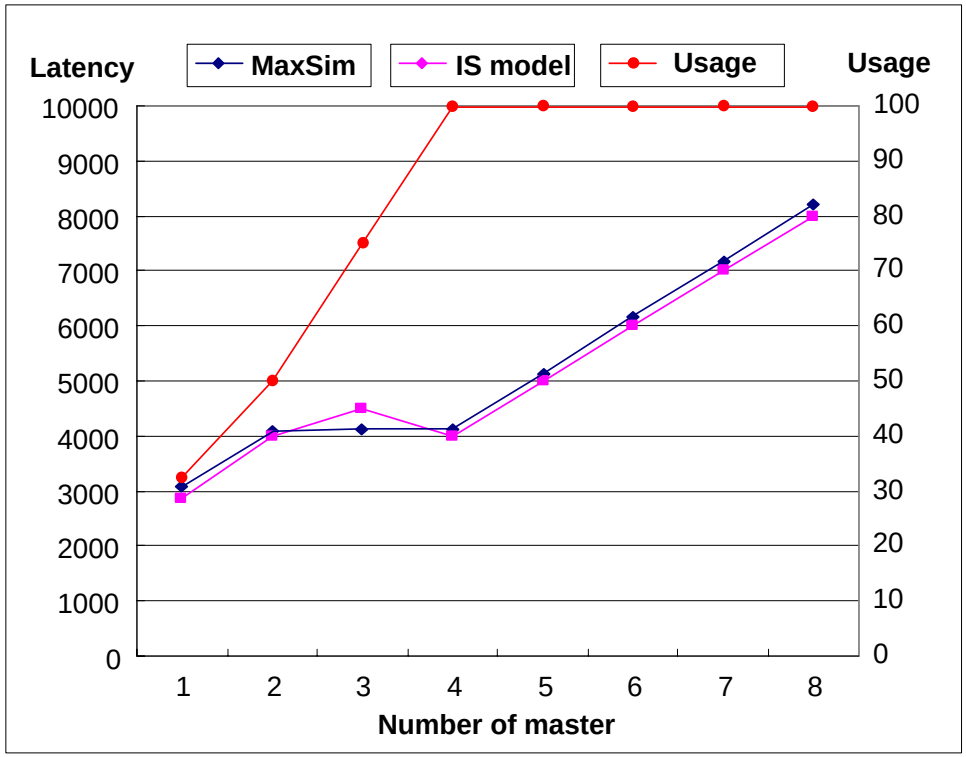
(b) multi-layer architecture

<The example of SoC on the MaxSim with single-layer and multi-layer>

Simulation & Result (5/9)

Single-layer results

- N_D = 1000
- 96% accuracy

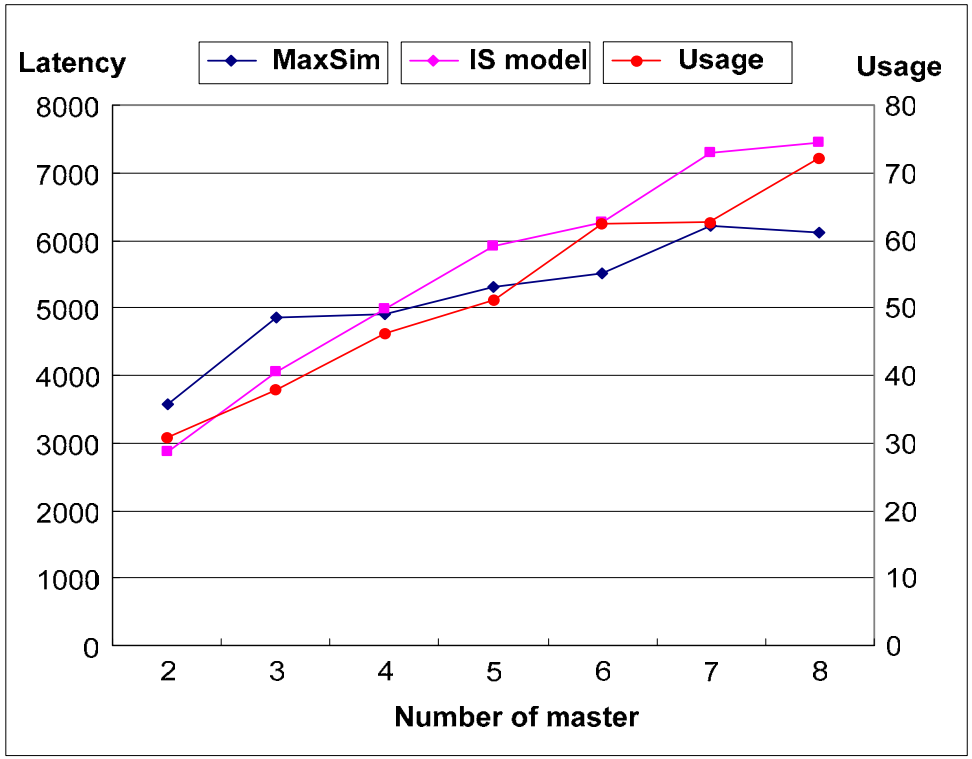


<The comparison of the results between IS model and MaxSim>

Simulation & Result (6/9)

2-layer results

- N_D = 1000, A = 20%
- 85% accuracy

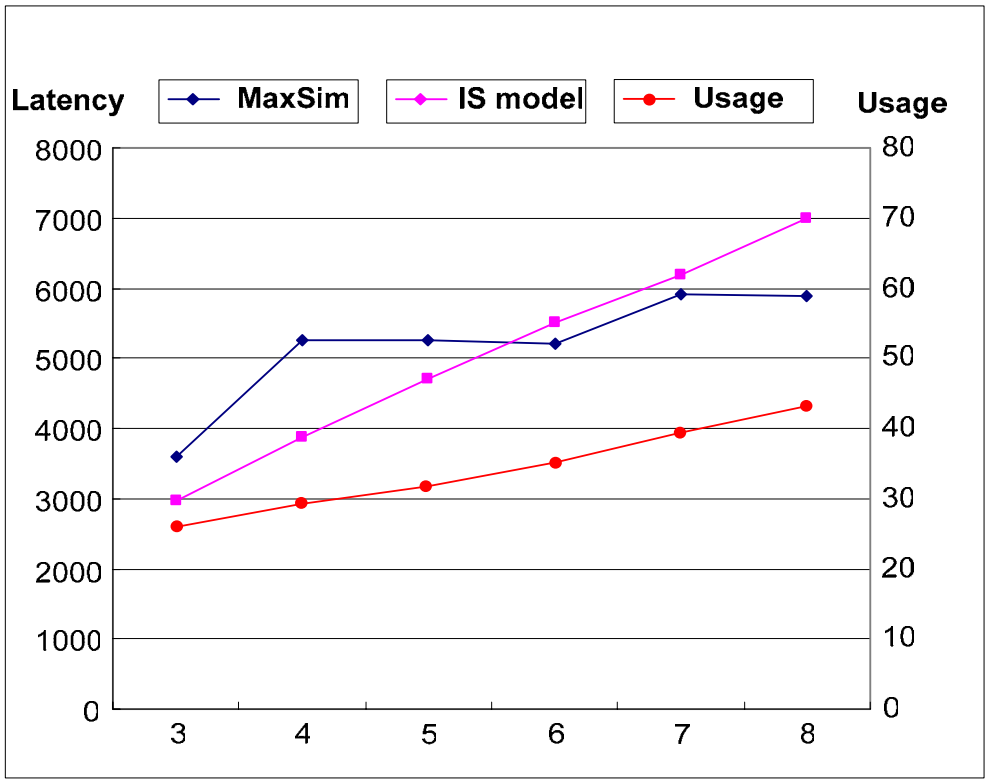


<The comparison of the results between IS model and MaxSim>

Simulation & Result (7/9)

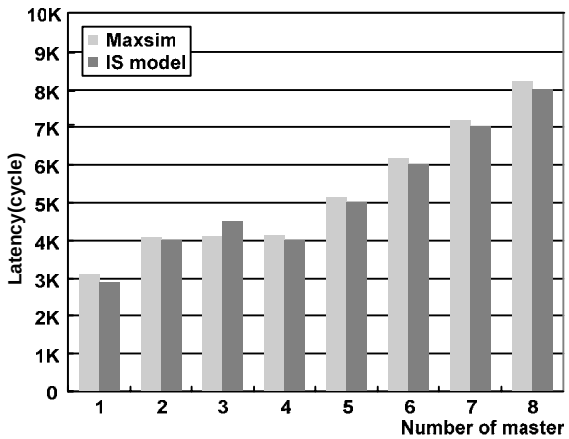
3-layer results

- N_D = 1000, A = 20%
- 85% accuracy



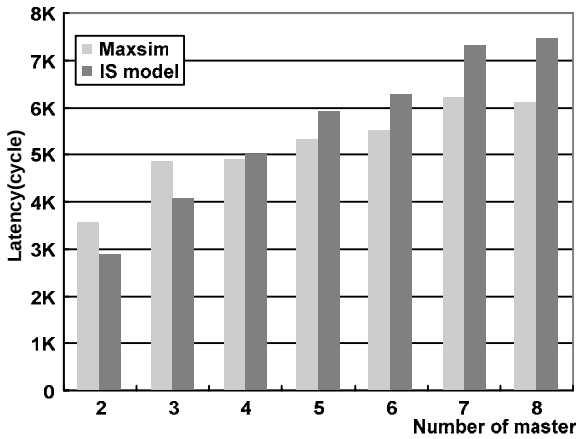
<The comparison of the results between IS model and MaxSim>

Simulation & Result (8/9)



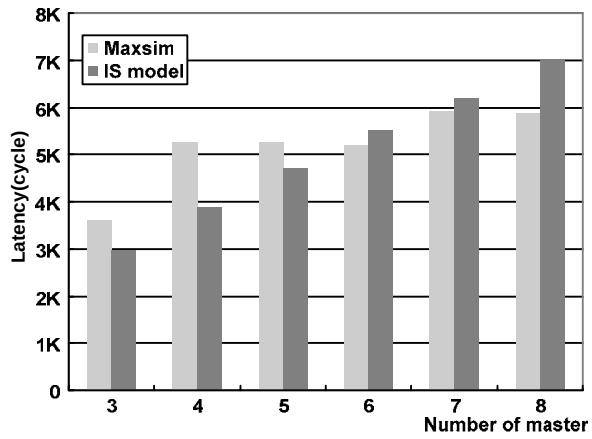
Number of master	Latency of IS model	Latency of Maxsim	Accuracy (%)
1	2856	3076	92.9
2	4000	4100	97.6
3	4500	4102	90.4
4	4000	4108	97.4
5	5000	5132	97.4
6	6000	6156	97.5
7	7000	7180	97.5
8	8000	8204	97.5
			Average = 96

(a) single-layer



Number of master	Latency of IS model	Latency of Maxsim	Accuracy (%)
2	2856	3576	79.9
3	4050	4856	83.4
4	4992	4908	98.3
5	5910	5304	88.6
6	6264	5504	86.2
7	7308	6208	82.3
8	7440	6108	78.8
			Average = 85.4

(b) 2-layer

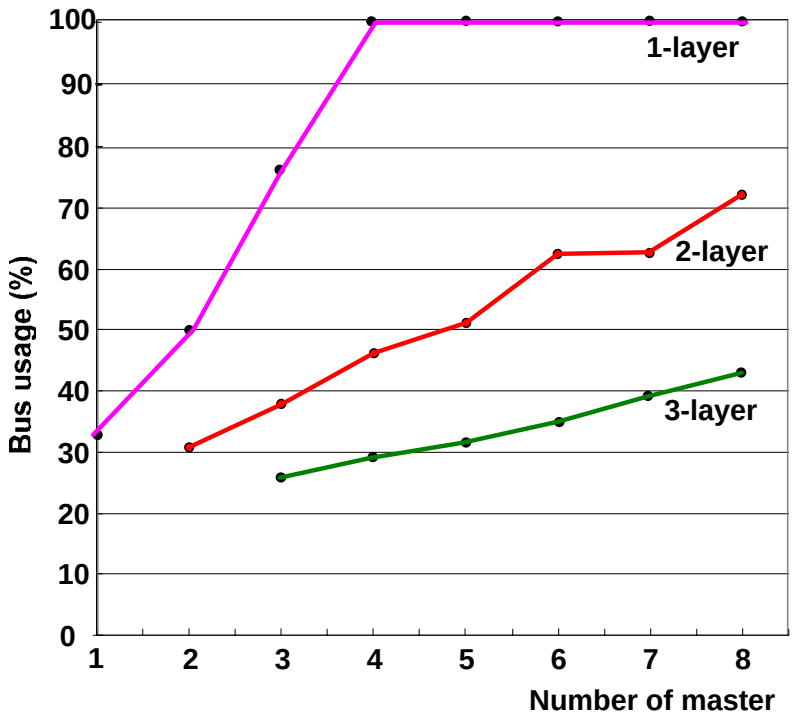


Number of master	Latency of IS model	Latency of Maxsim	Accuracy (%)
3	2966	3604	82.3
4	3862	5267	73.3
5	4708	5266	89.4
6	5510	5204	94.1
7	6188	5904	95.2
8	6990	5888	81.3
			Average = 85.9

(c) 3-layer

- The accuracy of the proposed latency model are over 96% for single-layer and 85% for multiple layers.

Simulation & Result (9/9)



- The bus usage indicates an average utilization of the bus as function of number of master IPs

Contents

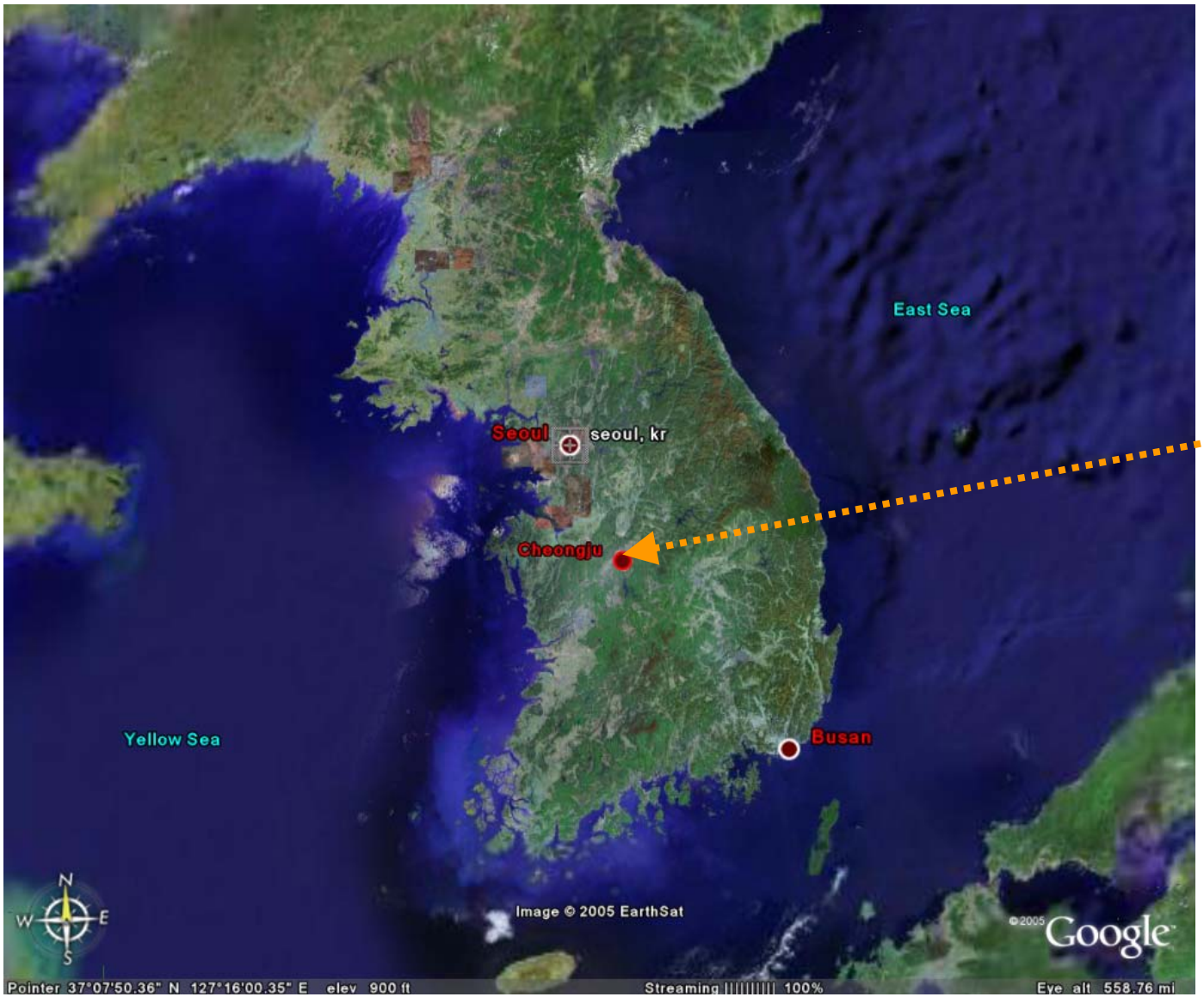
- SoC platform & shared-bus
- Design issue
- Proposed latency model
- Simulation & result
- **Conclusions**

Conclusions

- We propose a latency model (IS model) which to estimate a performance of system bus before actual design.

- Simulation & result
  - Analyze the parameters of shared bus latency
  - Analyze number of masters affecting to bus throughput
  - Find out an appropriate number of layers on specific SoCs
  - Compare the results with that of MaxSim

Thank you !



Chungbuk Nat'l Univ.