A Hierarchical Management Strategy Based on Folded Torus-Like NoC for Dark Silicon Many-Core Systems

Speaker: Lei Yang
Weichen Liu
Weiwen Jiang
Mengquan Li
Juan Yi and Edwin H.M Sha

Chongqing University

CACS: Center of Advanced Computing Systems
cacs.cqu.edu.cn

leiyang@cqu.edu.cn

Jan. 28. 2016

Content

Dark Silicon has come……

Utilization wall:

With each successive process generation, the percentage of a chip that can actively switch drops exponentially due to power constraints.

- **System scaling theory**
- **Experimental result**
- **What we observed everyday**

Utilization wall:

With each successive process generation, the percentage of a chip that can actively switch drops exponentially due to power constraints.

- **System scaling theory**
  - Slow voltage/capacitance scaling
  - Non-balanced transistor/power budget scaling
  - System exponentially scaling
- **Experimental result**
- **What we observed everyday**

Traditional Scaling	Δ Quantity	S^2
	Δ Frequency	S
	Δ Capacitance	$1/S$
	V_{DD}^2	$1/S^2$
	→ Δ Power = Δ $QFCV^2$	1
	→ Δ Utilization = 1/Power	1

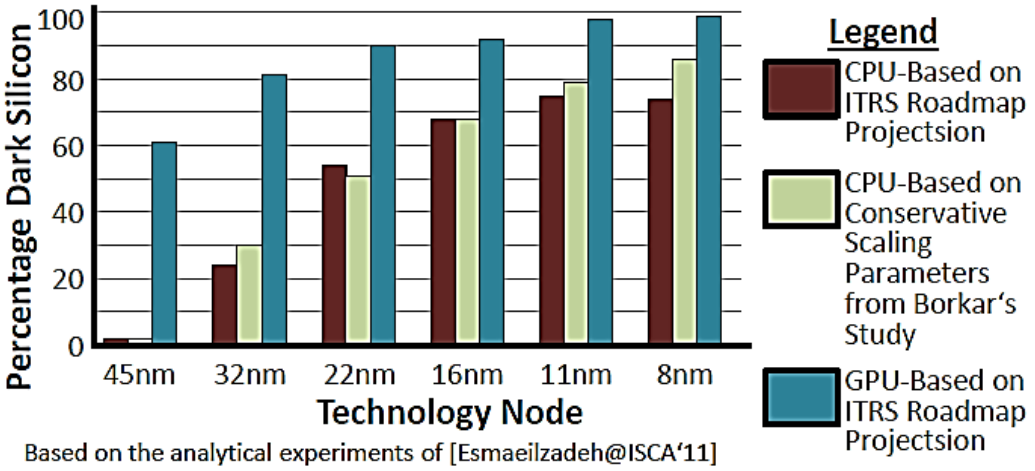
Leakage-limited Scaling	Δ Quantity	S^2
	Δ Frequency	S
	Δ Capacitance	$1/S$
	V_{DD}^2	1
	→ Δ Power = Δ $QFCV^2$	S^2
	→ Δ Utilization = 1/Power	$1/S^2$

Utilization wall:

With each successive process generation, the percentage of a chip that can actively switch drops exponentially due to power constraints.

- **System scaling theory**
- **Experimental result**
  - Thermal limited system
  - Cooling ability limited systems
- **What we observed everyday**

Dark silicon trends for different technology nodes



Based on the analytical experiments of [Esmaeilzadeh@ISCA'11]

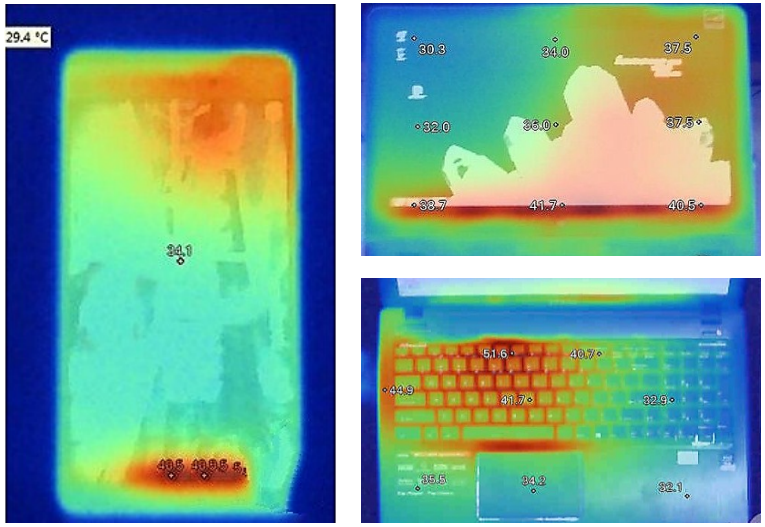
8-nm node:
>50% of the chip area will be dark

[Ref] M. Shafique, etc. ***The eda challenges in the dark silicon era: Temperature, reliability, and variability perspectives***. In Proceedings of the 51st Annual De-sign Automation Conference, DAC '14, pages 185:1–185:6, New York, NY, USA, 2014. ACM.

Utilization wall:

With each successive process generation, the percentage of a chip that can actively switch drops exponentially due to power constraints.

- **System scaling theory**
- **Experimental result**



- **What we observed everyday**
  - Battery limited system
  - Increasing power density
  - Chip hotspot

What is dark silicon ?

avoid overheating

cannot be powered-on all at the nominal operating voltage for a given TDP constraint

fraction of cores should be ‘Dark’ or ‘Dim’

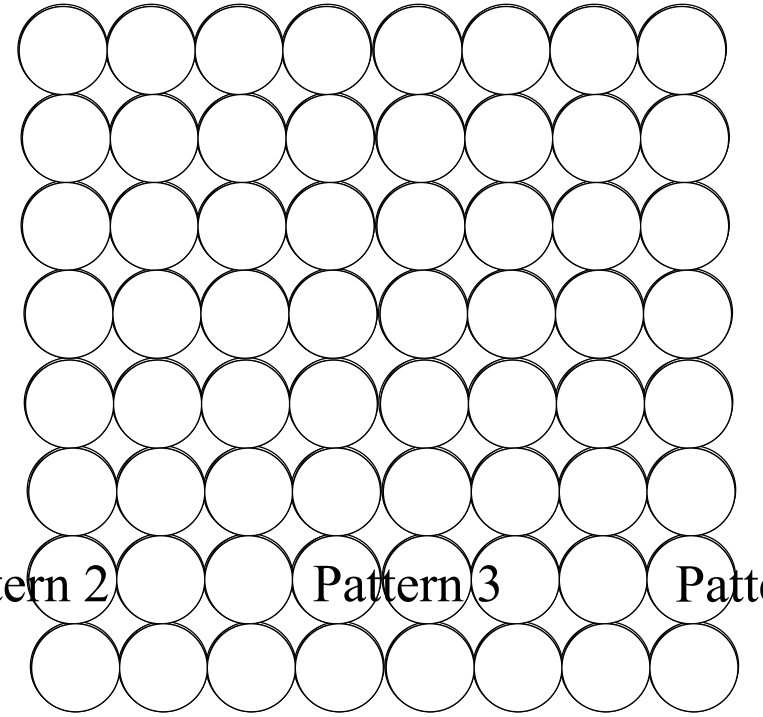
How to deal with dark silicon?

Leverage dark silicon to fight the utilization wall

When & Which

Motivation

Dark Silicon is 50%



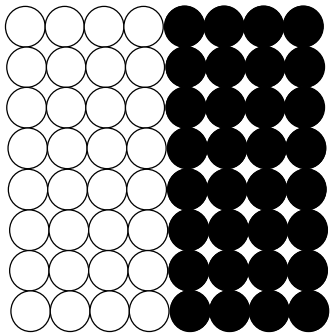
Pattern 1 Pattern 2 Pattern 3 Pattern 4 Random

8*8 Alpha 21264 core array

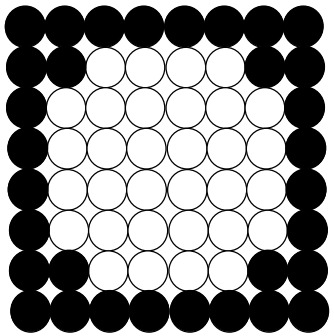
Motivation

Dark Silicon is 50%

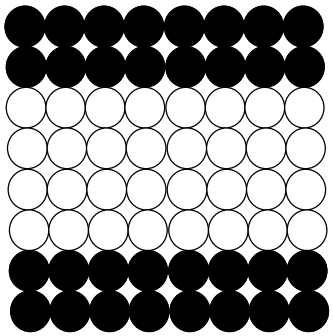
8*8 Alpha 21264 core array



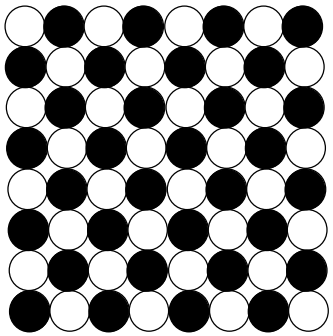
Pattern 1



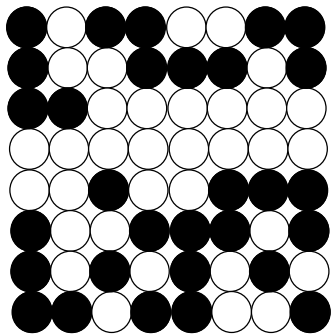
Pattern 2



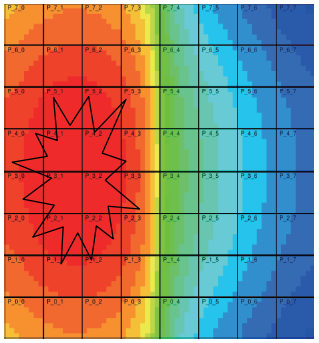
Pattern 3

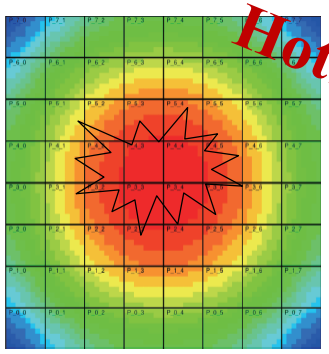


Pattern 4

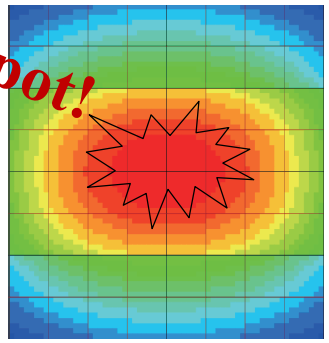


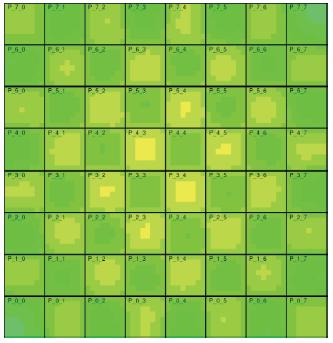
Random

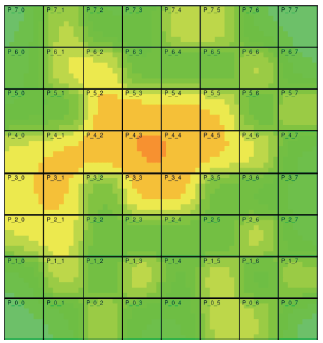




Hotspot!

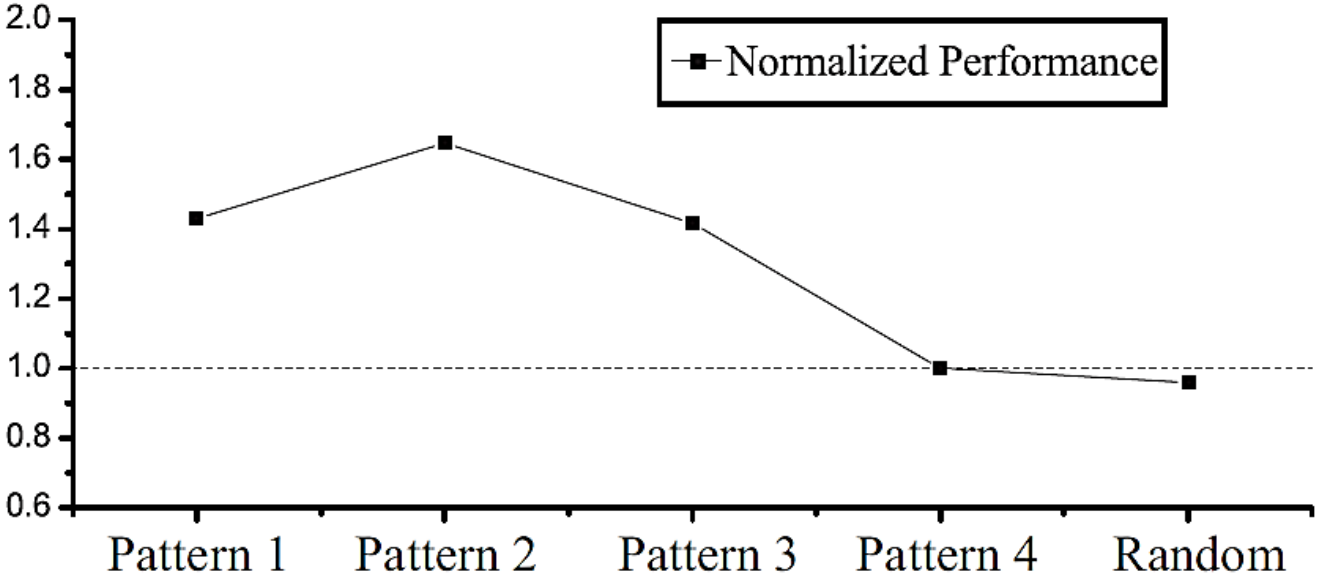






VS.1 Peak Temperature：

80 ℃ 82.17 ℃ 82.15 ℃ 72.81 ℃ 75.67 ℃

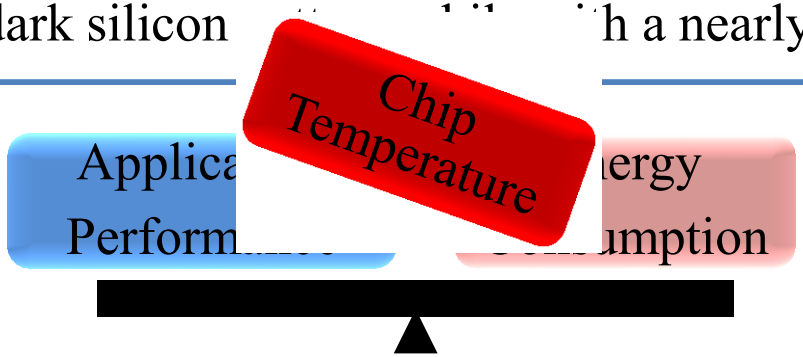


VS.2 Application Performance: **Pattern 4 is 37.4% worse**

Contradictions appeared:

-- Pattern 1, 2, 3 have better performance than 4

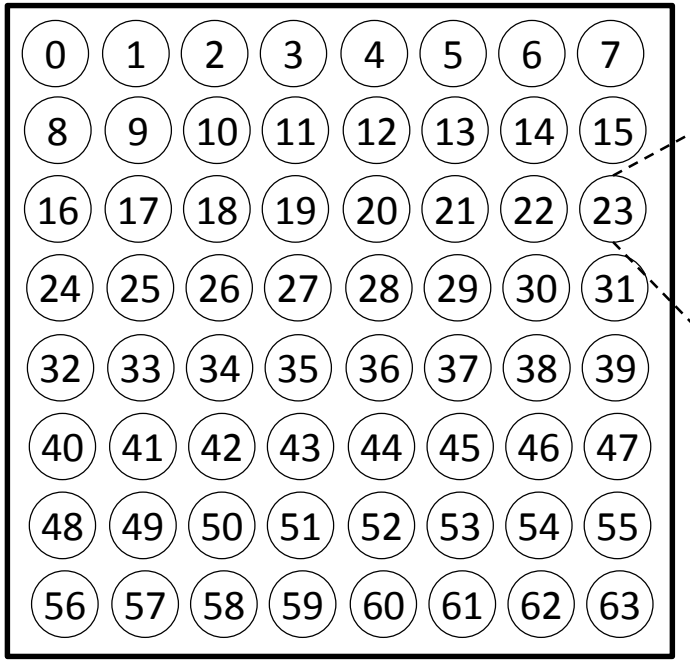
-- Pattern 4 is the best dark silicon [illegible] a nearly worst performance



Goals:

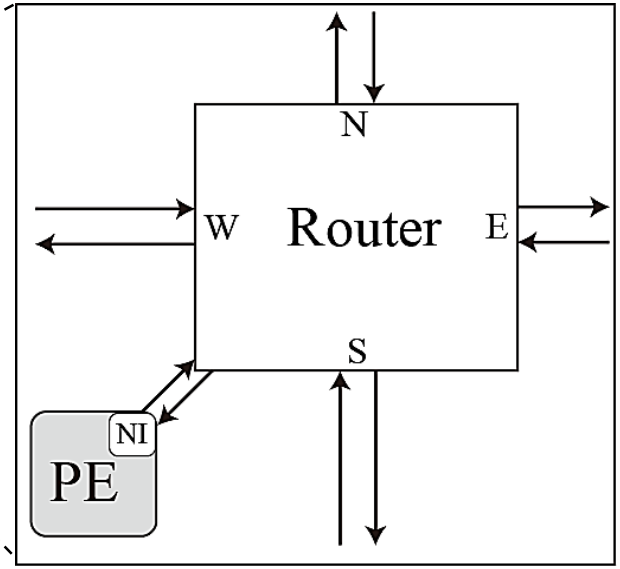
- Reduce long-distance communication and Achieve high performance
- Guarantee low power consumption and safe temperature

Physical arrangement on the chip



8*8 Alpha 21264 core array

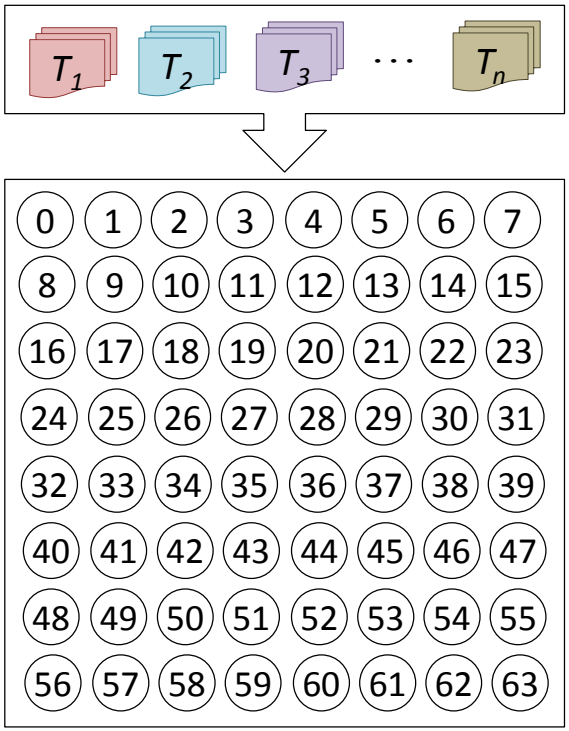
Logical Interconnection



Goals:

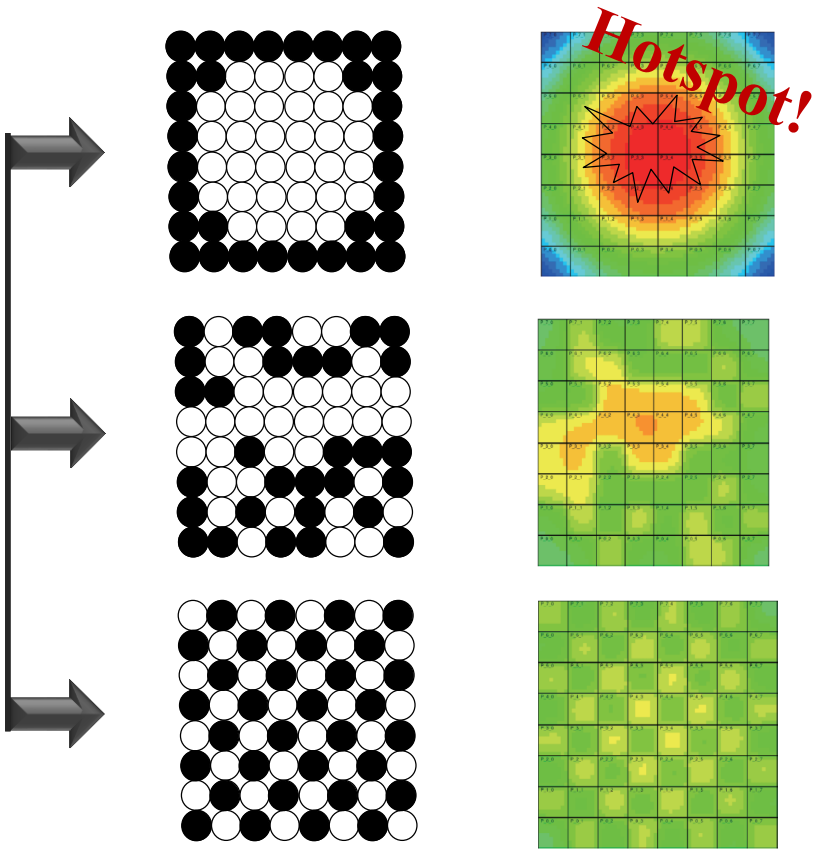
- Reduce long-distance communication and Achieve high performance
- Guarantee low power consumption and safe temperature

Physical arrangement on the chip



8*8 Alpha 21264 core array

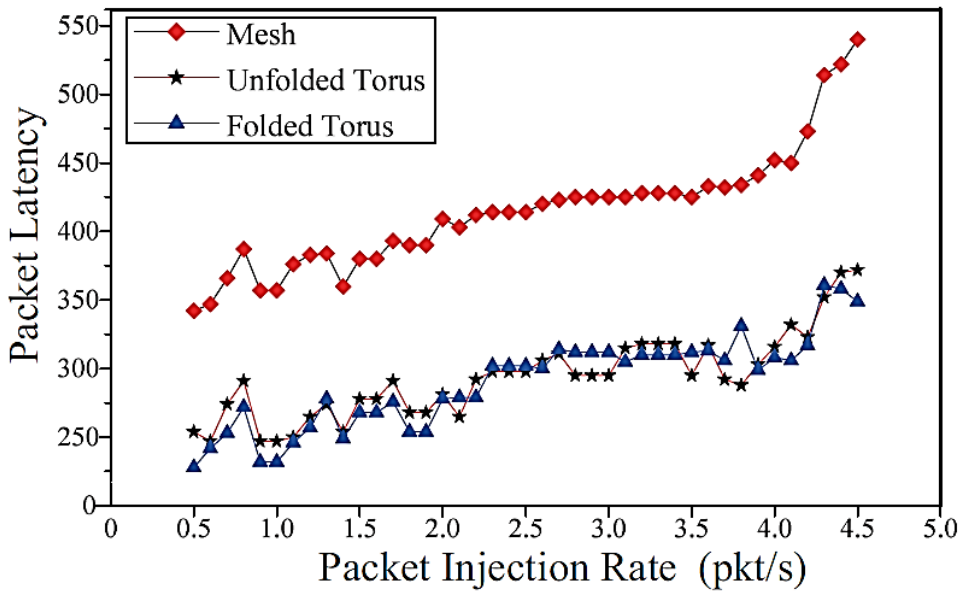
Logical Interconnection



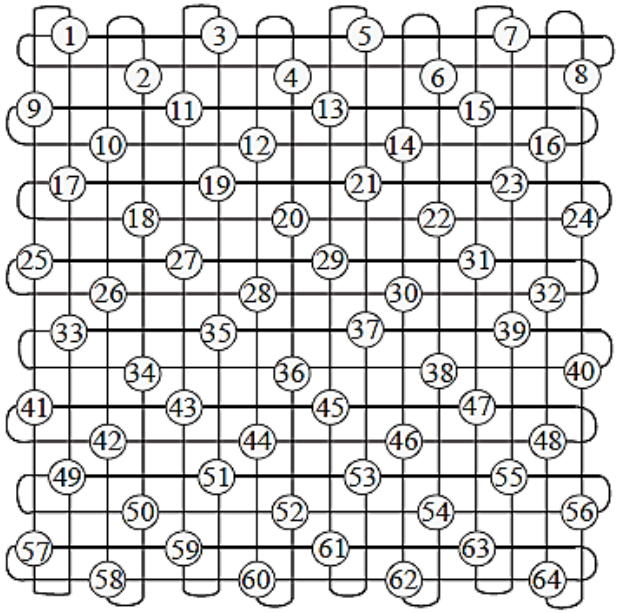
Goals:

- Reduce long-distance communication and Achieve high performance
- Guarantee low power consumption and safe temperature

Physical arrangement on the chip

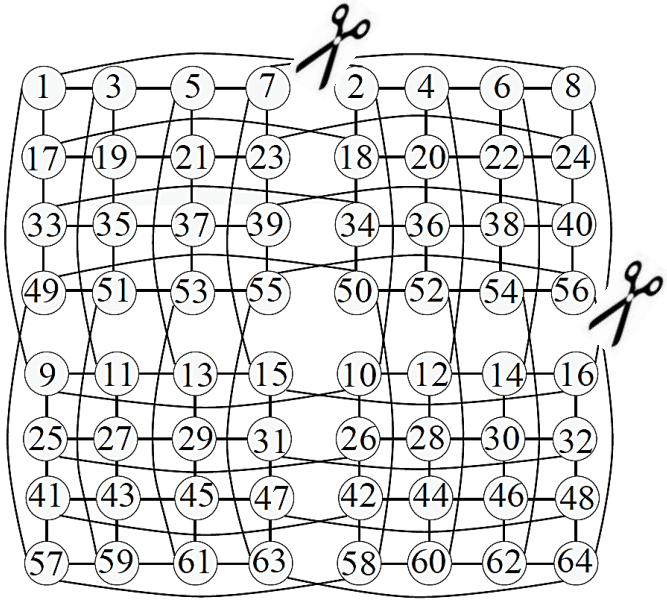


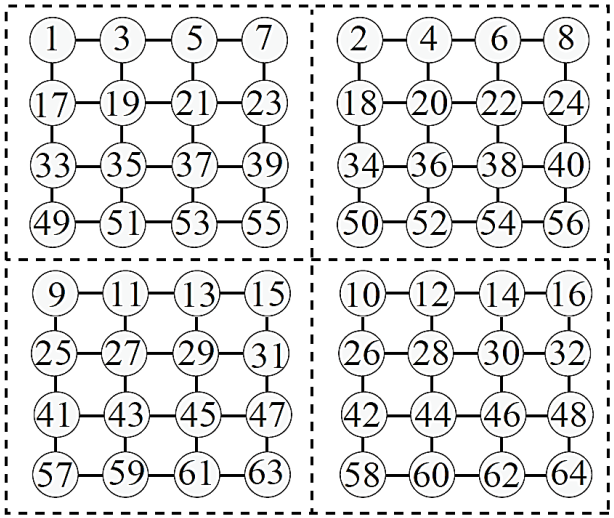
Logical Interconnection



FoToNoC : collaboration on clustered *Fo*lded *To*rus-like *NoC*

Logical Interconnection Transformation



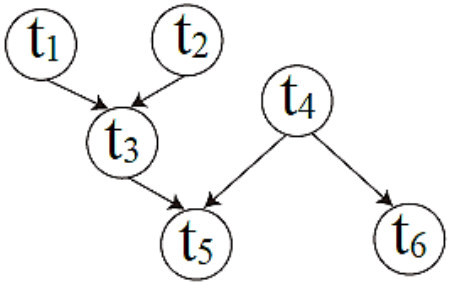


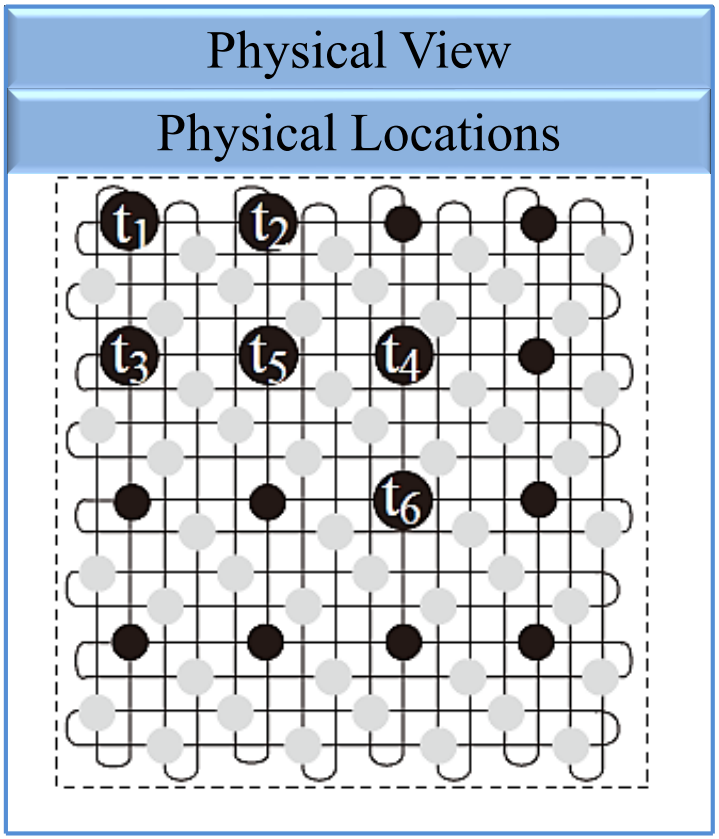
- Step 1: Equivalent topology transform
  - -- Stretching
  - -- Conversion
- Step 2: Virtually remove curve links
  - --(1, 2), (9, 10), (7,8)...,(63,64), (57, 58)
  - --(1, 9), (3, 11), (2,10)...,(54,62), (56,64)
- Step 3: Four clustered
  - -- Little (L): Low energy consumption
  - -- Big (b): High performance

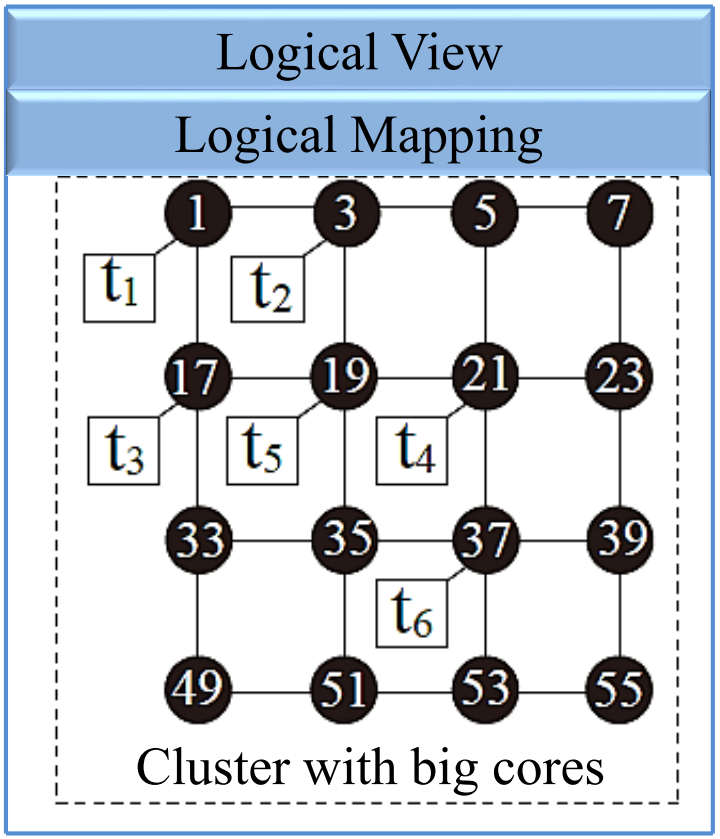
Cluster type	Freq (GHz)	Cache level	Power (W)	Area (mm^2)
L	3.0	L2	3.715	0.816
M_1	3.5	L2	6.544	1.046
M_2	4.0	L2	19.691	1.086
b	4.5	L2	28.808	1.178

Technology node: Alpha 21264 cores in 22-nm technology.
Power: Per-core power from McPAT [4].

Physical-logical view







Main CMS:

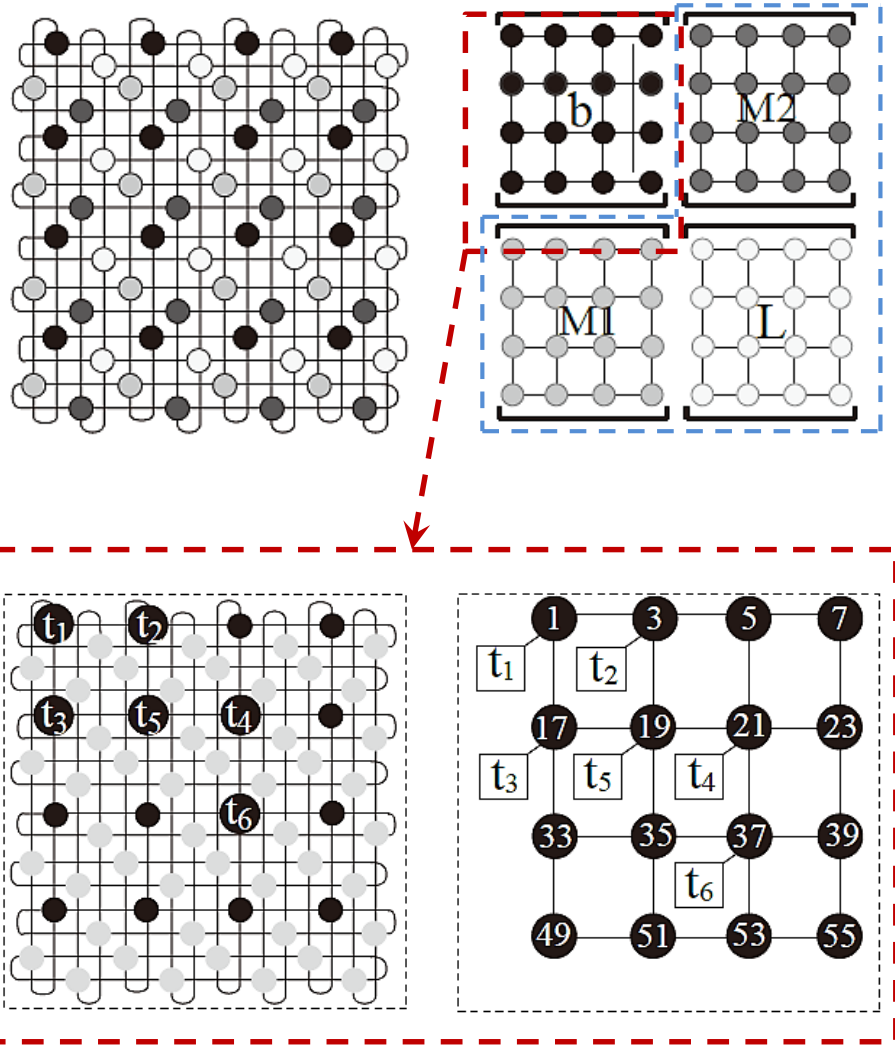
1) **Application *T* arrives.**

2) **Application *T* finishes on *cs* at *freq*.**

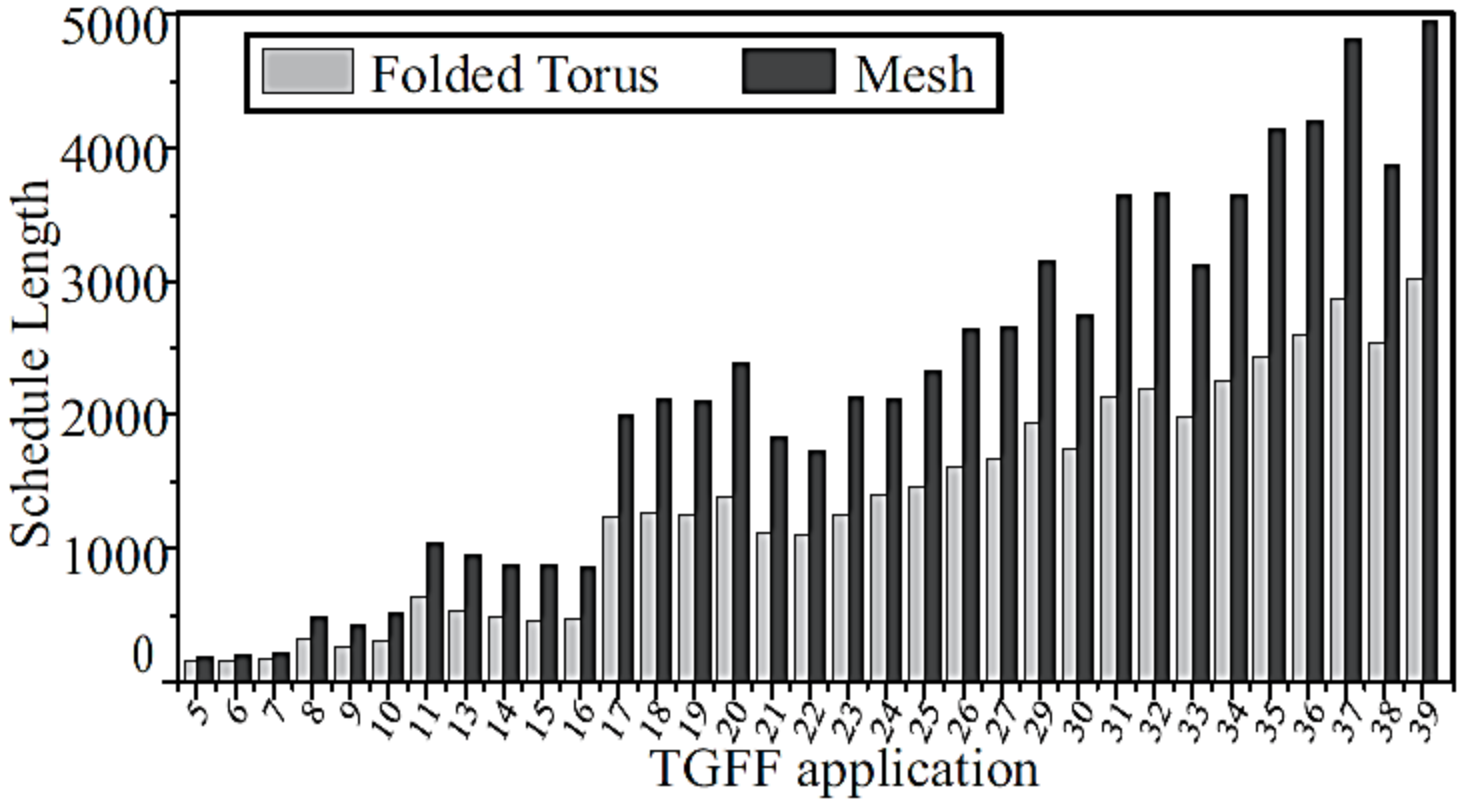
Algorithm 1 Cluster Management Strategy (CMS)

$(bool, ID, freq, t_{start}) = Cluster\_Manager(AC, \mathcal{T}, \Gamma)$

1. Sort active cluster set AC in ascending order by frequency;
2. $cs_{min} \leftarrow$ the cluster in $\mathcal{CS} - \mathcal{AC}$ with the lowest frequency;
3. for cs in ordered list AC do
4. $(flag, f_{req}, t_{start}) = Map\_to\_Cluster(\mathcal{T}, id(cs), \Gamma)$;
5. if $flag = 1$ or ($flag = 0$ and $f_{req} < f_{id(cs)}$) then
6. return $(TRUE, id(cs), \max\{f_{id(cs)\_min}, f_{req}\}, t_{start})$;
7. else if $flag = 0$ then
8. $TMP\_ID = id(cs)$; $break$;
9. end if
10. end for
11. for cs in ordered list $\mathcal{CS} - \mathcal{AC}$ do
12. $(flag', f'_{req}, t'_{start}) = Map\_to\_Cluster(\mathcal{T}, id(cs), \Gamma)$;
13. if $flag' = 0$ then
14. $break$;
15. else if $flag' = 1$ then
16. return $(TRUE, id(cs), f'_{req}, t'_{start})$;
17. end if
18. end for
19. if $flag = 0$ then
20. return $(TRUE, TMP\_ID, f_{ID\_min}, t_{start})$;
21. else
22. return (*FALSE*, -1, -1, -1);
23. end if



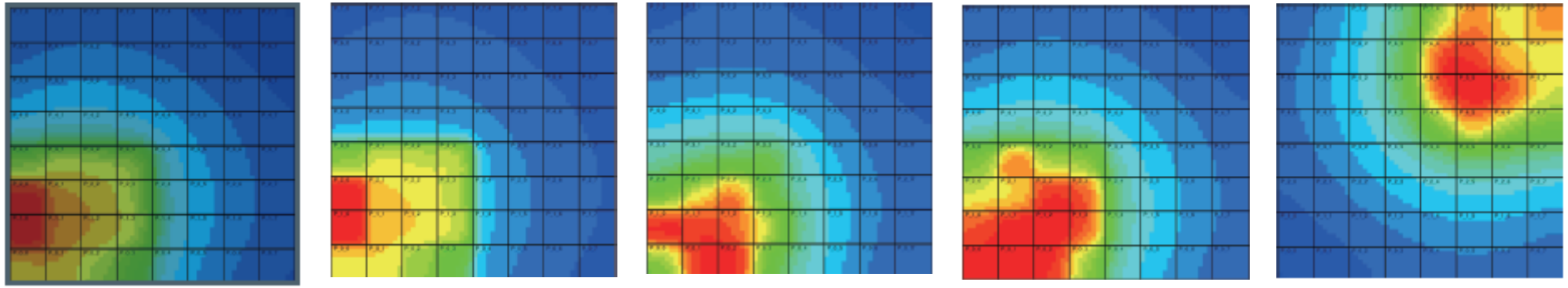
VS.1 Application Performance：



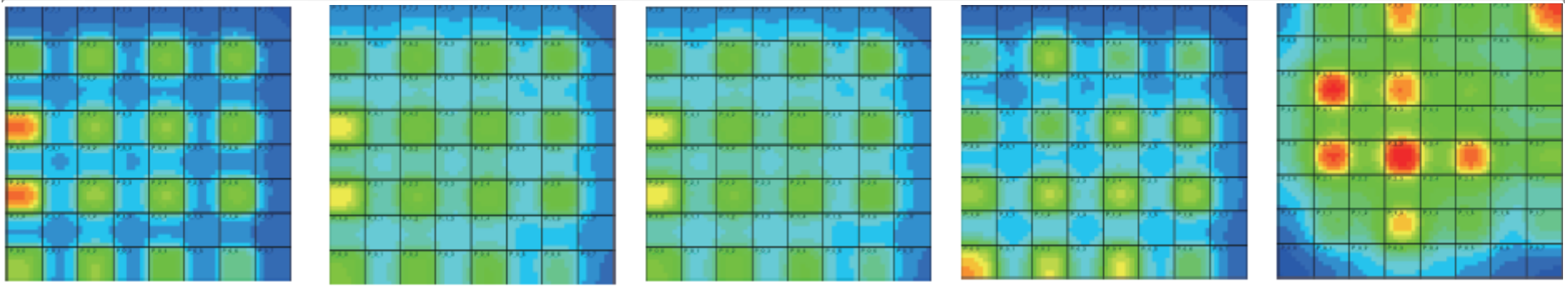
Application performance: 39.44% higher

VS.2 Thermal distribution：

Thermal states on Mesh



Thermal states on folded torus



VS.3 System Energy Consumption：

Task For Free Applications

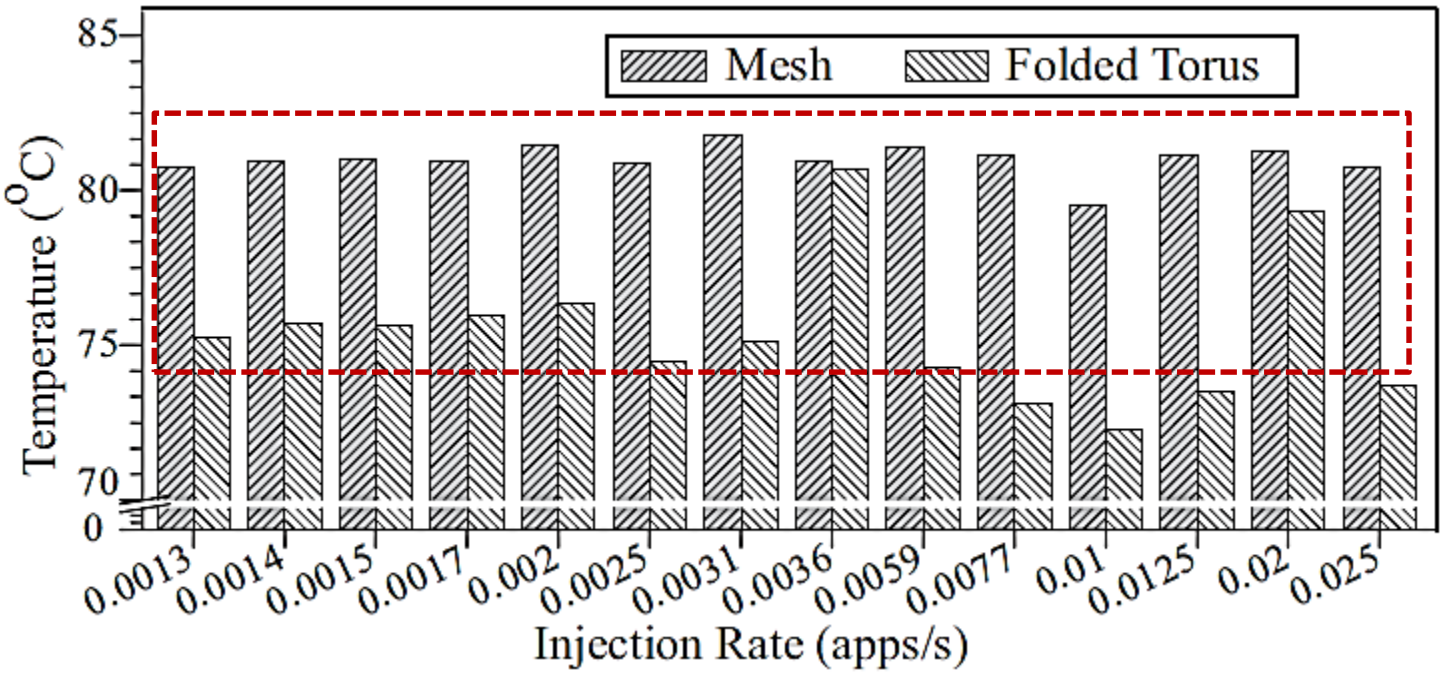
Rate_in (apps/s)	*NCMS*		*CMS*		*E_Impv (%)*
	R_{NC}	$E_{NC}/\mathcal{J}$	R_C	$E_C/\mathcal{J}$	
0.0017	0.00%	17988.72	0.00%	10844.15	39.72%
0.0020	0.00%	25361.99	0.00%	11893.03	53.11%
0.0025	0.00%	20313.29	0.00%	12337.59	39.26%
0.0031	5.88%	23527.44	0.00%	12482.4	46.95%
0.0036	5.88%	22876.37	0.00%	17523.37	23.40%
0.0040	11.76%	24390.85	0.00%	9626.71	60.53%
0.0045	11.76%	19487.82	5.88%	16675.62	14.43%
0.0050	11.76%	22117.87	5.88%	15743.38	28.82%
0.0056	23.53%	21357.88	5.88%	14129.43	33.84%
0.0059	5.88%	26297.28	5.88%	10278.83	60.91%
0.0077	23.53%	18347.37	5.88%	14645.48	20.18%
0.0100	35.29%	21134.16	23.53%	14653.93	30.66%
0.0125	35.29%	22930.42	29.41%	7054.49	69.24%
0.0200	47.06%	20199.93	23.53%	9867.29	51.15%
0.0250	52.94%	14456.66	41.18%	7895.71	45.38%
0.0500	70.59%	12052.38	47.06%	8460.49	29.80%
0.1000	70.59%	14001.3	52.94%	8751.58	37.49%
Average energy consumption reduction					45.02%

Realistic H.264 Benchmark Applications

Rate_in (apps/s)	*NCMS*		*CMS*		*E_Impv (%)*
	R_{NC}	$E_{NC}/\mathcal{J}$	R_C	$E_C/\mathcal{J}$	
0.0017	13.33%	55627.40	0.00%	35240.47	36.65%
0.0020	6.67%	50405.13	0.00%	35677.66	29.22%
0.0025	13.33%	48461.13	6.67%	30517.30	37.03%
0.0031	26.67%	45303.41	13.33%	38423.93	15.19%
0.0036	20.00%	53926.83	13.33%	48683.68	9.72%
0.0040	20.00%	39837.72	13.33%	30454.75	23.55%
0.0045	33.33%	49068.54	0.00%	19022.54	61.23%
0.0050	53.33%	21497.72	13.33%	13117.73	38.98%
0.0056	26.67%	45303.41	0.00%	29721.64	34.39%
0.0059	40.00%	40445.12	0.00%	30607.98	24.32%
0.0077	40.00%	40445.12	0.00%	25209.63	37.67%
0.0100	53.33%	35586.83	6.67%	20859.35	41.38%
0.0125	60.00%	18340.00	20.00%	12417.68	32.29%
0.0200	60.00%	25262.84	60.00%	19610.16	22.38%
0.0250	60.00%	26963.41	33.33%	12290.21	54.42%
0.0500	66.67%	22105.12	60.00%	11370.27	48.56%
0.1000	73.33%	22105.12	60.00%	10369.75	53.09%
Average energy consumption reduction					35.29%

System energy consumption: 45.02%, 35.29% lower

VS.4 Chip Peak Temperature：



Peak temperature : 5.61 ◦C lower

Conclusion

➢ Major Problem

- Fast increasing number of transistors
- Slow voltage/capacitance scaling
- Limited physical device cooling technology and heat dissipation ability
- Thermal management problem with long-distance communication

➢ We Solved

- Communication overhead reduction
- Application performance improvement
- System energy consumption reduction
- Chip temperature reduction
- Safe temperature reliability enhancing
- Trade-offs by Physical- Logical Isolation

➢ Major Contribution

- Physical- Logical isolation in the dark silicon era:

  -- Isolate and addressed the mixed design concerns of hierarchical H/W co-design

- *FoToNoC* architecture organization

  -- Keep safe regarding temperature reliability via distributed core activation

  -- Reduce communication overhead through “dark silicon-friendly” architecture

- Efficient clustered management strategies

  -- Maximize the benefits of *FoToNoC* in the dark silicon era

  -- Manage heterogeneous cores in clusters with DVFS

  --Optimized chip temperature, application performance , system power consumption.

➢ Future Work

- Long-range links:

-- Communication delay analysis

-- Power consumption for on-chip communication

- Heterogeneous cores arrangement

-- Inter-cluster Heterogeneity instead of homogeneity

-- Number of the types of processing cores

-- Fine-grained run-time Voltage/Frequency scaling

-- Run-time task migration within/among clusters

Thanks

Q & A

Lei Yang
leiyang@cqu.edu.cn
Chongqing University