MonTM: Monitoring-Based Thermal Management for Mixed-Criticality Systems

Marcel Mettler*, Martin Rapp[†], Heba Khdr[†], Daniel Mueller-Gritschneder*, Jörg Henkel[†] and Ulf Schlichtmann*

*Chair of EDA, Technical University of Munich, Germany

[†] Chair of ES, Karlsruhe Institute of Technology, Germany

Toulouse, 17. January 2023

Challenges of Mixed-Criticality Systems

Mixed-Criticality Systems

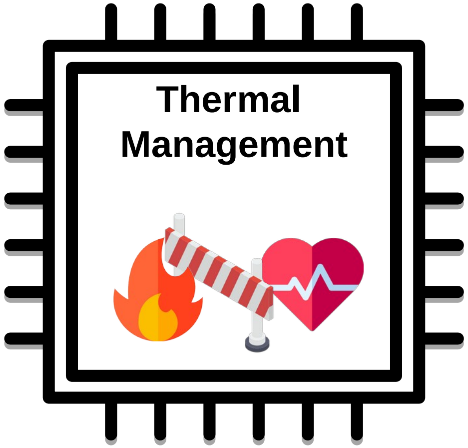
- Integrate tasks of different safety integrity levels (SILs)
- ➢ Common platform reduces cost, power, space…
- ➢ **Require isolation of SILs**

1. Architectural Resources

- Interference via cores, memory, etc.
- ➢ Virtualization techniques

2. Thermal Manager

- Thermal coupling of neighboring cores
- ➢ ?



How can we limit the thermal interference between SILs?

State-of-the-Art Thermal Management

Reactive Thermal Management

Intel‘s Turbo Boost [1]

AMD‘s Turbo Core

- Change frequency based on current temperature and power consumption
- Are not predictable

➢ **Fully utilize thermal headroom**

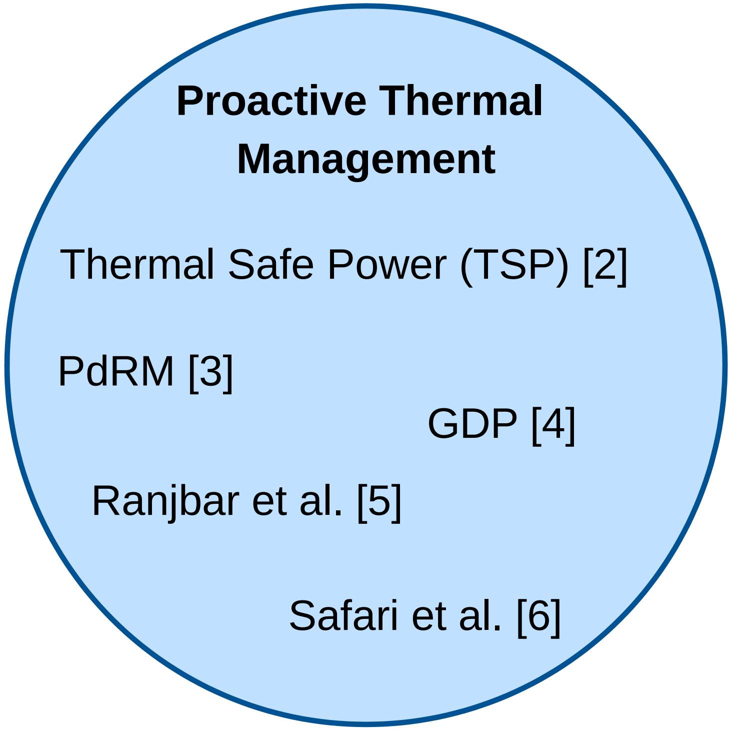
➢ **Cannot give timing guarantees for safety-critical tasks**

State-of-the-Art Thermal Management

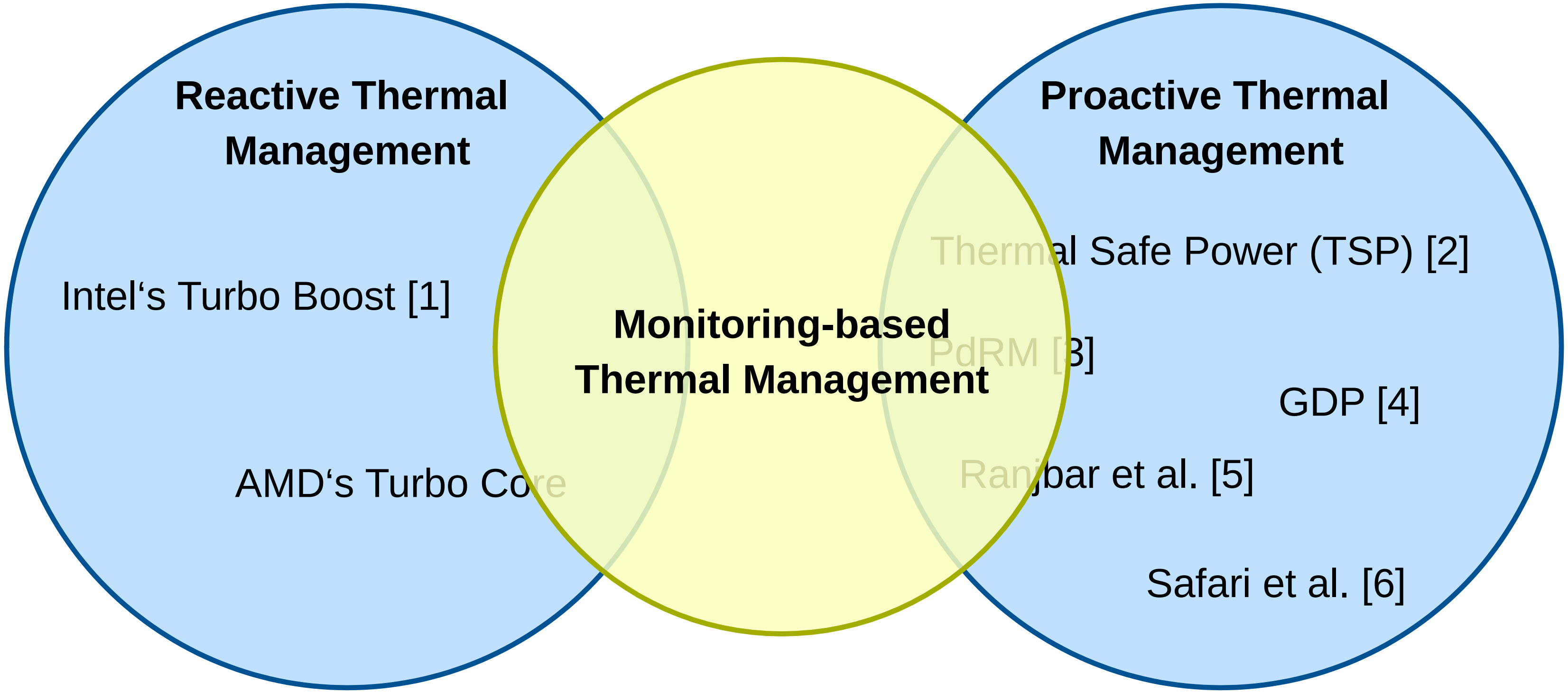
- Assign thermally safe power budgets
- Rely on maximal power consumption

➢ **Predictable execution times**

➢ **Overly pessimistic if power consumption shows high variance**



State-of-the-Art Thermal Management



- Reactive thermal management for best-effort tasks
- Proactive thermal management for safety-critical tasks

Contributions

MonTM: A decentralized thermal management strategy

- Prevents best-effort tasks from inducing thermal violations into safety-critical tasks

Light-weight DTM interconnect

- Enable DTMs to communicate thermal status

Slack Monitor

- Statically assigned V/f levels of safety-critical tasks may be pessimistic if they run faster than WCET
- Determines minimal V/f requirement based on slack

Contributions

MonTM: A decentralized thermal management strategy

- Prevents best-effort tasks from inducing thermal violations into safety-critical tasks

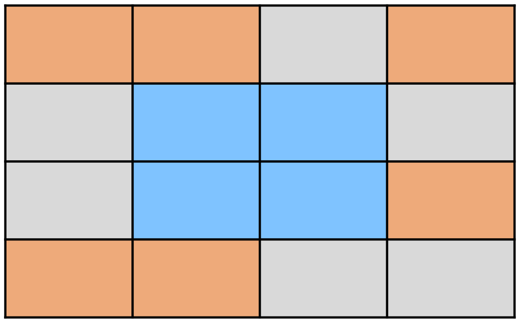
Light-weight DTM interconnect

- Enable DTMs to communicate thermal status

Slack Monitor

- Statically assigned V/f levels of safety-critical tasks may be pessimistic if they run faster than WCET
- Determines minimal V/f requirement based on slack

Problem Formulation



Floorplan

Safety-Critical Tasks

- Service Level Agreements (SLAs)
  - Deadline
  - WCET
  - Exclusive resource, i.e. core

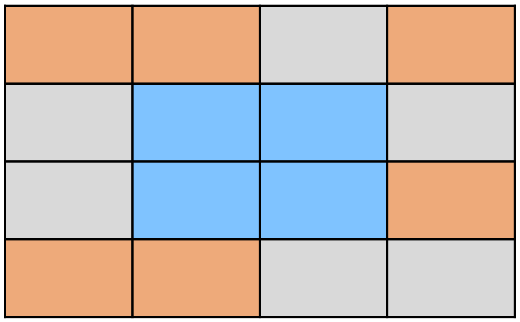
Best-Effort Tasks

- No service level agreements (SLAs)

Objective

- Minimize the latency of best-effort jobs s.t.
  - All critical jobs meet their deadline
  - Thermal requirements of all cores are satisfied

Thermal Management Strategy



Floorplan

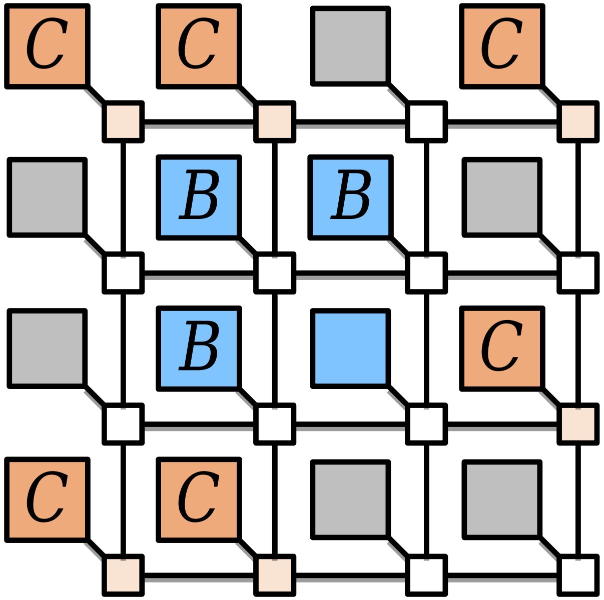
Safety-Critical Tasks

- Service Level Agreements (SLAs)
  - Deadline
  - WCET
  - Exclusive resource, i.e. core

Best-Effort Tasks

- No service level agreements (SLAs)
- Must **not** induce thermal violations in safety-critical tasks

Thermal Pre-error Interconnect



Thermal Pre-error Interconnect

- Communicates imminent thermal violations of safety-critical tasks
- Supports four pre-error levels
  - ✉ no action
  - ✉ throttle in hop distance of 1
  - ✉ throttle in hop distance of 2
  - ✉ halt all best-effort tasks

Comparison to State of the Art (1)

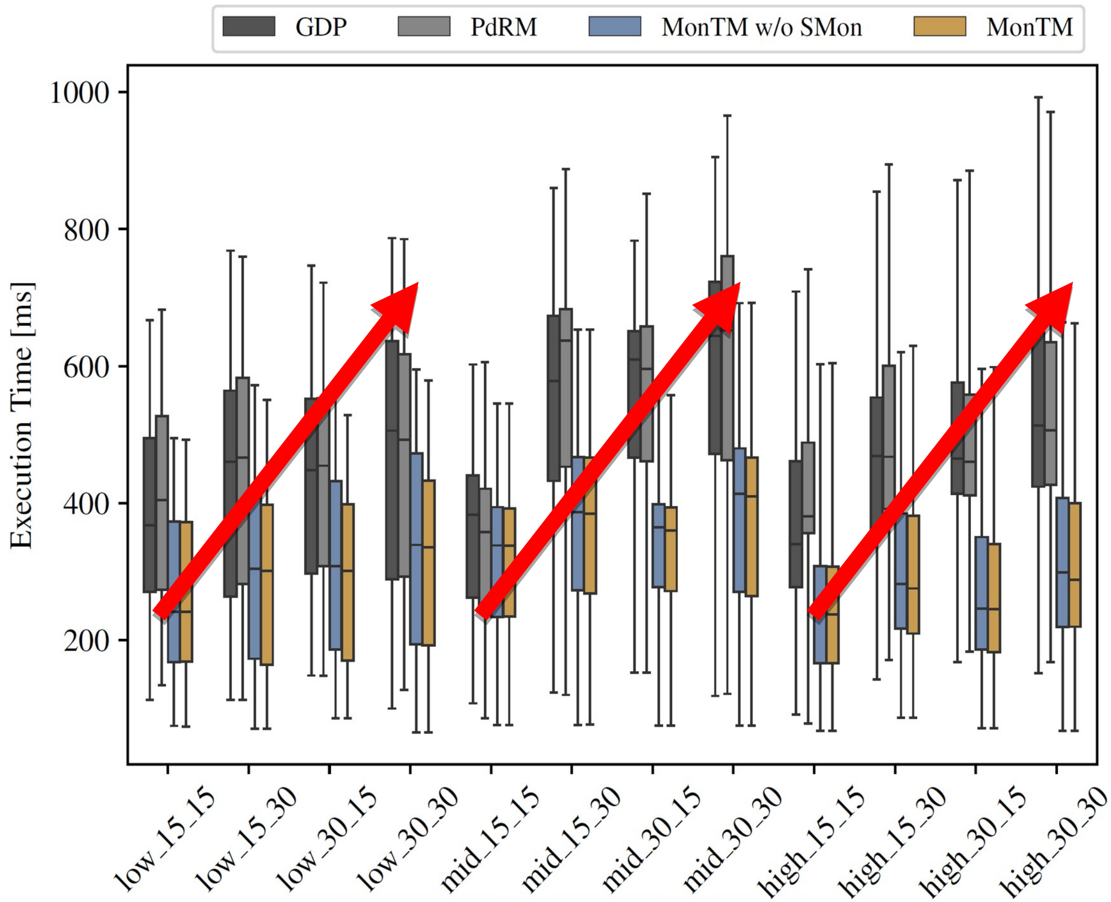
Evaluation Setup

- FPGA prototype of 80-core processor
- Per-core power, temperature emulation
- DVFS emulation with 2 locktime

Synthetic Workloads: <>_<>_<>

- Variance of maximal power consumption
  - Low
  - Mid
  - High
- Number of safety-critical tasks
- Number of best-effort tasks

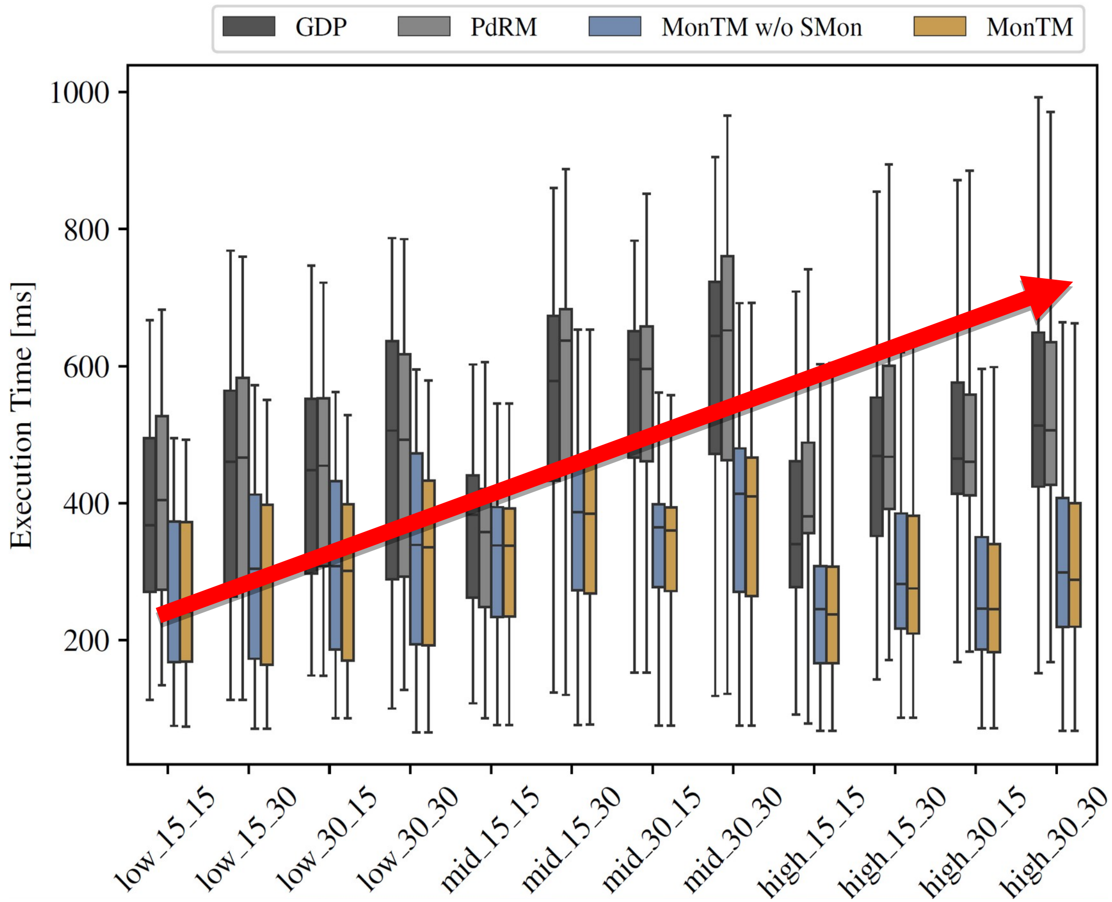
Comparison to State of the Art (2)



Execution times

- ➢ Increase with system load

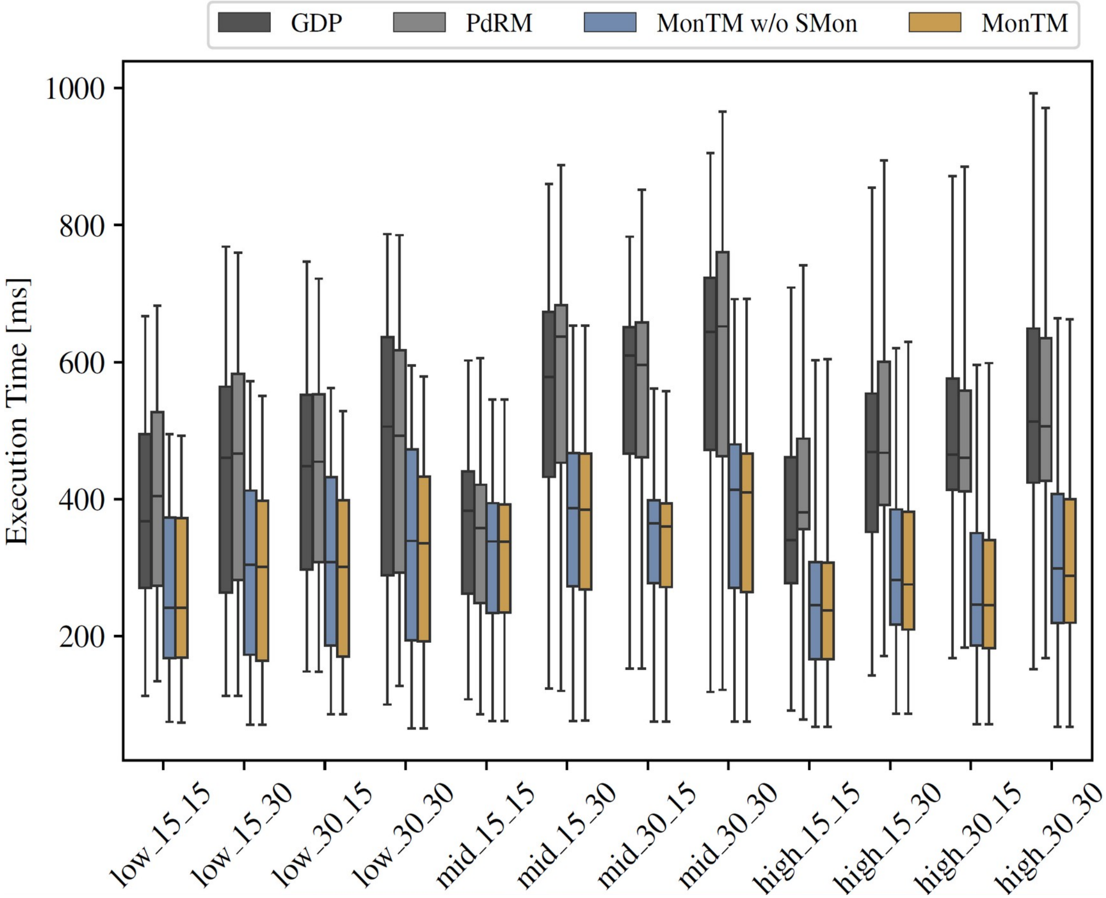
Comparison to State of the Art (3)



Execution times

- Increase with system load
- Improvement increases with variance in power consumption

Comparison to State of the Art (4)



Execution times

- Increase with system load
- Improvement increases with variance in power consumption

- 7-44% improvement without slack monitor
- Additional 1-6% improvement with slack monitor

Conclusion

Monitoring-based thermal management

- Thermal pre-error interconnect
  - ➢ Communicates imminent thermal violations
  - ➢ Provides sufficient thermal isolation
- Slack monitor
  - ➢ Safely reduce the frequency of safety-critical tasks

➢ Reduces run-time of best-effort tasks by up to 45%

Questions?

Sources

[1] J. Casazza. Intel turbo boost technology in intel core microarchitecture (nehalem) based processors. Technical report, Intel Corporation, 11 2008.

[2] S. Pagani, H. Khdr, J.-J. Chen, M. Shafique, M. Li, and J. Henkel. Thermal Safe Power (TSP): Efficient Power Budgeting for Heterogeneous Manycore Systems in Dark Silicon. IEEE Trans. Computers (TC), 66(1):147–162, 2017.

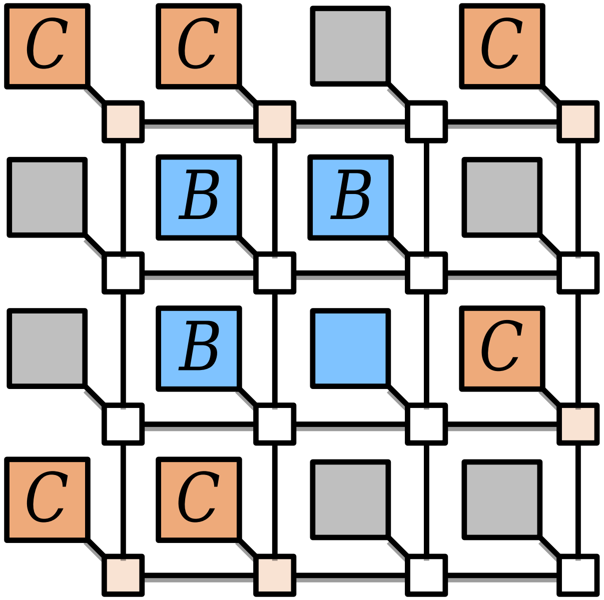
[3] H. Khdr, S. Pagani, É. Sousa, V. Lari, A. Pathania, F. Hannig, M. Shafique, J. Teich, and J. Henkel. Power density-aware resource management for heterogeneous tiled multicores. IEEE Trans. Computers (TC), 66(3):488–501, 2017

[4] H. Wang, D. Tang, M. Zhang, S. X.-D. Tan, C. Zhang, H. Tang, and Y. Yuan. Gdp: A greedy based dynamic power budgeting method for multi/many-core systems in dark silicon. IEEE Trans. Computers (TC), 68(4):526–541, 2019.

[5] B. Ranjbar, A. Hosseinghorban, M. Salehi, A. Ejlali, and A. Kumar. Toward the design of fault-tolerance-aware and peak-power-aware multicore mixed-criticality systems. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 41(5):1509–1522, 2022.

[6] S. Safari, H. Khdr, P. Gohari-Nazari, M. Ansari, S. Hessabi, and J. Henkel. Therma-mics: Thermal-aware scheduling for fault-tolerant mixed-criticality systems. IEEE Transactions on Parallel and Distributed Systems, 33(7):1678–1694, 2022.

Thermal Pre-error Interconnect



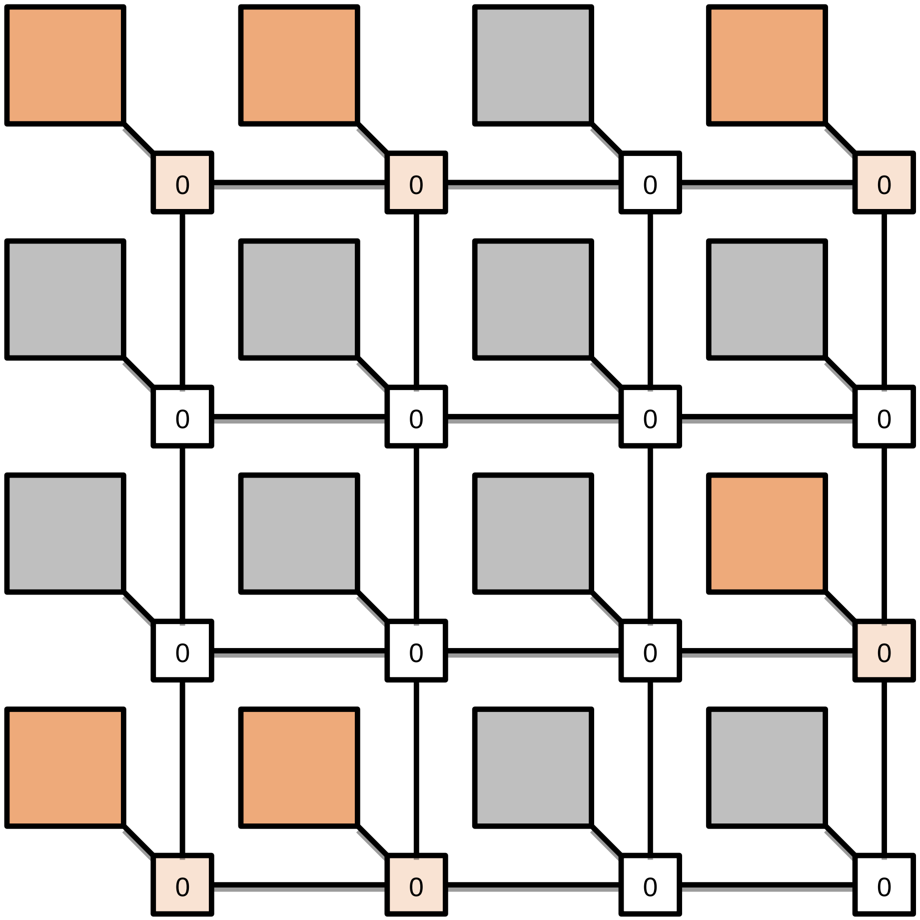
Thermal Pre-error Interconnect

- Communicates imminent thermal violations of safety-critical tasks
- Supports four pre-error levels
  - no action
  - throttle in hop distance of 1
  - throttle in hop distance of 2
  - halt all best-effort tasks

Routers at safety-critical tasks

Other routers

Thermal Pre-error Interconnect – Example (1)

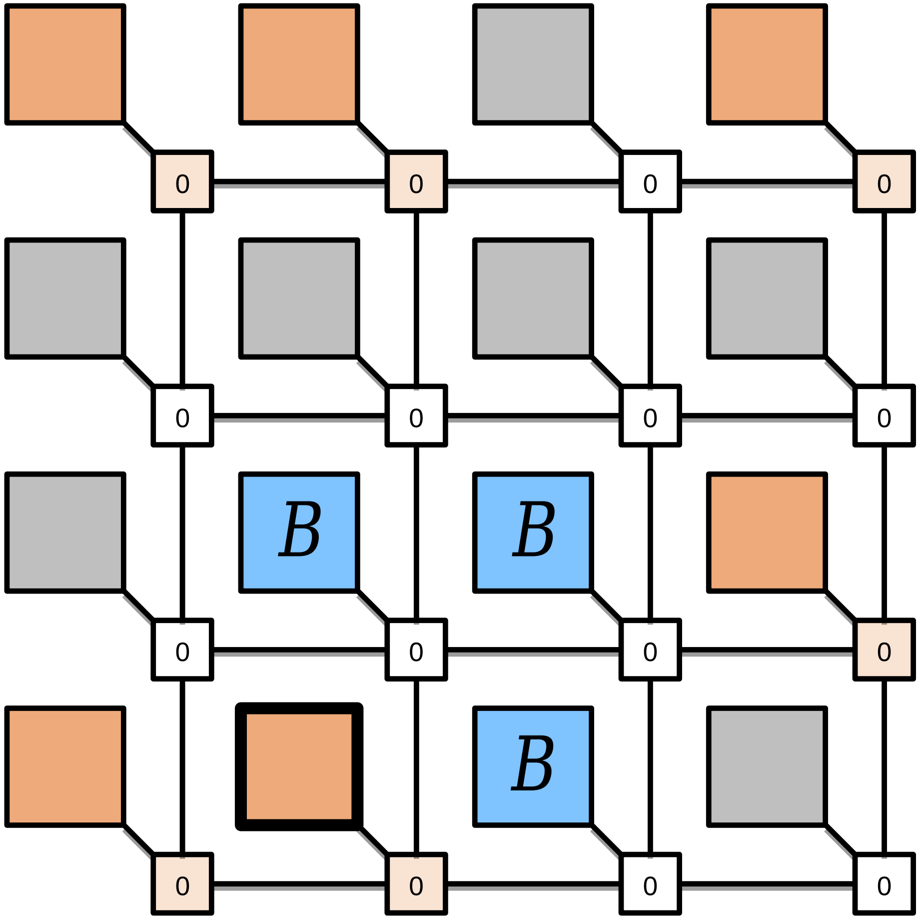


State t=0
- No best-effort tasks launched yet

Routers at safety-critical task

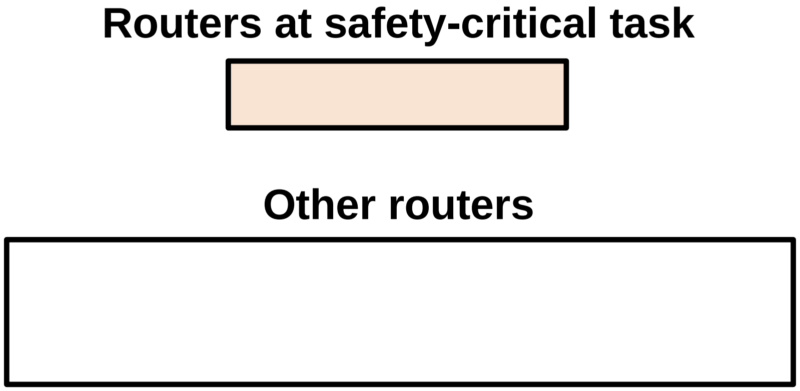
Other routers

Thermal Pre-error Interconnect – Example (2)

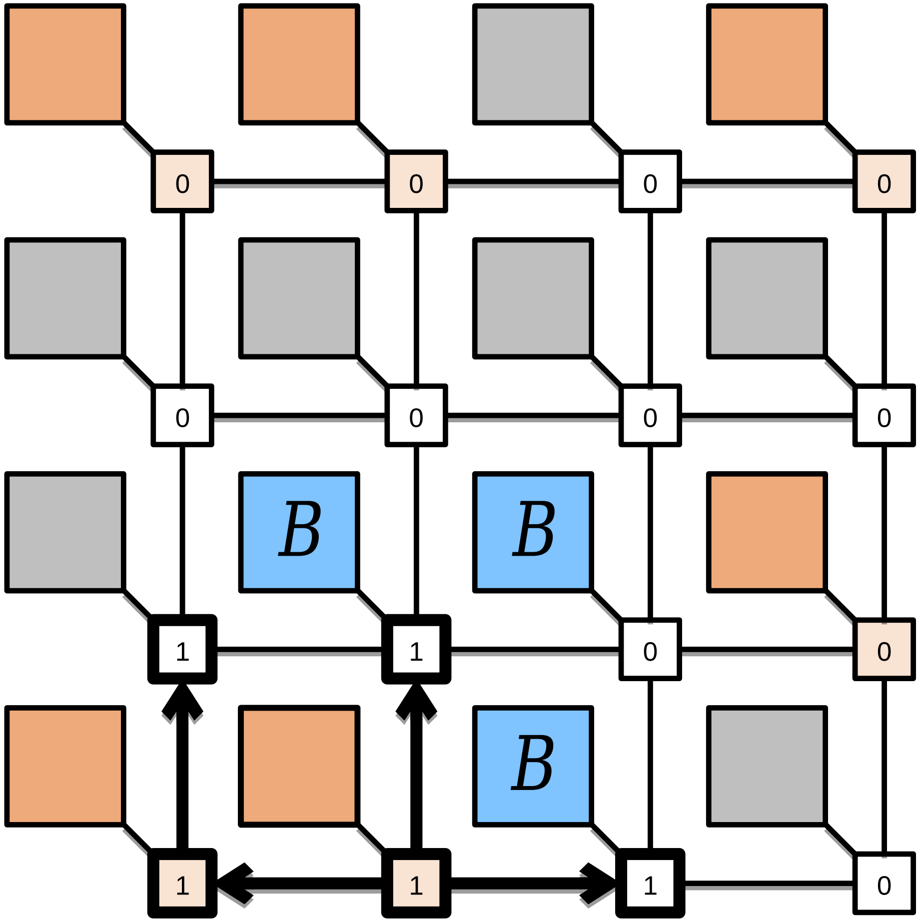


State t=1

- Launch 3 best-effort tasks
- Thermal coupling results in a thermal pre-error



Thermal Pre-error Interconnect – Example (3)



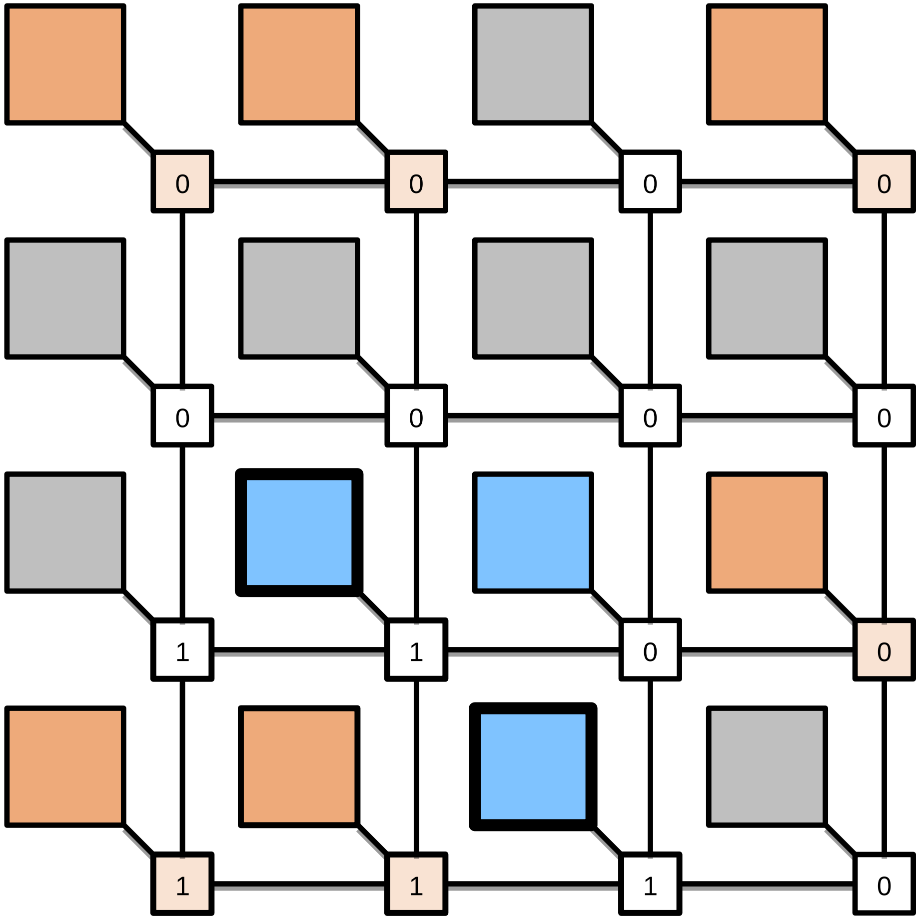
State t=2

- Router broadcasts thermal pre-error

Routers at safety-critical task

Other routers

Thermal Pre-error Interconnect – Example (4)



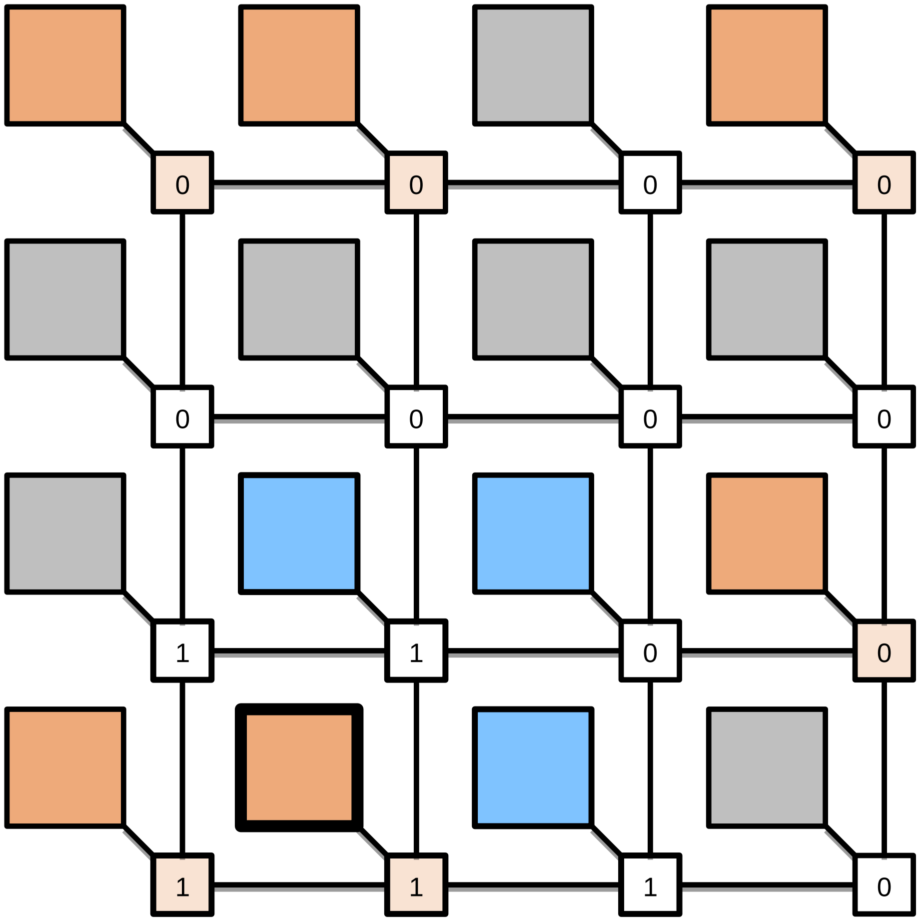
State t=2

- Router broadcasts thermal pre-error

Routers at safety-critical task

Other routers

Thermal Pre-error Interconnect – Example (5)



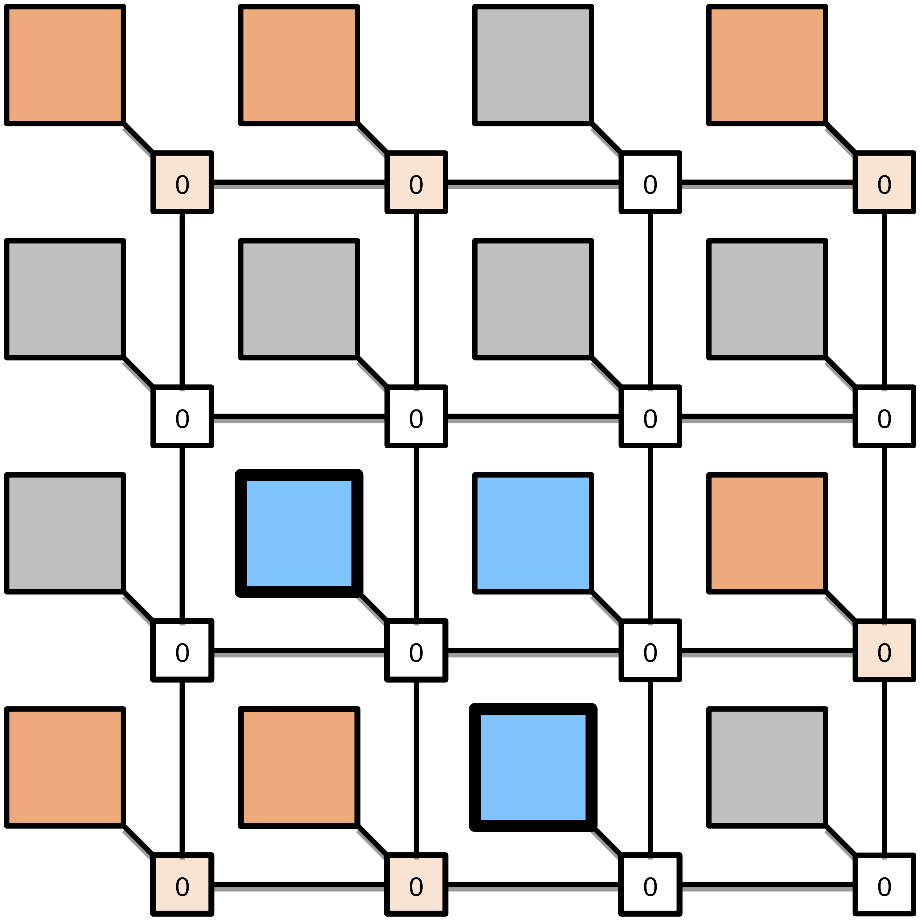
State t=2

- Router broadcasts thermal pre-error

Routers at safety-critical task

Other routers

Thermal Pre-error Interconnect – Example (6)



State t=2

- Router broadcasts thermal pre-error

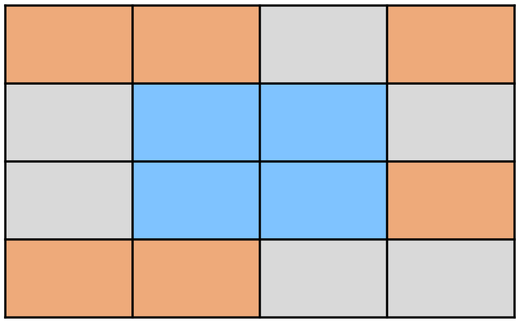
Routers at safety-critical task

Other routers

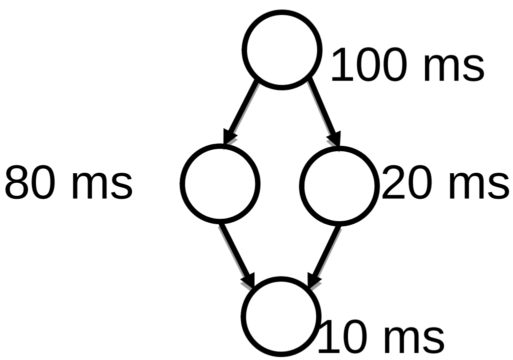
Hardware Overhead

	Slice LUTs		Slice Registers	
	Absolute	Relative	Absolute	Relative
Router	101	< 0.1%	208	0.2%
Thermal Manager	70	< 0.1%	60	< 0.1%
Slack Monitor	1,465	1.0%	3,176	3.3%
Probe	356	0.2%	830	0.9%
Total	**1,997**	**1.3%**	**4274**	**4.4%**

Slack Monitor



Floorplan



Control Flow Graph (CFG)

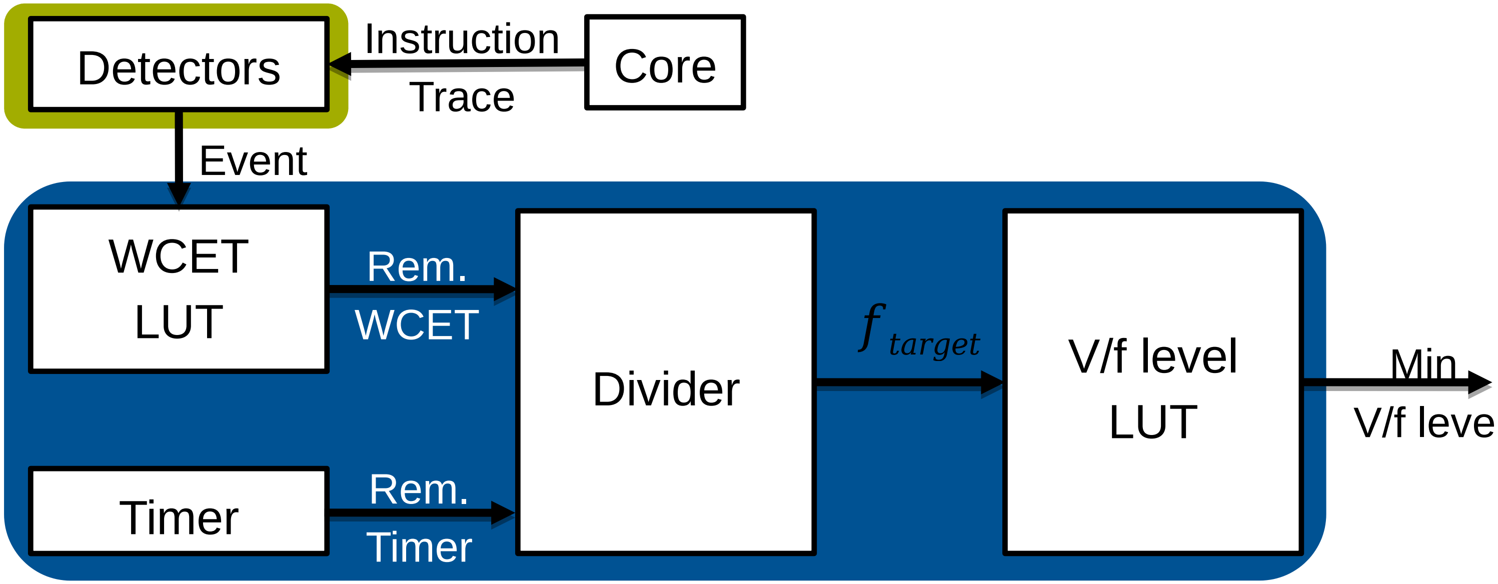
Critical Tasks

- Static V/f levels are assigned based on WCET
- If task finishes faster, static V/f levels are overly pessimistic

- Run-time Monitoring
  - Identify basic blocks in CFG
  - Map basic block to remaining WCET

➢ **Boosts best-effort tasks by an increased thermal headroom**

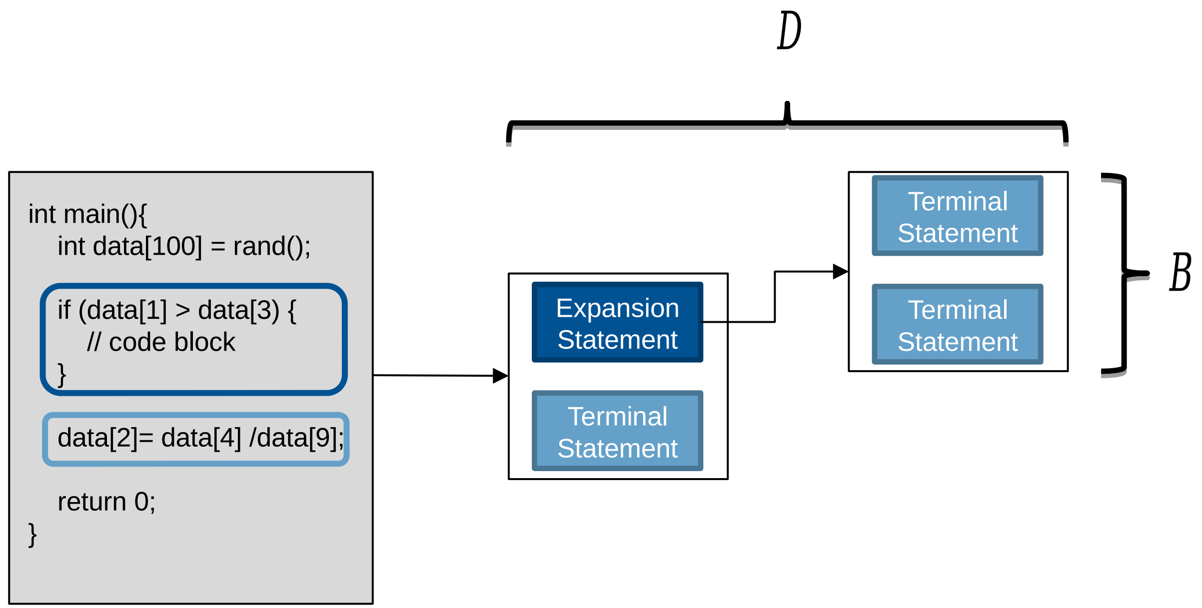
Implementation of Slack Monitor



Benchmark Generation

Recursive Expansion (REX) Process [1] using

- maximal level of nesting (depth)
- number of statements per code block (breadth)



Terminal Statement — Simple arithmetic statement, e.g. variable assignment

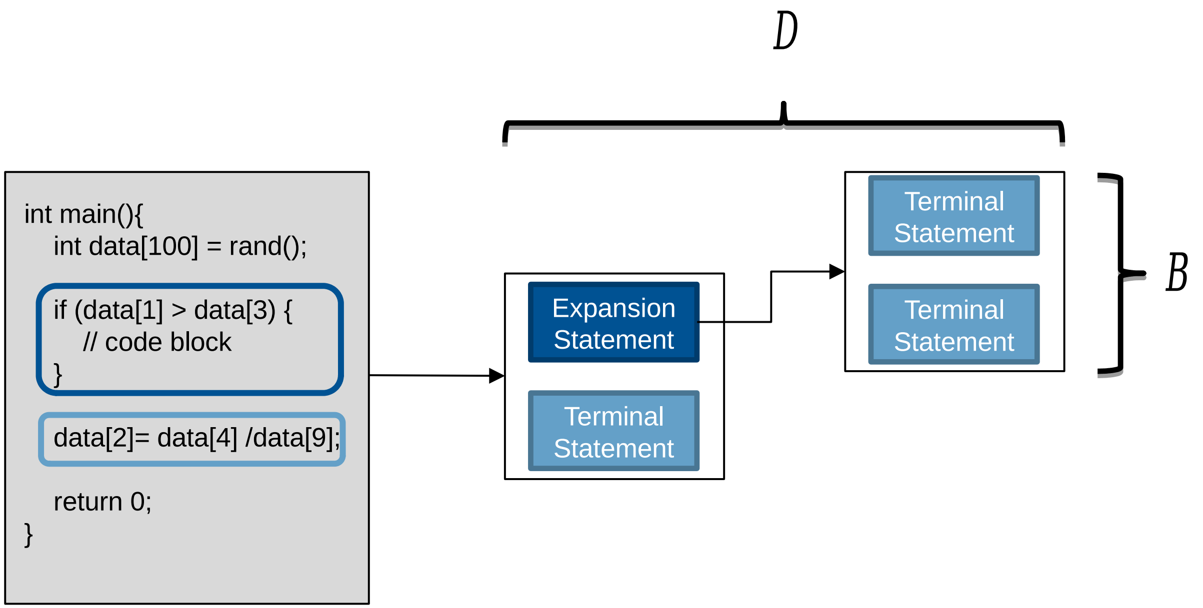
Expansion Statement — Expandable frame of statements e.g. if or loop clauses

[1] Jozo Dujmović. 2010. Automatic generation of benchmark and test workloads. In *Proceedings of the first joint WOSP/SIPEW international conference on Performance engineering* (*WOSP/SIPEW '10*). Association for Computing Machinery, New York, NY, USA, 263–274. DOI:https://doi.org/10.1145/1712605.171265

Benchmark Generation

Recursive Expansion (REX) Process [1] using

- maximal level of nesting (depth)
- number of statements per code block (breadth)



Simple arithmetic statement, e.g. variable assignment

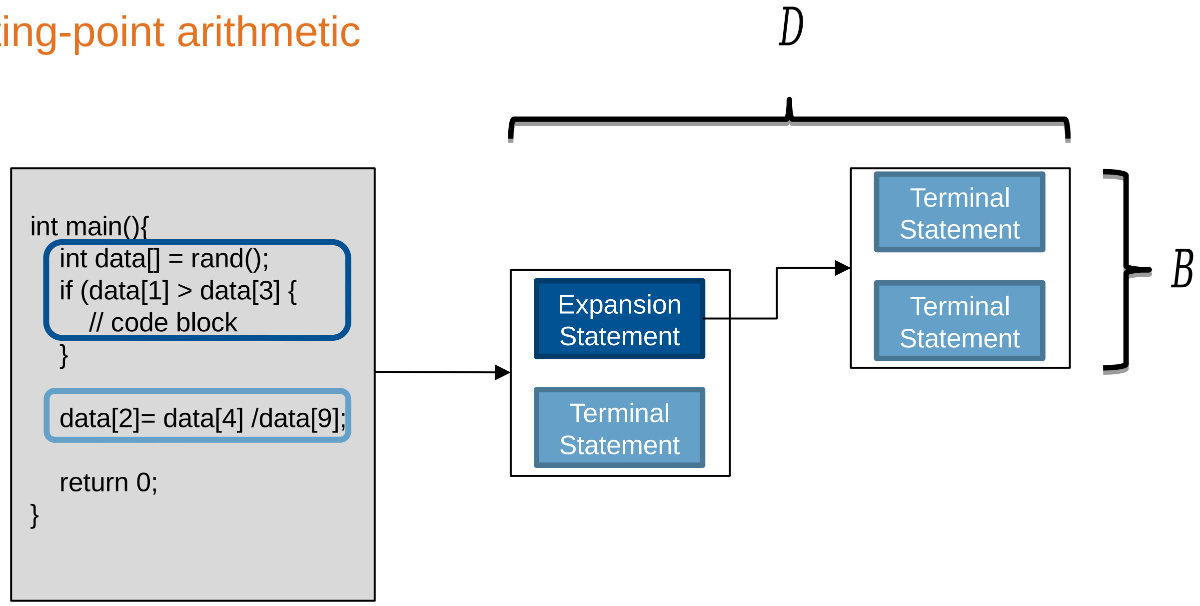
Expandable frame of statements e.g. if or loop clauses

[1] Jozo Dujmović. 2010. Automatic generation of benchmark and test workloads. In *Proceedings of the first joint WOSP/SIPEW international conference on Performance engineering* (*WOSP/SIPEW '10*). Association for Computing Machinery, New York, NY, USA, 263–274. DOI:https://doi.org/10.1145/1712605.171265

Benchmark Generation

Recursive Expansion (REX) Process [1] using

- maximal level of nesting (depth)
- number of statements per code block (breadth)
- memory size on which the application operates
- probability to use floating-point arithmetic



Simple arithmetic statement

Expandable frame of statements e.g. if or loop clauses

[1] Jozo Dujmović. 2010. Automatic generation of benchmark and test workloads. In *Proceedings of the first joint WOSP/SIPEW international conference on Performance engineering* (*WOSP/SIPEW '10*). Association for Computing Machinery, New York, NY, USA, 263–274. DOI:https://doi.org/10.1145/1712605.1712654