



ASAP 2026



International
Institute of Information
Technology Bangalore

QXL - A SCALABLE HARDWARE FRAMEWORK FOR TABULAR Q-LEARNING INFERENCE ON RISC-V

AUTHORS:

CHINMAY KRISHNA R

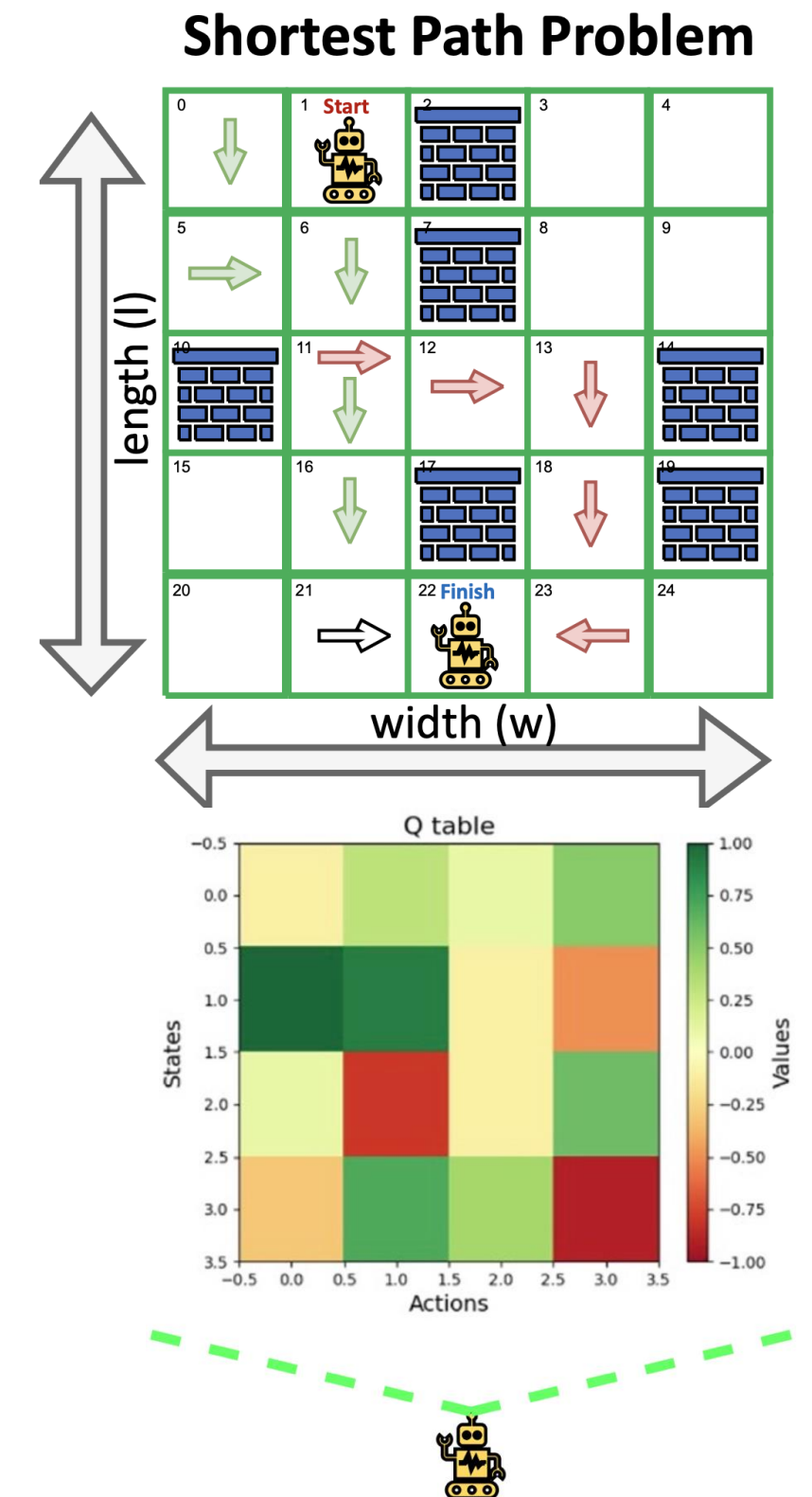
PRUTHVI PARATE

PHANI PAVAN K

PROF. MADHAV RAO

Motivation & Background

- **Tabular Q-Learning at the Edge:** Foundational reinforcement learning algorithm with strict convergence guarantees and simplicity.
- Highly effective for **robotics** and **network optimization** without the overhead of neural network training.
- Uses a knowledge matrix called Q-table, of dimensions $S_n \times A_n$.
- **The Inference Phase:** During deployment, learning ceases, and the agent strictly exploits the converged Q-table.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore

The Research Gap (The Memory Wall)

- Existing Accelerators: Frameworks like **QTAcel** and **MemOReL** primarily optimize the continuous training phase.
- The Inference Bottleneck: **Inference dominates the operational lifecycle** of deployed systems (e.g., autonomous navigation, industrial scheduling).
- **Memory Bound Execution:** Selecting an optimal action requires fetching all An state-action values from memory to compute $\max Q(s,a)$.
- These latency-sensitive, bandwidth-heavy accesses **destroy real-time responsiveness**.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore

Key Contribution: The QXL Framework

- QXL Architecture: A hardware-software co-designed framework **exclusively optimized for tabular Q-learning inference**.
- Core Mechanism: Uses an on-chip **firmware profiling** routine to identify frequently accessed state-action pairs.
- **Static Scratchpad**: Pre-loads optimal actions into a lightweight, configurable register scratchpad.
- Impact: Intercepts high-traffic queries to completely **bypass main memory lookups** during inference.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore

RISC-V ISA Extension & Interface

- Integrated into the **VEER EL2 RISC-V core** using the RVfpga framework.
- Custom Instruction: Implements the **q.inference** instruction using standard R-type format (Opcode: **1010011**).
- Operands: Uses two source registers (rs1, rs2) for state variables and one destination register for the optimal action.
- Efficiency: **Bypasses standard ALU**. Replaces multiple load, compare, and branch instructions with a single deterministic operation.

funct7	funct3	opcode	Mnemonic
0000000	000	1010011	q.inference rd, rs1, rs2



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

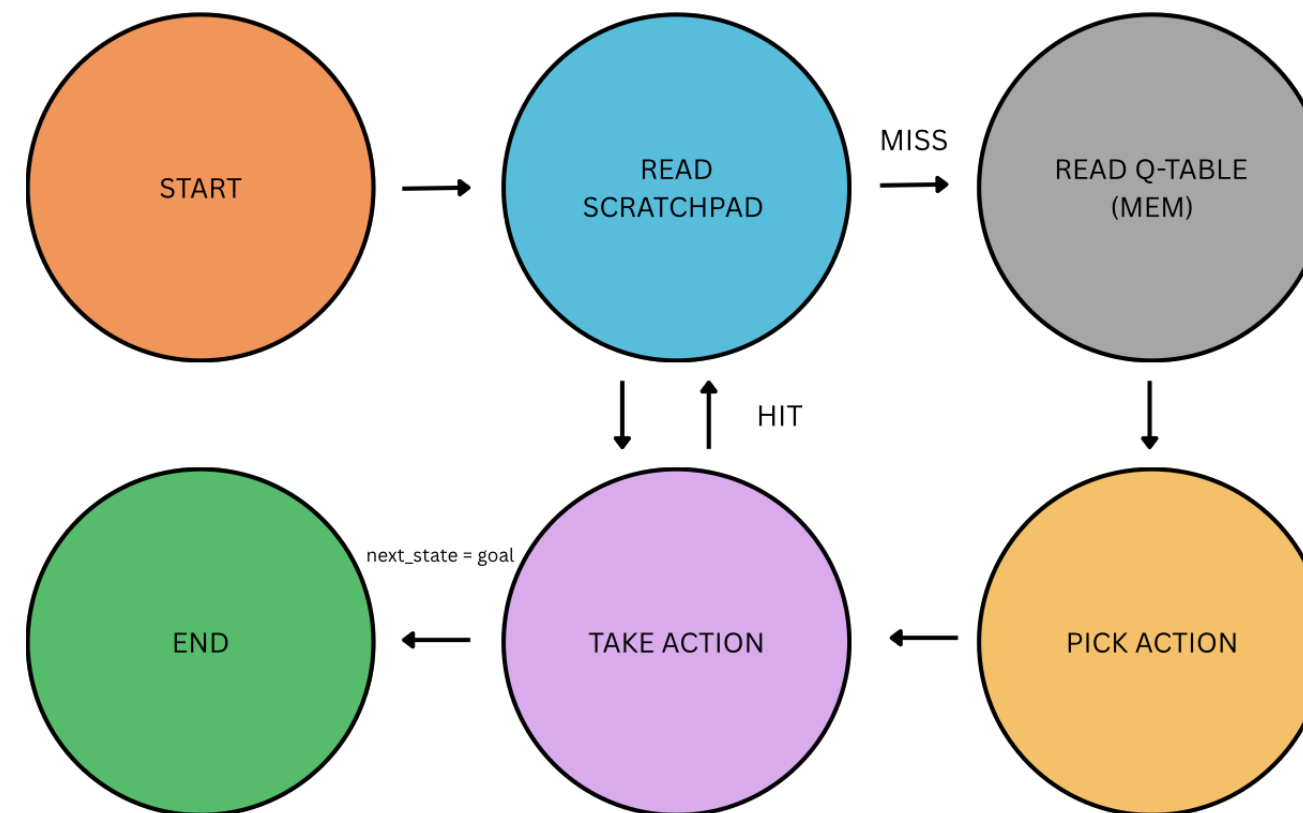
Conclusion



International
Institute of Information
Technology Bangalore

Hardware Architecture

- **FSM-based Control:** Sequential, Finite State Machine control design paired with a custom action-selection datapath.
- Data Storage: Unlike conventional caches that store raw data, **QXL stores decisions** (state identifier and one precomputed optimal action).



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore

Multi-Level Memory & Cycle Timing

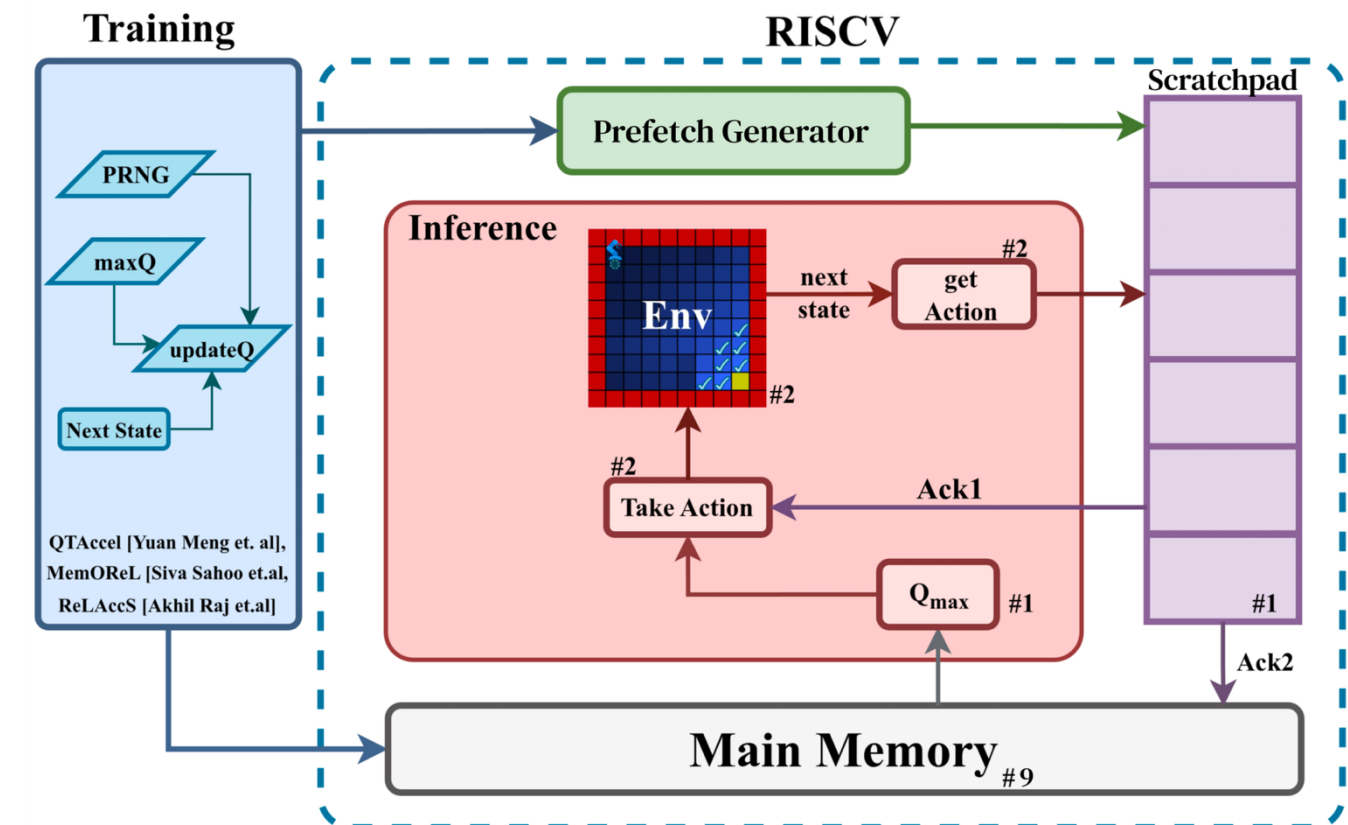
- **Memoization Depth (c%)**: The percentage of total environment states allocated to fast memory, defined by the designer.

1. Hit (Ack1 Event): Current state matches scratchpad.

- Optimal action retrieved immediately.
- latency: **7 cycles**.

2. Miss (Ack2 Event): State not in scratchpad.

- Fallback to baseline datapath (1 cycle)
- Retrieves all n Q-values from main memory (9 cycles).
- Routes through comparator tree (7 cycles).
- Total latency: **17 cycles**.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore

Experimental Setup & Workloads

Hardware Synthesis: **Nexys FPGA** board and ASIC synthesis via Yosys toolchain with **Nangate45 45 nm** standard-cell library.

Evaluated Domains:

- **Traffic Signal Control (SUMO)**: Single crossroad, 4 lanes, minimizing waiting times.
- **Stochastic Inventory Management**: State space defined by on-hand/on-order inventory (max cap 10 units), minimizing penalties.
- **Spatial Pathfinding**: Empty 2D grids of varying sizes, finding shortest paths.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore

Performance

- Overall Gains: **Achieves up to a 60% reduction** in inference cycle count.
- **Traffic Control**: Over 50% cycle count gains with under 5% memoization depth due to highly skewed state-visitation.
- **Inventory Management**: Achieves ~35% to 48% cycle count gains at low memoization depths (10% to 20%), due to moderate state-visitation skew across its dynamic action space.
- **Pathfinding**: Yields 10% to 30% gains at similar depths for more spread-out state visitation.

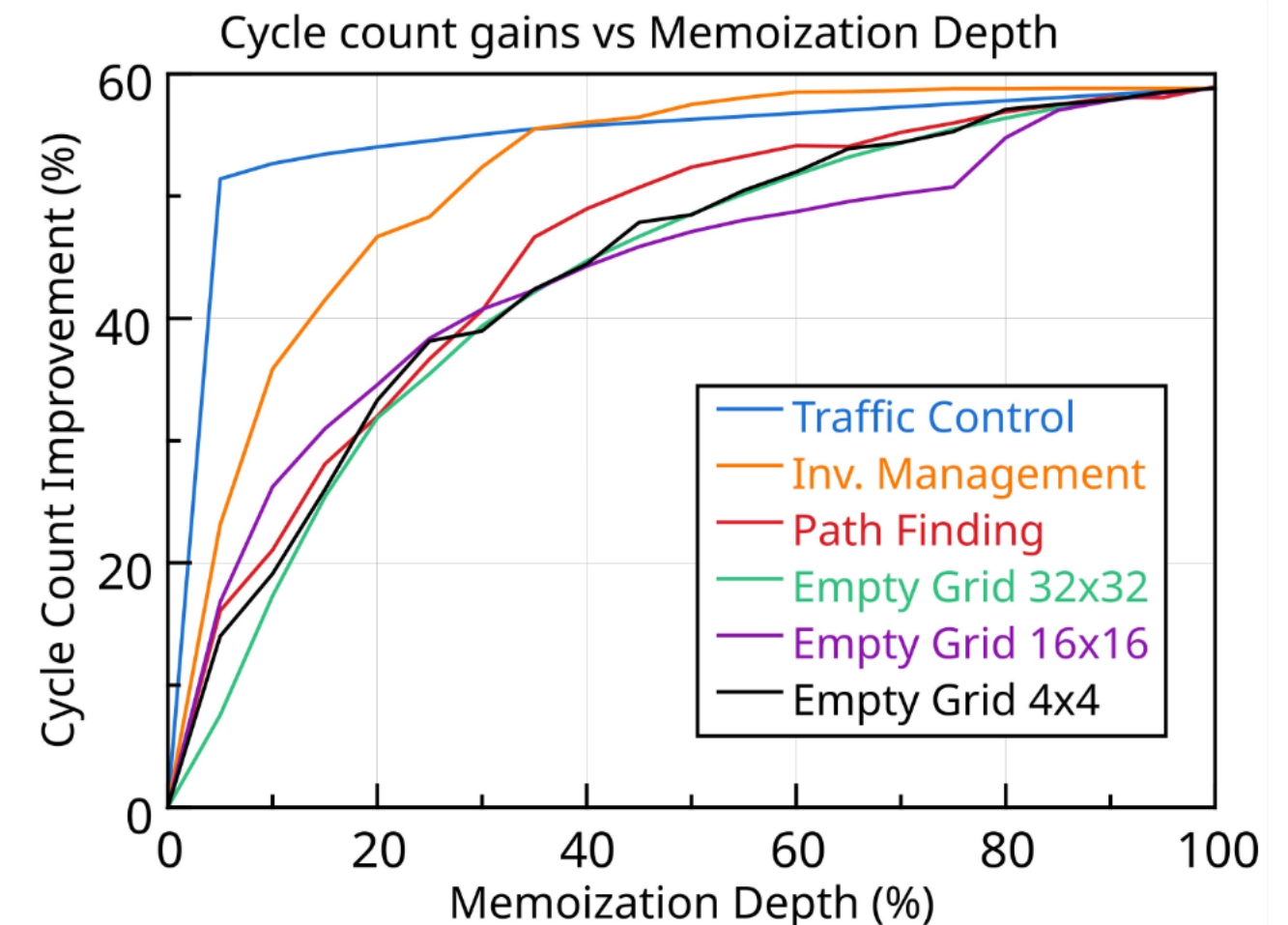


Fig. 2. Cycle Count Improvement versus Memoization Depth.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore

Hardware Metrics & Scalability (FPGA & ASIC)

FPGA			
Grid Size (N)	LUTs	Latency (ns)	Power (mW)
4	306	5.25	109
8	318	5.26	109
16	345	5.30	110
32	382	5.35	110
64	430	5.42	111
128	495	5.51	112

ASIC			
Grid Size (N)	Area (μm^2)	Latency (ns)	Power (mW)
4	2882.3	2.96	75
8	2915.6	2.96	78
16	2980.1	2.98	78
32	3065.4	2.98	79
64	3190.8	3.01	80
128	3355.1	3.05	84

*Results shown only for custom datapath overhead. RISC-V and memory overhead are not included.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

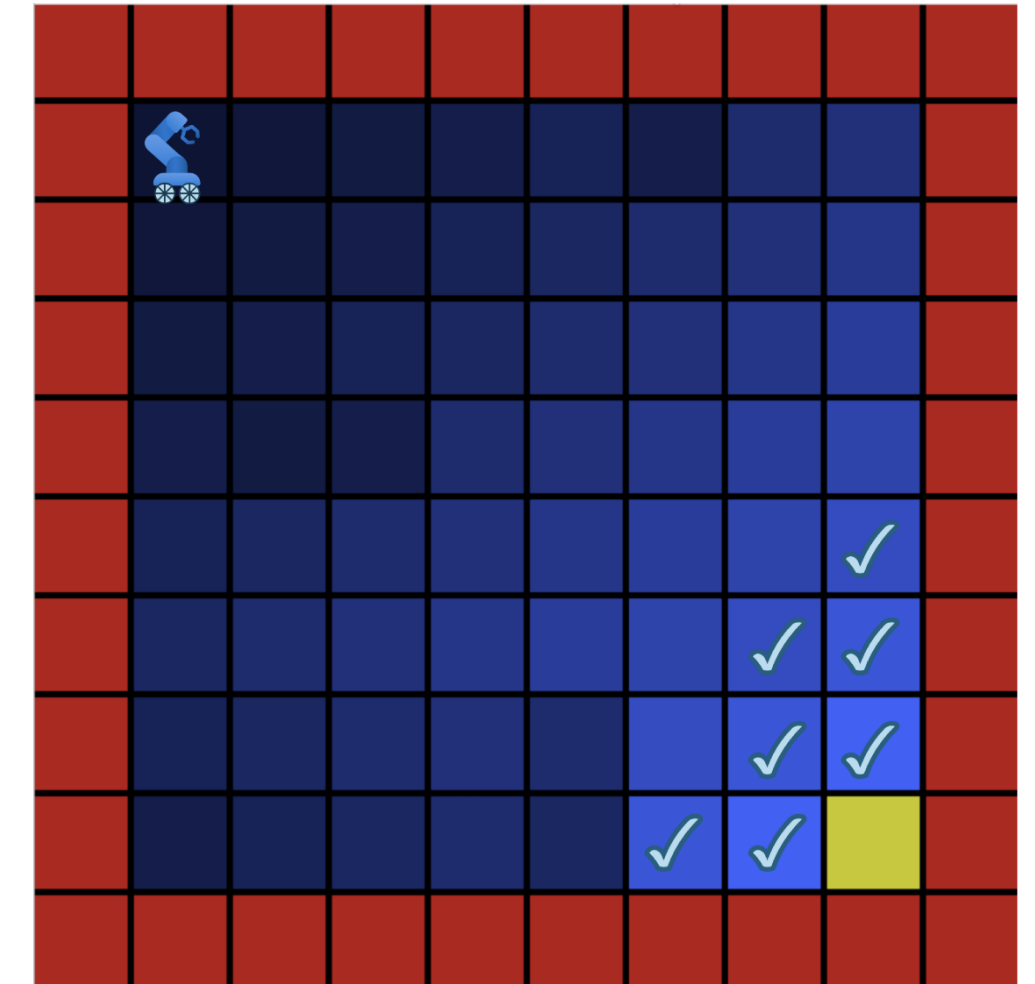
Conclusion



International
Institute of Information
Technology Bangalore

Conclusion

- This work presents an accelerator for **tabular Q-learning inference on RISC-V processors**.
- This idea relies on **skewed state-visitation distributions** to maximize hit rates.
- QXL offloads the decision-making loop to custom hardware, achieving up to a **60% reduction in cycle count**, creating a **low-power architecture for edge RL**.



ASAP 2026

Outline

Introduction

Proposed Methodology

Results

Conclusion



International
Institute of Information
Technology Bangalore



ASAP 2026



International
Institute of Information
Technology Bangalore

Thank You

A Big Thanks to ASAP Organizing Team

Contact @

Chinmay.Krishna@iiitb.ac.in

OR

scan the QR

