RP2040 SIMT ARCHITECTURE (SINGLE LANE)

Frontend: PIO State Machine

Fetch 32-bit Instr (ESP32 @ 50MB/s) → Local FIFO Decode

Register File (SRAM Bank)

```
R0-R31  : 32-bit GP [High:BF16 | Low:BF16]
P0-P7   : Predicate Masking
   SR   : System State (LANE_ID)
```

32-bit Wide

EXECUTION DATAPATH

Integer ALU

IADD / IMUL
Logic (AND/OR)
CMP / ISETP
Address Calc

BF16 Tensor Core

BFADD2
BFMUL2
BFMA2
CVT.BF16

SFU (Trans.)

EXP2 (Softmax)
SIN/COS (RoPE)
RSQRT (Attn)
GELU

Load / Store Unit (LSU)

Logic: Addr = Base + Offset + (LaneID * 4)

- **LDG/STG**: Global Coalesced Access
- **LDL/STL**: Local Thread-Private Access

Local SRAM

(Banked / Stack)

Shared VRAM

(Weights / KV Cache)

MEMORY INTERFACE