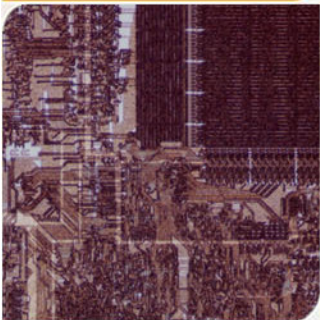


# CS6290

Fall 2009

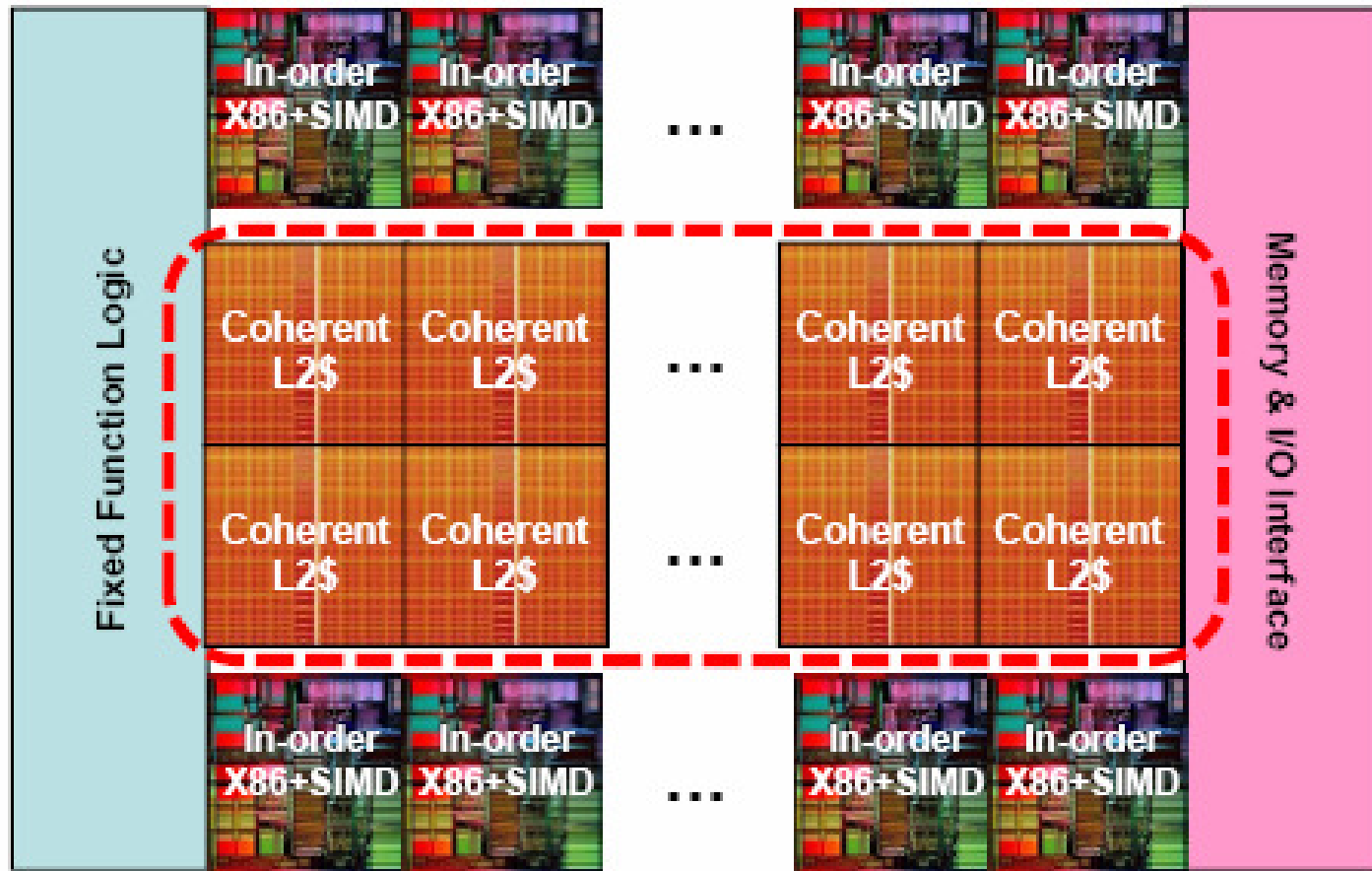
Prof. Hyesoon Kim





# CASE STUDY

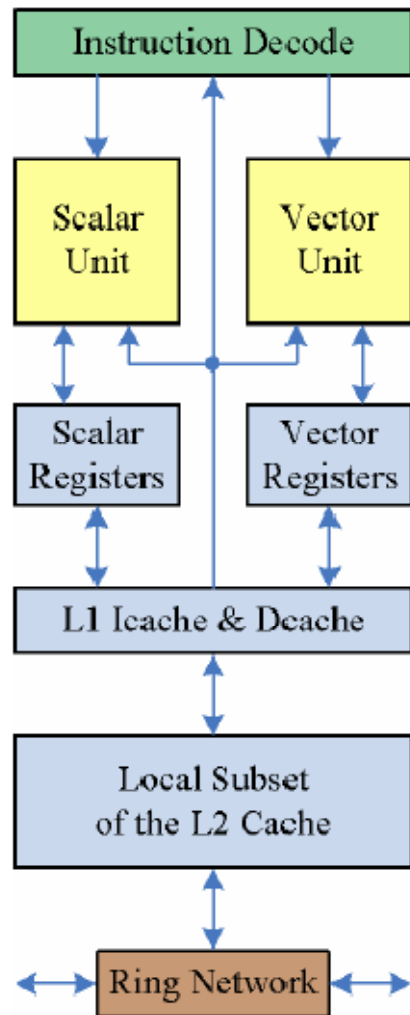
# Larrabee Architecture



- 4-way SMT in order processor with cache coherence
- Extended X86 ISA
- Fixed functions: texture filtering



# Core



- Pentium processor in-order
- Extended X86 (64bit, new instructions)
- 4-way SMT
- 32KB I-cache, 32KB D-cache (statically partitioned)
- 256KB Local L2 cache (subset of L2 cache)



# Dual issue

- U-pipe V-pipe
- Primary pipeline: All instructions (U-pipe)
- Secondary: Limited instructions (V-pipe)  
loads, stores, simple ALU operations, branches, cache manipulation instructions, and vector stores.
- Reply on compiler's pairing
  - VLIWish again

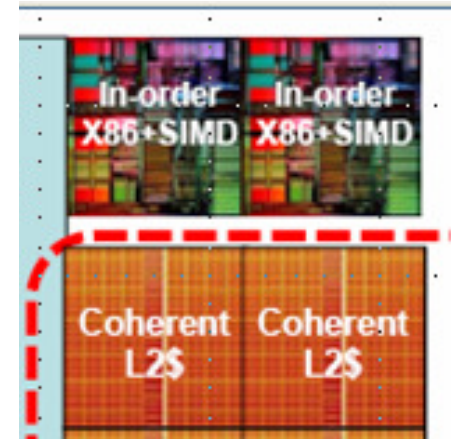


# Cache managements

- Use cache as extended register file storage
- Target for Stream applications
- Each core can
  - Fast-access its local subset of L2 (256KB) – Access other's L2 shares too
  - Control for non-temporal streaming data (SSE)
  - Prefetch to L1, or L2 only
  - Mark a streaming cache line for early eviction
  - Render target kept in L2 (e.g., FB, ZB, SB, etc)

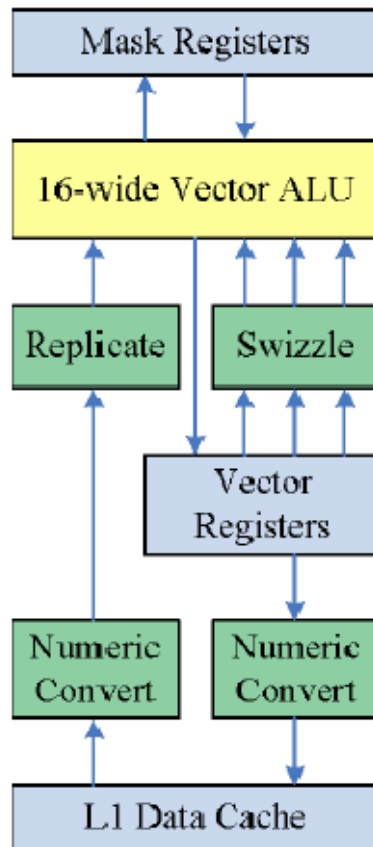
# L2-Cache and Ring network

- Global L2 cache is divided into 256KB  
Local L2 per core
- Data written by a CPU core is stored in its own L2 cache subset and is flushed from other subsets, if necessary
- Bi-direction Ring network (<16)
  - Even cycle, odd cycle: one clock per one hop
- Each ring data-path is 512-bits wide per direction
- L2 cache Insertion requires cache coherence checking
- Memory and fixed function access





# VPU (Vector Processor Units)

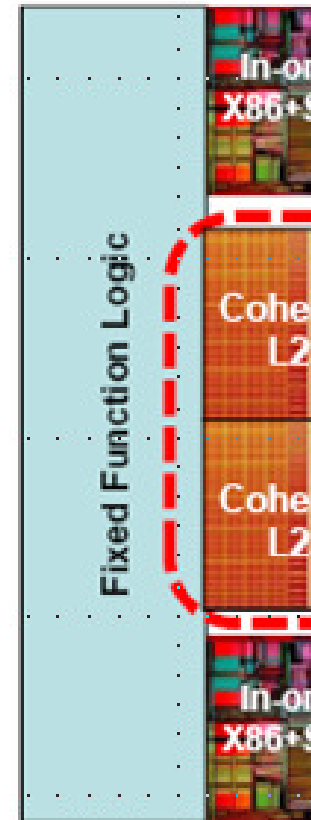


- 16-wide SIMD unit
  - 16 wide Single precision
  - 8-wide double precision
- Hardware scatter/gathering operations : 16 elements are loaded from or stored to up to 16 different addresses that are specified in another vector register.
- New instructions: fused multiply-add, and the standard logical operations, including instructions to extract non-byte-aligned
- Data can be replicated from L2 cache directly
- Free numeric type conversion and data replication while reading from memory
- Mast registers: predicated

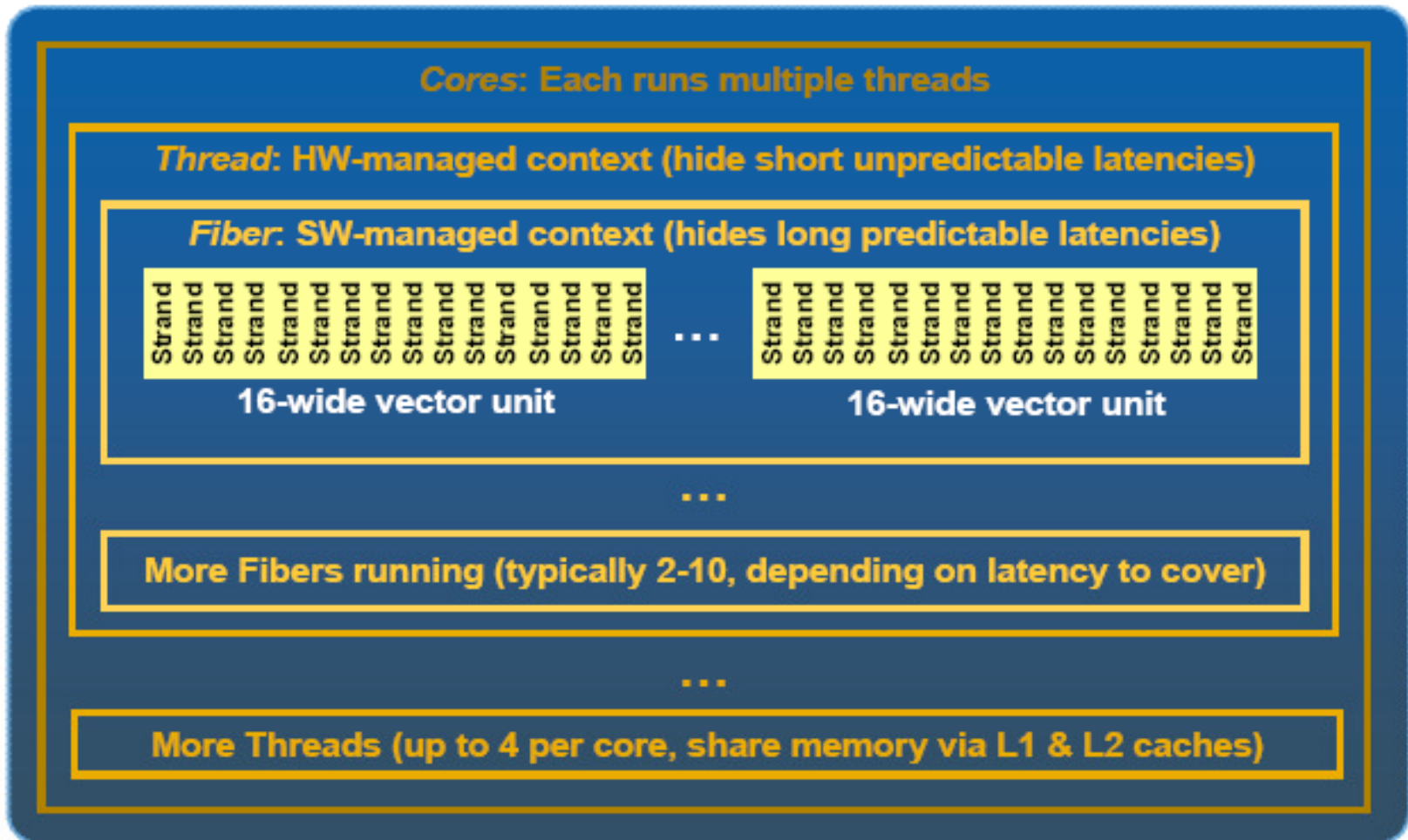


# Fixed Functions

- Use FIFO for load balancing
- No rasterizations
- Texture filtering
  - 32KB texture cache per each core
  - Core passes commands through L2 cache
- Texture unites perform **virtual to physical page** translation



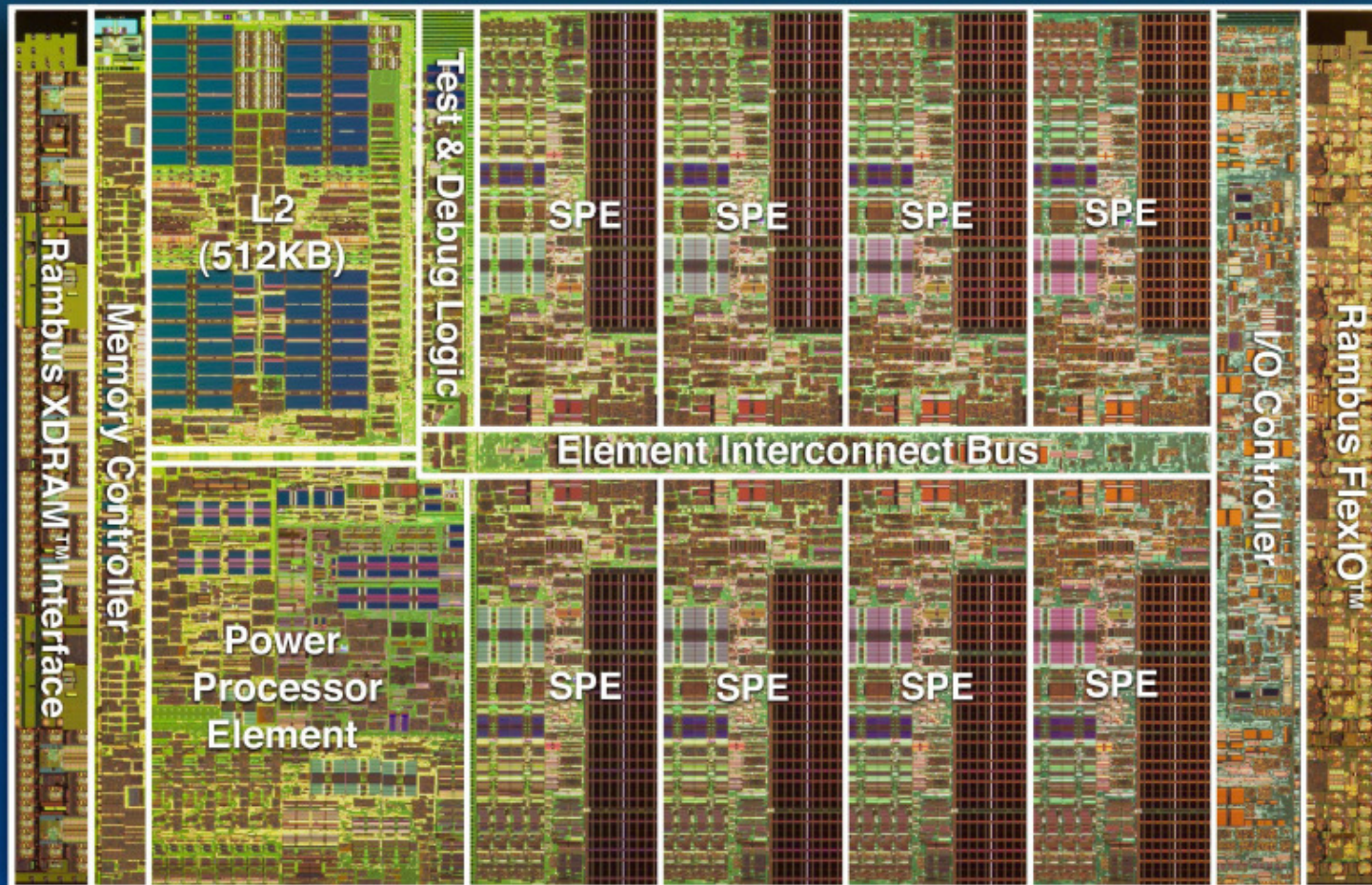
# Larrabee Data Parallelism: Threads



# Cell

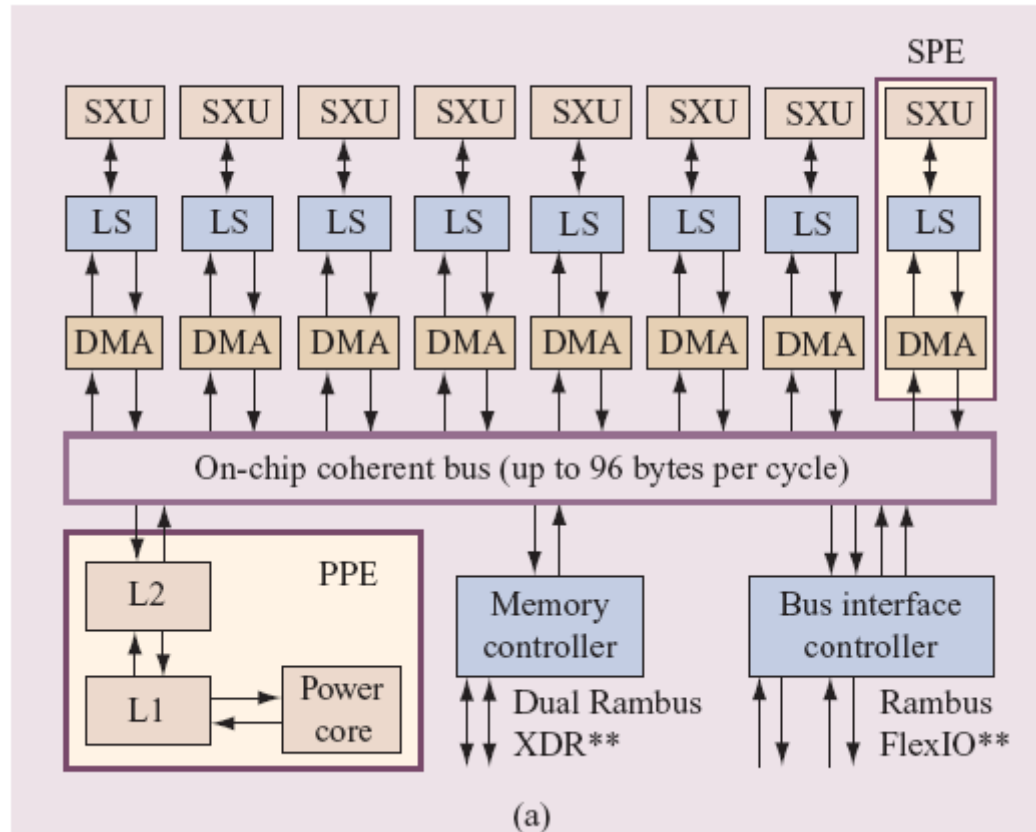


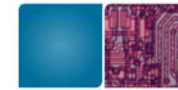
## Cell Broadband Engine Processor



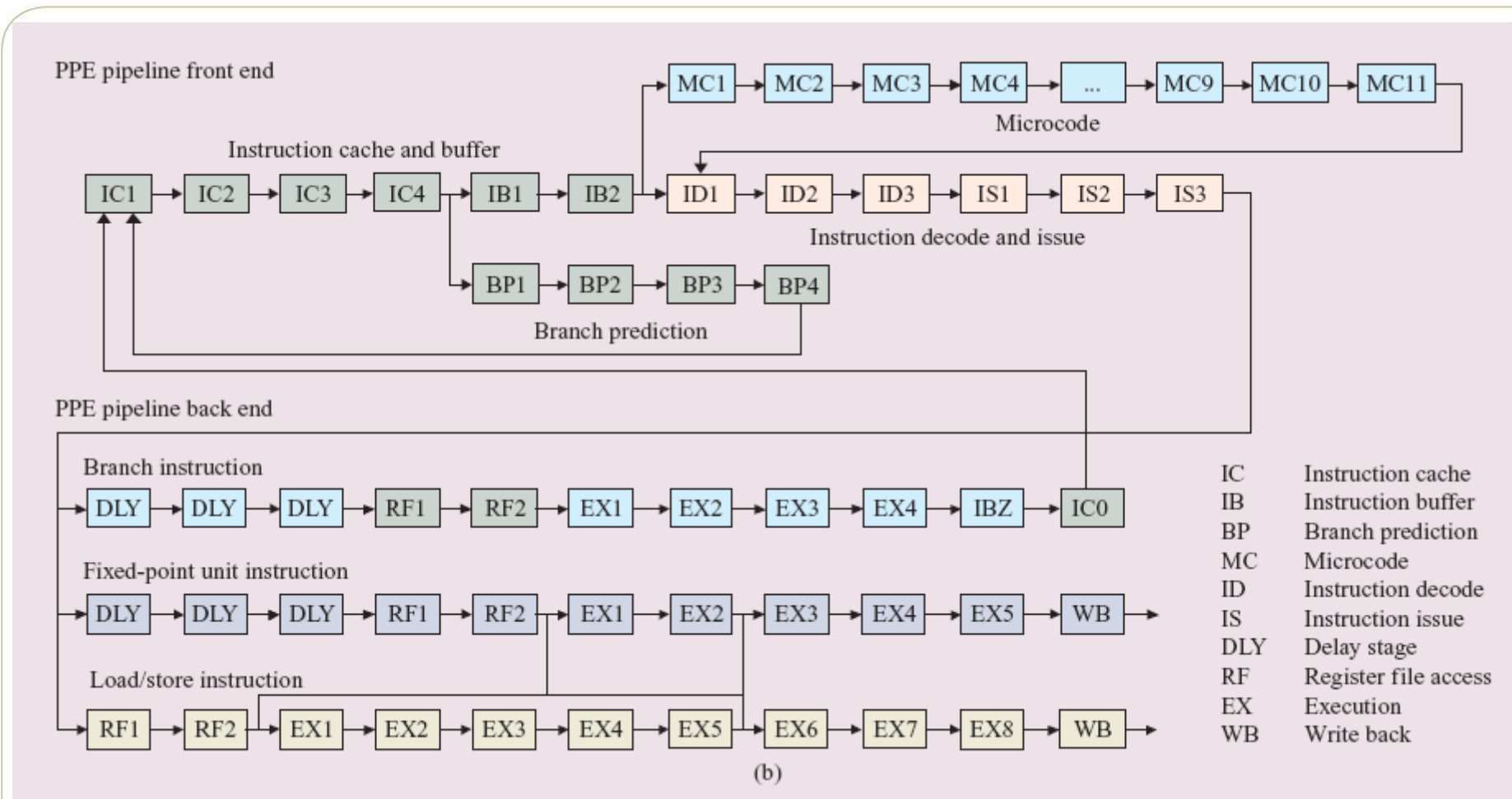


# Cell Processor Block Diagram

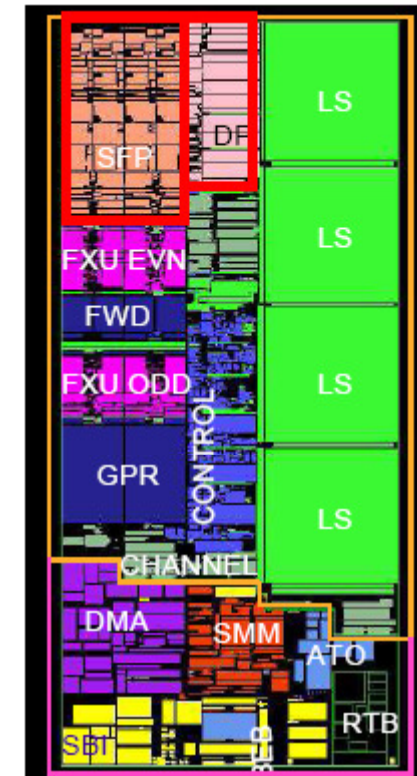
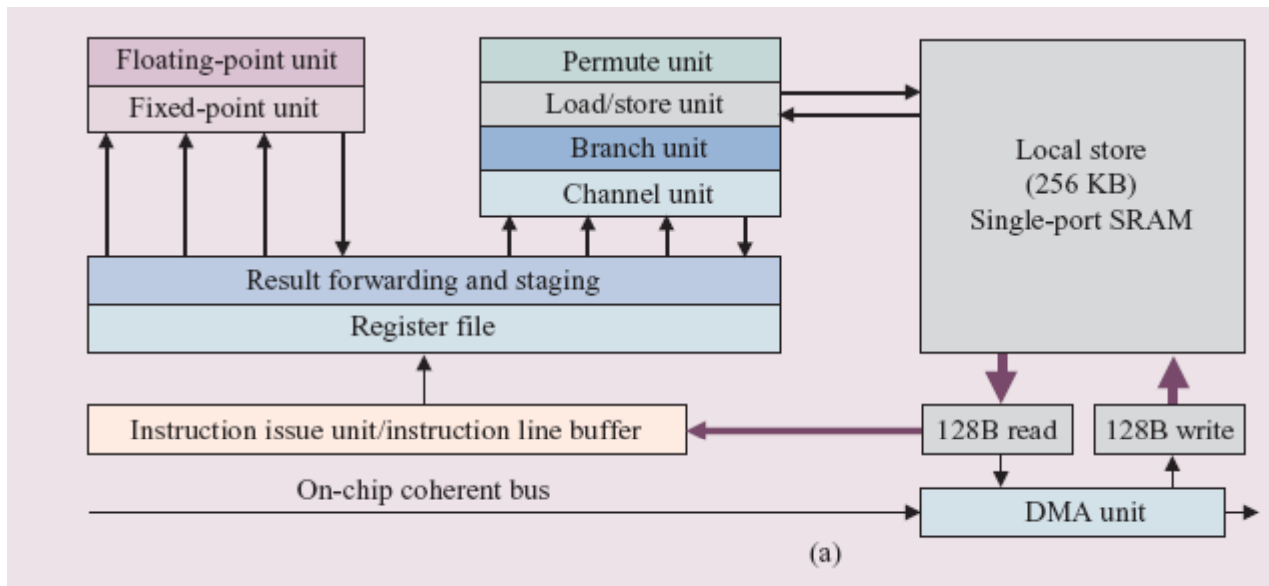




# PPE Pipeline



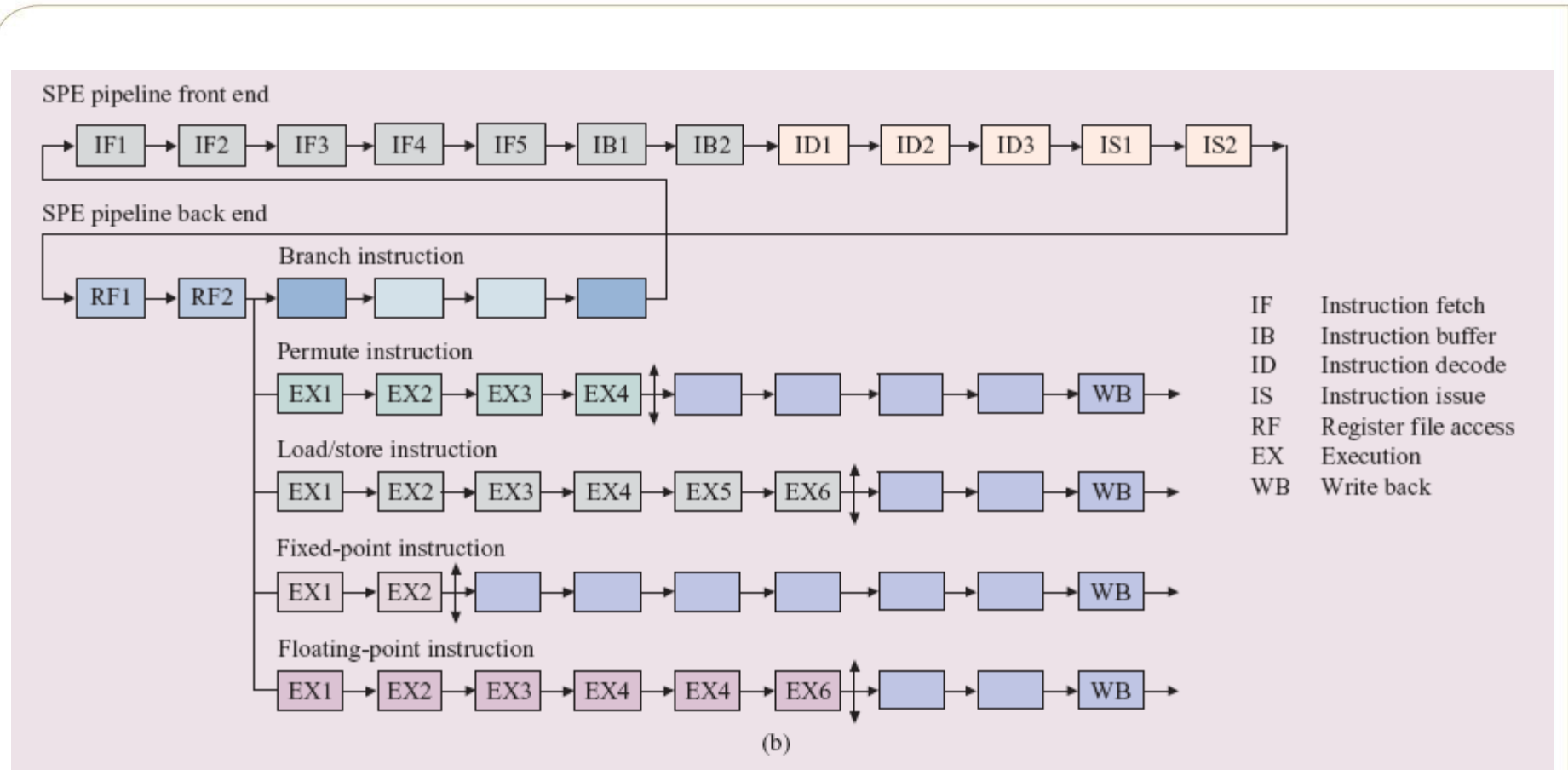
# SPE Major Units



14.5mm<sup>2</sup> (90nm SOI)



# SPE Pipeline





# Final Exam Preparation

- 30% (before mid-term)
- 70% (after mid-term)
- Homework questions. Extra homework questions (more might come so keep eye on it!)
- Buzzwords (some explanation questions)
- Programming assignments related question



# MUST KNOW

- Pipeline, OOO
- Performance metrics (IPC/CPI...)
- Power
- Tolerance metrics (MTTF)
- Cache/Cache coherence
- Memory/DRAM/Bandwidth
- Performance enhancements
  - Branch prediction, prefetcher,
- Static code improvements
- SMT/CMP