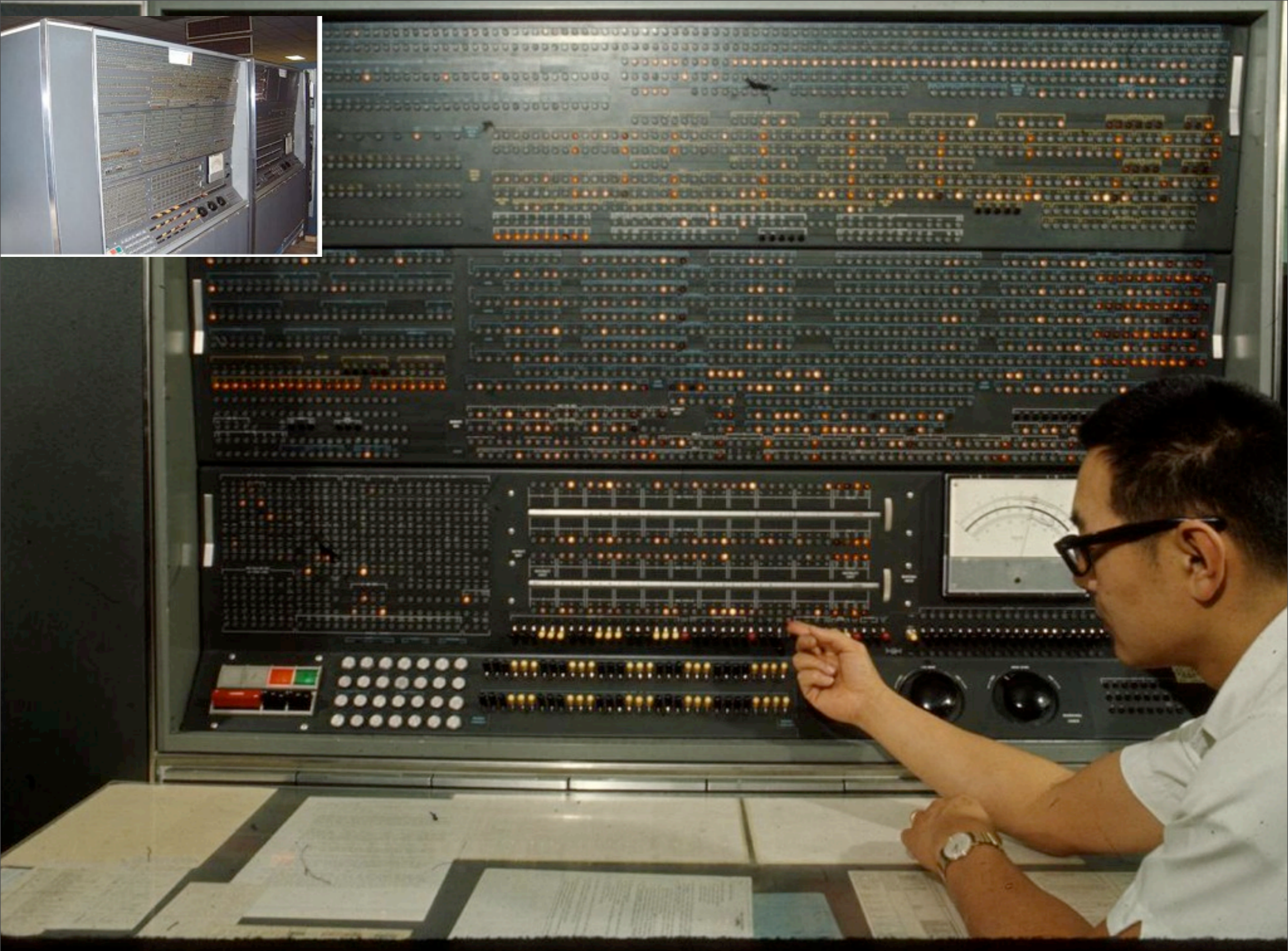
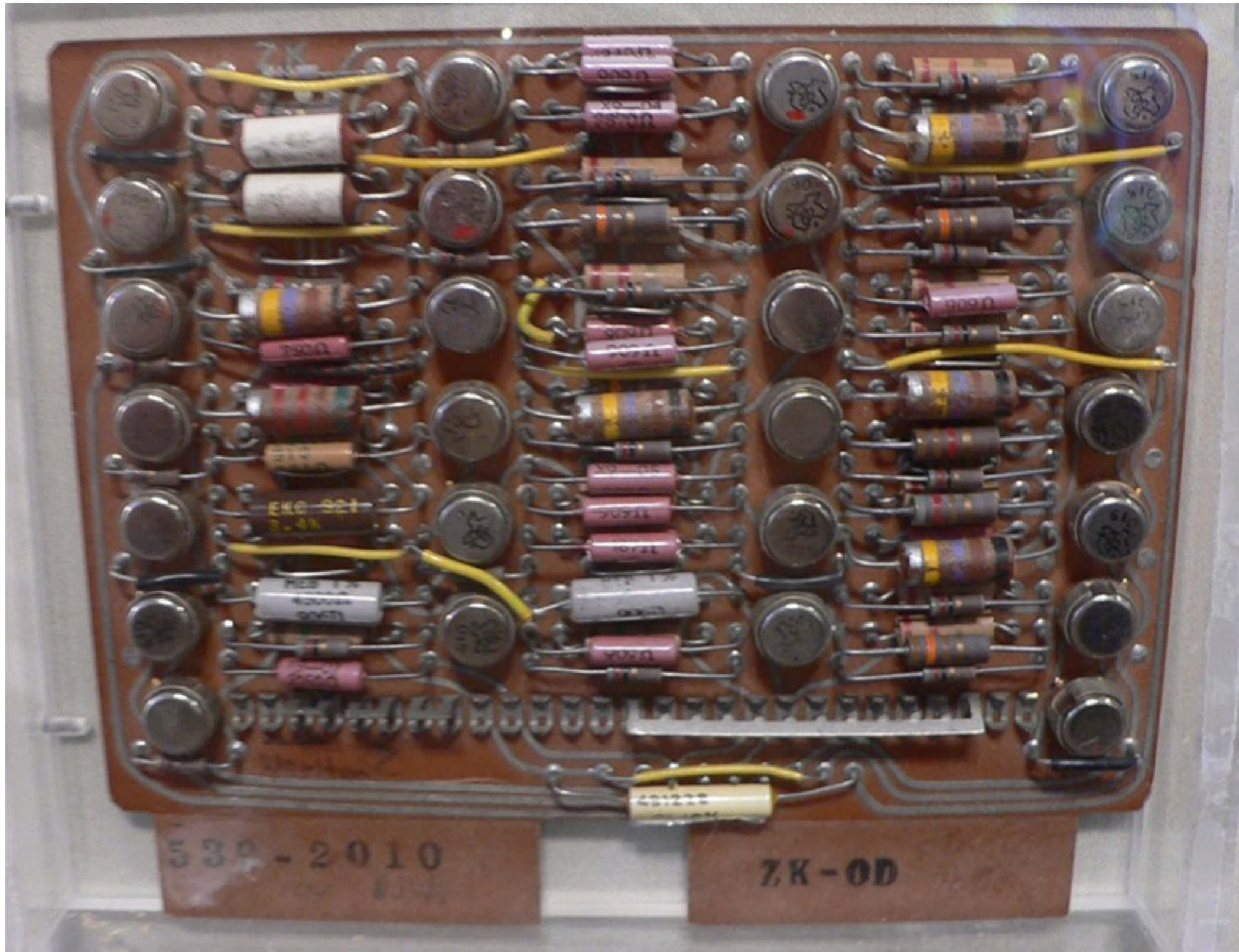
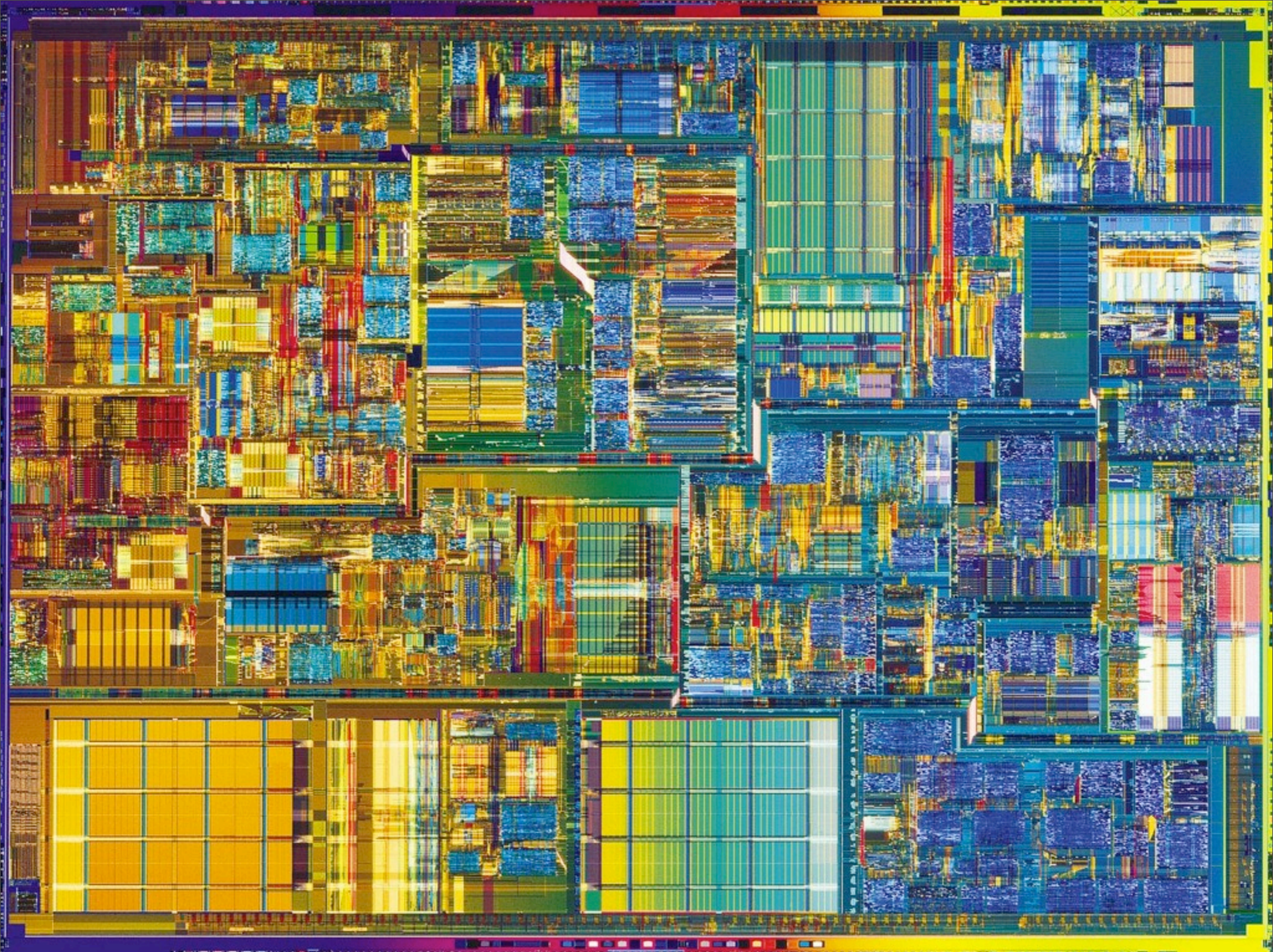


# 548 - Pipelining

## Lecture 3







# Intel Pentium 5 Prescott

Trace Cache Access,  
next Address Predict

Instruction Trace Cache

Execution Pipeline Start

Buffer Allocation &  
Register Rename

Trace Cache Branch Prediction  
Table (BTB), 1024 entries.  
Return Stacks (4 x 16 entries)  
Trace Cache next IP's (4x)

Instruction Decoder

Up to 4 decoded uOps/cycle out.  
(from max. one x86 instr/cycle)  
Instructions with more than four  
are handled by Micro Sequencer

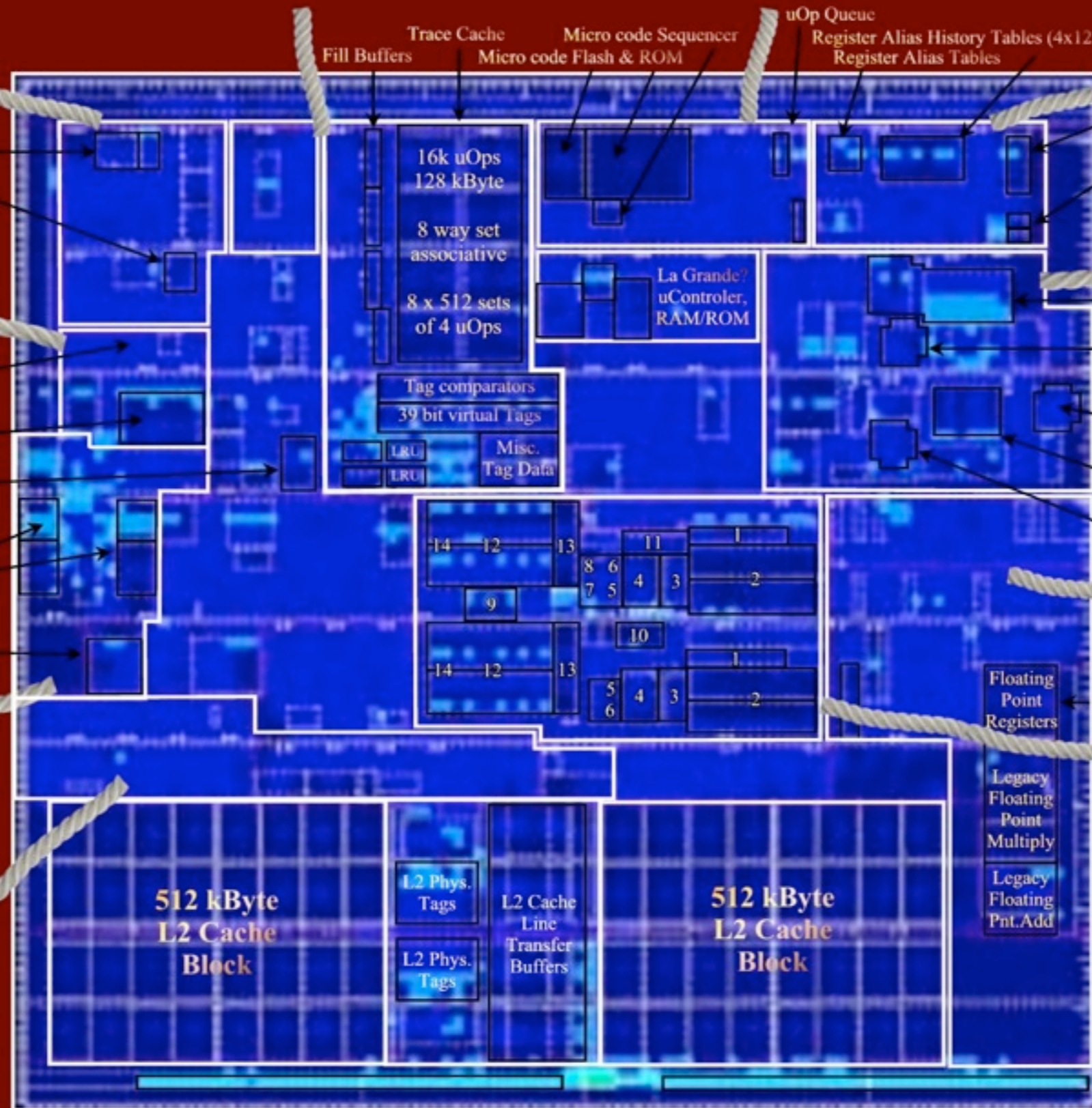
Raw Instruction Bytes in  
Data TLB, 64 entry fully  
associative, between threads  
dual ported (for loads and stores)

Front End Branch Prediction  
Tables (BTB), shared, 4096  
entries in total

Instruction TLB's 128 entry,  
fully associative for 4k and 4M  
pages. In: Virtual address [47:12]  
Out: Physical address [39:12] +  
2 page level bits

Instruction Fetch  
from L2 cache and  
Branch Prediction

Front Side Bus Inter-  
face, 533..800 MHz



Instruction Queue (for less critical  
fields of the uOps )  
General Instruction Address Queue &  
Memory Instruction Address Queue  
(queues register entries and latency  
fields of the uOps for scheduling)

uOp Schedulers

Parallel (Matrix) Scheduler  
for the two double pumped ALU's

General Floating Point and  
Slow Integer Scheduler:  
(8x8 dependency matrix)

FP Move Scheduler:  
(8x8 dependency matrix)

Load / Store Linear Address  
Collision History Table

Load / Store uOp Scheduler:  
(8x8 dependency matrix)

FP, MMX, SSE1.3

Floating Point, MMX, SSE1.3  
Renamed Register File  
256 entries of 128 bit.

Integer Execution Core

- (1) uOp Dispatch unit & Replay Buffer  
Dispatches up to 6 uOps / cycle
- (2) Integer Renamed Register File  
256 entries of 32 bit (+ 6 status flags)  
12 read ports and six write ports
- (3) Databus switch & Bypasses to and  
from the Integer Register File.
- (4) Flags, Write Back
- (5) Double Pumped ALU 0
- (6) Double Pumped ALU 1
- (7) Load Address Generator Unit
- (8) Store Address Generator Unit
- (9) Load Buffer ( 96 entries )
- (10) Store Buffer ( 48 entries )

- (13) Databus multiplexing
- (14) Cache Line Read / Write Transferbuffers and  
256 bit wide bus to and from L2 cache

- (11) ROB Reorder Buffer 4x64 entries
- (12) 16 kByte Level 1 Data cache  
four way set associative. 1R/1W

# Pipelining?

- Where: Breaking computation down into smaller pieces that can execute concurrently with other pieces of other instructions
- Where: front-end of processor, execution, retirement; the memory system

# What is a basic processor pipeline?

- fetch, decode, reg, exec, mem, writeback

# Pros?

- Utilize hardware more efficiently
- Increases throughput / Increases instruction level parallelism
- Design compartmentalization



# Cons?

- Increases overall complexity
  - Hazards: forwarding, stall (WaW, ...)
- Increases size
  - Forwarding networks, hazard detection,
- Increases latency
  - Branch misspeculation penalty
  -

# Complexities?

- Can be difficult to split the logic (design complexity)
- Rollback (miss-speculation)
- Speculative scheduling
-

# “Limits”?

- Fixed cost: latch overhead (area, power)
- Loops : “Lose Loops Sink Chips”
  - Load-use, branch resolution, high-latency ops
- Balancing the whole-system design