Performance Design of Streaming Systems

by *Gerrit Muller* HSN-NISE
e-mail: gaudisite@gmail.com
www.gaudisite.nl

Abstract

Video and audio content is a continuous stream of data. Video and audio systems have to be designed in such a way that these streams are processed and delivered continuously. We discuss the pipelining of multiple functions and the impact on bus bandwidth, memory use and CPU overhead.

August 21, 2020
status: preliminary draft
version: 0

Case Video Streaming

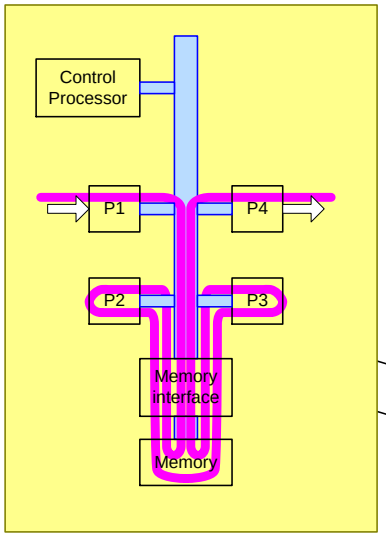
Video Streaming

Hard real-time performance for distributed system with memory-bus

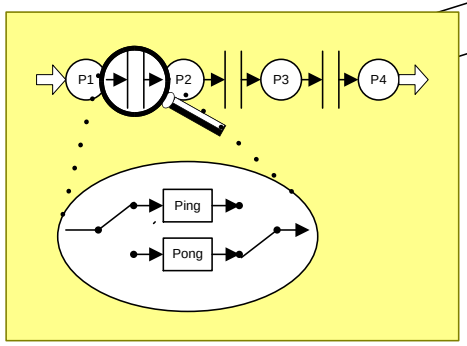
Trade-off of between latency, memory and overhead

Performance consideration in increasing detail

Case Video Streaming: Performance Design



Fixed HW diagram



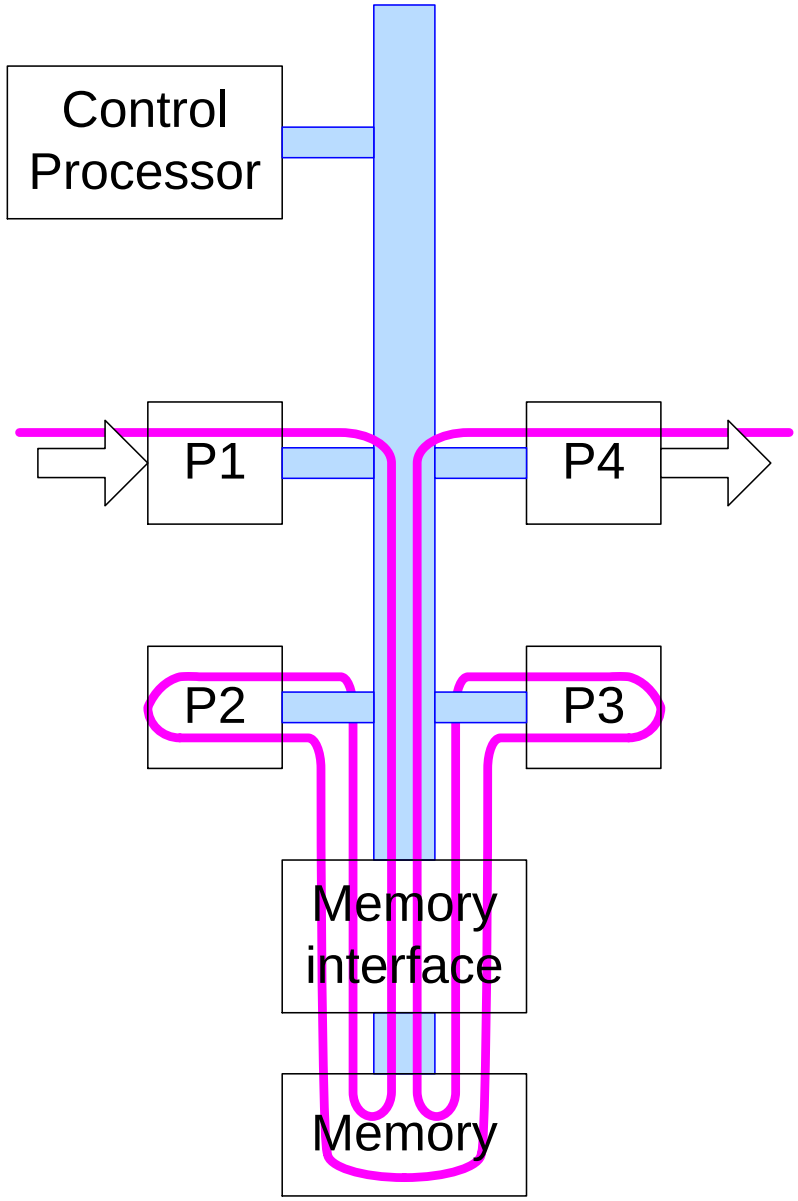
Fixed algorithmic flow

Design process

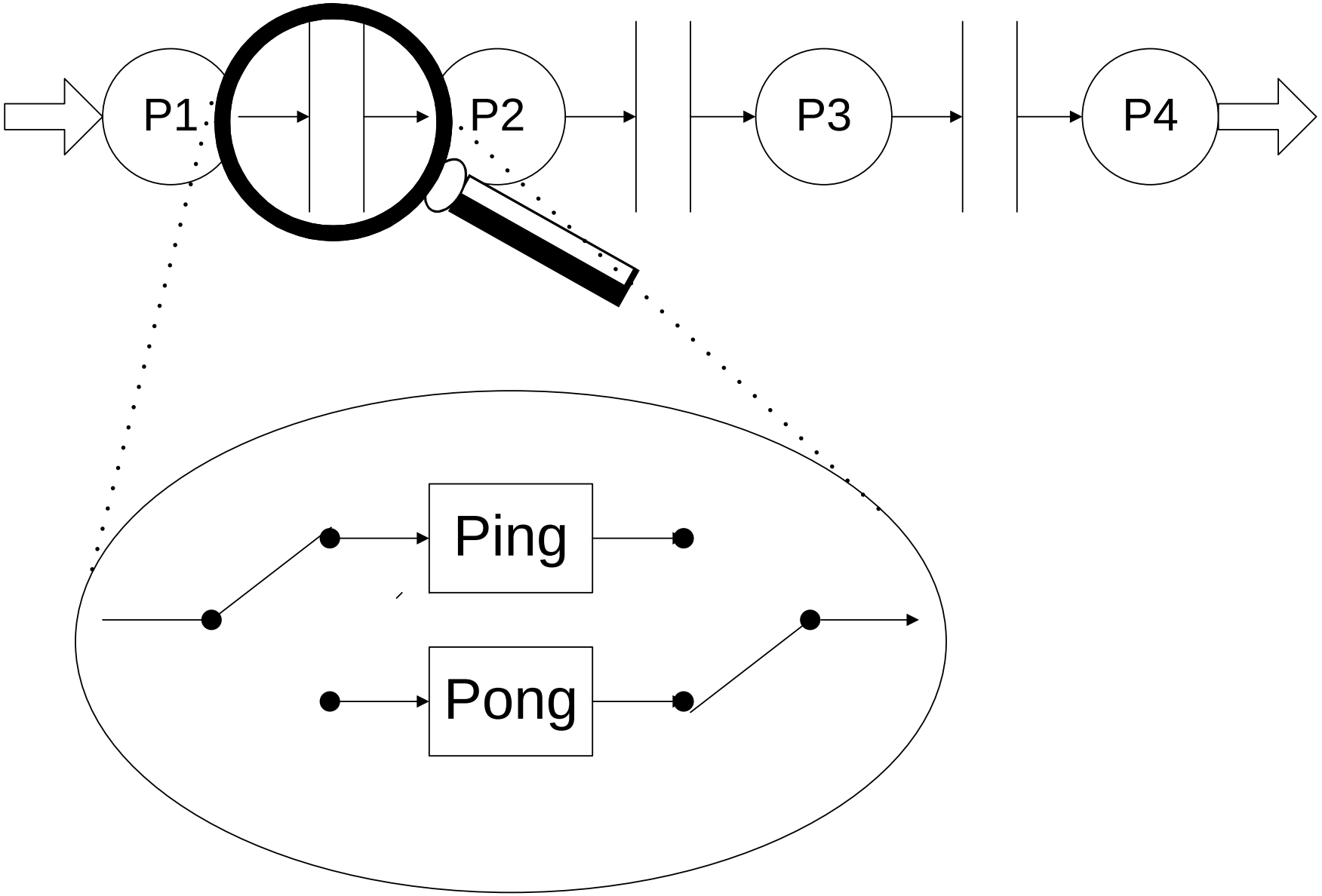
- Latency
- CPU Overhead
- Memory use
- Bus load

- Granularity
- Synchronisation strategy

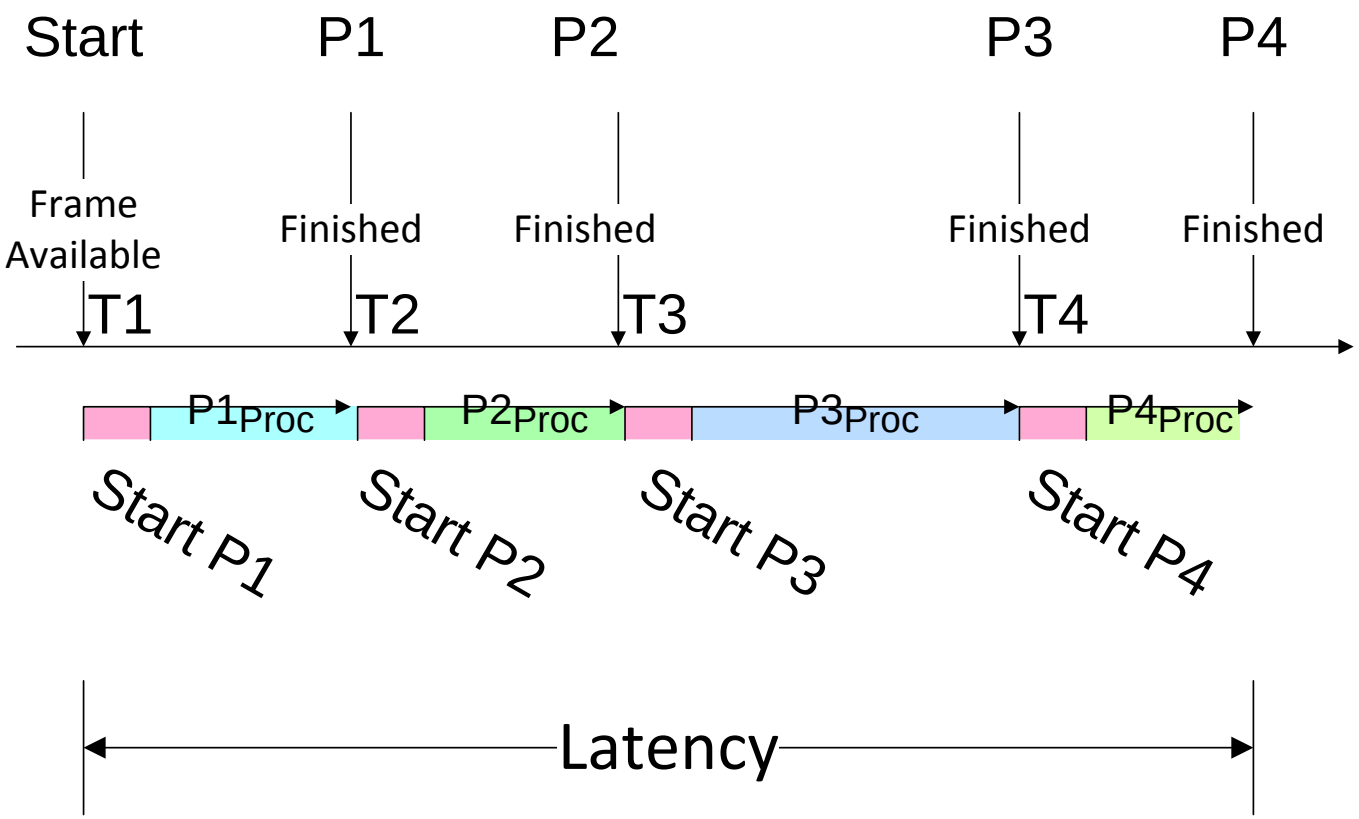
Video Streaming: HW Diagram



Video Streaming Pipeline



Video Streaming: Latency



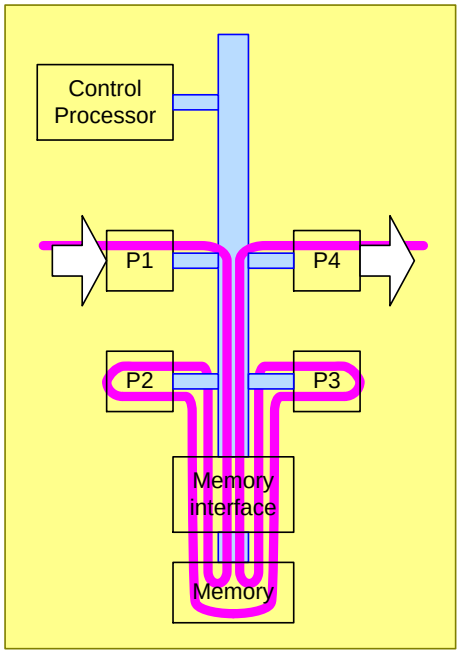
$P3_{Proc} \lesssim \frac{1}{2} T_{frame}$

$Latency \approx 2\ T_{frame}$

Legenda

- Task switch
- Process 1
- Process 2
- Process 3
- Process 4

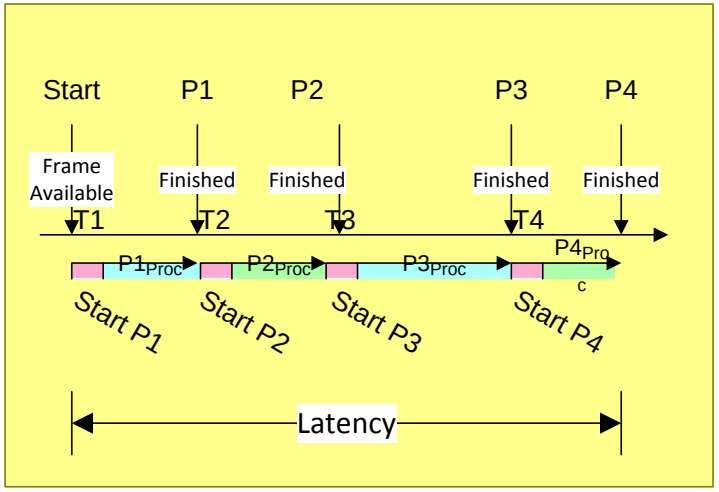
Video Streaming: Resources



$$\text{Overhead} = (T1 + T2 + T3 + T4) * \text{Frame rate}$$

$$\text{Memory usage} = 3 * 2 * \text{Frame size}$$

$$\text{Bus load} = \frac{3 * 2 * \text{Frame size} * \text{Frame rate}}{\text{Bus capacity}} \quad \%$$



T1 .. T4 = Overhead to start P1 .. P4

Latency Calculation

Latency = Nr. of Proc. blocks (4) * processing time per block (0.01s) * frame fragment (1)

Memory = (Nr. of Proc. blocks (4) - 1) * 2 * pixels per frame (414720) * frame fragment (1)

Overhead = Nr. of Proc. blocks (4) * task switch time (10 µs)

Overhead (%) = Overhead (40 µs) / Latency (40 ms)

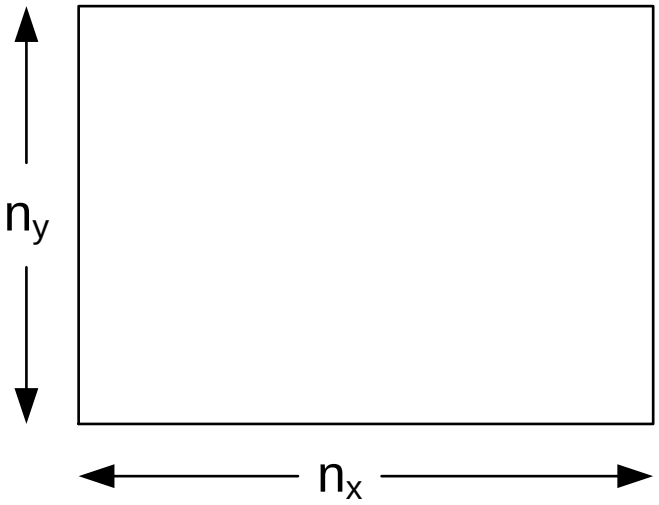
Busload = Memory usage (2430 kB) * frame fragment (1) * (frames/s) (25) / BusCapacity (500MB/s)

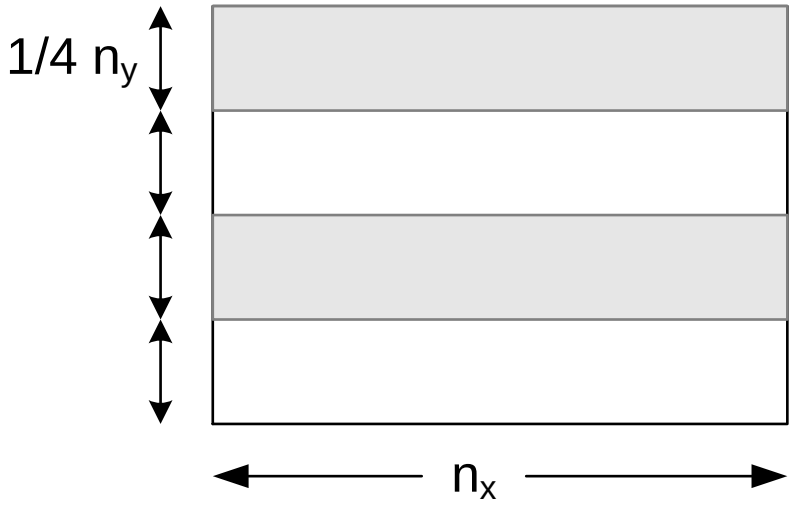
(mind the units, ms vs. µs and kB vs MB!)

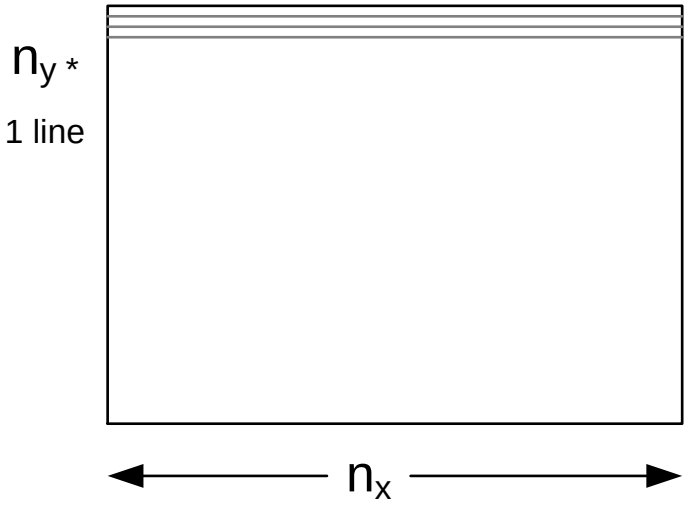
lines	576		pixels per frame	414720
pixels per line	720		Memory in kB	405
			Memory in MB	0.40
frame time	0.04		frame time in µs	40000
task switch time (µs)	10			
Processing per block	0.01		Processing in µs	10000
Bus capacity (MB/s)	500			
Line time (µs)	69			
Frame fragment	Full frame : 1			

Nr of Processing Blocks	4
Latency (ms)	40
Memory (kB)	2430
Overhead (µs)	40
Overhead (%)	0
Busload (%)	12.15

Exercise







Calculate:
Processing time
Overhead
Memory Use
Latency
for buffer size = 1/4 frame size
and for
buffer size = 1 video line

Exercise Worksheet

Nr of Processing Blocks		4		20
Block size				
	Latency (ms)	40		200
Frame	Memory (kB)	2430		15390
1	Overhead (µs)	40		200
	Overhead (%)	0		0
	Busload (%)	12.15		76.95
	Latency (ms)			
½ Frame	Memory (kB)			
2	Overhead (µs)			
	Overhead (%)			
	Busload (%)			
	Latency (µs)			
Line	Memory (kB)			
576	Overhead (µs)			
	Overhead (%)			
	Busload (%)			

lines	576
pixels per line	720
pixels per frame	414720
Memory in kB	405
Memory in MB	0.395508
frame time	0.04
frame time in µs	40000
task switch time (µs)	10
Processing per block	0.01
Processing in µs	10000
Bus capacity (MB/s)	500
Line time (µs)	69

Changing the Buffer Size

buffersize = ¼ frame

Processing time = ¼ * original (per fragment)
Latency ~ ¼ * original
Overhead = 4 * original
Memory use = ¼ original

buffersize = 1 line

Processing time = $^{1}/_{576}$ * original (per fragment)
Latency ~ $^{1}/_{576}$* original **+ overhead**
Overhead = 576 * original
Memory use = $^{1}/_{576}$ original

Summary Case Video Streaming

Video Streaming

Properly designing distributed HRT systems requires trade-off between latency, overhead, and memory needs

Performance model detailing dependent on significance of impact factors

Colofon

The ASP™ course is partially derived from the EXARCH course developed at *Philips CTT* by *Ton Kostelijk* and *Gerrit Muller*.

Extensions and additional slides have been developed at *ESI* by *Teun Hendriks*, *Roland Mathijssen* and *Gerrit Muller*.