Modeling and Simulation of a Manycore PCR-CE Lab on Chip for DNA Sequencing using SystemC/SystemC-AMS

amr.habib@soc.lip6.fr

francois.pecheux@lip6.fr

BMAS 2010 - September 24, 2010 – San Jose

Presentation Overview

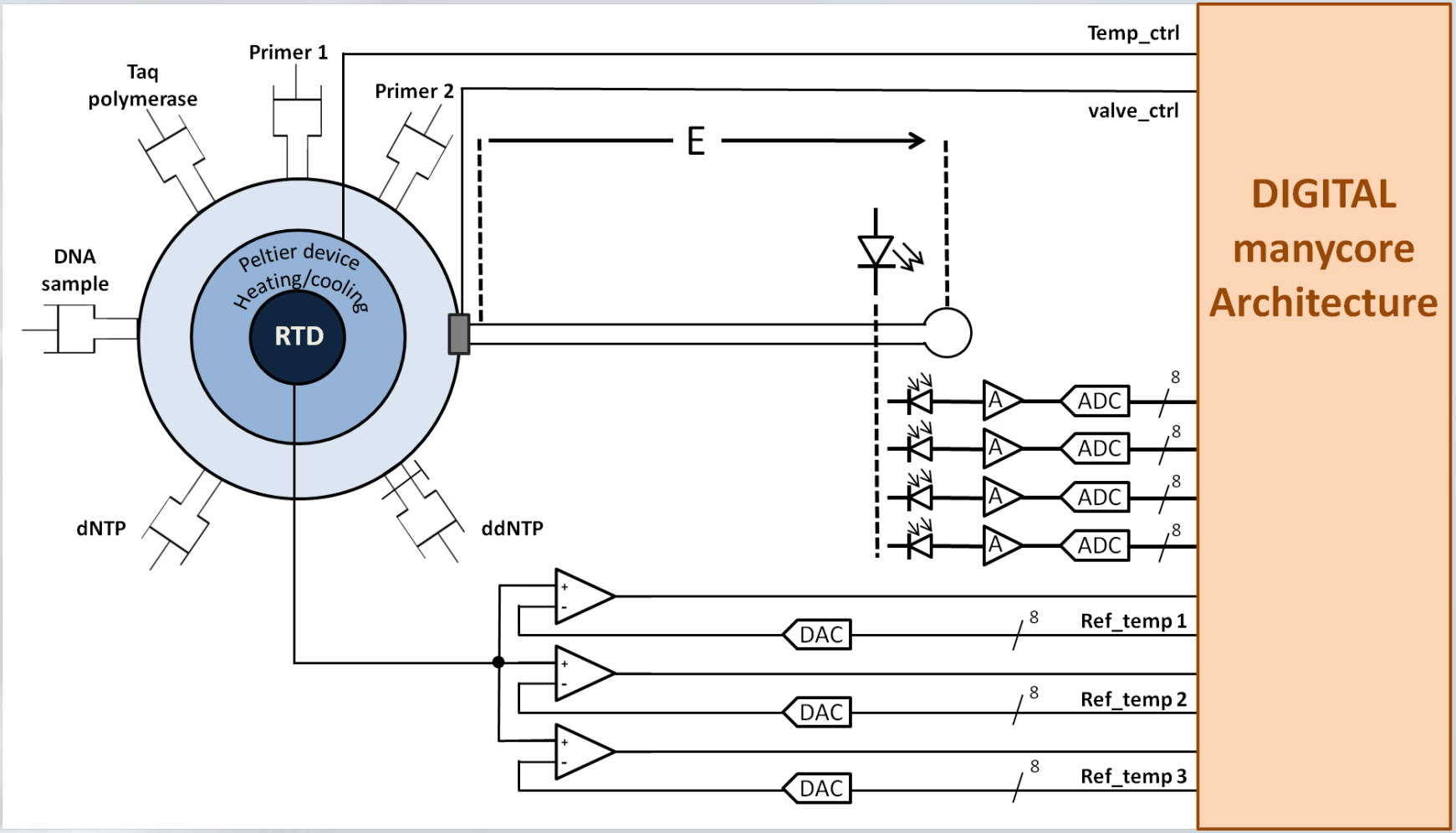
-Part I: System analysis
- Lab-on-Chip for DNA analysis
- DNA Amplification by PCR (Polymerase Chain reaction)
- Separation by CE (Capillary electrophoresis)
- Optical detection and shaping
- Multicore processing and PCR control
- Multithreaded Smith-Waterman

Part II: Mathematical Modeling (lightspeed presentation)

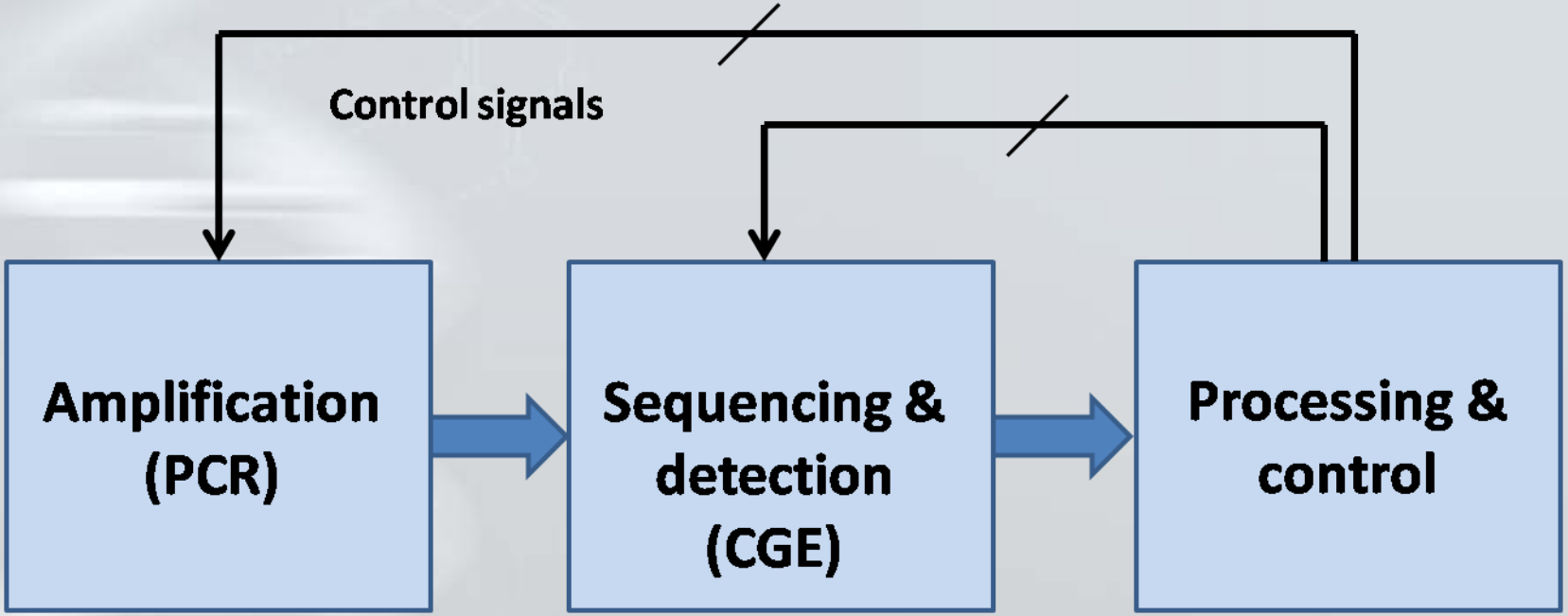
Part III: Hardware/Software Modeling
- SystemC-AMS for AMS parts
- SystemC + SOCLIB for digital parts
- Embedded application: Multithreaded Smith-Waterman

Results & conclusion

Part I: System Analysis

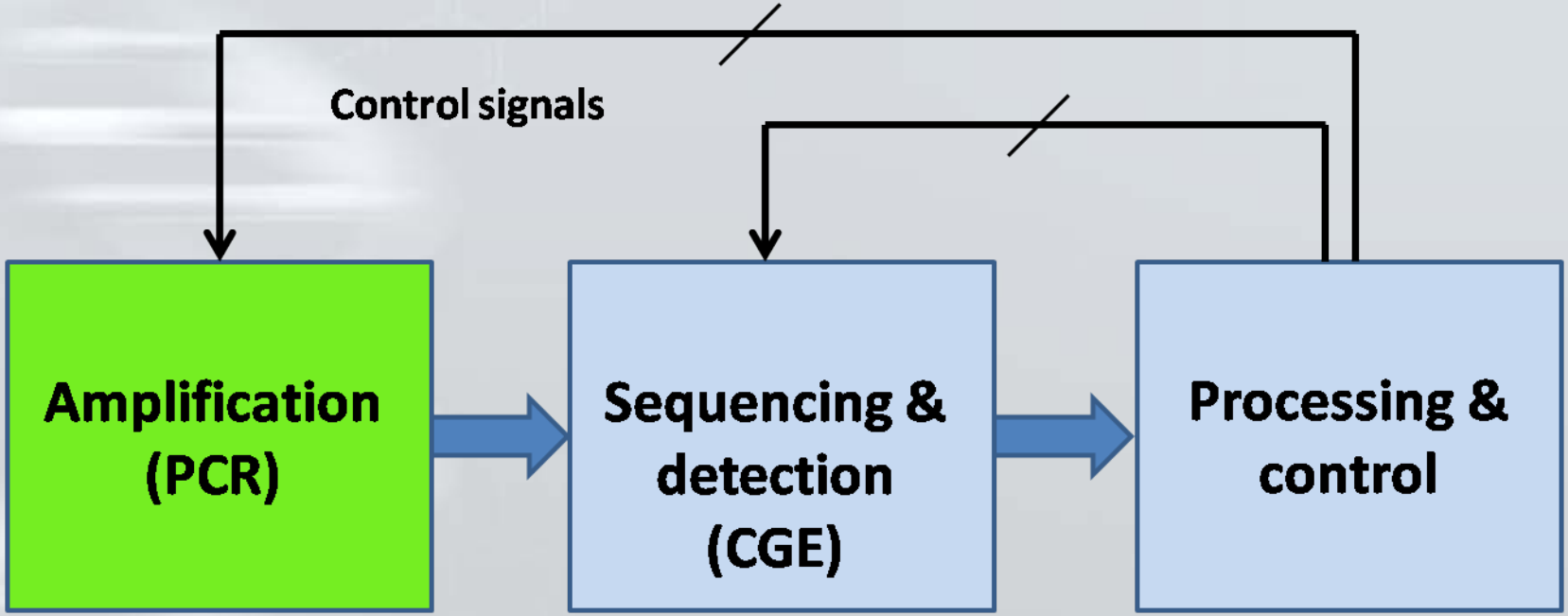


Part I: System Analysis



Part I: System Analysis

Amplification by Polymerase Chain Reaction (PCR)



Polymerase Chain Reaction

Principal reagents:

- DNA template
- Two primers
- Taq polymerase
- dNTPs

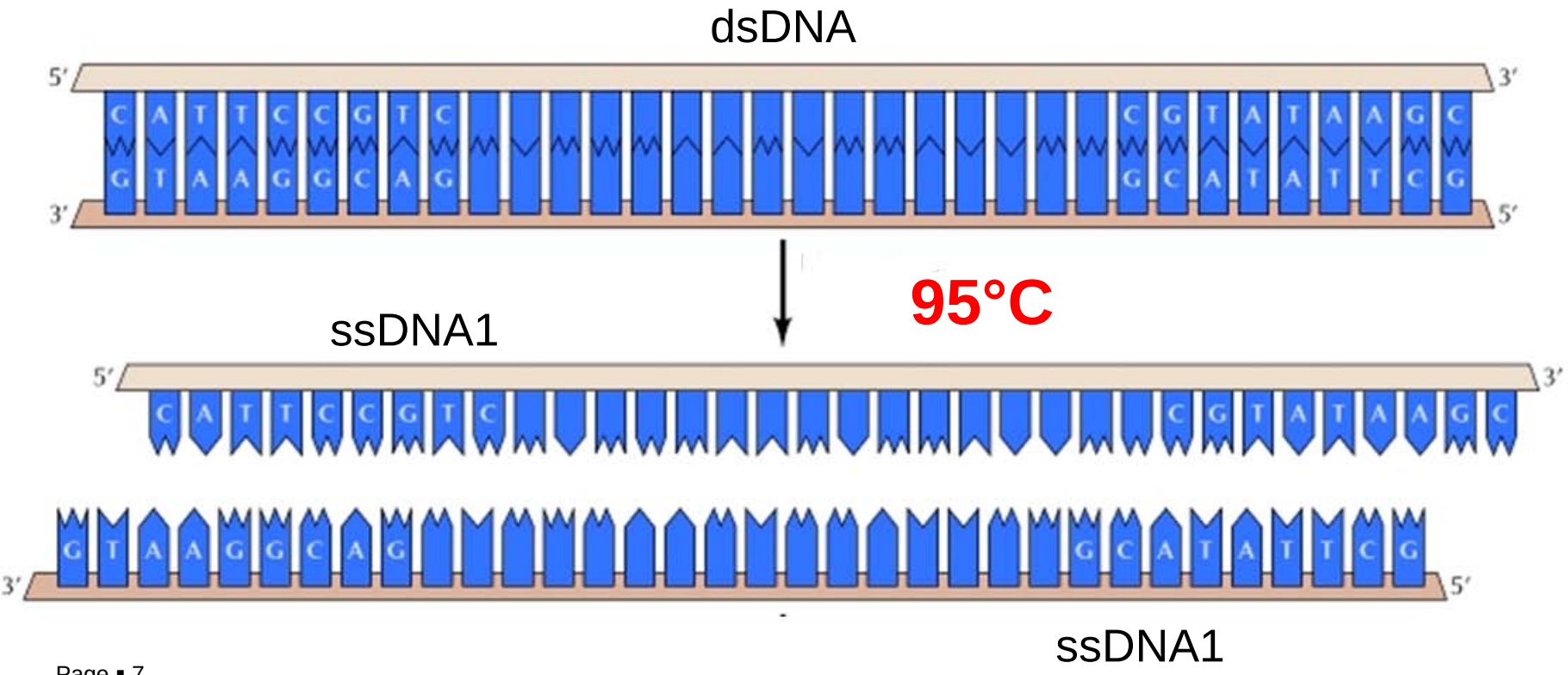
Each PCR Cycle is composed of three phases:

1. Denaturation phase
2. Annealing phase
3. Extension phase

Part I: System Analysis

Polymerase Chain Reaction

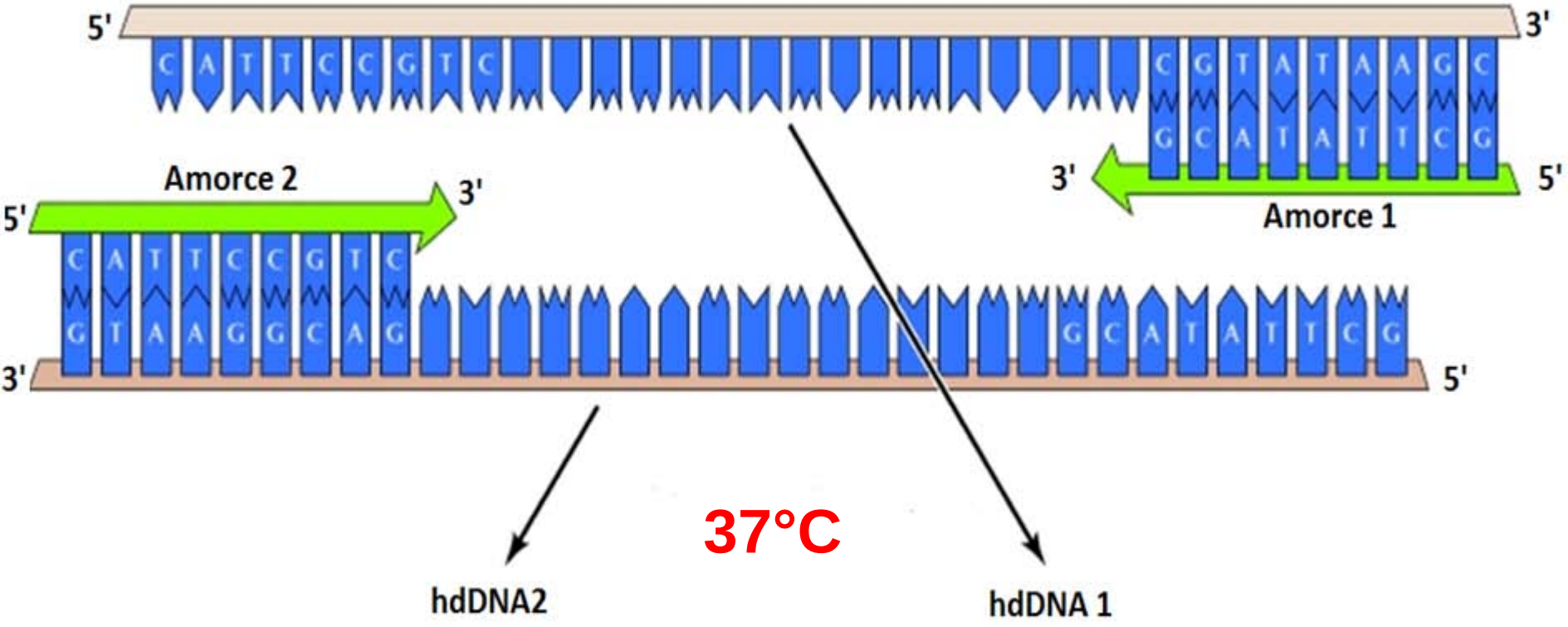
1. Denaturation



Part I: System Analysis

Polymerase Chain Reaction

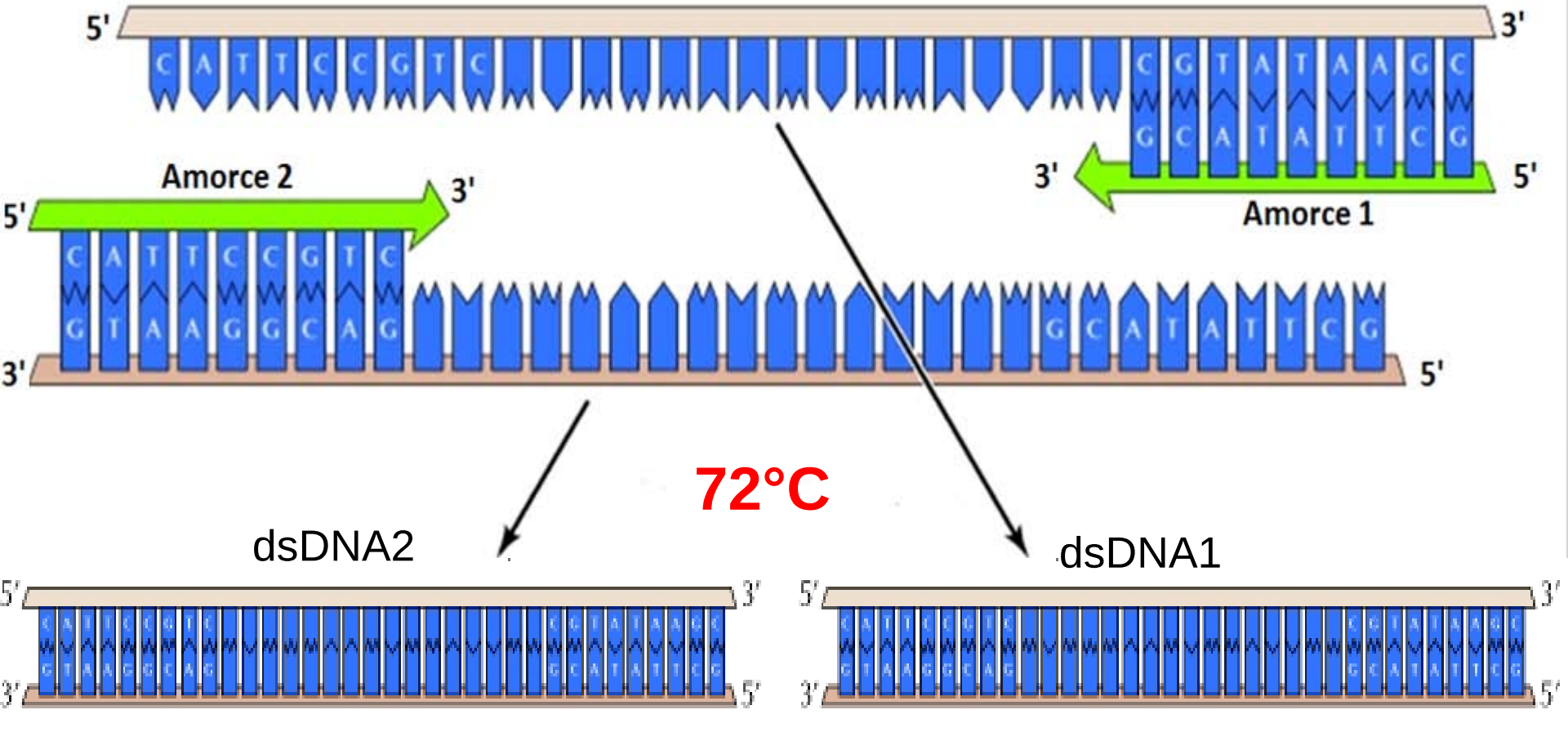
2. Annealing



Part I: System Analysis

Polymerase Chain Reaction

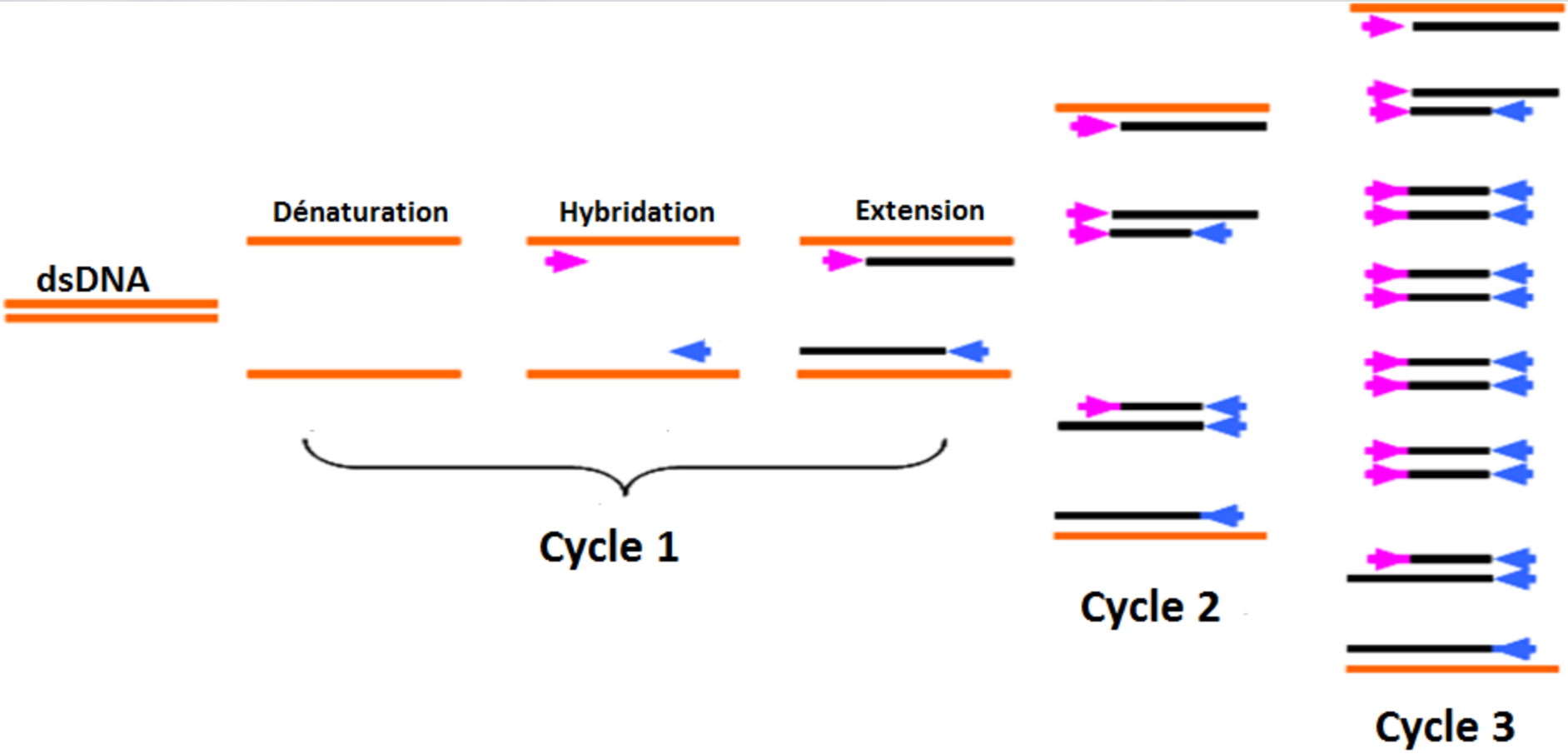
3. Extension



Part I: System Analysis

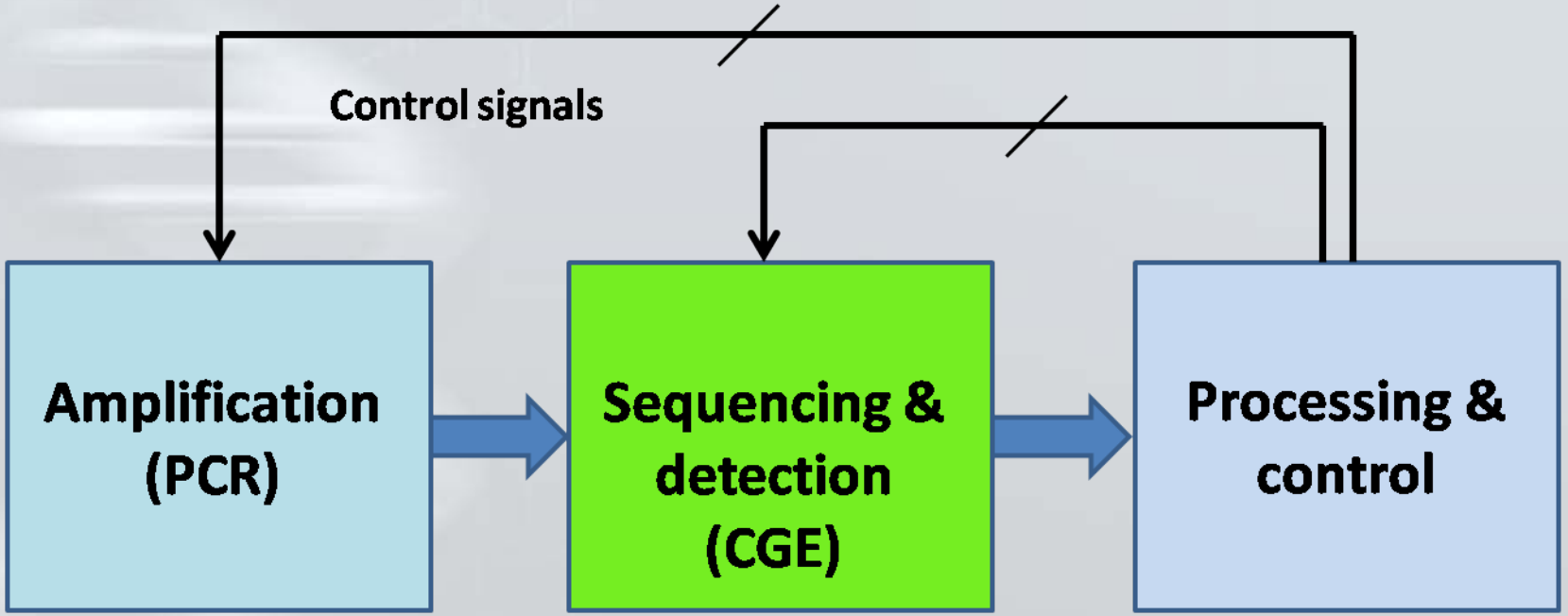
Polymerase Chain Reaction

Cycles of PCR



Part I: System Analysis

2. Sequencing and detection of DNA

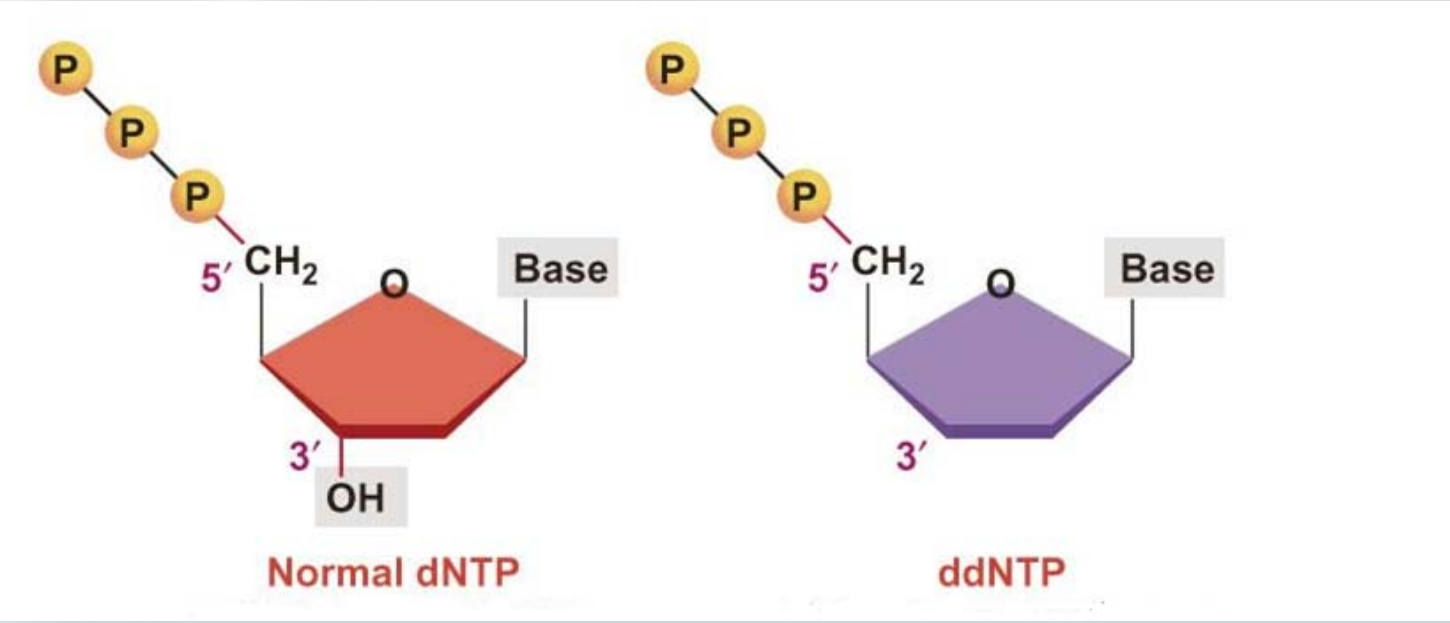


Part I: System Analysis

2. Sequencing and detection of DNA

Sanger method

- Use of « ddNTP » chain terminators

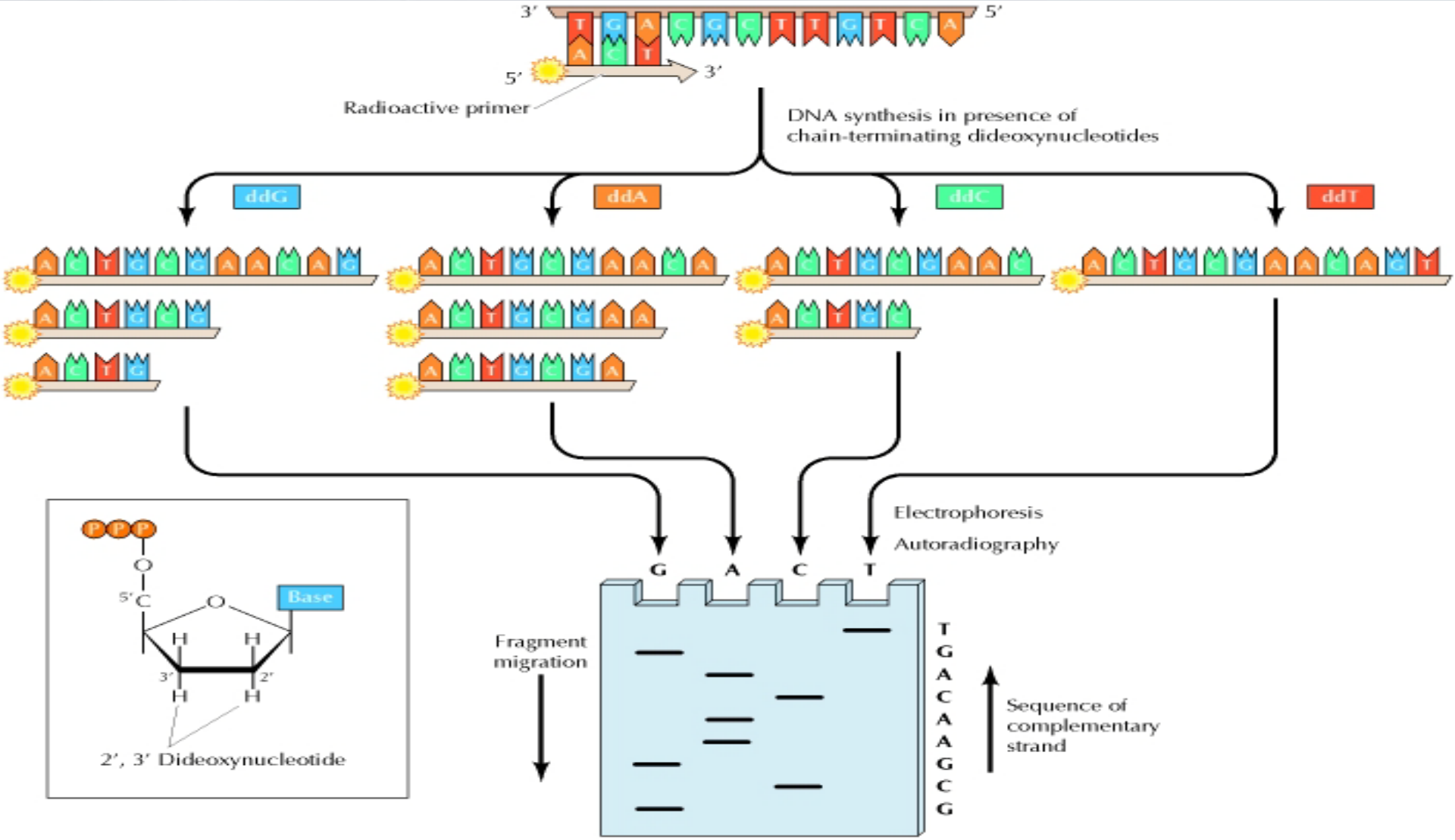


Extends the chain

terminates the chain

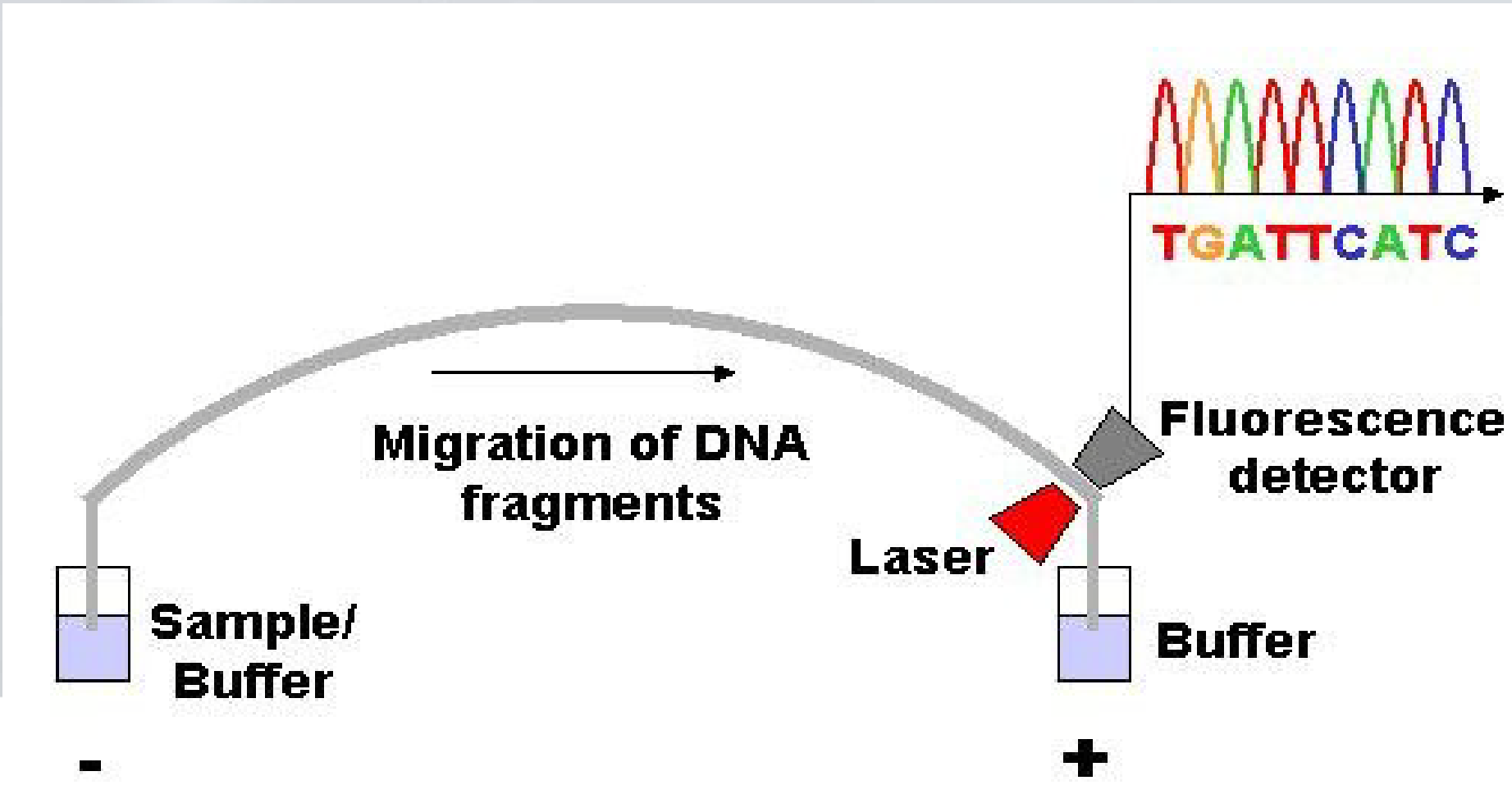
Part I: System Analysis

Sanger method



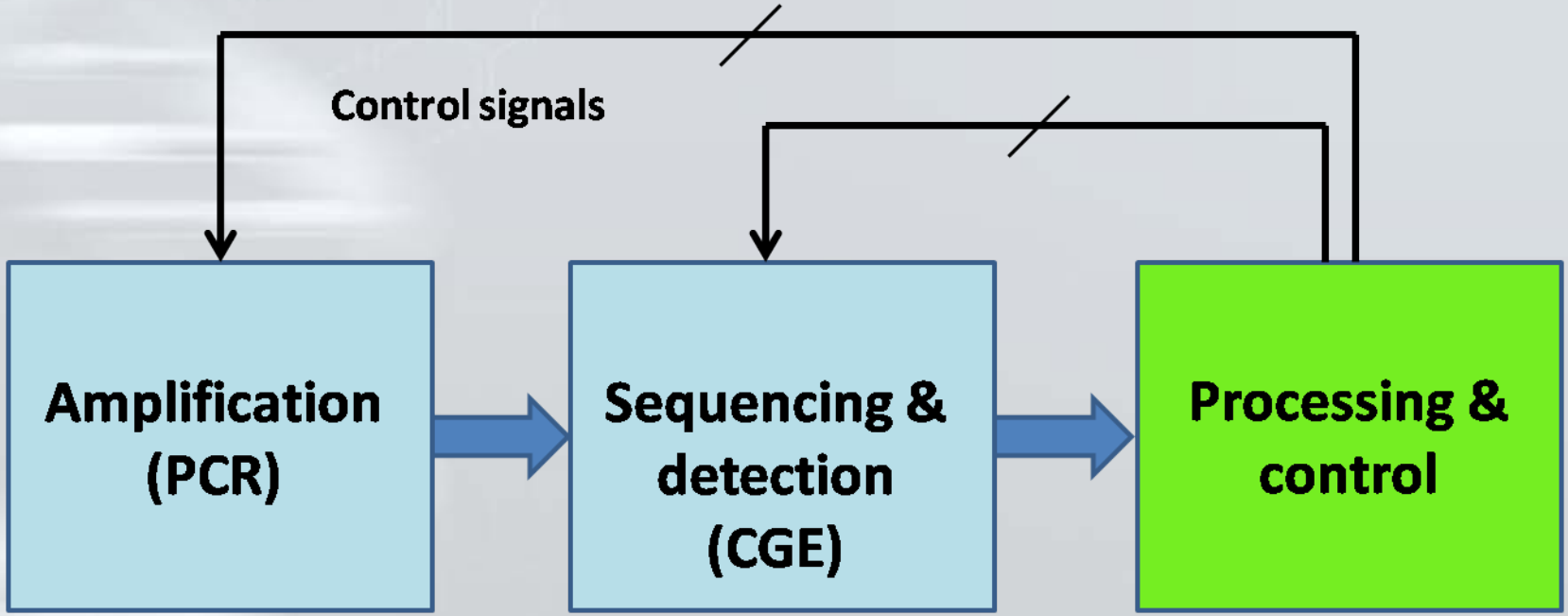
Part I: System Analysis

Capillary electrophoresis

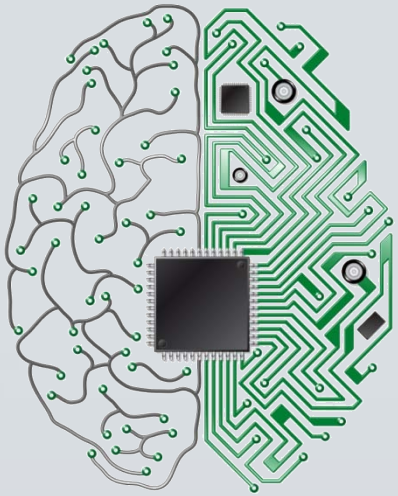


Part I: System Analysis

Processing and control



3. Multicore Processing and control



MultiProcessor System on chip (MPSoC):

- Control temperature
- Control cycle duration
- Signal processing
- Information processing
    - Align DNA sequences
    - Compare acquired sequence with detected sequences
    - Search within a database
    - etc …

Part II: Modeling of PCR

Modeling the denaturation phase

60 seconds at T = 95°C -> Enzyme deactivation

$$(-r_A) = (-\frac{da}{dt}) = K_d\, a^{\alpha}$$

$$a = \frac{E(t)}{E_0}$$

$$K_d = K_{do}\, e^{-\frac{E_d}{RT}}$$

Part II: Modeling of PCR

Modeling of extension phase

The rate of generation of DNA in the first cycle:

$$\frac{dC_D}{dt} = K_1^* \cdot \frac{C_{D0}}{y}$$

CD0 = Initial concentration of target DNA in M
CD = Concentration of target ADN after time "t" in M
K_1^* = Incorporation rate of nucleotides in one DNA strand in the first cycle
y = Number of nucléotides constituting one target sequence

Part II: Modeling of PCR

Modeling of extension phase

The amplification rate of DNA in the first cycle is given by:

$$\boxed{\frac{dN}{dt} = K_1}$$

$$N = \frac{C_D}{C_{D0}}$$

$$K_1 = \frac{K_1^*}{y}$$

Part II: Modeling of PCR

Modeling of extension phase

The amplification rate of ADN in the n^{th} cycle is given by:

$$\frac{dN}{dt} = 2^{n-1} K1$$

$$K_1 = K_0 \, e^{-\frac{E_a}{RT}}$$

Exponential growth

Modeling of PCR

Modeling of extension phase

Enzyme deactivation effect :

$$\frac{dN}{dt} = K_r \,.\, \beta_r$$

$$\beta_n = \frac{E_0 \; a_n}{C_{D0} \; 2^n}$$

quasi-linear growth

Part II: Modeling of Sequencing

Modeling of sanger termination

Probability of termination at base 'N':

$$p_n = R_N \frac{[ddNTP]}{[dNTP]}$$

If sequence begins with A,C,G:

$$d_{1a} = p_a D_0$$

$$d_{1c} = p_c D_0 (1 - p_a)$$

$$d_{1g} = p_g D_0 (1 - p_a)(1 - p_c)$$

...

Part II: Modeling of Sequencing

Modeling of sanger termination

The molar amount terminating at base number ‘i’ :

$$d[i] = X[i]D_0$$

D_0: inital molar amount, the result of the PCR stage

$$X[i] = \frac{X[i-1](1-p_{i-1})p_i}{p_{i-1}}$$

Part II: Modeling of Sequencing

Modeling of capillary electrophoresis

Assuming that diffusion is the only peak broadening mechanism that is present:
Gaussian concentration profile with variance:

$$\sigma^2 = 2Dt_m$$

D: diffusion coefficient
t_m: migration time from injection to the detection point

Part II: Modeling of Sequencing

Modeling of capillary electrophoresis

Migration time

$$t_m = \frac{l}{v}$$

Velocity

$$v = \mu_e E$$

Full width

$$w_t = \left(\frac{t_m}{l} \right) w_s$$

Modeling of Sequencing

Modeling of optical detection

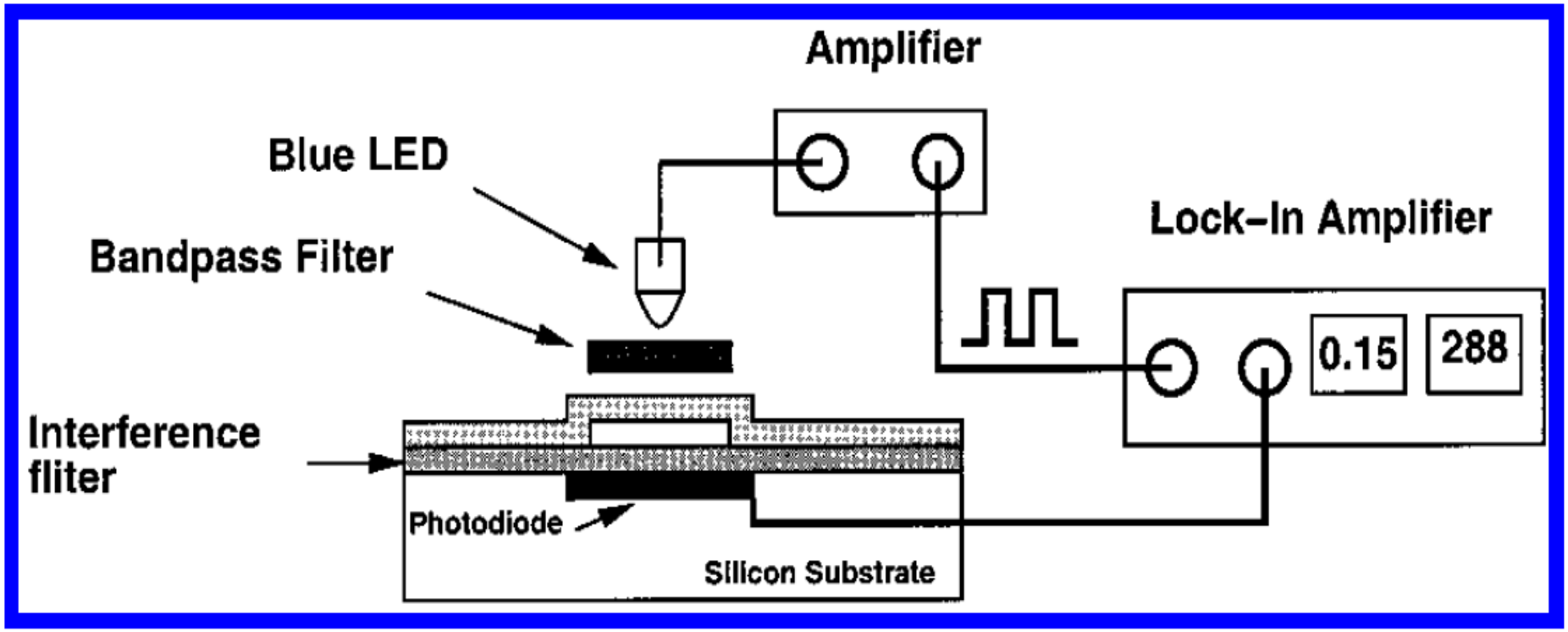
- Excitation from a source at 488nm
fluorescence from labels at: 500, 540, 570, 590 nm

- 4 photodiodes + dichroic filters

- Amplification and filtering

Modeling of Sequencing

Modeling of optical detection



Modeling of Sequencing

Modeling of optical detection

Band pass filters: Fabry-Perot

$$T = \frac{(1-R)^2}{1+R^2-2\mathrm{R}\cos\delta}$$

$$\delta = \left(\frac{2\pi}{\lambda}\right) 2\,nl\cos\theta$$

Part II: Modeling of Sequencing

Modeling of optical detection

Photodiode

Photocurrent

$$I_{out} = RP_{in}$$

Noise current
(not implemented)

$$\left\langle i_s^2(t) \right\rangle = 2q(I_{dark} + RP)B$$

$$\left\langle i_{Th}^2(t) \right\rangle = \frac{4K_B T}{R} B$$

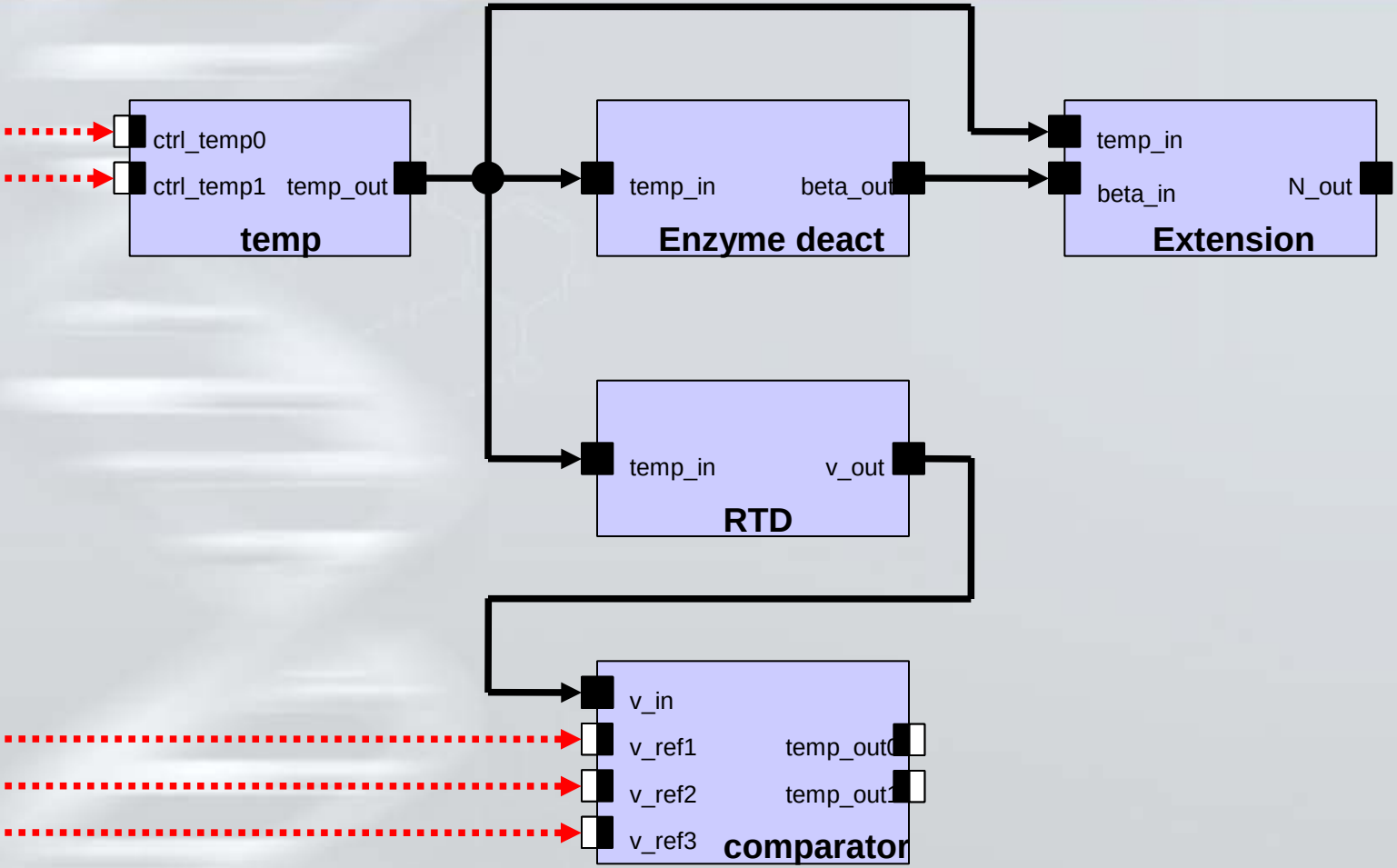
Part III: Hardware/Software modeling

- How to model and simulate multidomain AMS ?
  - SystemC-AMS
- How to model and simulate MPSoC architectures ?
  - SystemC + SoCLib library (www.soclib.fr)
- How to develop massively multithreaded applications on these architectures ?
  - Kahn Process Network paradigm and Smith-Waterman DNA sequencing application

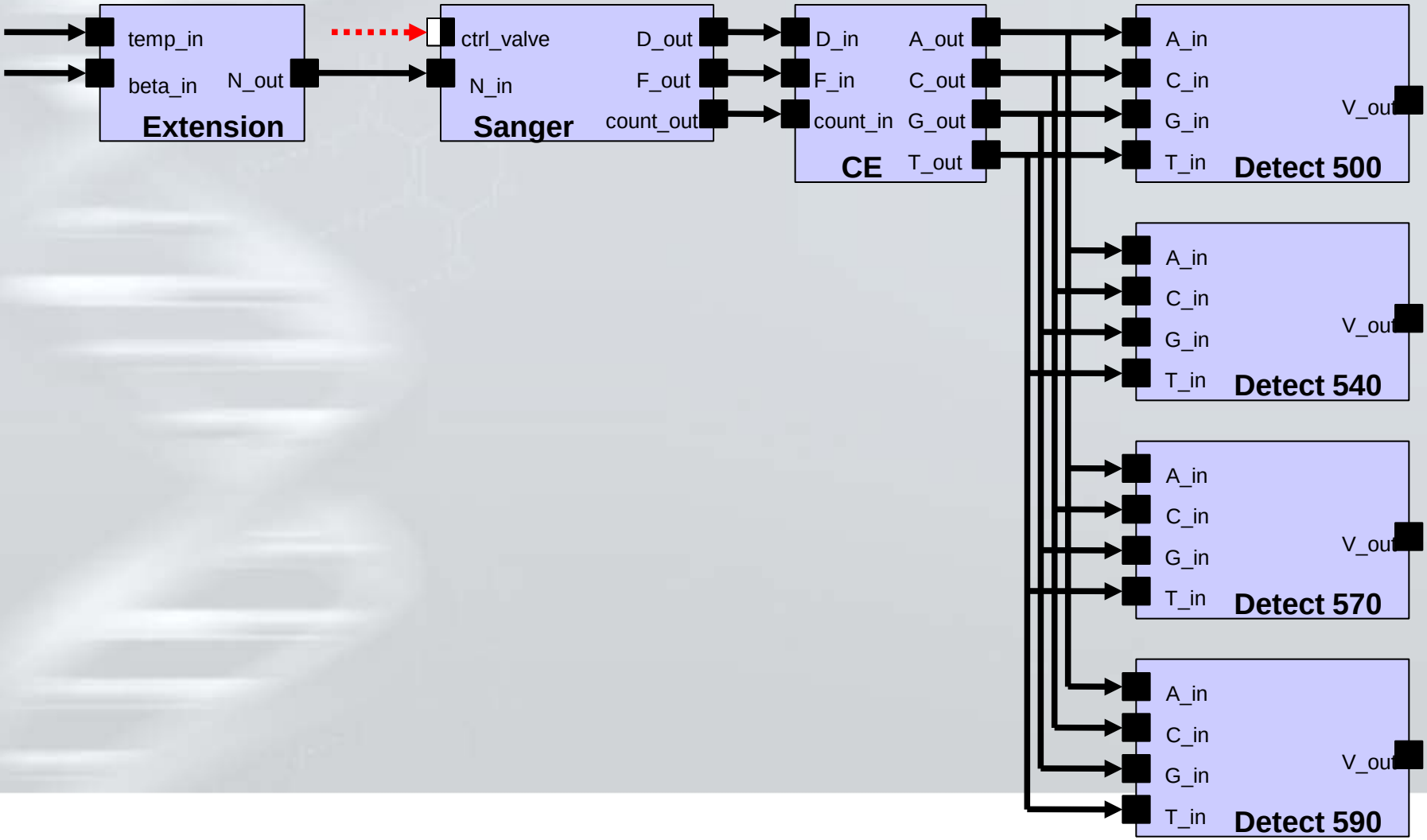
AMS & Analog: SystemC-AMS TDF

- Developed by Fraunhofer IIS/EAS Dresden, Contribution by TU Vienna
- Supported by NXP, STMicroelectronics, Bosch, Infineon, ...
- Supported by OSCI: LRM, User's Guide available
- Is a standard
- Proof of Concept freely available, access via OSCI website

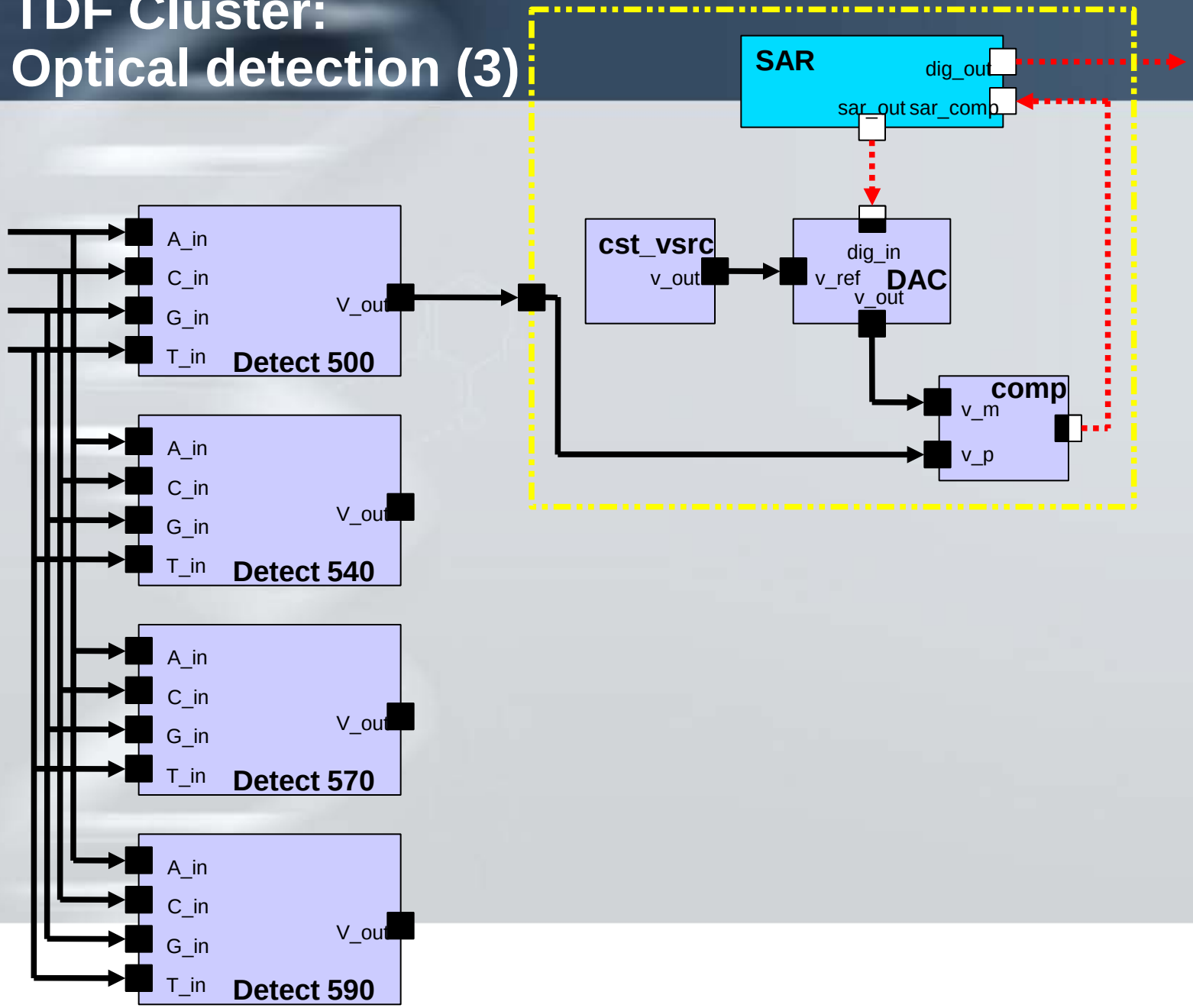
TDF Cluster: PCR part (1)



TDF Cluster: Sanger & CE part (2)

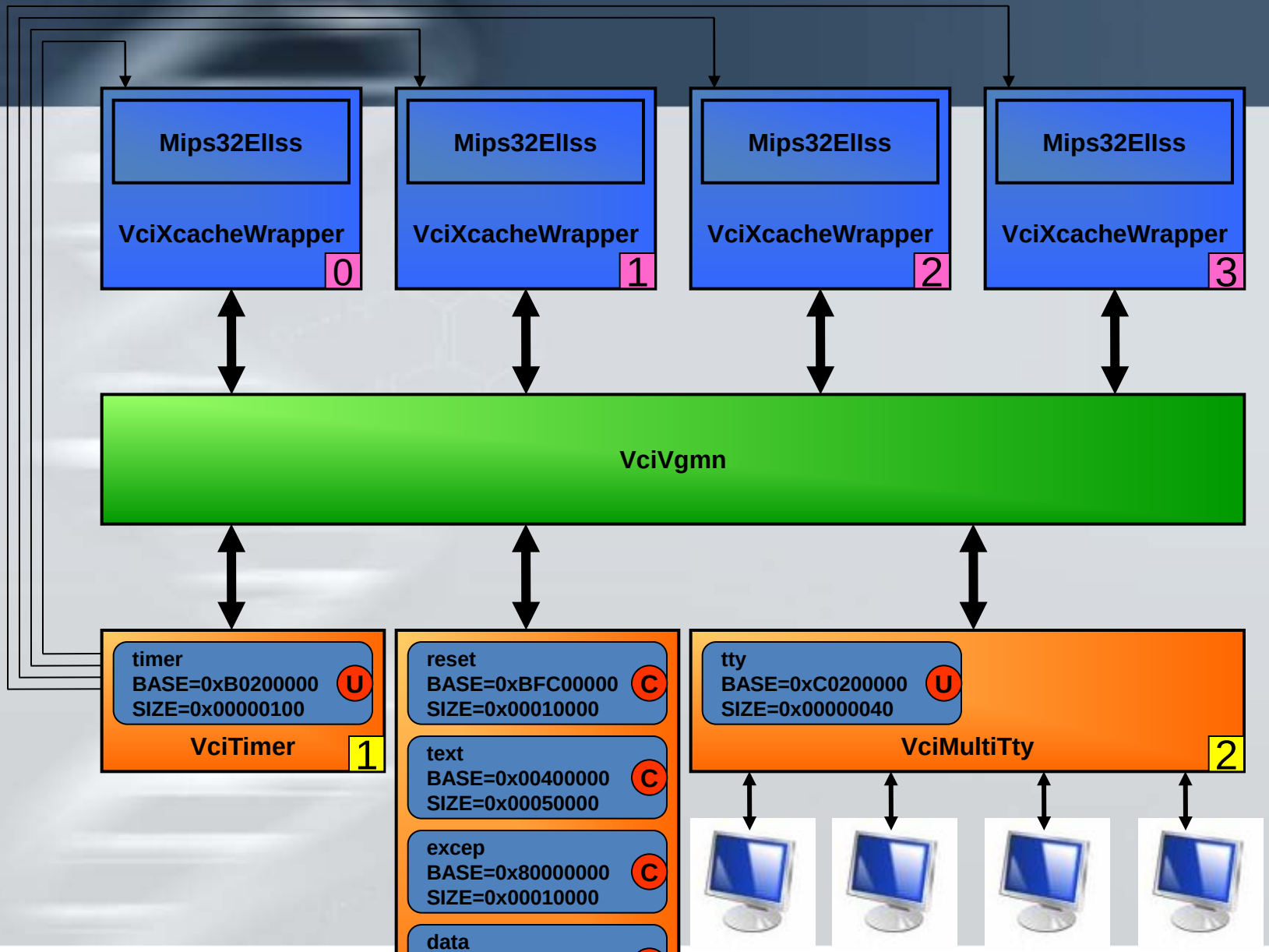


TDF Cluster: Optical detection (3)

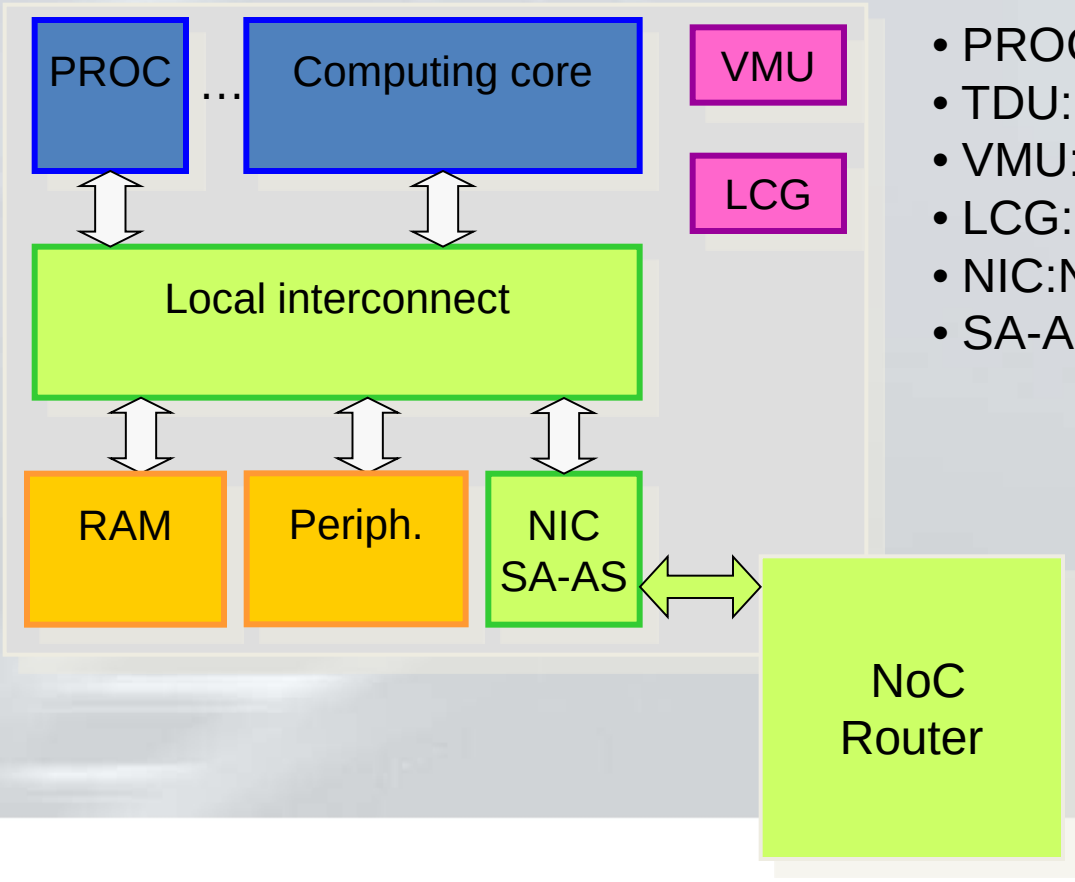


Digital: SystemC + SocLib Library

- National project, ended March 2010
- 11 Labs, 6 companies
- LGPL
- More than 100 BCA and TLM models
  - Based on SystemC
  - Processors, memories, Noc, peripherals
  - VCI/OCP compliant, interoperable components
  - Dedicated to MPSoCs, but proven useful for Monoproc systems
- SocLib Live-CD available on the website

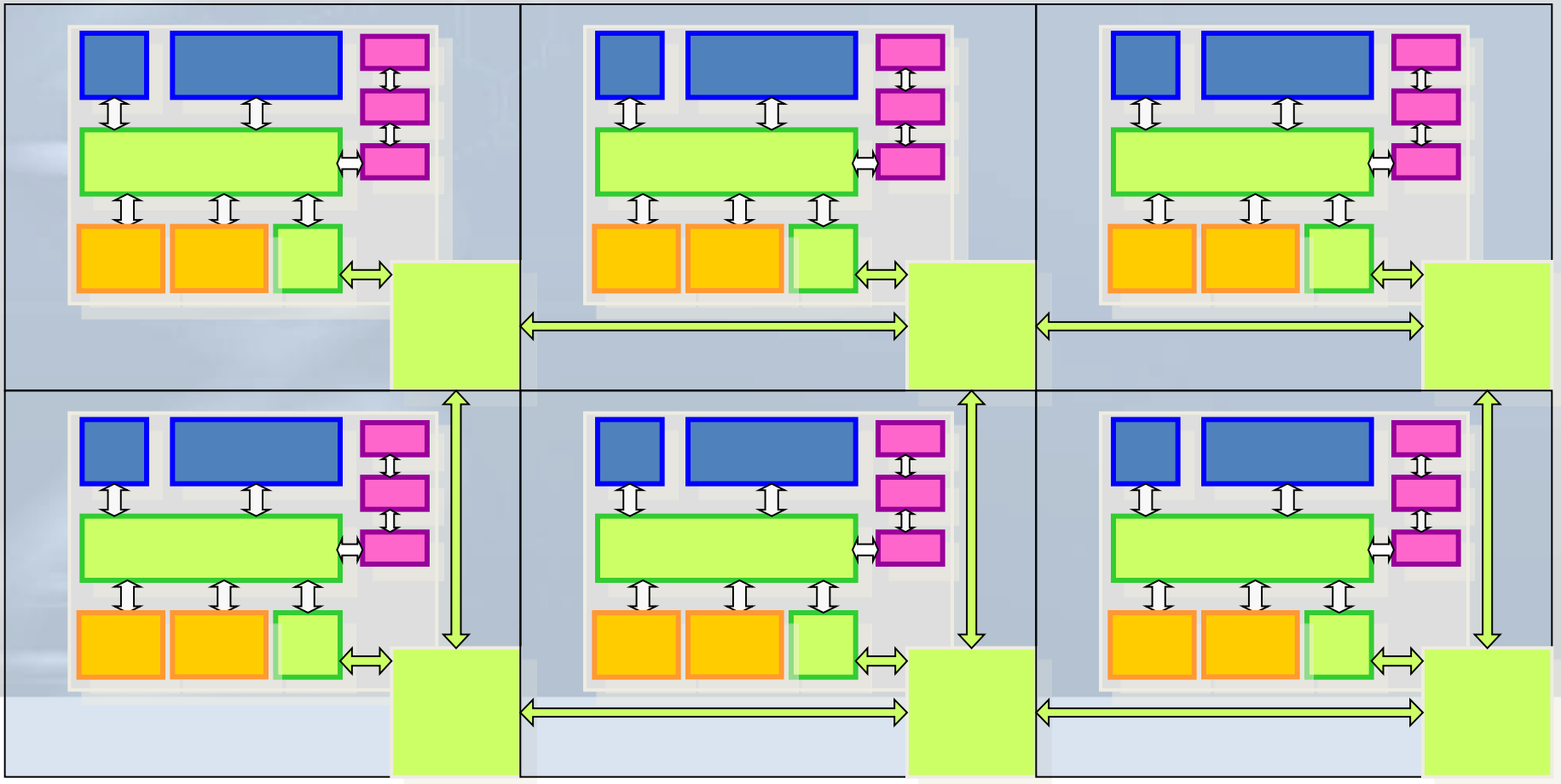

Mips32Ellss
VciXcacheWrapper
0
Mips32Ellss
VciXcacheWrapper
1
Mips32Ellss
VciXcacheWrapper
2
Mips32Ellss
VciXcacheWrapper
3
VciVgmn
timer
BASE=0xB0200000
SIZE=0x00000100
U
VciTimer
1
reset
BASE=0xBFC00000
SIZE=0x00010000
C
text
BASE=0x00400000
SIZE=0x00050000
C
excep
BASE=0x80000000
SIZE=0x00010000
C
data
BASE=0x10000000
SIZE=0x00020000
C
VciRam
0
tty
BASE=0xC0200000
SIZE=0x00000040
U
VciMultiTty
2

MPSoC generic tile



- PROC (MIPS32)
- TDU: Test & Decision Unit
- VMU: Voltage Management Unit
- LCG: Local Clock Generator
- NIC:Network Interface Controller
- SA-AS: Synchronous/Async converter

2D-Mesh shared memory NUMA architecture with NoC



Smith -Waterman

Smith-Waterman Algorithm

-Seq1 = $X_1,\ldots,X_M$

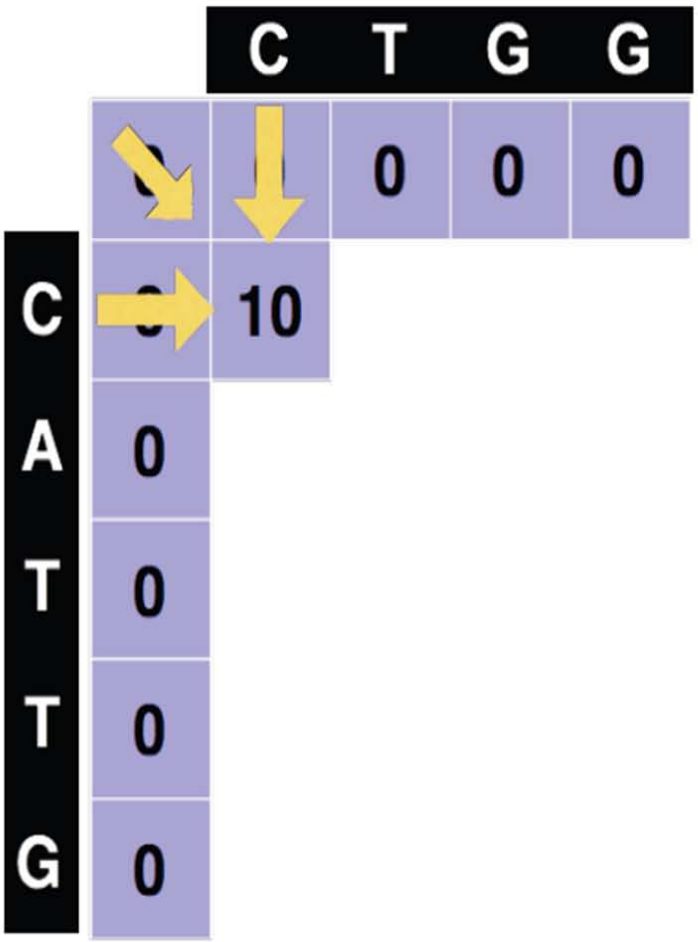
-Seq2 = $Y_1,\ldots,Y_N$

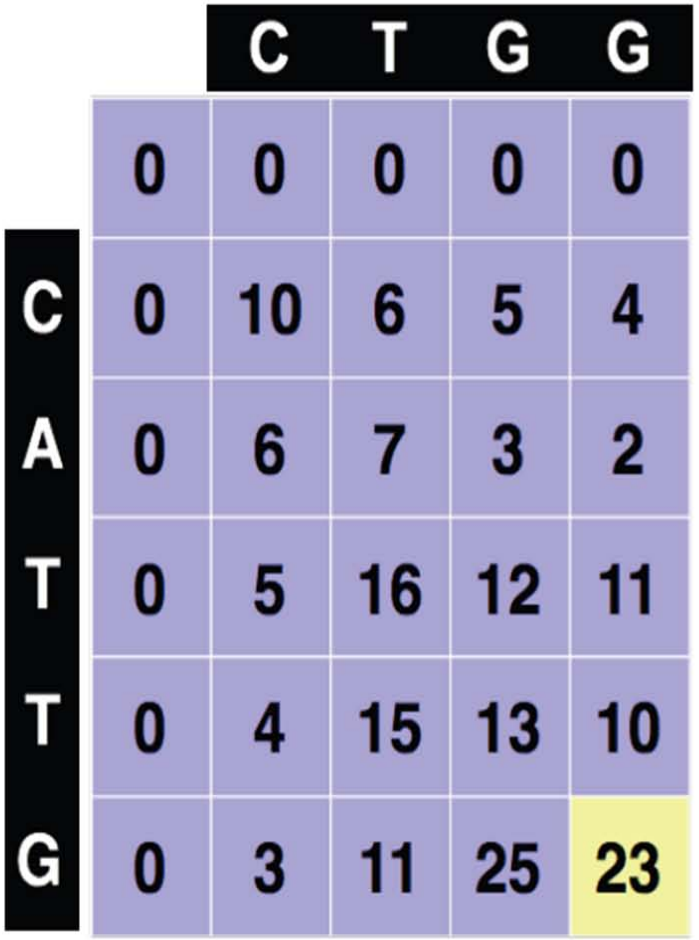
Score Matrix (4x4) for (A,C,T,G)

'F' Matrix (MxN)

$$F(i,j) = Max \begin{cases} F(i-1,j-1) + S(x_i, y_j) \\ F(i-1,j) + S(x_i, -) \\ F(i,j-1) + S(-, y_j) \\ 0 \end{cases}$$

Smith-Waterman





Smith Waterman

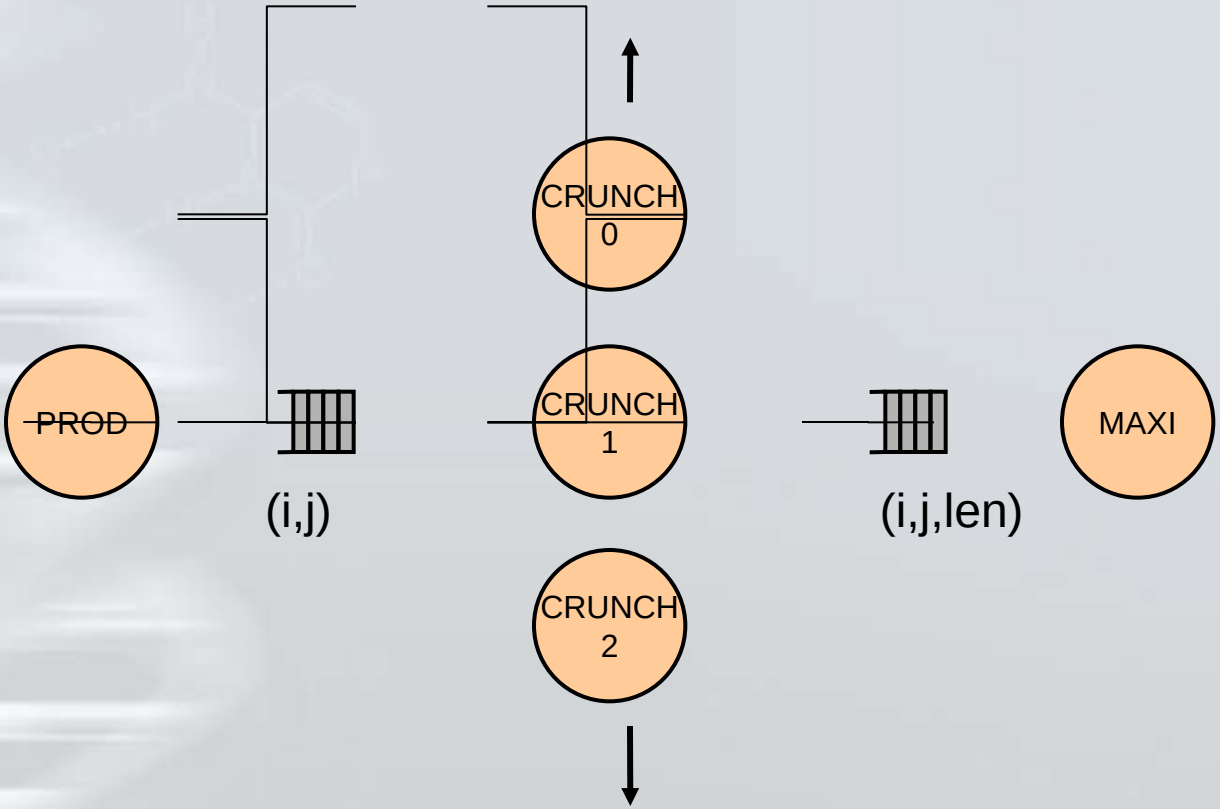
Sequence alignment

		C	T	G	G
	0	0	0	0	0
C	0	10	6	5	4
A	0	6	7	3	2
T	0	5	16	12	11
T	0	4	15	13	10
G	0	3	11	25	23

Optimal Alignment

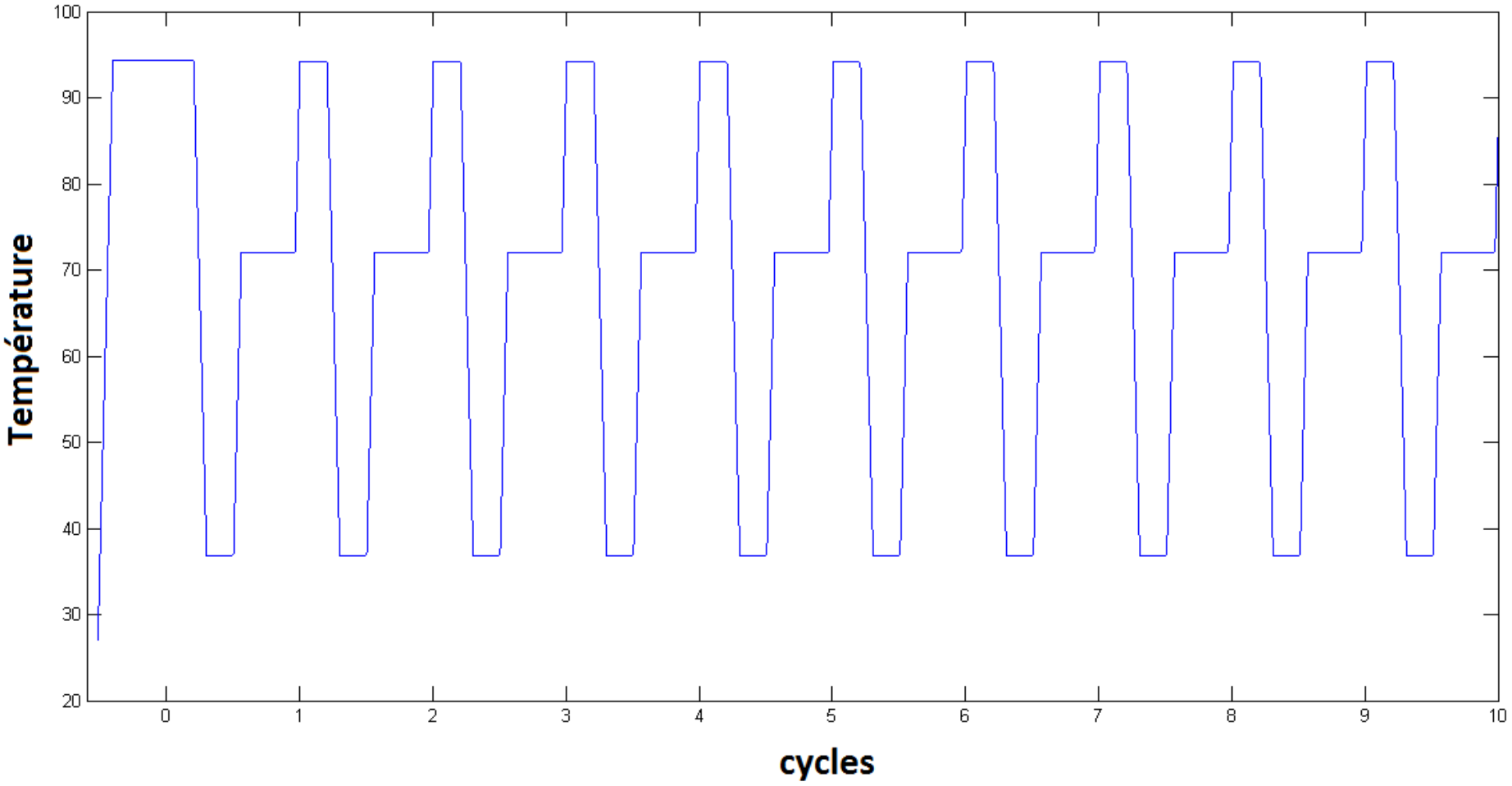
C A T T
G
C - - T
G

Part III: KPN-like multithreaded embedded application



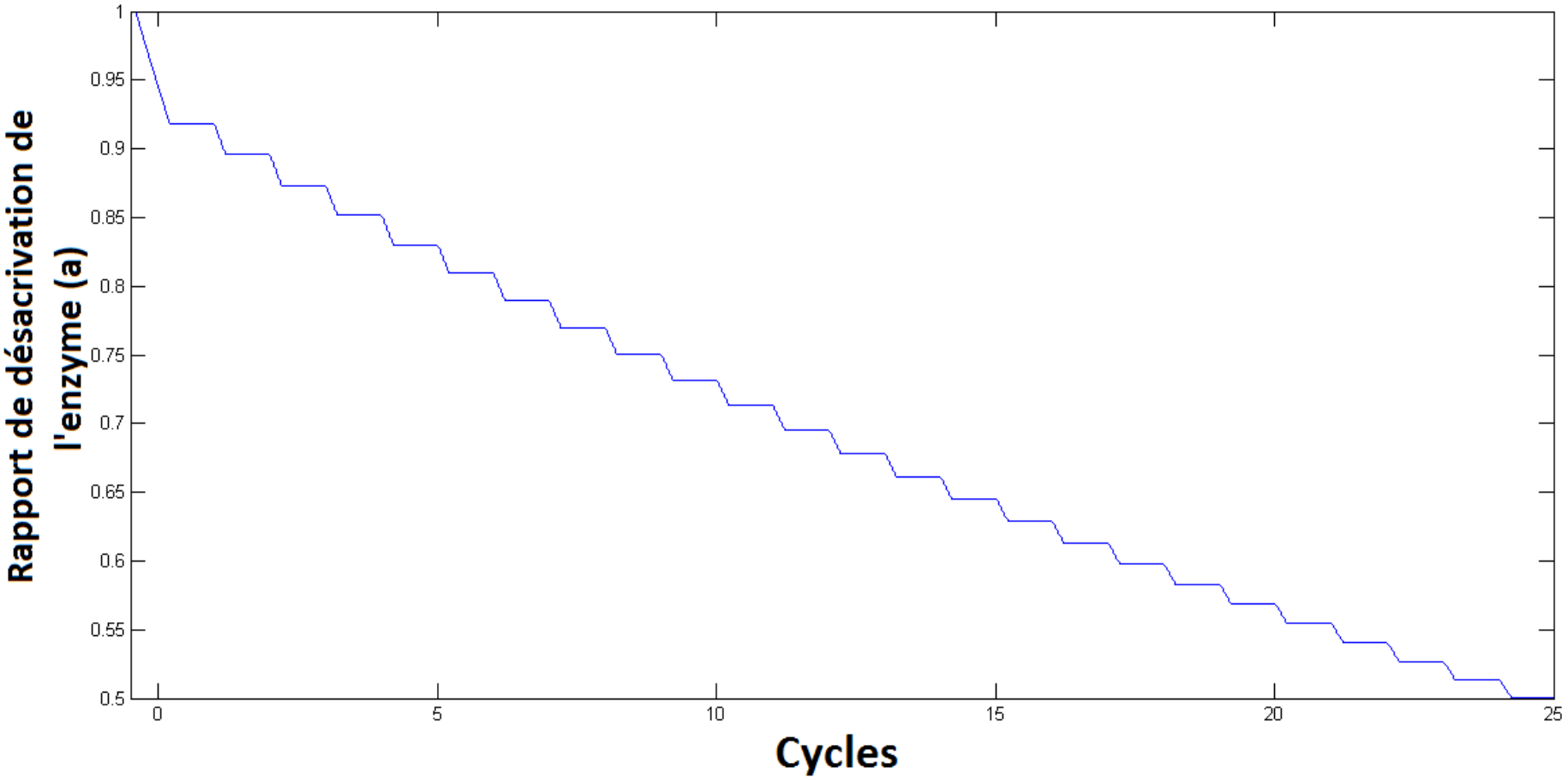
Results

Thermo-cycling



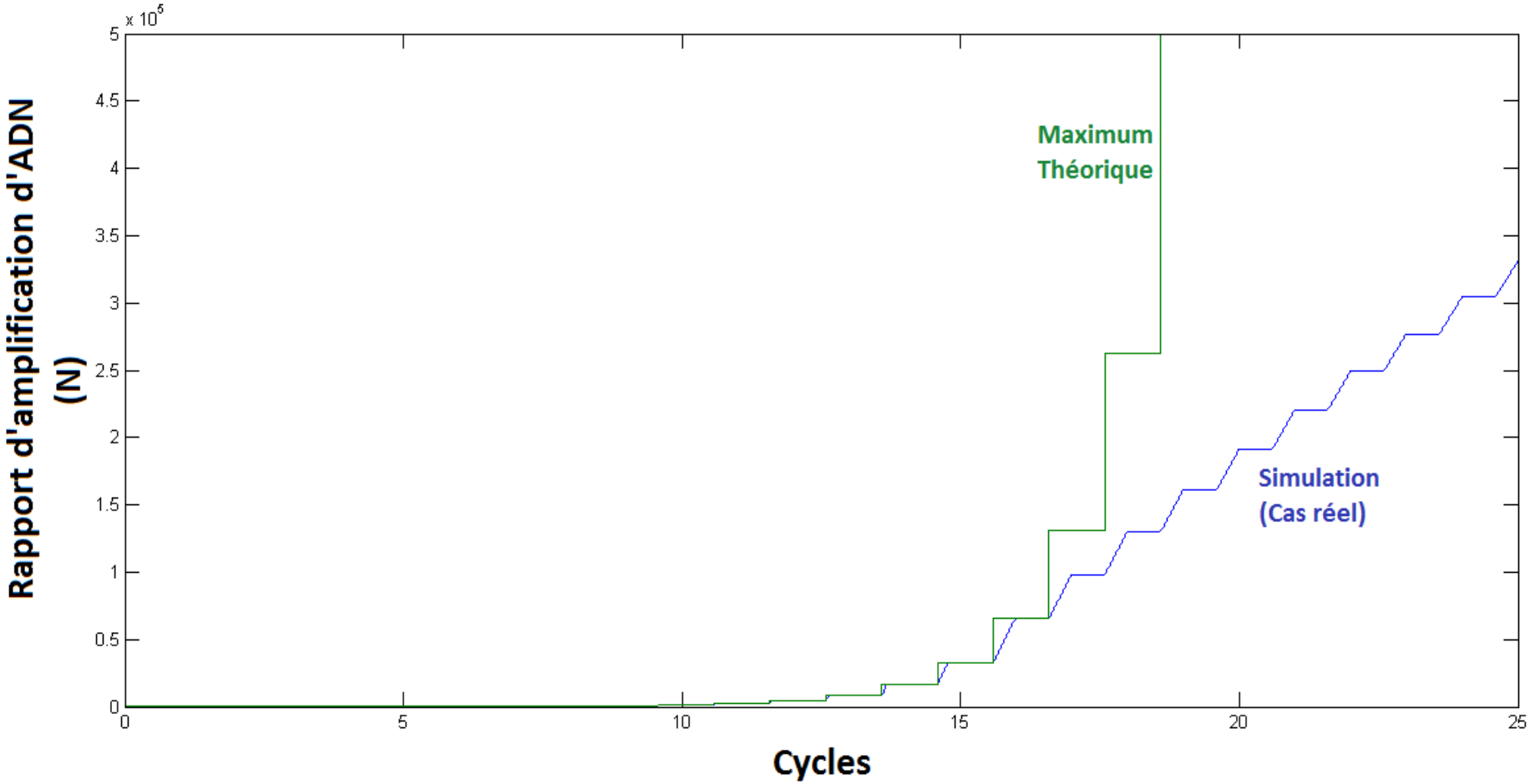
PCR Results

Deactivation of polymerase



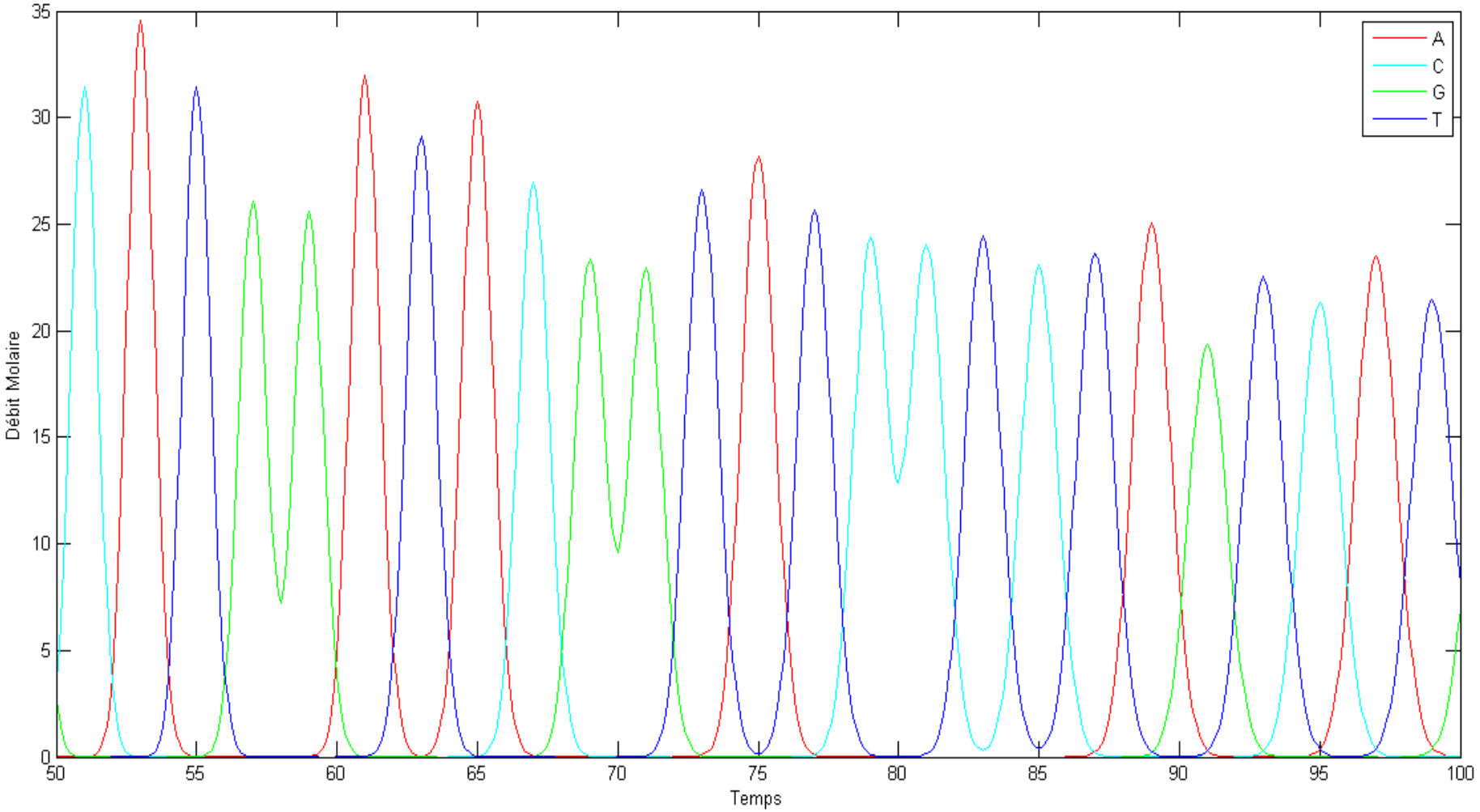
PCR Results

Amplification ratio of DNA



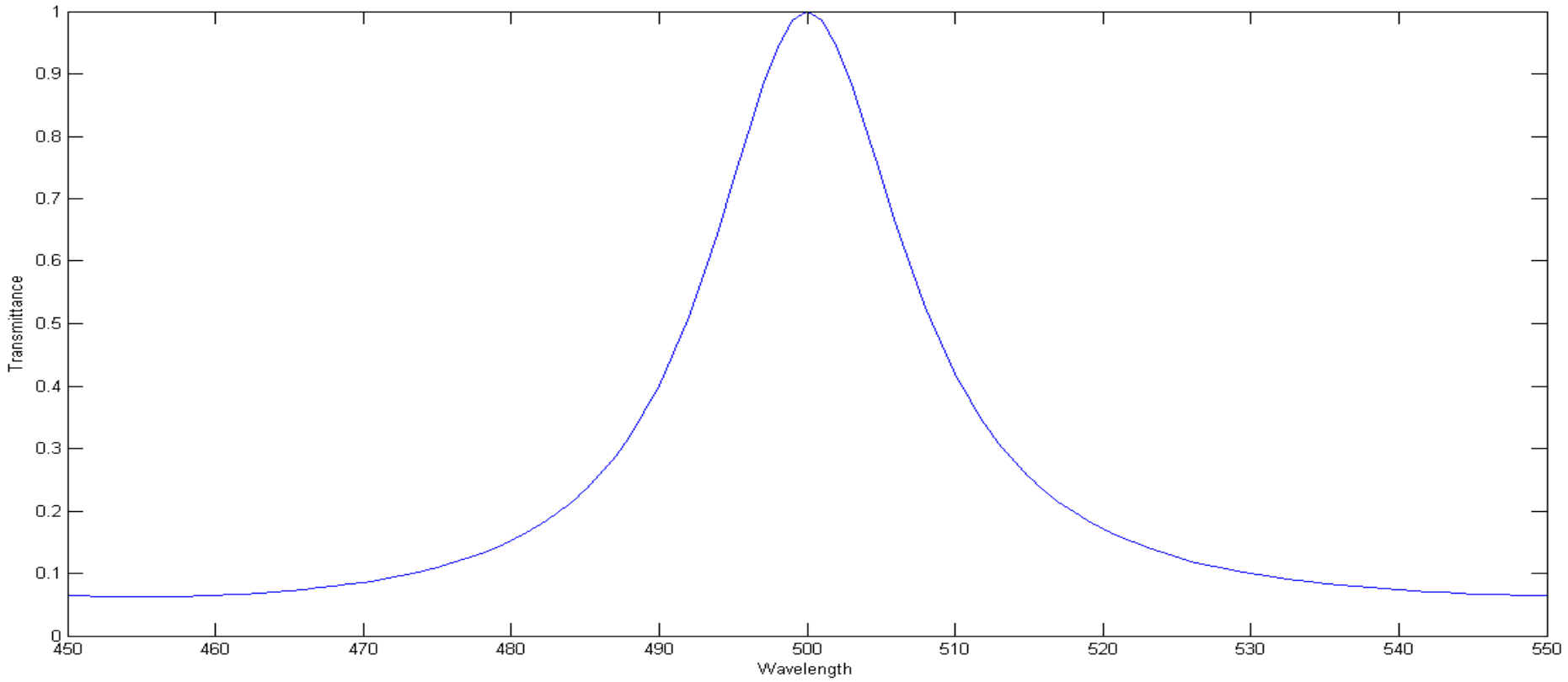
Modeling of Sequencing

Modeling of capillary electrophoresis



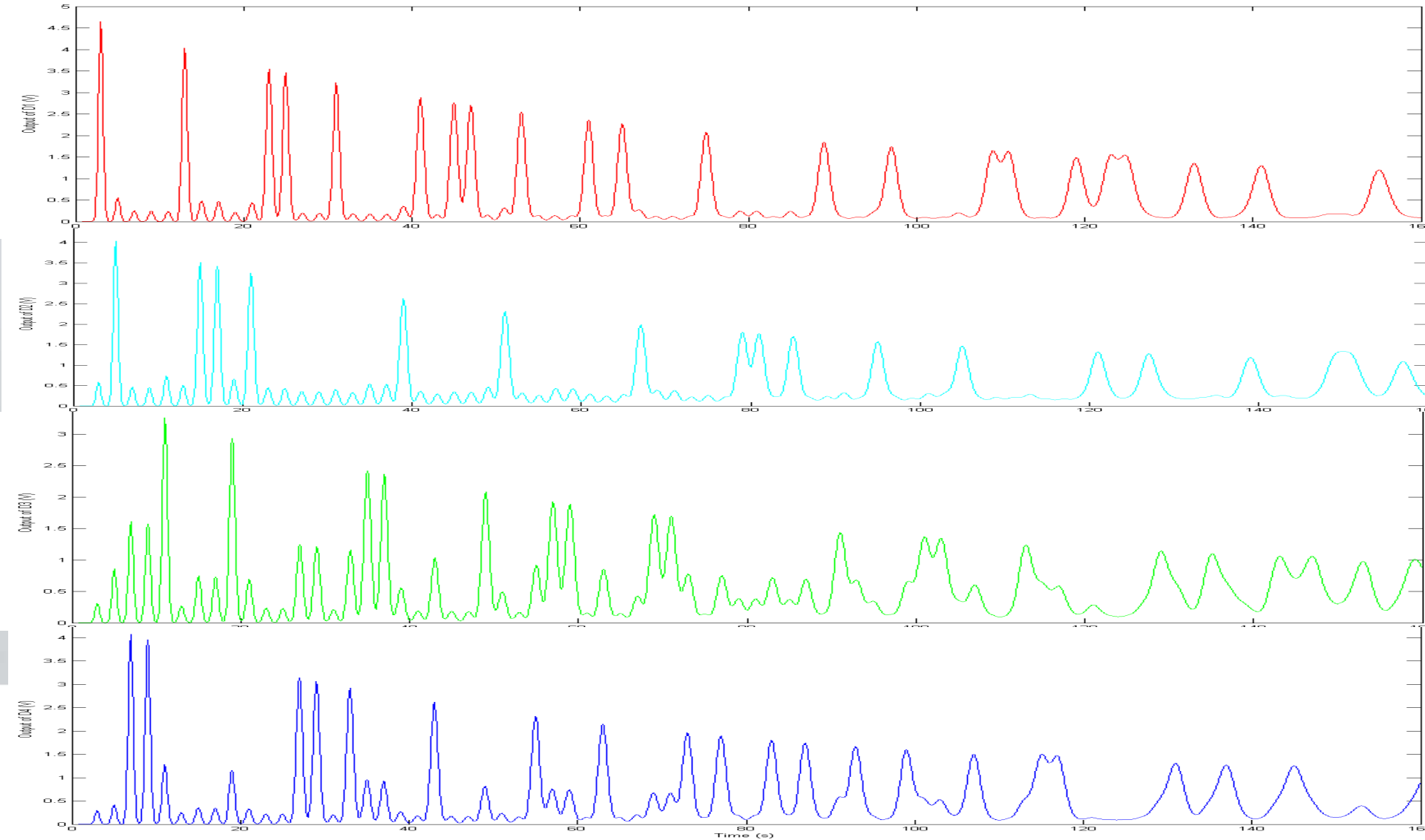
Modeling of Sequencing

Modeling of optical detection



Modeling of Sequencing

Output of detection stage:



Conclusion

- Success story !
  - Complex application captured with SystemC AMS extensions
  - Chemical kinetic reactions, electrophoresis, optics, ADC, DAC in AMS-TDF (digitally assisted)
  - 4 minutes to simulate PCR-CE and 1000 samples comparison
- Changing MPSoC operating frequency during simulation !
- Wish: deactivate TDF cluster simulation once DNA sequence acquired to speed up simulation
- Living project, architecture exploration, design refinement