High Level Synthesis

Binding

Virendra Singh

Computer Architecture and Dependable Systems Lab

Department of Electrical Engineering

Indian Institute of Technology Bombay

http://www.ee.iitb.ac.in/~viren/

E-mail: viren@ee.iitb.ac.in

EE-677: Foundations of VLSI CAD

Lecture 10 on 17 August 2021

CADSL

Cost

α · Area + $(1-\alpha)$ latency

a_1 = # multipliers

a_2 = # ALUs.

$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ ⇐

2M / 4A ⇓ 2M + 2A.

$\begin{bmatrix} 5a_1 \\ a_2 \end{bmatrix}$

$\begin{bmatrix} 5a_1 \\ a_2 \end{bmatrix}$.

5

Max latency

λ ⇐.

$\sum l \cdot x_i$ //

ASAP

ASAP ALAP

λ_{max} λ_{min} slack mobility

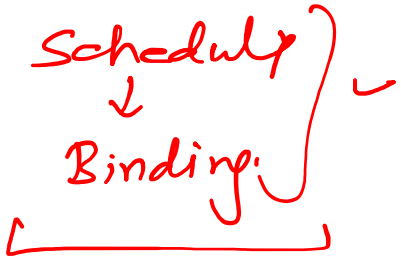
Pareto optimal solution

Concurrent Scheduling and Binding

Concurrent Scheduling and Binding

Concurrent Scheduling and Binding

- Define upper bound on resource usage and latency



Problems

- Intractable problem

Concurrent Scheduling and Binding

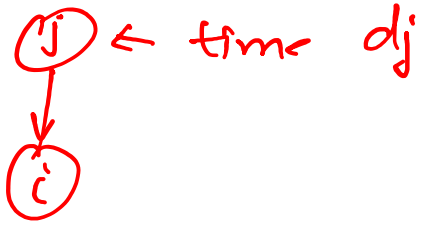
ILP Formulation

1. Start time of each operation is unique

   $\Sigma_l x_{il} = 1$

2. Sequencing relations represented by $G_s(V,E)$ must be satisfied

   $\Sigma_l l. x_{il} \geq \Sigma_l l. x_{jl} + d_j$



3. Resource bound must be met at every schedule step

   $\Sigma_k \Sigma_m x_{im} \leq a_k$

Concurrent Scheduling and Binding

4. Operation has to bound one and only one resource

$$\Sigma_r b_{ir} = 1$$

5. Operation bound to same resource must not be concurrent

$$\Sigma_i b_{ir} \sum_{m = l-d_i+1}^{l} x_{im} \leq 1$$

Latency: $\lambda = \sum_l l.\, x_{nl} - \sum_l l.\, x_{0l}$

Minimize area and latency simultaneously

ILP Formulation

All operation must start only once

$x_{0,1} = 1$

$x_{1,1} = 1$

$x_{2,1} = 1$

$x_{3,2} = 1$

$x_{4,3} = 1$

$x_{5,4} = 1$

$x_{6,1} + x_{6,2} = 1$

$x_{7,2} + x_{7,3} = 1$

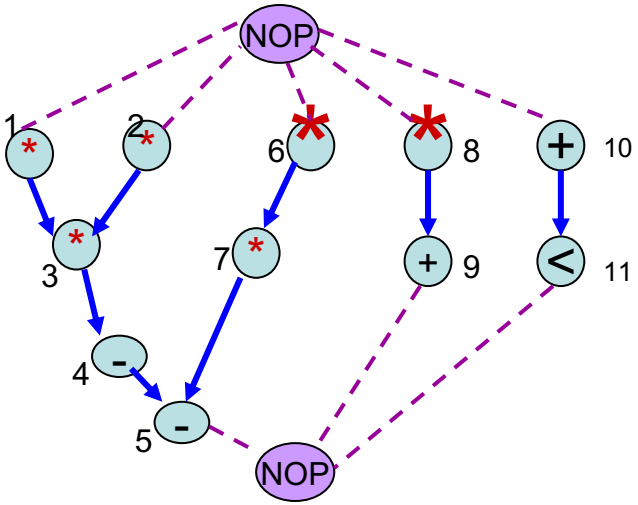
$x_{8,1} + x_{8,2} + x_{8,3} = 1$

$x_{9,2} + x_{9,3} + x_{9,4} = 1$

$x_{10,1} + x_{10,2} + x_{10,3} = 1$

$x_{11,2} + x_{11,3} + x_{11,4} = 1$

$x_{n,5} = 1$



ILP Formulation

Constraints – based on sequencing

(more than one starting time for at least one operation)

$2\,x_{7,2} + 3\,x_{7,3} - x_{6,1} - 2\,x_{6,2} - 1 \geq 0$

$2\,x_{9,2} + 3\,x_{9,3} + 4\,x_{9,4} - x_{8,1} - 2\,x_{8,2} - 3\,x_{8,3} - 1 \geq 0$

$2\,x_{11,2} + 3\,x_{11,3} + 4\,x_{11,4} - x_{10,1} - 2\,x_{10,2} - 3\,x_{10,3} - 1 \geq 0$

$4\,x_{5,4} - 2\,x_{7,2} - 3\,x_{7,3} - 1 \geq 0$

$5\,x_{n,5} - 2\,x_{9,2} - 3\,x_{9,3} - 4\,x_{9,4} - 1 \geq 0$

$5\,x_{n,5} - 2\,x_{11,2} - 3\,x_{11,3} - 4\,x_{11,4} - 1 \geq 0$

Concurrent Scheduling and Binding

4. Operation has to bound one and only one resource

$$\Sigma_r \, b_{ir} = 1$$

5. Operation bound to same resource must not be concurrent

$$\Sigma_i \, b_{ir} \sum_{m = l-d_i+1}^{l} x_{im} \leq 1$$

Latency: $\lambda = \sum_l l. \, x_{nl} - \sum_l l. \, x_{0l}$

Minimize area and latency simultaneously

ILP Formulation

Resource Constraints

$x_{1,1} + x_{2,2} + x_{6,1} + x_{8,1} \leq a1$

$x_{3,2} + x_{6,2} + x_{7,2} + x_{8,2} \leq a1$

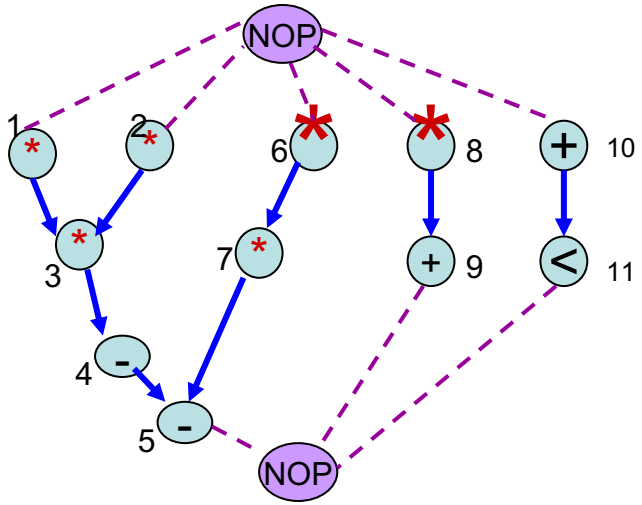
$x_{7,3} + x_{8,3} \leq a1$

$x_{10,1} \leq a2$

$x_{9,2} + x_{10,2} + x_{11,2} \leq a2$

$x_{4,3} + x_{9,3} + x_{10,3} + x_{11,3} \leq a2$

$x_{5,4} + x_{9,4} + x_{11,4} \leq a2$



ILP Formulation

Resource Bind

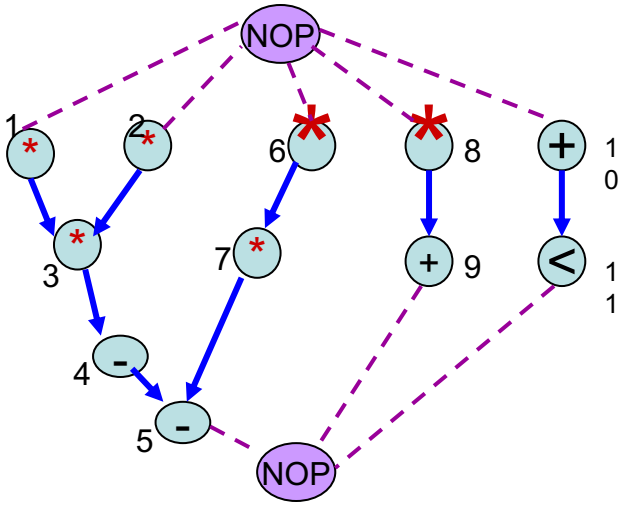
$$b_{1,1} + b_{1,2} + b_{1,3} + \ldots + b_{1,a1} = 1$$

$$b_{2,1} + b_{2,2} + b_{3,3} + \ldots + b_{4,a1} = 1$$

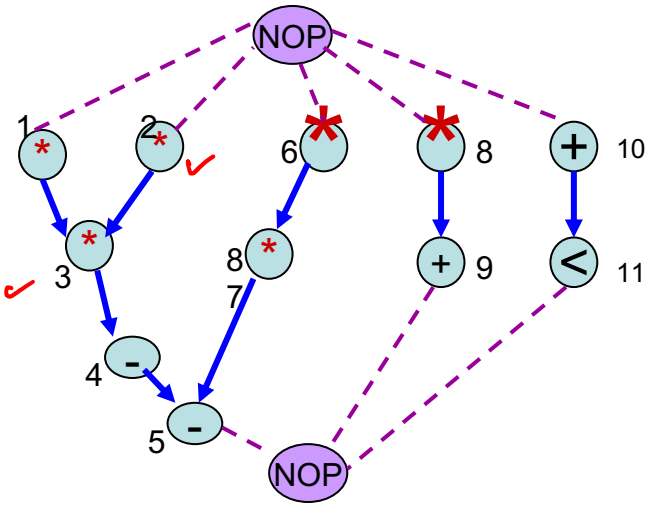
$$b_{10,1} + b_{10,2} + b_{10,3} + \ldots + b_{10,a2} = 1$$

Multi

ALU



ILP Formulation

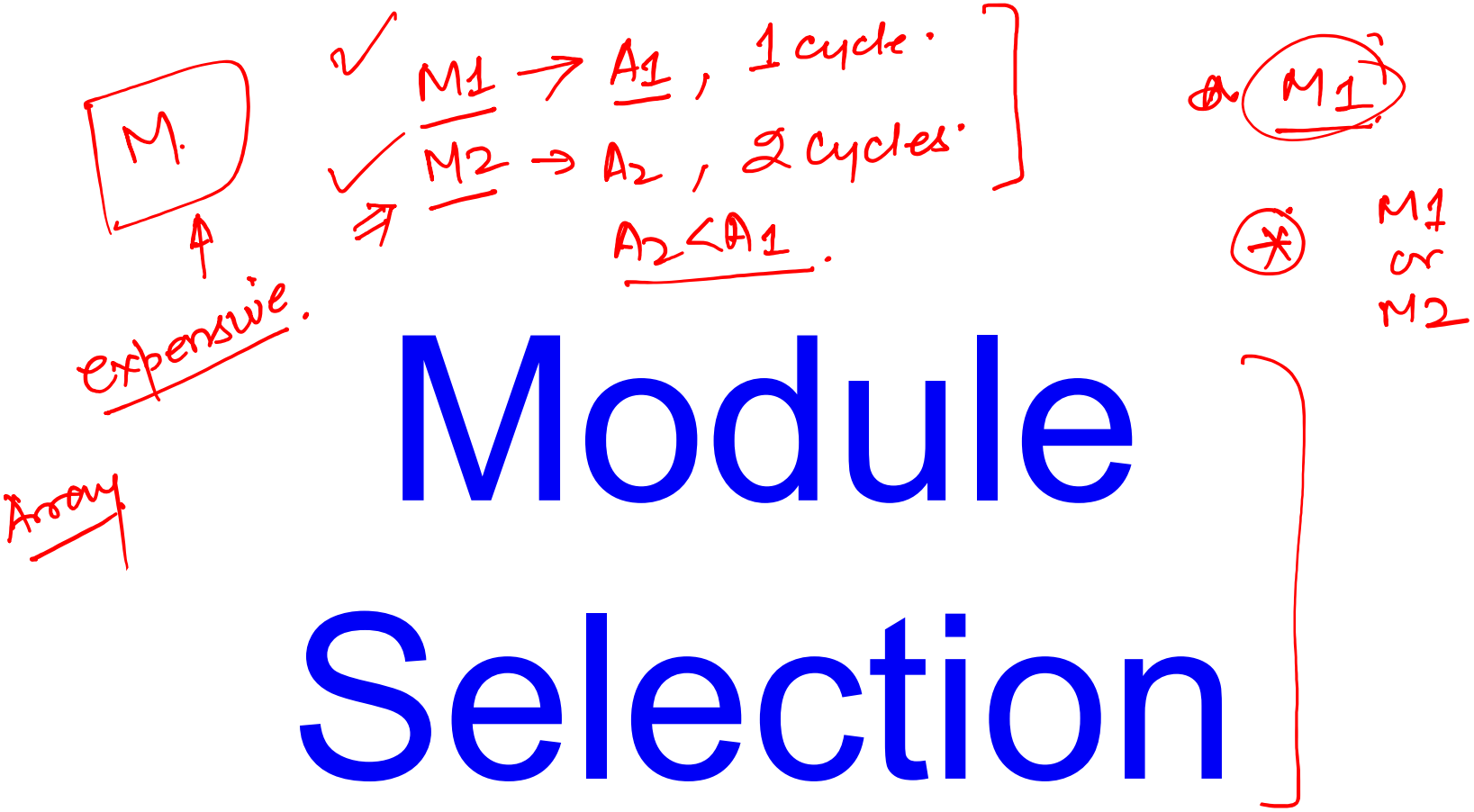


Resource Bind

$$b_{1,1} . x_{1,1} + b_{2,1} x_{2,1} + b_{3,1} x_{3,1} + b_{6,1} x_{6,1} + b_{7,1} x_{7,1} + b_{8,1} x_{8,1} \leq 1$$

$$b_{1,2} . x_{1,1} + b_{2,2} x_{2,1} + b_{3,2} x_{3,1} + b_{6,2} x_{6,1} + b_{7,2} x_{7,1} + b_{8,2} x_{8,1} \leq 1$$

$$b_{1,a1} . x_{1,1} + b_{2,a1} x_{2,1} + b_{3,a1} x_{3,1} + b_{6,a1} x_{6,1} + b_{7,a1} x_{7,1} + b_{8,a1} x_{8,1} \leq 1$$



Module Selection

Module Selection

- Resource type selection ⟹ Module selection
- Generalization of resource binding
- We assume more than one resource type can match the functional requirement of an operation type
  - Resource types are compatible if they can perform the same operation
  - Resource types are characterized by the pair
    - {(area_type$_k$, delay_type$_k$); k = 1,2, .. ,n_{res})}

A_1 d_1 $A_1 = 5a_1$ $d_1 = 1$
A_2 d_2 $A_2 = 3a_1$ $d_2 = 2$

Module Selection

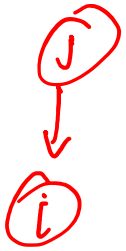
ILP Formulation

1. Start time of each operation is unique

   $\Sigma_l x_{il} = 1$

2. Sequencing relations represented by $G_s(V,E)$ must be satisfied

   $\Sigma_l l. x_{il} \geq \Sigma_l l. x_{jl} + d_j$



3. Operation has to bound one and only one resource

   $\Sigma_r b_{ir} = 1$

Module Selection

4. Operation bound to same resource must not be concurrent

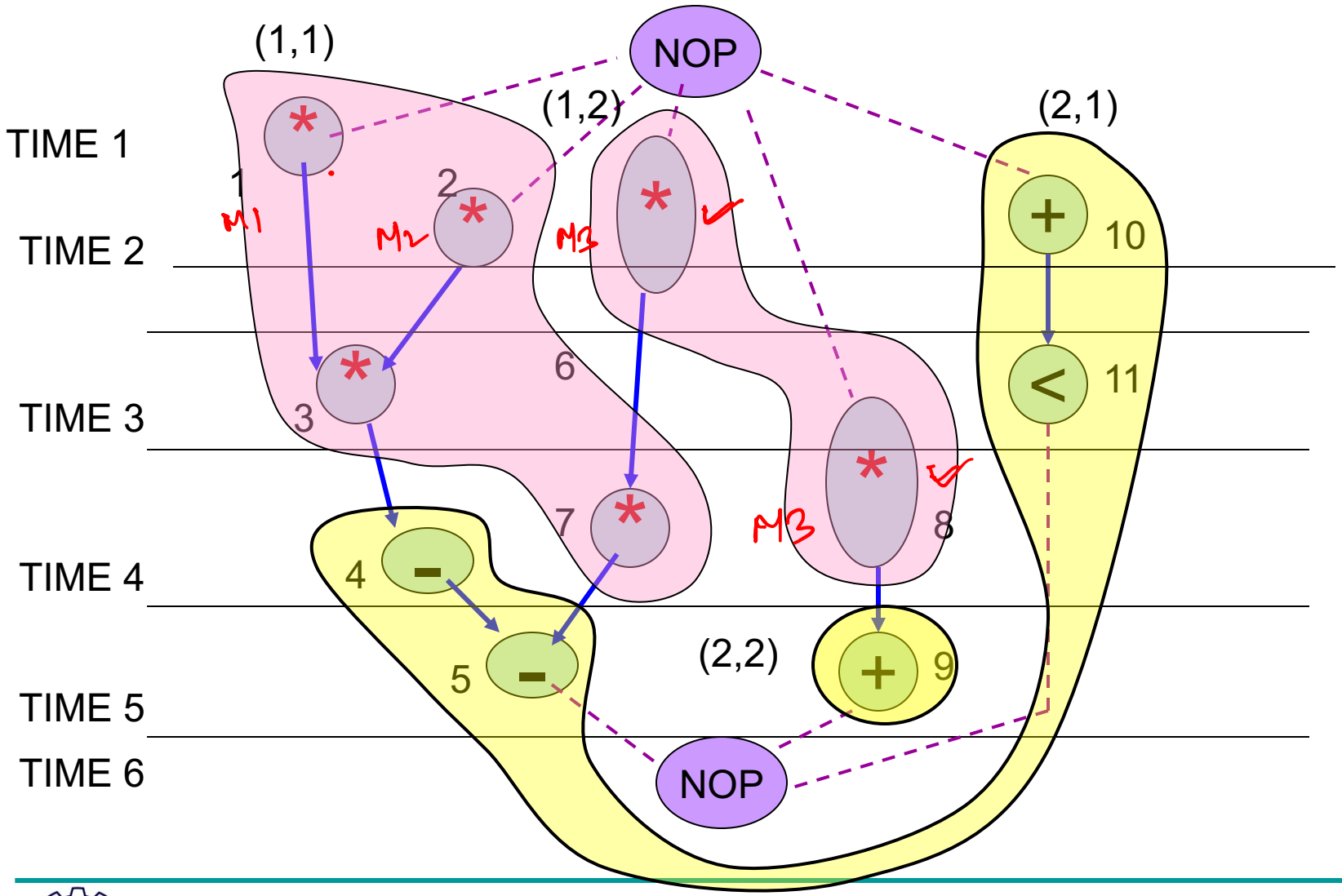
$$\Sigma_i \, b_{ir} \sum_{m = l-d_i+1}^{l} x_{im} \leq 1$$

5. Execution delay for each bound operation

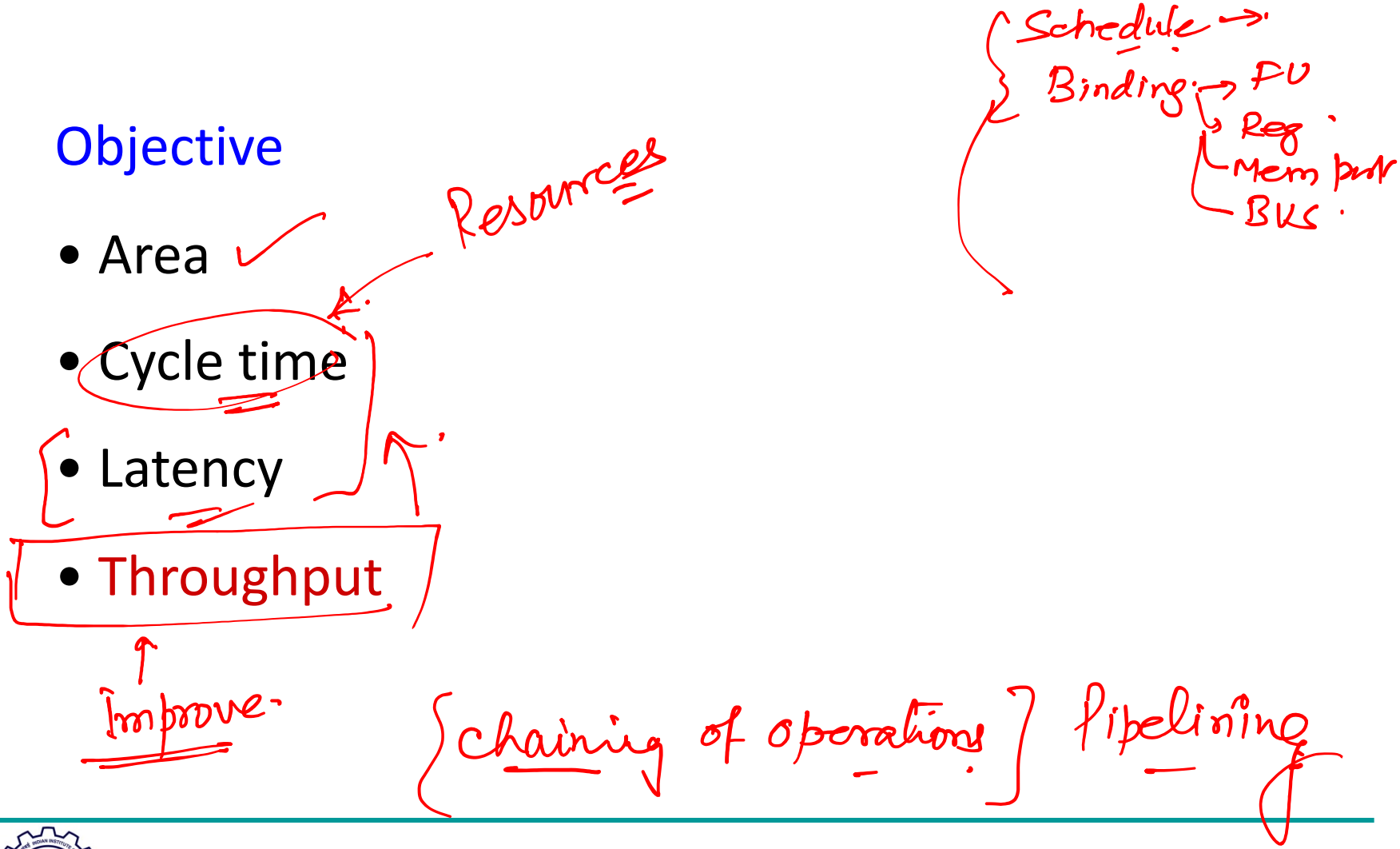
$$d_j = \sum_{r=1}^{a} b_{jr} . \, delay_r \; ; \; j = 1,2, \ldots, n_{op}$$

Minimize area and latency simultaneously

Module Selection



Architectural Synthesis



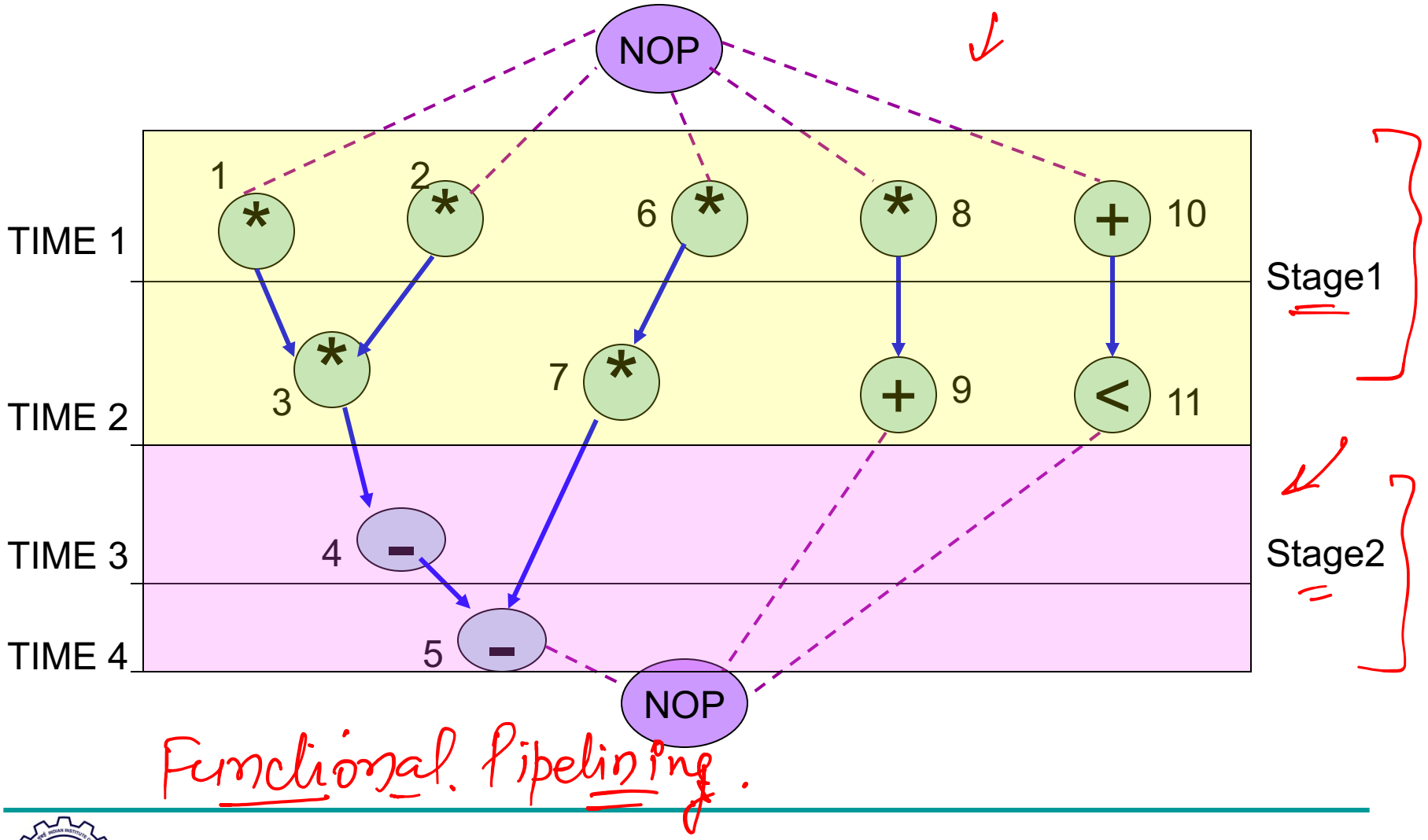
Pipelined Circuit Scheduling

Specification

- Sequencing Graph model
- Data Rate

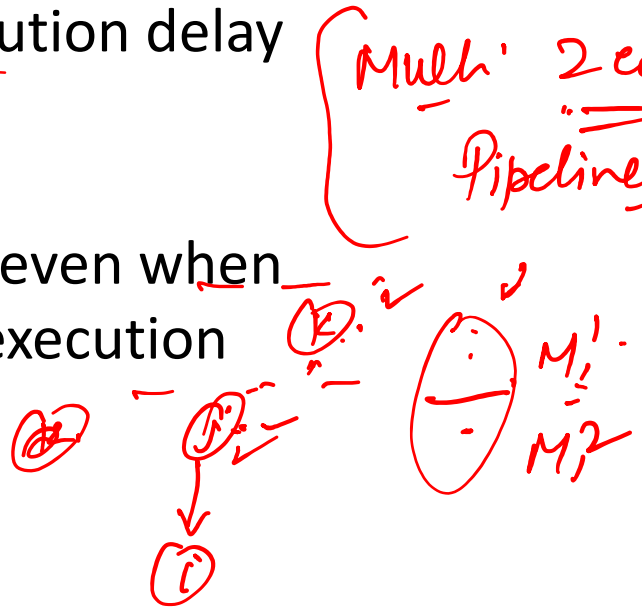
1. Non-pipelined Sequencing graph
   - Operations can be bound to pipelined model
2. Non hierarchical pipelined model with non-pipelined resources

Pipelined Sequencing Graph



Scheduling with Pipelined Resources

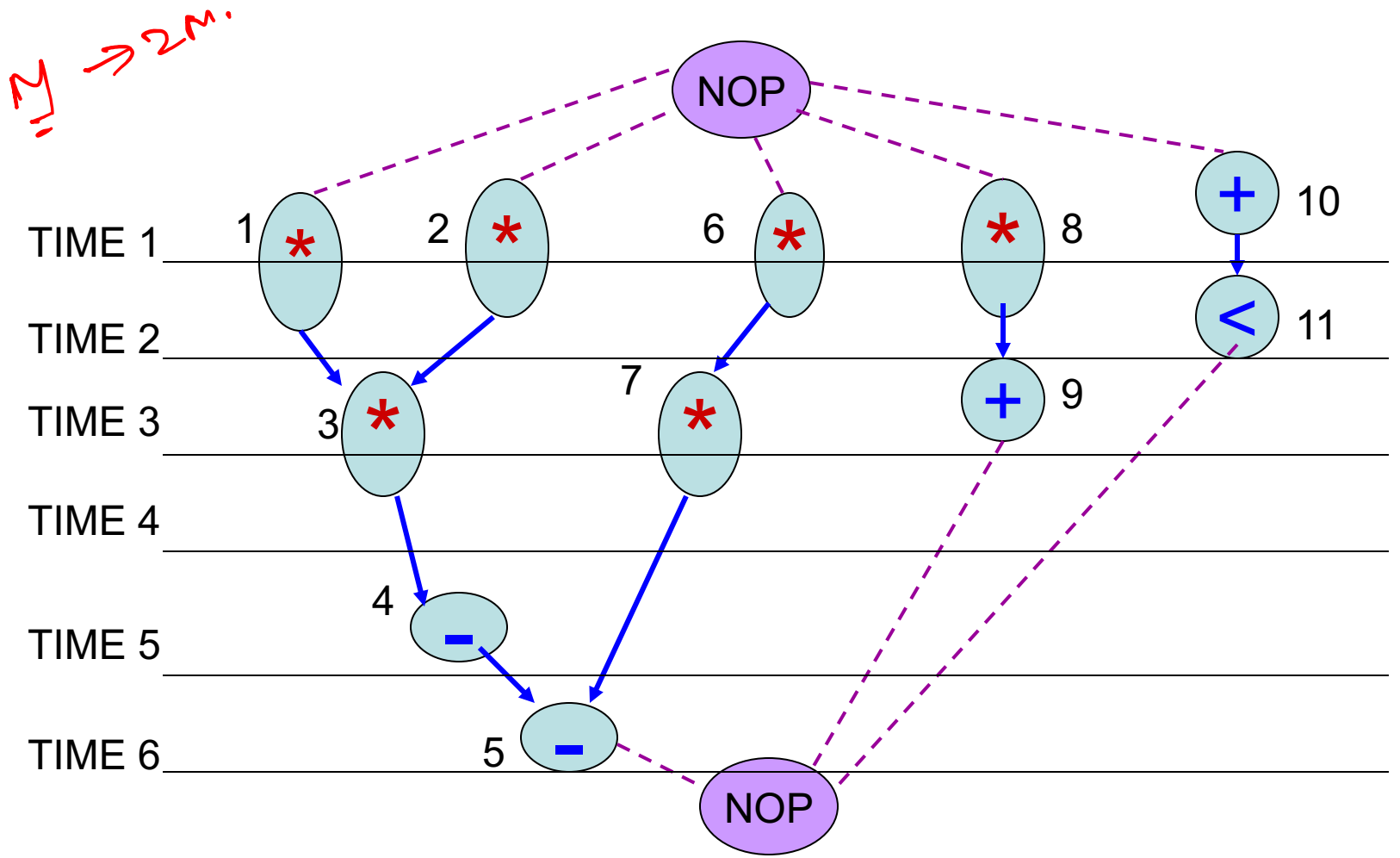
- Pipelined resources consume and produce data at time intervals that are smaller than execution delay
  - Data introduction interval
- Pipelined resources can be shared, even when corresponding operations overlap execution
  - Necessary requirement
    - no data dependency
    - Operations do not start in the same time step



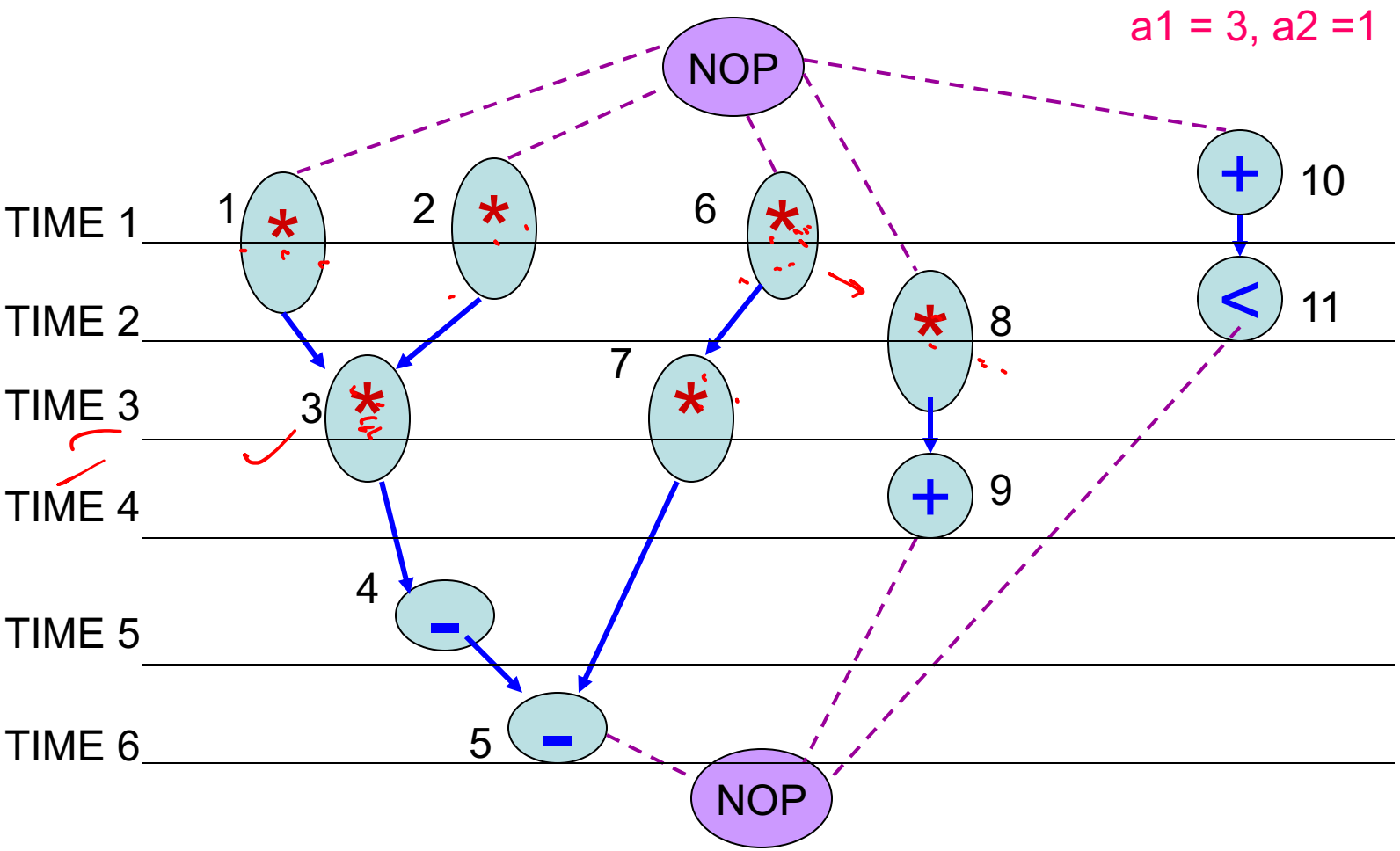
Scheduling with Pipelined Resources

- List scheduling algorithm can be extended to handle pipelined resources by allowing
  - Scheduling of overlapping operations with different start time, and
  - No data dependencies

Scheduled Sequencing Graph



Scheduled Seq. Graph with Pipelined Resources



Functional Pipelining

Assumption

- Resources are not pipelined
- Graph model is not hierarchical

data introduction interval.

- The number of pipelined resources depend on δ_0
- The higher δ_0 , larger the operations executing concurrently
- Upper bound on resource uses implies lower bound on δ_0

Functional Pipelining

ILP Formulation

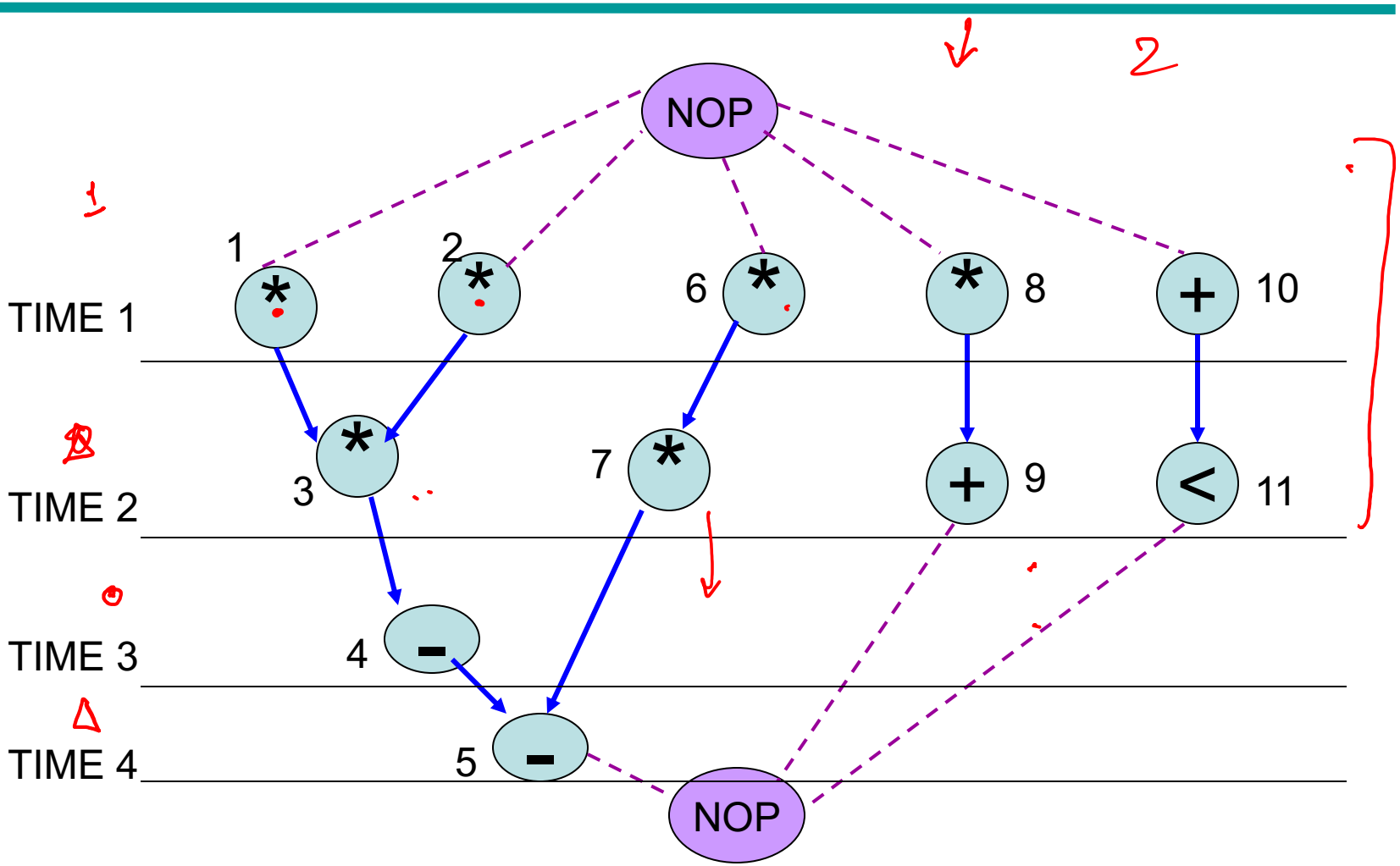
- Uniqueness constraints
- Sequencing constraints
- Resource constraints

$$\Sigma_p \Sigma_k x_{i,l+p\delta 0} \leq a_k$$

$\delta_0 = 2$

$x_{1,1}$ $x_{1,3}$

Sequencing Graph



Functional Pipelining

All operation must start only once

$x_{0,1} = 1$

$x_{1,1} = 1$

$x_{2,1} = 1$

$x_{3,2} = 1$

$x_{4,3} = 1$

$x_{5,4} = 1$

$x_{6,1} + x_{6,2} = 1$

$x_{7,2} + x_{7,3} = 1$

$x_{8,1} + x_{8,2} + x_{8,3} = 1$

$x_{9,2} + x_{9,3} + x_{9,4} = 1$

$x_{10,1} + x_{10,2} + x_{10,3} = 1$

$x_{11,2} + x_{11,3} + x_{11,4} = 1$

$x_{n,5} = 1$

Functional Pipelining

Constraints – based on sequencing

(more than one starting time for at least one operation)

$2\,x_{7,2} + 3\,x_{7,3} - x_{6,1} - 2\,x_{6,2} - 1 \geq 0$

$2\,x_{9,2} + 3\,x_{9,3} + 4\,x_{9,4} - x_{8,1} - 2\,x_{8,2} - 3\,x_{8,3} - 1 \geq 0$

$2\,x_{11,2} + 3\,x_{11,3} + 4\,x_{11,4} - x_{10,1} - 2\,x_{10,2} - 3\,x_{10,3} - 1 \geq 0$

$4\,x_{5,4} - 2\,x_{7,2} - 3\,x_{7,3} - 1 \geq 0$

$5\,x_{n,5} - 2\,x_{9,2} - 3\,x_{9,3} - 4\,x_{9,4} - 1 \geq 0$

$5\,x_{n,5} - 2\,x_{11,2} - 3\,x_{11,3} - 4\,x_{11,4} - 1 \geq 0$

Functional Pipelining

Resource constraints

$$\Sigma_p \Sigma_k x_{i,l+p\delta 0} \leq a_k$$

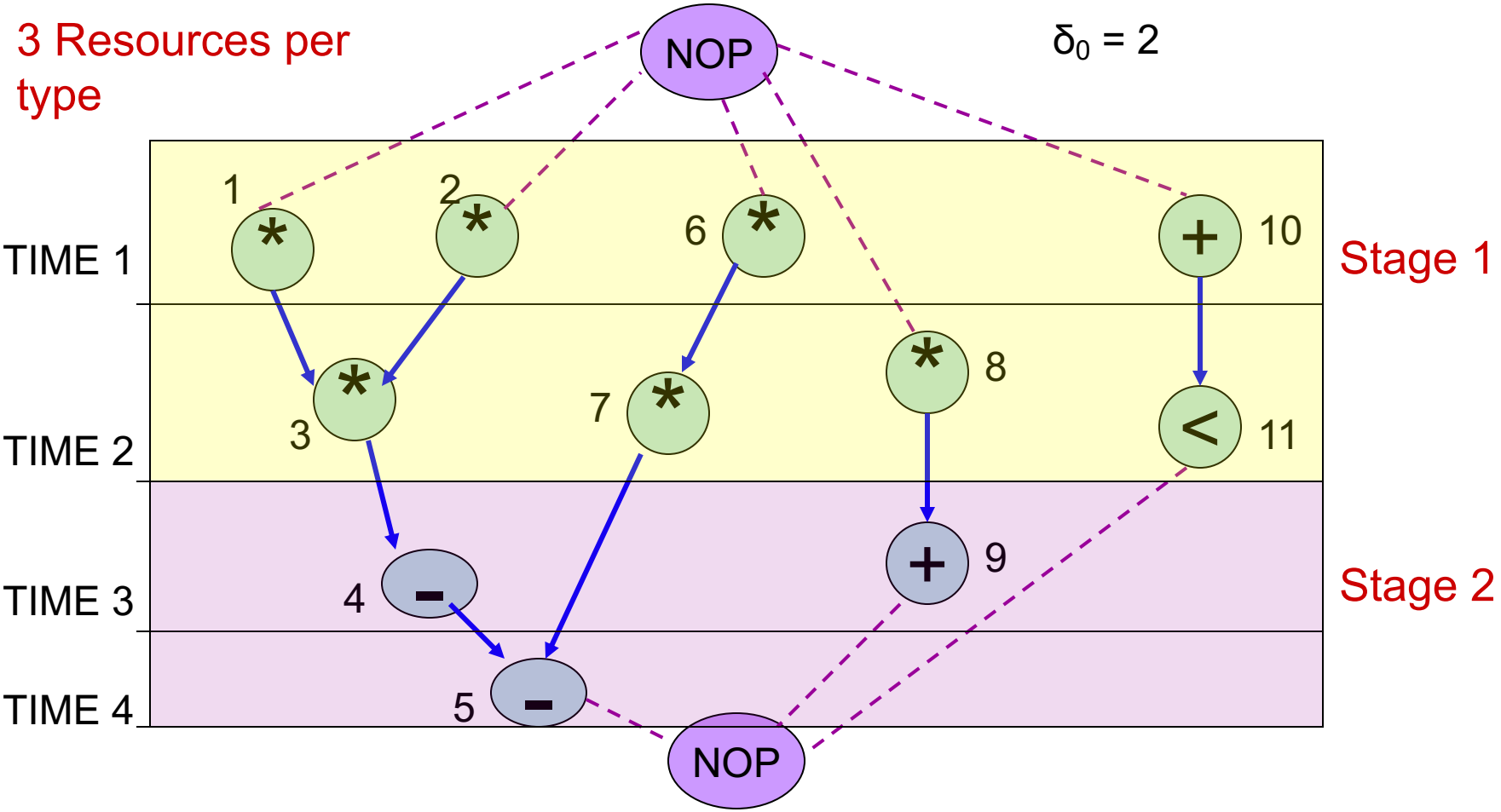
$$x_{1,1}+x_{2,1}+x_{6,1}+x_{8,1}+x_{7,3}+x_{8,3} \leq 3$$

$$x_{3,2}+x_{6,2}+x_{7,2}+x_{8,2} \leq 3$$

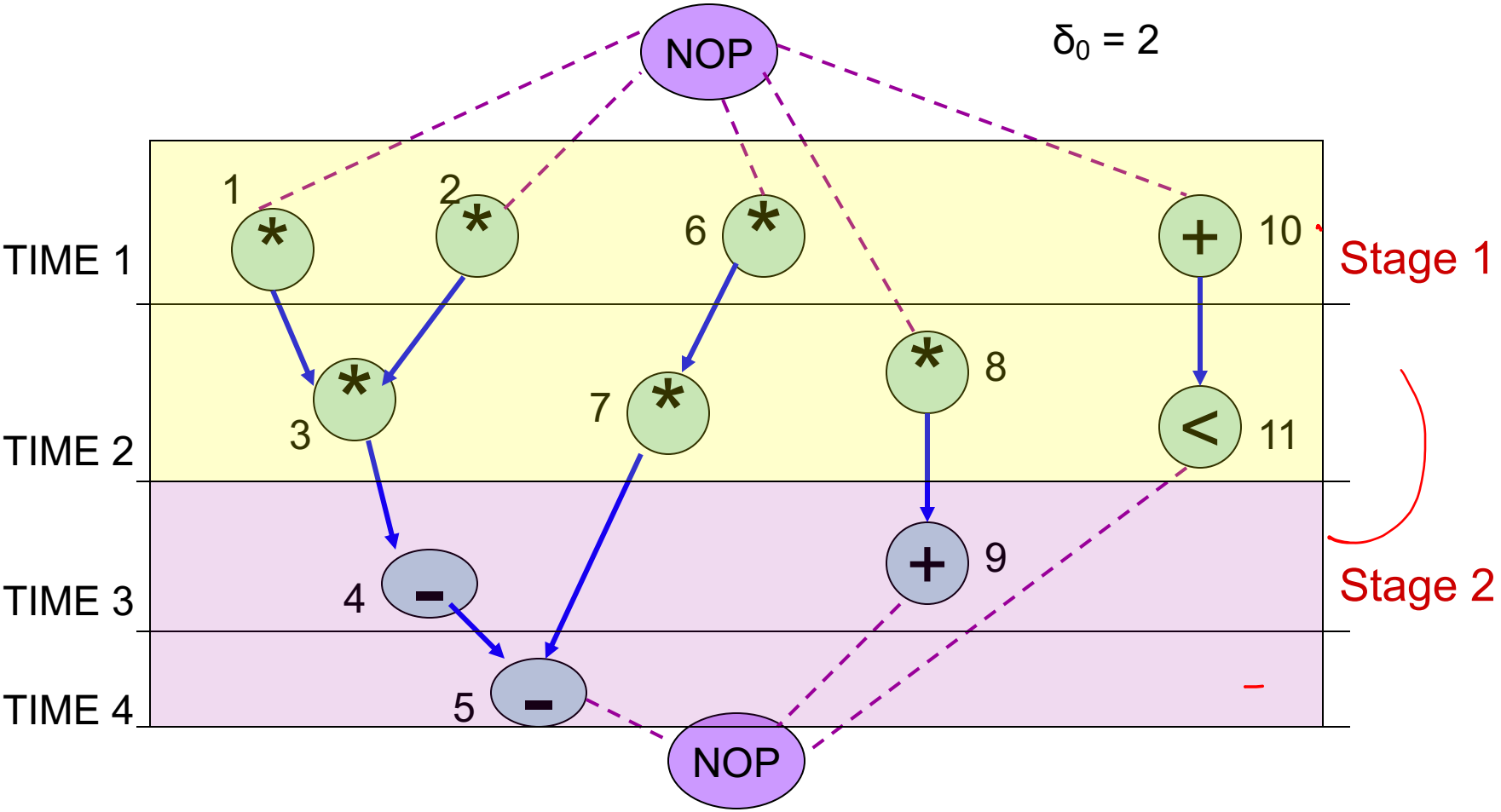
$$x_{10,1}+x_{4,3}+x_{9,3}+x_{10,3}+x_{11,3} \leq 3$$

$$x_{9,2}+x_{10,2}+x_{11,2}+x_{5,4}+x_{9,4}+x_{11,4} \leq 3$$

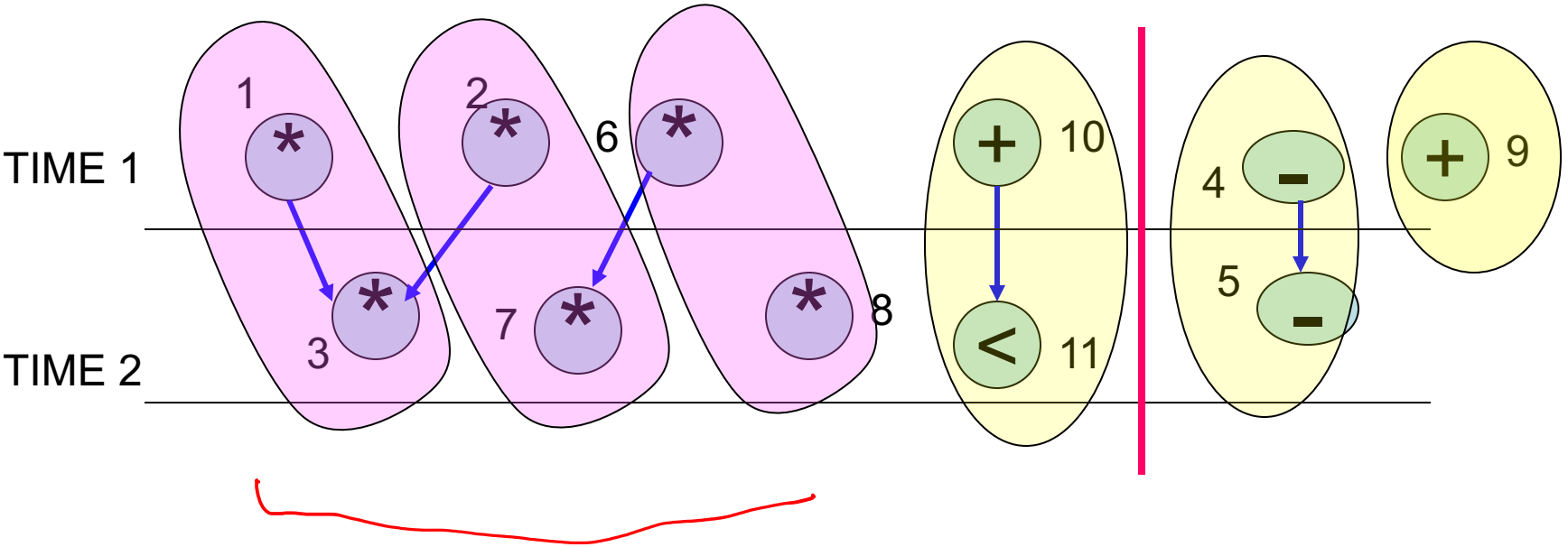
Scheduled Sequencing Graph



Scheduled Sequencing Graph

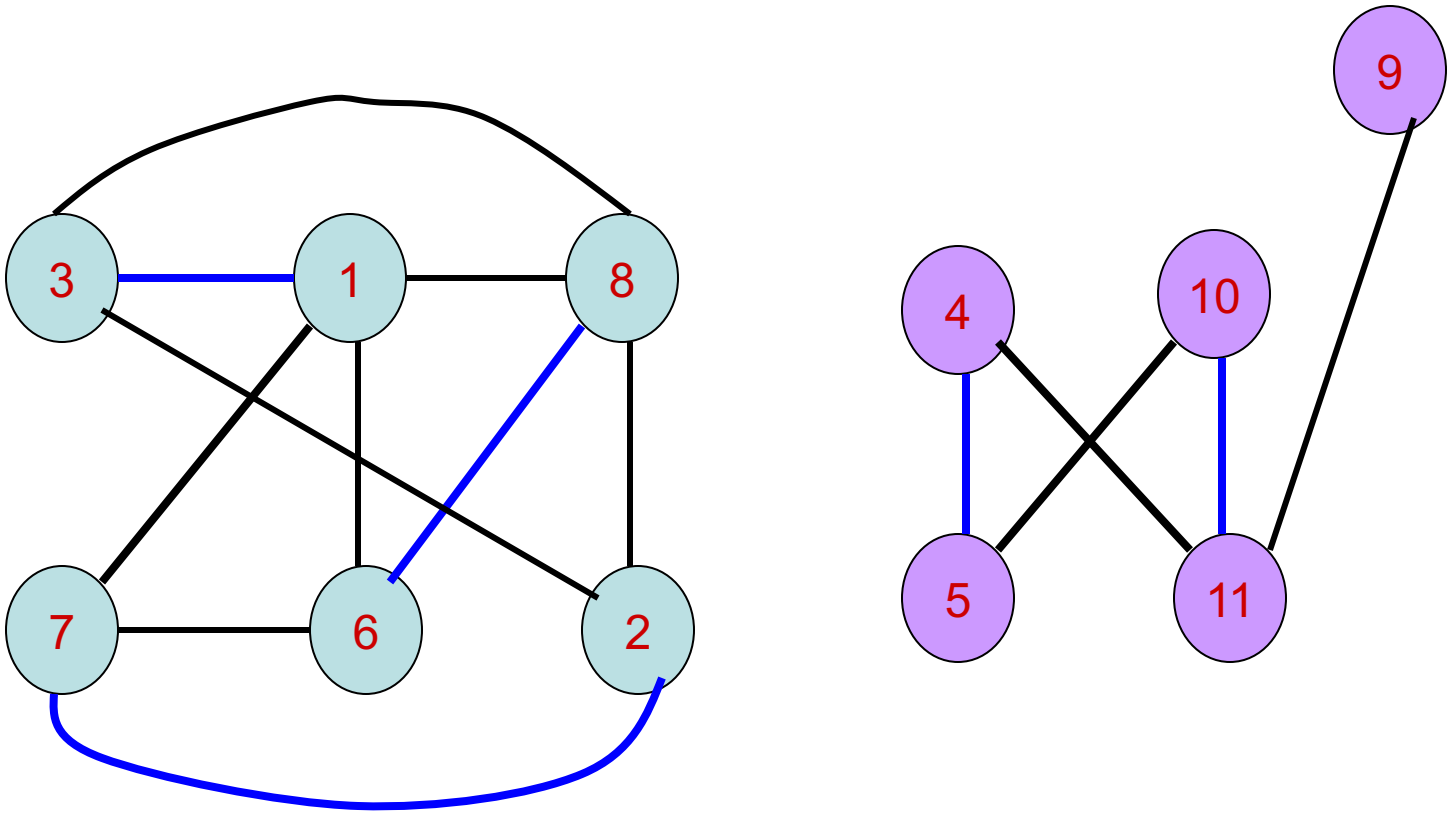


Sequencing Graph



Compatibility Graph

ZOLP.



Thank You