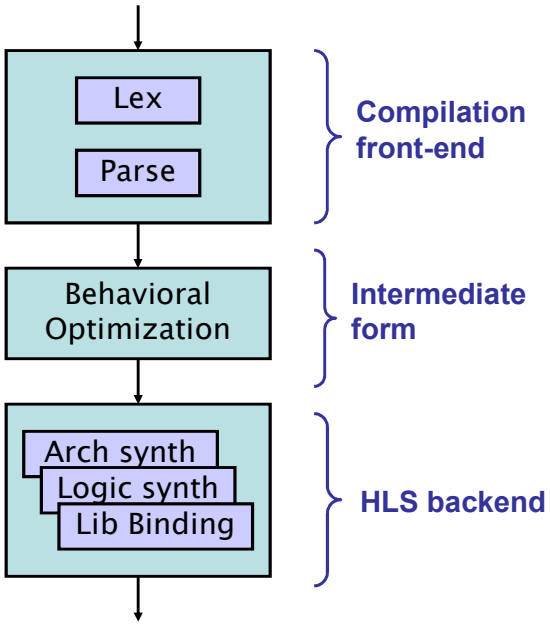
ECE382M.20: System-on-Chip (SoC) Design

Lecture 8 – High-Level Synthesis Algorithms

Sources:
Rajesh Gupta, UC San Diego

Andreas Gerstlauer
Electrical and Computer Engineering
The University of Texas at Austin
gerstl@ece.utexas.edu

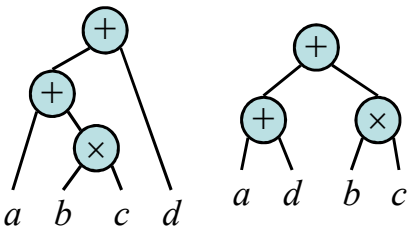
High-Level Synthesis Flow



- **Expressions**

$$x = a + b \times c + d$$

- **Parse tree**

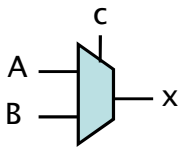


- **Data-flow graph (DFG)**
- **Control/data-flow of basic blocks (CDFG)**

Source: R. Gupta

Behavioral Optimization

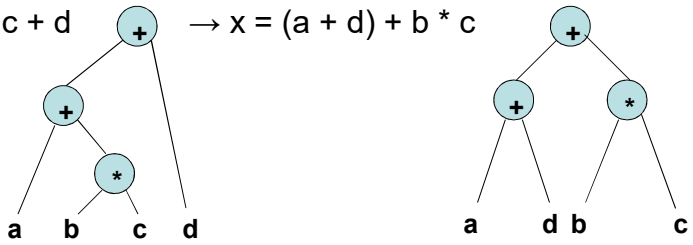
- **Data-flow transformations from software compilation**
  - Tree height reduction
    - Balance expression tree, expose parallelism
  - Constant and variable propagation (a = 1; c = 2 * a; → c = 2;)
  - Common sub-expression elimination (a=x+y; c=x+y; → c = a;)
  - Dead-code elimination
  - Operator strength reduction (e.g., *4 → << 2)
- **Control-flow transformations for hardware**
  - Conditional expansion
    - If (c) then x=A else x=B
    - ➢ compute A and B in parallel, x=(C)?A:B
  - Loop expansion/unrolling
    - Instead of three iterations of a loop, replicate the loop body three times



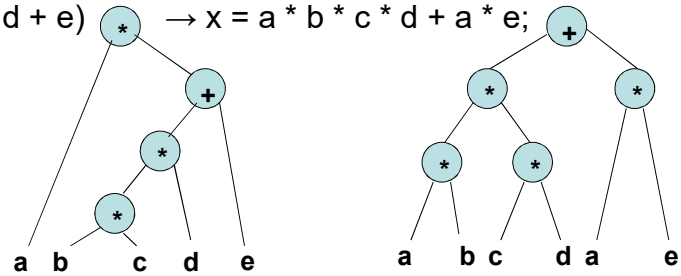
Source: R. Gupta

Tree-Height Reduction

- **Commutativity and associativity**
  - x = a + b * c + d → x = (a + d) + b * c



- **Distributivity**
  - x = a * (b * c * d + e) → x = a * b * c * d + a * e;



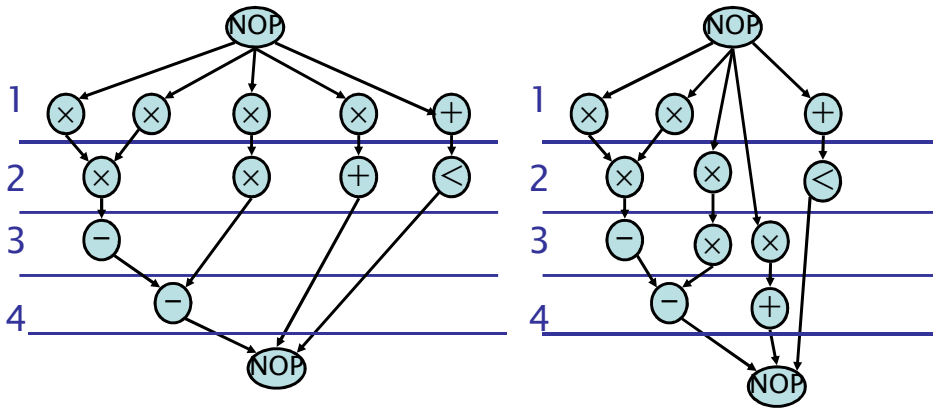
Architectural Synthesis

- **Deals with "computational" behavioral descriptions**
  - Behavior as sequencing graph (called dependency graph, or data flow graph DFG)
  - Hardware resources as library elements
    - Pipelined or non-pipelined
    - Resource performance in terms of execution delay
  - Constraints on operation timing & clock period
  - Constraints on hardware resource availability
  - Storage as registers, data transfer using wires
- **Objective**
  - Generate a synchronous, single-phase clock circuit
  - Might have multiple feasible solutions (explore tradeoff)
  - Satisfy constraints, minimize objective:
    - Maximize performance subject to area constraint
    - Minimize area subject to performance constraints

Source: R. Gupta

Synthesis in Temporal Domain

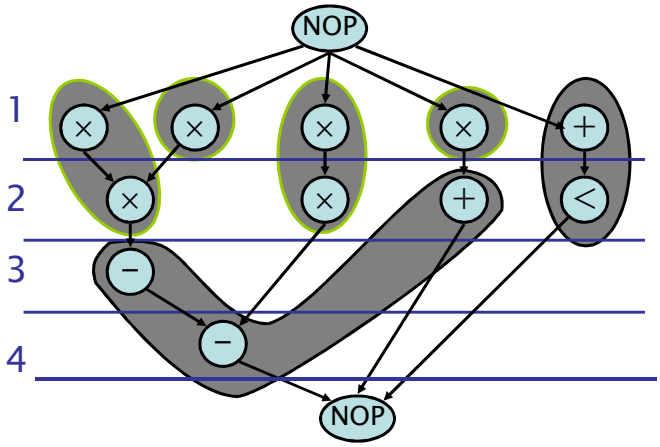
- **Scheduling**
  - Schedule is a mapping of operations to time slots (cycles)
  - Scheduled sequencing graph is a labeled graph



Source: R. Gupta

Synthesis in Spatial Domain

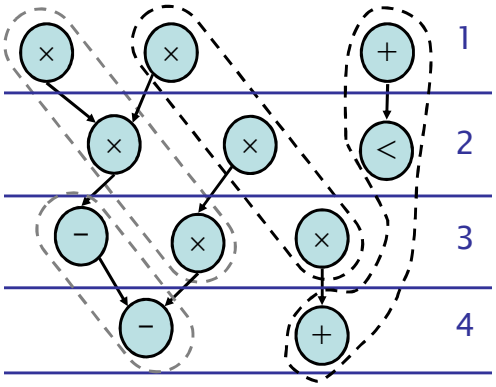
- **Resource binding and sharing**
  - Map each operation to a resource of the same type
  - More than one serial operation bound to same resource
  - Can be represented using hyperedges (vertex partition)



Source: R. Gupta

How Is the Datapath Implemented?

- **Assuming the following schedule and binding**
  - Wires between modules?
  - Input selection?
  - How does binding/ scheduling affect congestion?
  - How does binding/ scheduling affect steering logic?



Scheduling and Binding

- **Resource constraints:**
  - Mapping of operations into resource types $T(v_2) = adder$
  - Number of resource instances of each type $a_{adder} = 2$
- **Scheduling:**
  - Labeled vertices $\phi(v_2)=1$
- **Binding:**
  - Hyperedges (or vertex partitions) $\beta(v_2)=adder1$
- **Cost:**
  - Number of resources $\sum a_k \approx$ area
  - Registers, steering logic (Muxes, busses), wiring, control unit
- **Delay:**
  - Start time of the "sink" node $\phi(v_{sink})$
  - Might be affected by steering logic and schedule (control)
  - ➢ Resource-dominated vs. ctrl-dominated

Architectural Optimization

- **Optimization in view of design space flexibility**
- **A multi-criteria optimization problem**
  - Determine schedule ϕ and binding β.
  - Under area A, latency λ and cycle time τ objectives
- **Find non-dominated points in solution space**
  - Solution space tradeoff curves
  - Non-linear, discontinuous
- **Evaluate (estimate) cost functions**
- ➢ **Unconstrained optimization problems**
  - Min area: solve for minimal A binding
  - Min latency: solve for minimum λ scheduling
- ➢ **Single-objective optimization problems**
  - Minimize A under λ constraint
  - Minimize λ under A (≈ resource) constraint

Scheduling and Binding

- **Cost λ and A determined by both ϕ and β**
  - Also affected by floorplan and detailed routing
- **β affected by ϕ:**
  - Resources cannot be shared among concurrent ops
- **ϕ affected by β:**
  - Resources cannot be shared among concurrent ops
  - When register and steering logic delays added to execution delays, might violate cycle time
- **Order?**
  - Apply either one (scheduling, binding) first