Embedded Systems

REVIEW

TMS320C6x Datapath

Overview XILINX FPGA

- • All Xilinx FPGAs contain the same basic resources
	- – Slices grouped into Configurable Logic Blocks (CLBs)
		- Contain combinatorial logic and register resources
	- IOBs
		- Interface between the FPGA and the outside world
	- Programmable interconnect
	- Other resources
		- Memory
		- Multipliers
		- Global clock buffers
		- Boundary scan logic

CS - ESen andere en de statistike en de statistik

Embedded Processors in FPGAs REVIEW

- Hard Core
	- EP is a dedicated physical component of the chip separate from the programmable logic
	- E.g. Xilinx Virtex families (PowerPC 405)
- Soft Core
	- \blacksquare Embedded processor is also a synthesized to the FPGA to th programmable logic on the chip
	- E.g. Altera (NIOS), Xilinx (MicroBlaze)

Partial Reconfiguration

Technology and Benefits

- **Partial Reconfiguration enables:**
	- **System Flexibility**
		- Perform more functions while maintaining communication links
	- **Size and Cost Reduction**
		- Time-multiplex the hardware to require a smaller FPGA
	- **Power Reduction**
		- Shut down power-hungry tasks when not needed

Embedded System Hardware

 \blacksquare Embedded system hardware is frequently used in a loop (*"hardware in a loop"*):

Communication Requirements -

-

- Real-time behavior
- **Efficient, economical** (e.g. centralized power supply)
- Appropriate bandwidth and communication delay
- Robustness
- Fault tolerance
- **Maintainability**
- П **Diagnosability**
- **Security**
- **Safety**

Memory

- For the memory, efficiency is again a concern:
	- speed (latency and throughput); predictable timing
	- **E** energy efficiency
	- **size**
	- cost
	- other attributes (volatile vs. persistent, etc)

Memory hierarchy

 $_{\rm S}$ $\,$ (in terms of energy consumption, access times, size) $_{\rm I}$ 9- "Small is beautiful"

CS - ES

Static Timing Analysis

producing the input to schedulability analysis

Schedulability analysis has assumed the knowledge of the execution time of tasks.

So, the problem to solve:

- \blacksquare Given
	- 1 a software task to produce some reaction,
	- 2 a hardware platform, on which to execute the software,
	- 3 a required reaction time, e.g. the period of the task.

Derive:

 \triangleright a reliable (and precise) upper bound on the execution times.

Timing Analysis

- **Example 25 Architecture Synthesis**
	- HW/SW Codesign
	- **Power Aware Computing**
	- 3.2.2011 Lecture by Bernd Finkbeiner, Head of Reactive Systems Group at Saarland University(http://react.cs.uni-sb.de/

Architecture Synthesis

Design a hardware architecture that efficiently executes a given algorithm.

- \blacksquare *tasks*:
	- ٠ **allocation** (determine the necessary hardware resources)
	- \blacksquare **scheduling** (determine the timing of individual operations)
	- Ξ *binding* (determine relation between individual operations of the algorithm and hardware resources)

Classification of synthesis algorithms

Synthesis methods can often be applied *independently of granularity*

Synthesis in Temporal Domain

- \blacksquare Scheduling and binding can be done in different orders or together
- Schedule:
	- Mapping of operations to time slots (cycles)
	- A scheduled sequencing graph is a labeled graph

Schedule in Spatial Domain

- **Resource sharing**
	- **More than one operation bound to same resource**
	- **Serialized operations**

Source: Teich: Dig. HW/SW Systeme;Thiele ETHZ

Models

- Sequence graph $G_S = (V_S, E_S)$ where $V_{\rm s}$ denotes the operations of the algorithm and $E_{\rm s}$ the dependence relations.
- **Resource graph** $G_R = (V_R, E_R)$, $V_R = V_S \cup V_T$ where V_{τ} denote the resource types of the architecture and G_R is a bipartite graph. An edge $(v_s, v_t) \in E_R$ represents the availability of a resource type v_t for an operation $V_{\rm s}$.
- ► Cost function $c:V_T\to \mathbf{Z}$
- ► Execution times $w: E_R \rightarrow \mathbf{Z}^{\geq 0}$ are assigned to each edge $(v_s, v_t) \in E_R$ and denote the execution time of operation $v_s \in V_S$ on resource type $v_t \in V_T$.

Allocation and Binding

An allocation is a function $\alpha: V_T \to \mathbb{Z}^{\geq 0}$ that assigns to each resource type $v_t \in V_T$ the number $\alpha(v_t)$ of available instances.

A binding is defined by functions $\beta: V_S \to V_T$ and $\gamma: V_S \to \mathbb{Z}^{>0}$. Here, $\beta(v_s) = v_t$ and $\gamma(v_s) =$ r denote that operation $v_s \in V_S$ is implemented on the rth instance of resource type $v_t \in V_T$.

Scheduling

A schedule is a function $\tau: V_S \to \mathbb{Z}^{>0}$ that determines the starting times of operations. A schedule is feasible if the conditions

$$
\tau(v_j)-\tau(v_i)\geq w(v_i) \quad \forall (v_i,v_j)\in E_S
$$

are satisfied. $w(v_i) = w(v_i, \beta(v_i))$ denotes the execution time of operation v_i .

The latency L of a schedule is the time difference between start node v_0 and end node v_n : $L = \tau(v_n) - \tau(v_0)$.

Schedule

$$
L = \mathcal{I}(v_n) - \mathcal{I}(v_0) = 4
$$

 $\tau(v_1) = \tau(v_2) ... = 1$ $\tau(v_0) = 1$

 \bullet \bullet \bullet

Binding

Example $(\alpha(r_1) = 4, \alpha(r_2) = 2)$:

$$
\beta(v_1) = r1, \ \gamma(v_1) = 1
$$

$$
\beta(v_2) = r2, \ \gamma(v_2) = 1
$$

$$
\beta(v_3) = r1, \ \gamma(v_3) = 2
$$

$$
\beta(v_6) = r1, \gamma(v_3) = 3
$$

 \bullet \bullet \bullet

 \bullet \bullet \bullet

As soon as possible (ASAP) scheduling

ASAP: All tasks are scheduled as early as possible

- Loop over (integer) time steps:
	- Compute the set of unscheduled tasks for which all predecessors have finished their computation
	- Schedule these tasks to start at the current time step.

ASAP Schedules

```
ASAP(G_S(V_S, E_S), w) {
   \tau(v_0) = 1;REPEAT {
      Determine v_i whose predec. are planed;
      \tau(v_i) = \max{\{\tau(v_j) + w(v_j) \; \forall (v_j, v_i) \in E_S\}}} UNTIL (v_n is planned);
   RETURN (\tau);
\}
```


As-late-as-possible (ALAP) scheduling

ALAP: All tasks are scheduled as late as possible

Start at last time step*:

Schedule tasks with no successors and tasks for which all successors have already been scheduled.

^{*} Generate a list, starting at its end

ALAP Schedules

4

```
ALAP(G<sub>S</sub>(V<sub>S</sub>, E<sub>S</sub>), w, L<sub>max</sub>) {
     \tau(v_n) = L_{max} + 1;REPEAT {
          Determine v_i whose succ. are planed;
         \tau(v_i) = \min{\{\tau(v_j) \; \forall (v_i, v_j) \in E_S\}} - w(v_i)} UNTIL (v_0 is planned);
     RETURN (\tau);
\}NOP1\times) (\times\times\bm{\times}2
                                       \overline{\mathsf{X}}\overline{\mathsf{X}}3
                                                                                 +\overline{\mathsf{X}}
```
 $\, +$

NOP

 $(+)$ \leq

Scheduling under Detailed Timing Constraints

- Motivation
	- Interface design.
	- Control over operation start time.
- Constraints
	- Upper/lower bounds on start-time difference of any operation pair.
- **Minimum timing constraints between two operations**
	- An operation follows another by *at least* a number of prescribed time steps
- **Maximum timing constraints between two operations**
	- An operation follows another by *at most* a number of prescribed time steps

Scheduling under Detailed Timing Constraints

- **Example**
	- Circuit reads data from a bus, performs computation, writes result back on the bus.
	- **Bus interface constraint: data written three cycles after read.**
	- Minimum and maximum constraint of 3 cycles between read and write operations.

Constraint graph model

- \blacksquare Start from a sequencing graph
- \blacksquare Model delays as weights on edges
- \blacksquare Add forward edges for minimum constraints
- \blacksquare Add backward edges for maximum constraints

Add this edge for min constraint

Weighted Constraint Graph

In order to represent a *feasible schedule*, we have one edge corresponding to each precedence constraint with

$$
d(v_i, v_j) = w(v_i)
$$

where $w(v_i)$ denotes the execution time of v_i .

- A consistent assignment of starting times $\tau(v_i)$ to all operations can be done by solving a *single source* **longest path problem.**
- A possible algorithm (**Bellman-Ford**) has complexity $O(|V_C| |E_C|)$:

Iteratively set $\tau(v_i) := \max{\{\tau(v_i), \tau(v_i)\} + d(v_i, v_i)}$: $(v_i, v_j) \in E_C$ for all $v_j \in V_C$ starting from $\tau(v_i) = -\infty$ for $v_i \in V_C \setminus \{v_0\}$ and $\tau(v_0) = 1$.

Weighted Constraint Graph

Example: $w(v1) = w(v3) = 2$

$$
w(v2) = w(v4) = 1
$$

 $\tau(v_j) := \max{\{\tau(v_j), \tau(v_i) + d(v_i, v_j)\}}$

en andere en de statistike en de statistik 31 - Source: Thiele, ETHZ

Solution - Constraint Graph Model

Mul delay = 2 ADD delay =1

(Resource constrained) List Scheduling

- List scheduling: extension of ALAP/ASAP method
- Preparation:
	- Greedy strategy (does NOT guarantee optimum solution)
	- Topological sort of task graph *G*=(*V,E*)
	- \blacksquare Computation of priority of each task:

Possible priorities *u*:

- Number of successors
- Longest path
- **Mobility** = τ (ALAP schedule)- τ (ASAP schedule)
	- Defined for each operation
	- Zero mobility implies that an operation can be started only at one given time step
	- Mobility greater than 0 measures span of time interval in
- 33 -which an operation may start Slack on the start time.

CS - ES

Mobility as a priority function

Mobility is not very precise

Algorithm

```
\blacksquareList(G(V,E), B, u){
i :=0;
  repeat {
    Compute set of candidate tasks Ai ;
    Compute set of not terminated tasks G_i ;
    Select S_i \subseteq A_i of maximum priority r such that
    |S_i| + |G_i| \leq B (*resource constraint*)
\blacksquareforeach (v_j \in S_i): \tau(v_j):=i; (*set start time*)
    i := i +1;
  }
  until (all nodes are scheduled);
```
may be repeated for different task/ processor classes

CS - ESen andere en de statistike en de statistik

return (τ);

} Complexity: *^O*(|*V*|)

Example

Assuming $B = 2$, unit execution time and *u* : path length

$$
u(a)=u(b)=4
$$

u(c)=u(f)=3

$$
u(d)=u(g)=u(h)=u(j)=2
$$

$$
u(e)=u(i)=u(k)=1
$$

$$
\forall i: G_i=0
$$

$$
\begin{array}{c|c}\n\downarrow \\
\hline\n\downarrow \\
\h
$$

does NOT guarantee optimum solution e.g.

List Scheduling

Integer linear programming models

- Ingredients:
	- Cost function L Involving linear expressions of
	- Constraints integer variables from a set *^X*

Cost function $C = \sum a_i x_i$ with $a_i \in R, x_i \in \mathbb{N}$ (1) $x_i \in X$ $=\sum a_i x_i$ with $a_i\in R, x_i\in$

$$
\text{Constraints: } \forall j \in J : \sum_{x_i \in X} b_{i,j} \ x_i \ge c_j \ \text{with } b_{i,j}, c_j \in \mathbb{R} \ (2)
$$

Def.: The problem of minimizing (1) subject to the constraints (2) is called an **integer linear programming (ILP) problem**.

If all x_i are constrained to be either 0 or 1, the IP problem said to be a **0/1 integer linear programming problem**.

Example

$$
C = 5x_1 + 6x_2 + 4x_3
$$

$$
x_1 + x_2 + x_3 \ge 2
$$

$$
x_1, x_2, x_3 \in \{0, 1\}
$$

x_1	x_2	x_3	C
0	1	1	10
1	0	1	9
1	1	0	11
1	1	1	15

Remarks on integer programming

- **Integer programming is NP-complete**
- Running times depend exponentially on problem size, but problems of >1000 vars solvable with good solver (depending on the size and structure of the problem)
- **ILP/LP models good starting point for modeling, even if heuristics are** used in the end.
- Solvers: lp_solve (public), CPLEX (commercial), …

ILP Formulation of ML-RCS

- **Minimize latency given constraints on area or** the resources (ML-RCS)
- Use binary decision variables
	- *i = 0, 1, ..., n*
	- $l = 1, 2, ..., \lambda'+1$ λ' given upper-bound on latency
	- x_{il} = 1 if operation *i* starts at step *l*, 0 otherwise.
- **Set of linear inequalities (constraints),** and an objective function (min latency)

ILP Formulation of ML-RCS

Dbservation

$$
x_{il} = 0 \quad \text{for} \quad l < t_i^S \quad \text{and} \quad l > t_i^L
$$
\n
$$
(t_i^S = ASAP(v_i), t_i^L = ALAP(v_i))
$$

\n- \n
$$
t_i = \sum_{l} l \cdot x_{il}
$$
\n
$$
t_i = \text{start time of op } i.
$$
\n
\n- \n
$$
\sum_{m=l-d_i+1}^{l} x_{im} = 1
$$
\n
$$
\sum_{l}^{l} x_{im} = 1
$$
\n
$$
\sum_{l}^{l} x_{il} = 1
$$
\n<math display="block

Start Time vs. Execution Time

- $\textcolor{red}{\bullet}$ For each operation v_i , only one start time
- \blacksquare If d_i =1, then the following questions are the same:
	- Does operation v_i **start** at step *l*?
	- \blacksquare Is operation v_i running at step *l*?
- But if d_i >1, then the two questions should be formulated as:
	- Does operation *v*_istart at step *l*?
		- Does *xil = 1* hold?
	- \blacksquare Is operation v_i running at step *l*?
		- Does the following hold?

$$
\sum_{\text{CS - ES}}^{l} x_{im} = 1 ?
$$

Operation *vi* **Still Running at Step** *l* **?**

Is v_g running at step 6?

• **Is**
$$
x_{9,6} + x_{9,5} + x_{9,4} = 1
$$
 ?

- Note:
	- **Only one (if any) of the above three cases can happen**
- $\mathsf{CS}\text{-}\mathsf{ES}$ that type $\mathsf{CS}\text{-}\mathsf{ES}$ $-44-$ ■ To meet resource constraints, we have to ask the same question for ALL steps, and ALL operations of

Operation *vi* **Still Running at Step** *l* **?**

 \blacksquare Is $v_{\boldsymbol{i}}$ running at step *l* ?

• Is
$$
x_{i,l} + x_{i,l-1} + ... + x_{i,l-di+1} = 1
$$
 ?

ILP Formulation of ML-RCS (cont.)

- Constraints:
	- **Unique start times:**

$$
\sum_{i} x_{i i} = 1, \quad i = 0, 1, \ldots, n
$$

Sequencing (dependency) relations must be satisfied

$$
t_i \ge t_j + d_j \ \forall (v_j, v_i) \in E \Longrightarrow \sum_l l \cdot x_{il} \ge \sum_l l \cdot x_{jl} + d_j
$$

Resource constraints

$$
\sum_{i:T(v_i)=k} \sum_{m=l-d_i+1}^{l} x_{im} \le a_k, \quad k=1,\ldots,n_{res}, \quad l=1,\ldots,\overline{\lambda}+1
$$

- Objective: min $c^T t$.
	- *t* =start times vector, *c* =cost weight

ILP Example

- **•** Assume λ = 4
- First, perform ASAP and ALAP
	- (we can write the ILP without ASAP and ALAP, but using ASAP and ALAP will simplify the inequalities)

ILP Example: Unique Start Times Constraint

■ Without using ASAP and ALAP values:

■ Using ASAP and ALAP:

 $x_{1,1} = 1$

....

 $x_{1,1} + x_{1,2} + x_{1,3} + x_{1,4} = 1$ $x_{1} + x_{2} + x_{3} + x_{4} =$ $x_{2,1} = 1$ $x_{2,1} + x_{2,2} + x_{2,3} + x_{2,4} = 1$ $x_{2} + x_{2} + x_{3} + x_{4} =$ $x_{3,2} = 1$ $x_{4,3} = 1$... $x_{5,4} = 1$ $_{6,1} + x_{6,2} = 1$ x_{c} , $+x_{c}$, $=$ \bigoplus_{v10} $\begin{array}{c} 1 \\ \hline \end{array}$ $\bigotimes_{v1} \bigotimes_{v2}$ (x) $(x)_{\nu\delta}$ $(\mathsf{x})_{\mathsf{v}\mathsf{c}}$ $\bigotimes vII$ $\overline{\mathbf{z}}$ $x_{7,3} + x_{7,3} = 1$ $x_{7,2} + x_{7,3} =$ $(\times)_{\nu\delta}$ $\bigoplus_{\nu\downarrow 0}$ $(\mathsf{x})_{\mathsf{v}}$ $x_{8,1} + x_{8,2} + x_{8,3} = 1$ x_{0} , $+ x_{0}$, $+ x_{0}$, $=$ $\bigotimes vII$ $v₂$ $s_{9,2} + x_{9,3} + x_{9,4} = 1$ x_{0} , $+ x_{0}$, $+ x_{0}$, $=$

CS - ESen andere en de statistike en de statistik

ILP Example: Dependency Constraints

 Using ASAP and ALAP, the non-trivial inequalities are: (assuming unit delay for + and *)

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

\n
$$
2.x_{9,2} + 3.x_{9,3} + 4.x_{9,4} - x_{8,1} - 2.x_{8,2} - 3.x_{8,3} - 1 \ge 0
$$

\n
$$
2.x_{11,2} + 3.x_{11,3} + 4.x_{11,4} - x_{10,1} - 2.x_{10,2} - 3.x_{10,3} - 1 \ge 0
$$

\n
$$
4.x_{5,4} - 2.x_{7,2} - 3.x_{7,3} - 1 \ge 0
$$

\n
$$
\underbrace{\bigotimes_{\mathcal{V}^1 \otimes \mathcal{V}^2} \bigotimes_{\mathcal{V}^2 \otimes \mathcal{V}^2} \bigotimes_{\mathcal{V}^3 \otimes \mathcal{V}^4} \bigotimes_{\mathcal{V}^4} \bigotimes_{\mathcal{V}^5 \otimes \mathcal{V}^4} \bigotimes_{\mathcal{V}^6 \otimes \mathcal{V}^7} \bigotimes_{\mathcal{V}^8 \otimes \mathcal{V}^8} \bigotimes_{\mathcal{V}^9} \bigotimes_{\mathcal{V}^9} \bigotimes_{\mathcal{V}^8} \bigotimes_{\mathcal{V}^9} \bigotimes
$$

ILP Example: Resource Constraints

 (x)

 \bigcirc v

 $v₂$

Resource constraints (assuming 2 adders and 2 multipliers)

 \bigoplus v10

 (x) v 8

- $x_{2,1} + x_{3,1} + x_{4,1} + x_{8,1} \leq 2$ $x_{1} + x_{2} + x_{3} + x_{4} + x_{9}$
- $x_{6,2} + x_{7,2} + x_{7,2} + x_{8,2} \leq 2$ $x_2 + x_3 + x_4 + x_2 + x_3 + x_4$

- $_{10,1}\leq 2$ $x_{10,1} \leq$
- $x_{10,2} + x_{11,2} \leq 2$ x_{0} , x_{10} , x_{21} , x_{31} , x_{4}
- $x_{9,4} + x_{9,4} + x_{11,4} \leq 2$ $x_{1,3} + x_{9,3} + x_{10,3} + x_{11,3} \leq 2$ $x_{5,4} + x_{2,4} + x_{11,4} \leq$ $x_{4,2} + x_{0,3} + x_{10,2} + x_{11,3} \leq$
- Objective: *Min X_{n,1}* + 2*X_{n,2}* +3*X_{n,3}* +4*X_{n,4}*

 $\bigotimes v1 \bigotimes v2$

 $\overline{\bigotimes_{vII}}$

 \overline{a}

 (x)

50 -

CS - ESen andere en de statistike en de statistik

(Time constrained) Force-directed scheduling

- Goal: balanced utilization of resources
- Based on spring model
- Originally proposed for high-level synthesis
- Force
	- Used as a priority function
	- Related to concurrency sort operations for least force
	- Mechanical analogy: Force = constant x displacement
		- Constant = operation-type distribution
		- Displacement = change in probability

* [Pierre G. Paulin, J.P. Knight, Force-directed scheduling in automatic data path synthesis, *Design Automation Conference* (DAC), 1987, S. 195-202]

Force-Directed Scheduling

The Force-Directed Scheduling approach reduces the amount of:

- Functional Units
	- Registers
	- Interconnect

This is achieved by balancing the concurrency of operations to ensure a high utilization of each unit.

Next: computation of "forces"

- Direct forces push each task into the direction of lower values of *D*(*i*)*.*
- Impact of direct forces on dependent tasks taken into account by indirect forces
- **Balanced resource usage** \approx **smallest forces**
- For our simple example and time constraint=6: result = ALAP schedule

1.Compute time frames *R***(***j***) 2. Compute "probability"** $P(j,i)$ **of assignment** $j \rightarrow i$

3. Compute "distribution" *D***(***i***) (# Operations in control step** *i***)**

Example

Scheduling – An example

Step 3 : Calculate the *force* (a new metric)

A metric called *force* is introduced. The force is used to optimize the utilization of units. A high positive force value indicates a poor utilization.

$$
Force(j) = DG(j) - \sum_{i=t}^{b} \frac{DG(i)}{h}
$$

Scheduling – An example

Step 3 : Calculate the *force* (a new metric) With the operation x' in control-step 1.

Scheduling – An example

By repeatedly assigning operations to various control-steps and calculating the force associated with the choice several force values will be available.

The Force-directed scheduling algorithm chooses the assignment with the lowest force value, which also balances the concurrency of operations most efficiently.

Overall approach

procedure forceDirectedScheduling; **begin**

AsapScheduling;

AlapScheduling;

while not all tasks scheduled **do**

begin

May be repeated for different task/ processor classes

select task *T* with smallest total force; schedule task *T* at time step minimizing forces; recompute forces;

end;

Not sufficient for today's complex, heterogeneous hardware platforms

end

Force-Directed Scheduling

The Force-Directed Scheduling approach reduces the amount of:

- Functional Units
- Registers
	- Interconnect

By introducing Registers and Interconnect as *storage operations*, the force is calcuted for these as well.

Force-Directed Scheduling

- **Example 25 Architecture Synthesis**
- HW/SW Codesign
	- **Power Aware Computing**
	- 3.2.2011 Lecture by Bernd Finkbeiner, Head of Reactive Systems Group at Saarland University (http://react.cs.uni-sb.de/

Codesign Definition and Key Concepts

- Codesign
	- The meeting of system-level objectives by exploiting the trade-offs between hardware and software in a system through their concurrent design
- Key concepts
	- **Concurrent:** hardware and software developed at the same time on parallel paths
	- \blacksquare Integrated: interaction between hardware and software development to produce design meeting performance criteria and functional specs

Typical Codesign Process

