

طراحی سیستم‌های تعبیه شده Embedded System Design

فصل پنجم - قسمت هشتم

پیاده‌سازی سیستم‌های تعبیه شده

Implementing Embedded Systems: Hardware / Software Codesign

کاظم فولادی
دانشکده‌ی مهندسی برق و کامپیوتر
دانشگاه تهران

kazim@fouladi.ir



جریان‌های طراحی واقعی و ابزارها Actual design flows and tools

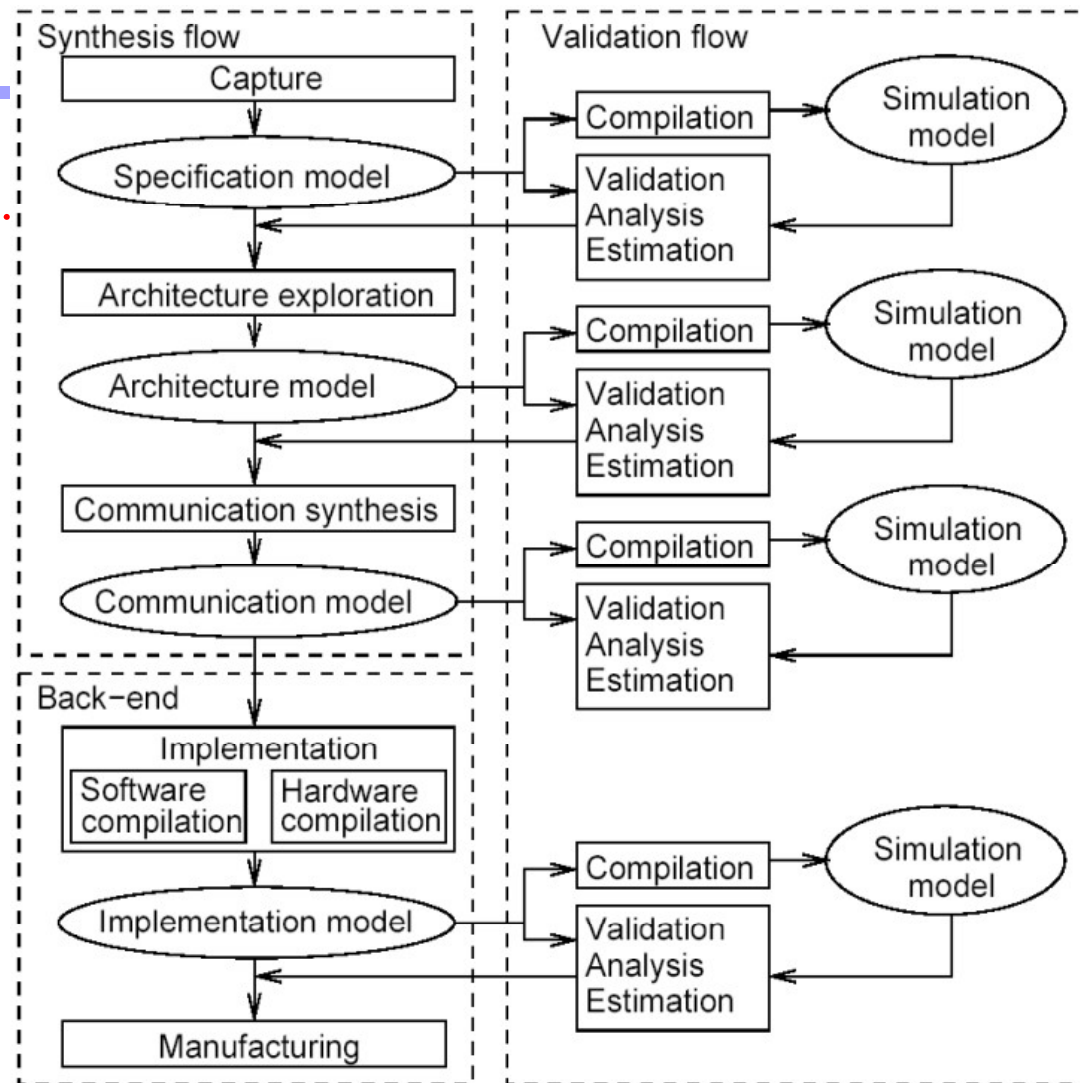
Design steps can be used in different combinations.
In the following, we will present some examples ...

1. SpecC [Gajski, Dömer, Gerstlauer]
Focus on intellectual property (IP) components



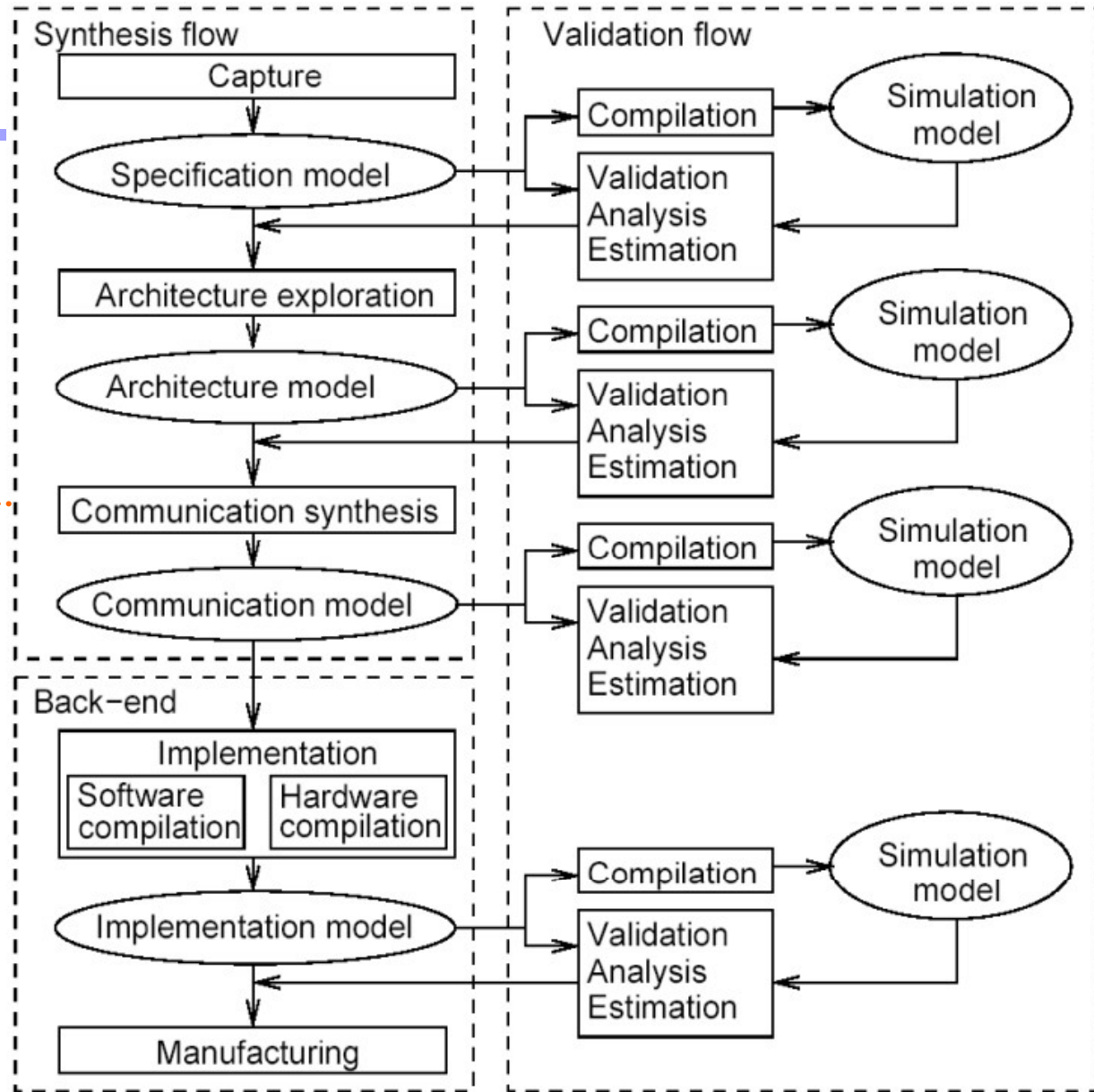
SpecC

Comprises allocation (selecting components from a library), partitioning (mapping of parts of the system specification onto the components) & scheduling (serialization of execution).



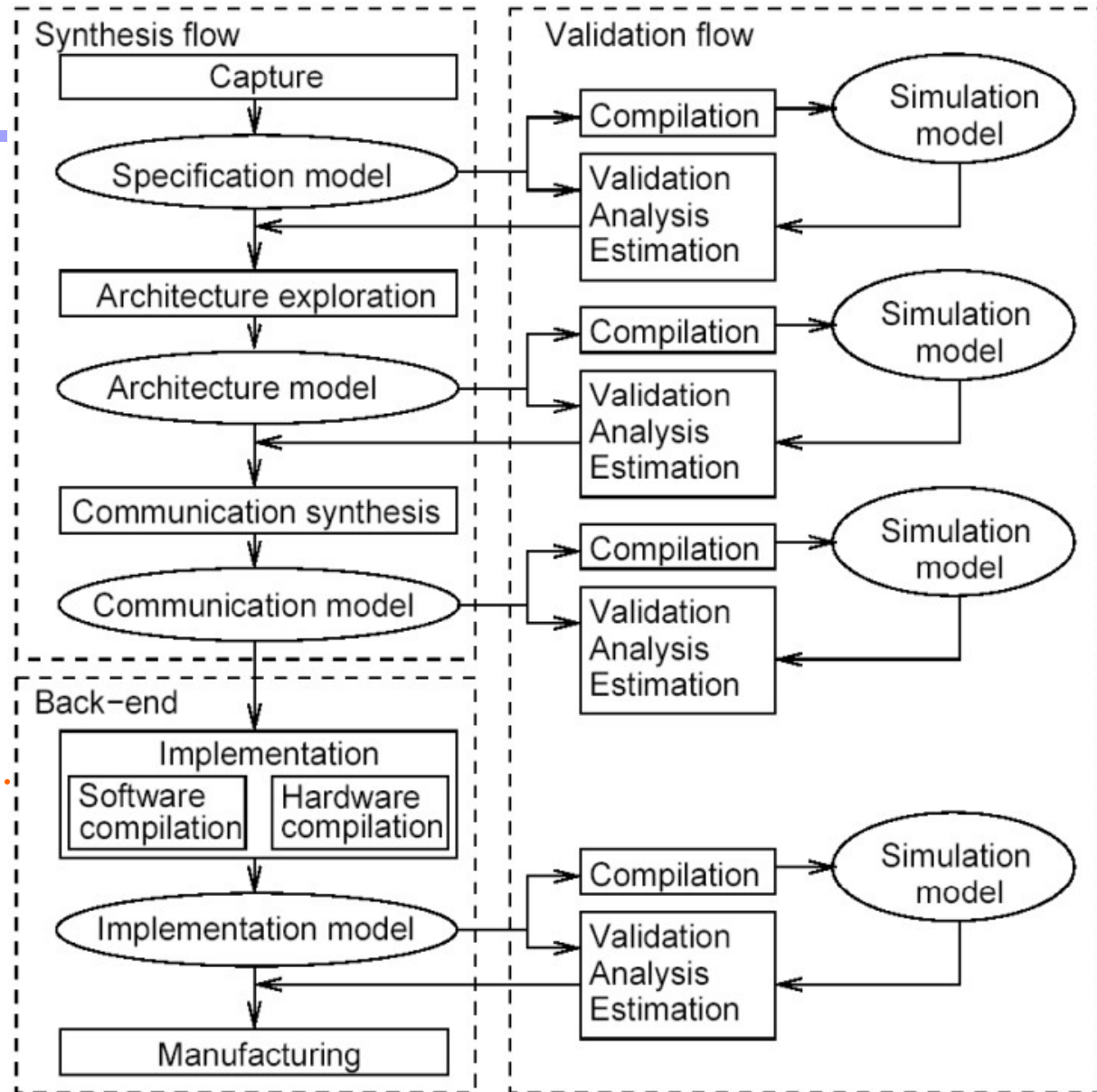
SpecC

abstract busses
replaced by
actual wires in
a series of
refinements



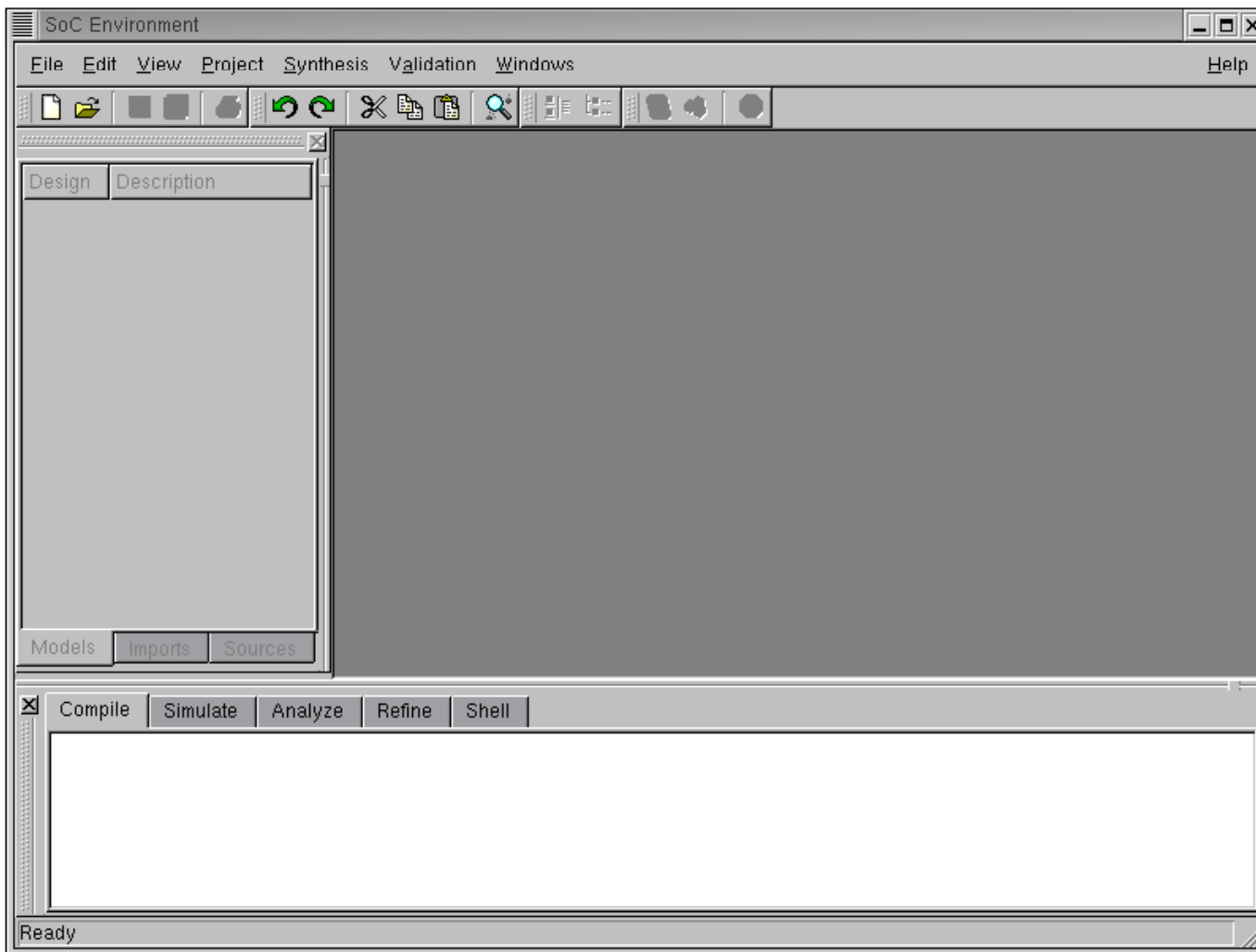
SpecC

Compilers are used to generate binary machine code and hardware synthesis tools are used to generate custom hardware.



An actual example

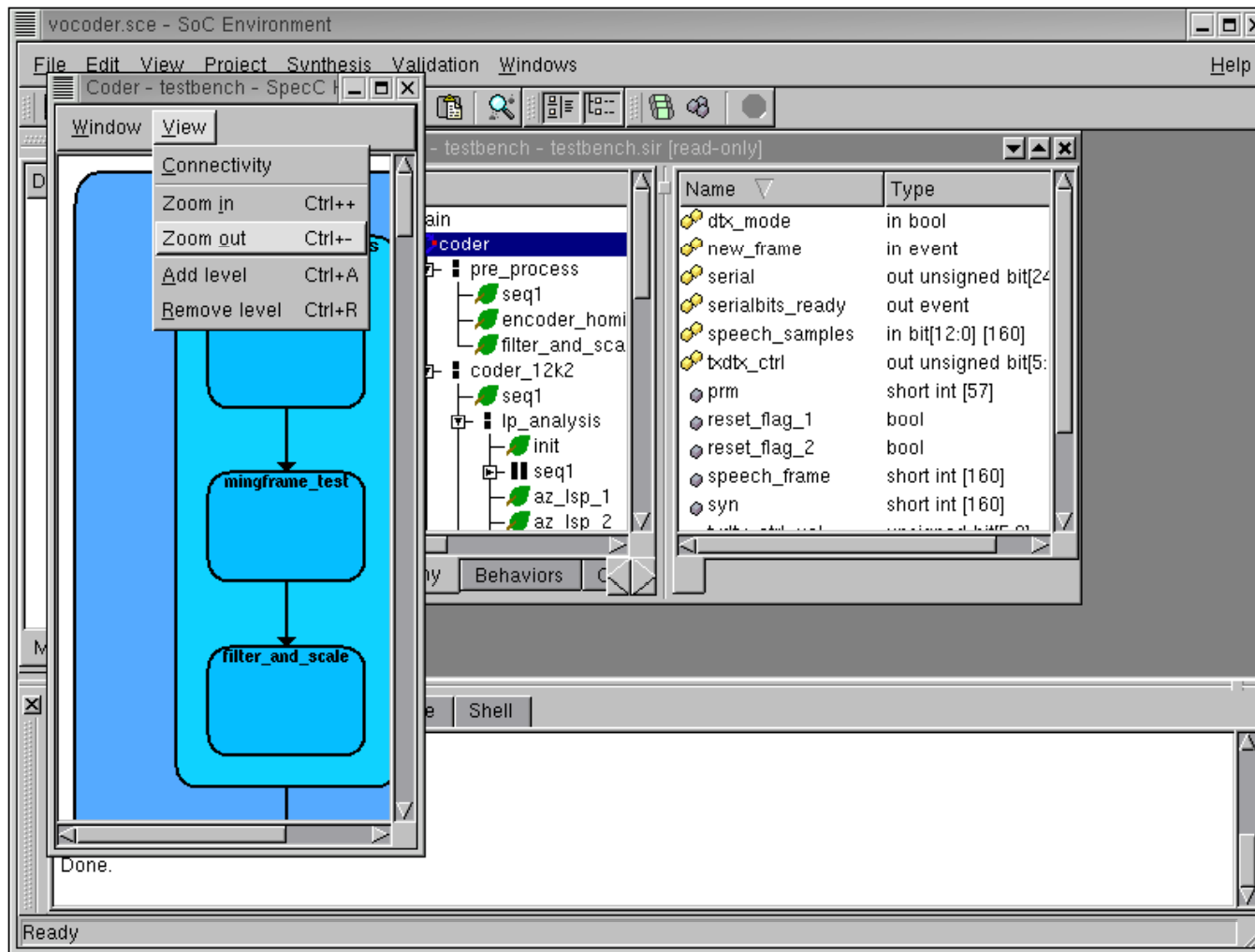
- Getting started with the SCE window -



Copyright © 2002 by Center of Embedded Computing Systems (CECS), UC Irvine

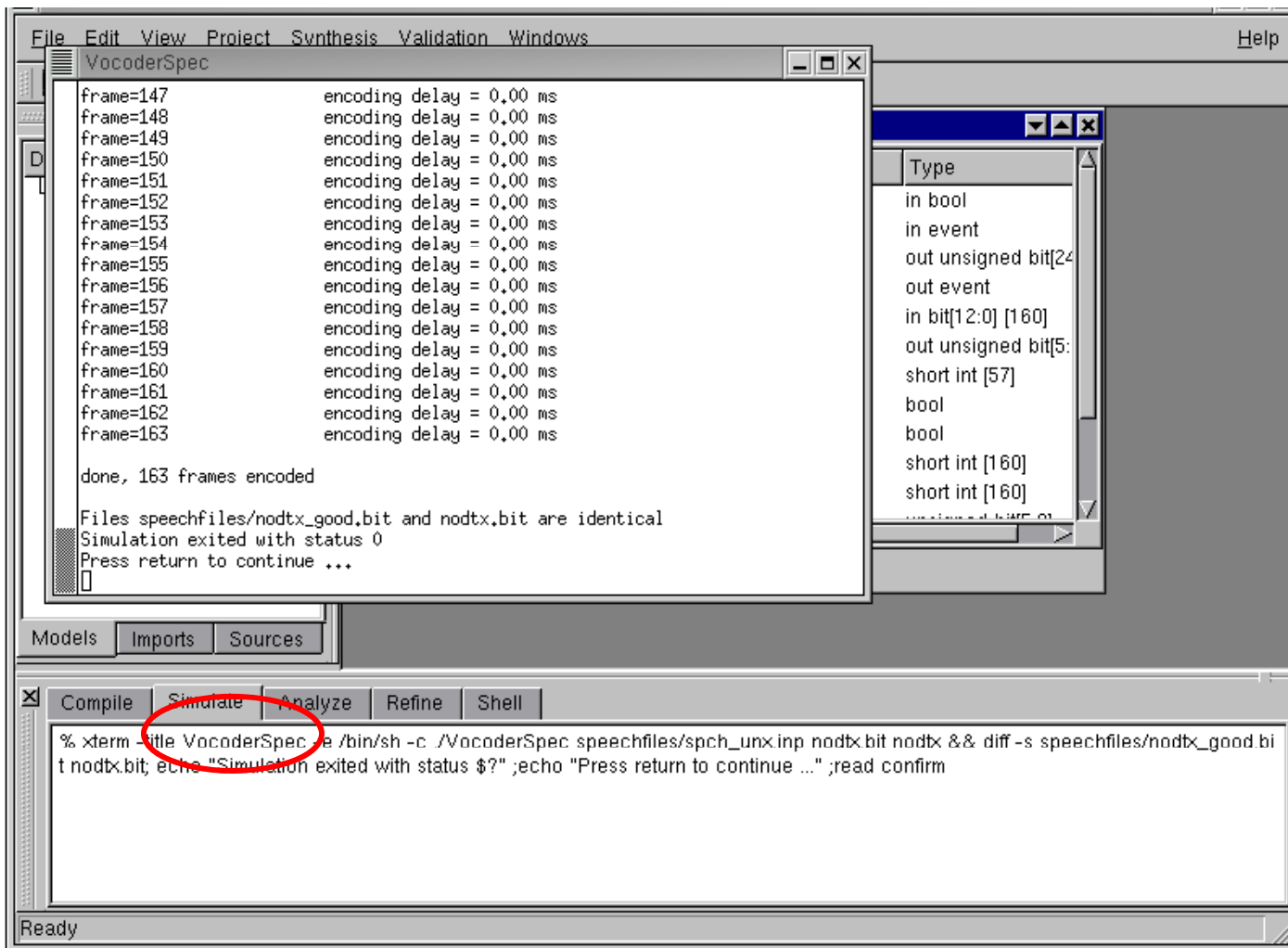


Browsing the specification

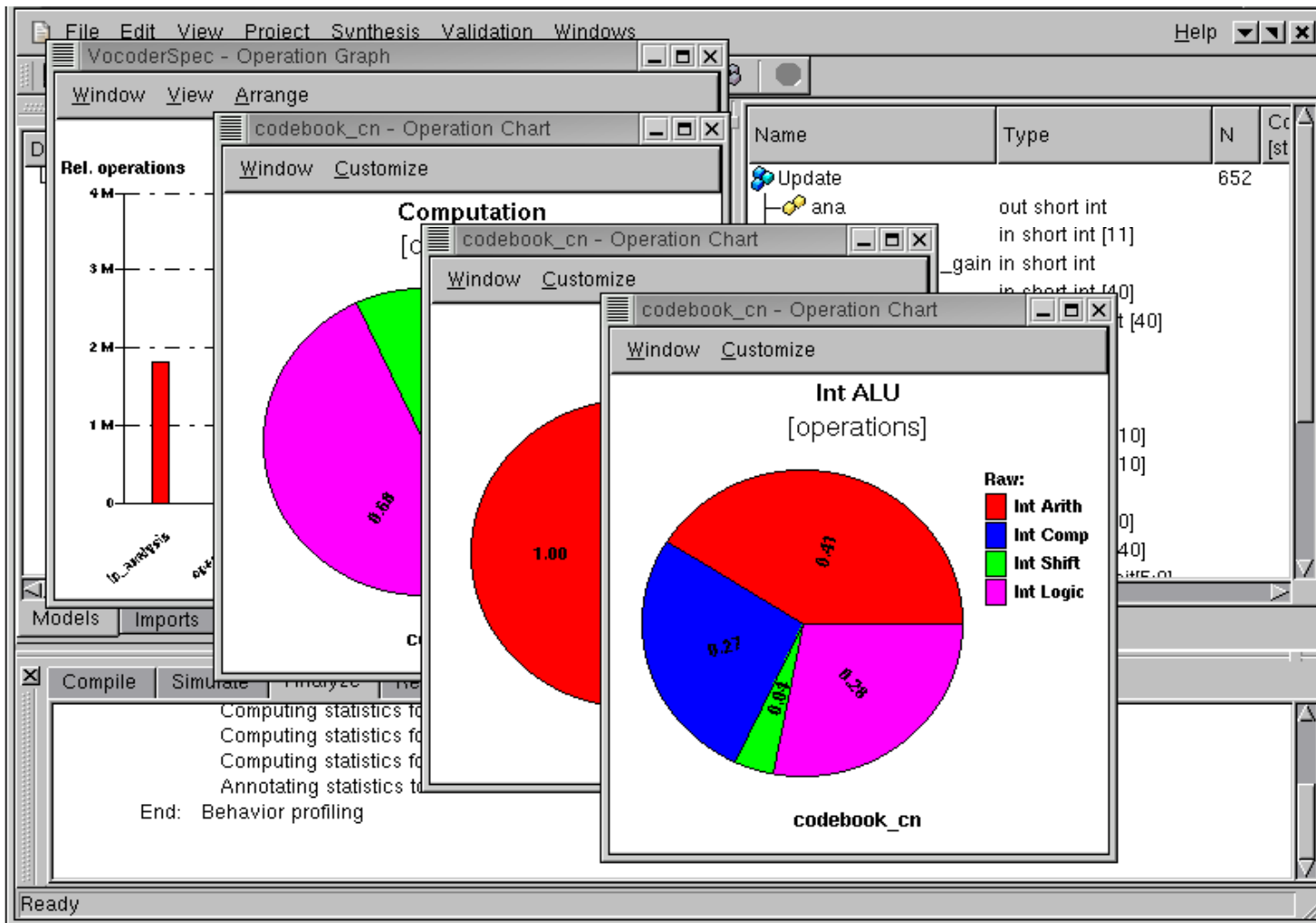


An actual example

- Validation by simulation -



Analyze profiling results



PE selection

vocoder.sce - SoC Environment - [Update - VocoderSpec - VocoderSpec.sir*]

PE Selection

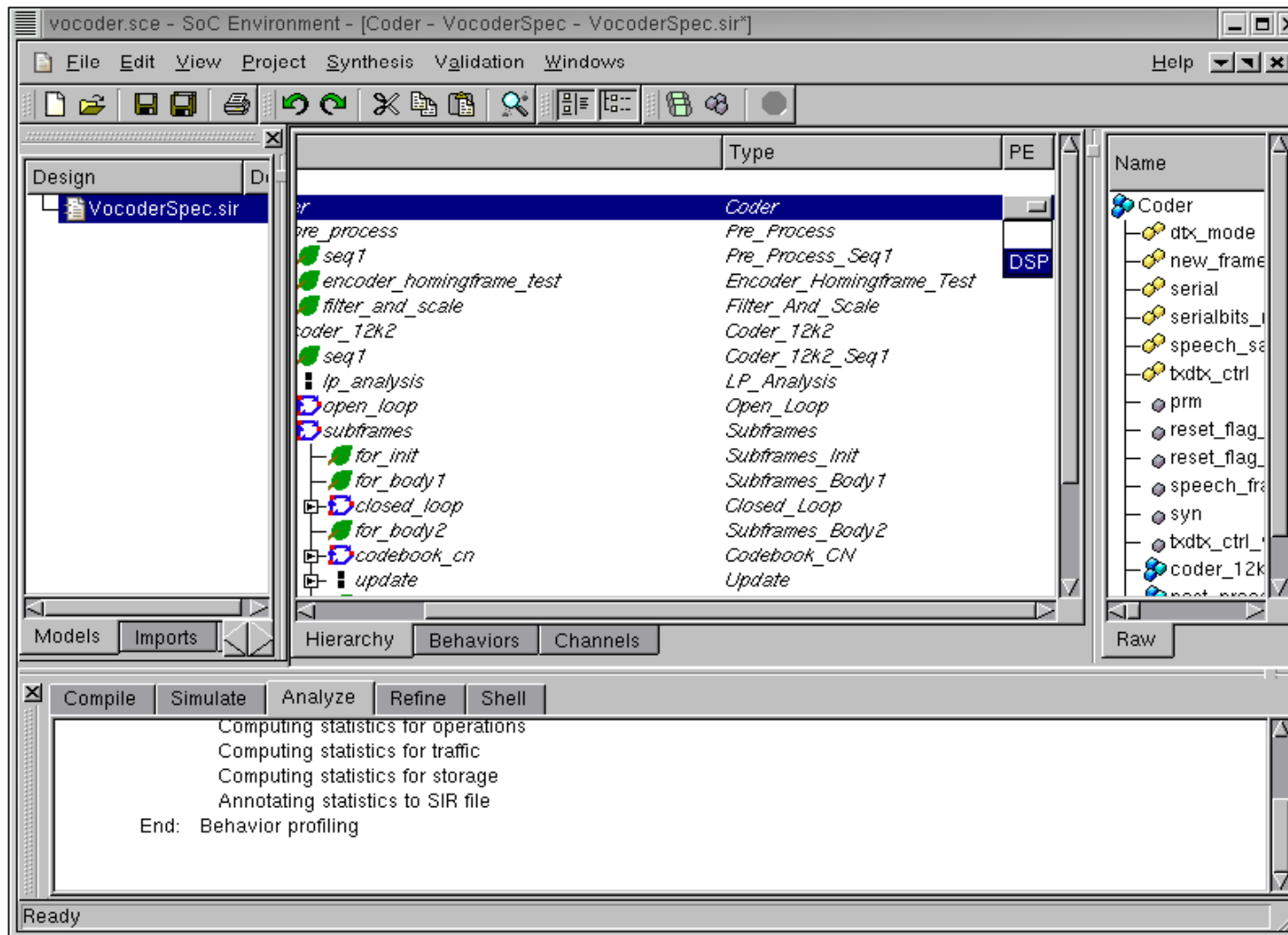
| Categories: | Component | Max. clock [MHz] | MIPS | Cost | Program [kB] | Data [kB] | Instruction [bits] | D: |
|-----------------|-------------------|------------------|-------|-------|--------------|-----------|--------------------|----|
| DSP | AMD_K6 | 400.0 | 200.0 | 100.0 | 64.0 | 64.0 | 32 | 32 |
| Processor | AMD_K7 | 700.0 | 350.0 | 120.0 | 64.0 | 64.0 | 32 | 32 |
| Mem | ARM1020 | 325.0 | 150.0 | 23.0 | 64.0 | 64.0 | 32 | 32 |
| Custom Hardware | ARM720 | 100.0 | 50.0 | 23.0 | 64.0 | 64.0 | 32 | 32 |
| Controller | ARM920 | 250.0 | 125.0 | 23.0 | 64.0 | 64.0 | 32 | 32 |
| | IDT_32300 | 100.0 | 50.0 | 5.0 | 64.0 | 64.0 | 32 | 32 |
| | Intel_P1 | 200.0 | 100.0 | 4.5 | 64.0 | 64.0 | 32 | 32 |
| | Intel_P2 | 550.0 | 200.0 | 23.0 | 64.0 | 64.0 | 32 | 32 |
| | Intel_P3 | 900.0 | 450.0 | 90.0 | 64.0 | 64.0 | 32 | 32 |
| | MIPS32 | 100.0 | 50.0 | 10.0 | 64.0 | 64.0 | 32 | 32 |
| | MIPS64 | 350.0 | 200.0 | 20.0 | 64.0 | 64.0 | 64 | 64 |
| | Motorola_68000 | 20.0 | 20.0 | 3.5 | 64.0 | 64.0 | 32 | 32 |
| | Motorola_68010 | 120.0 | 100.0 | 23.0 | 64.0 | 64.0 | 32 | 32 |
| | Motorola_Coldfire | 120.0 | 100.0 | 23.0 | 64.0 | 128.0 | 32 | 32 |
| | UltraSparcII | 480.0 | 250.0 | 100.0 | 64.0 | 64.0 | 64 | 64 |

Help OK Cancel

Ready



The top level behavior is mapped to the DSP



Estimate the performance (too slow)

The screenshot displays a software interface for a SoC Environment. The main window shows a table of components and their performance metrics. The 'Coder' component is highlighted, and its 'Computation' time is circled in red.

| Name | Type | N | Code | Computation | Data | Head |
|------------------|-------------------------|-----|------|--------------|---------|------|
| Coder | | 1 | | 4003994.7 us | 17396 B | |
| dtx_mode | in bool | | | | | |
| new_frame | in event | | | | | |
| serial | out unsigned bit[243:0] | | | | | |
| serialbits_ready | out event | | | | | |
| speech_samples | in bit[12:0] [160] | | | | | |
| txdtx_ctrl | out unsigned bit[5:0] | | | | | |
| prm | short int [57] | | | | 114 B | |
| reset_flag_1 | bool | | | | 2 B | |
| reset_flag_2 | bool | | | | 2 B | |
| speech_frame | short int [160] | | | | 320 B | |
| syn | short int [160] | | | | 320 B | |
| txdtx_ctrl_val | unsigned bit[5:0] | | | | 0 B | |
| coder_12k2 | Coder_12k2 | 163 | | 24280.4 us | 16268 B | |
| post_process | Post_Process | 163 | | 0.0 us | 322 B | |

Below the table, a log window shows the following text:

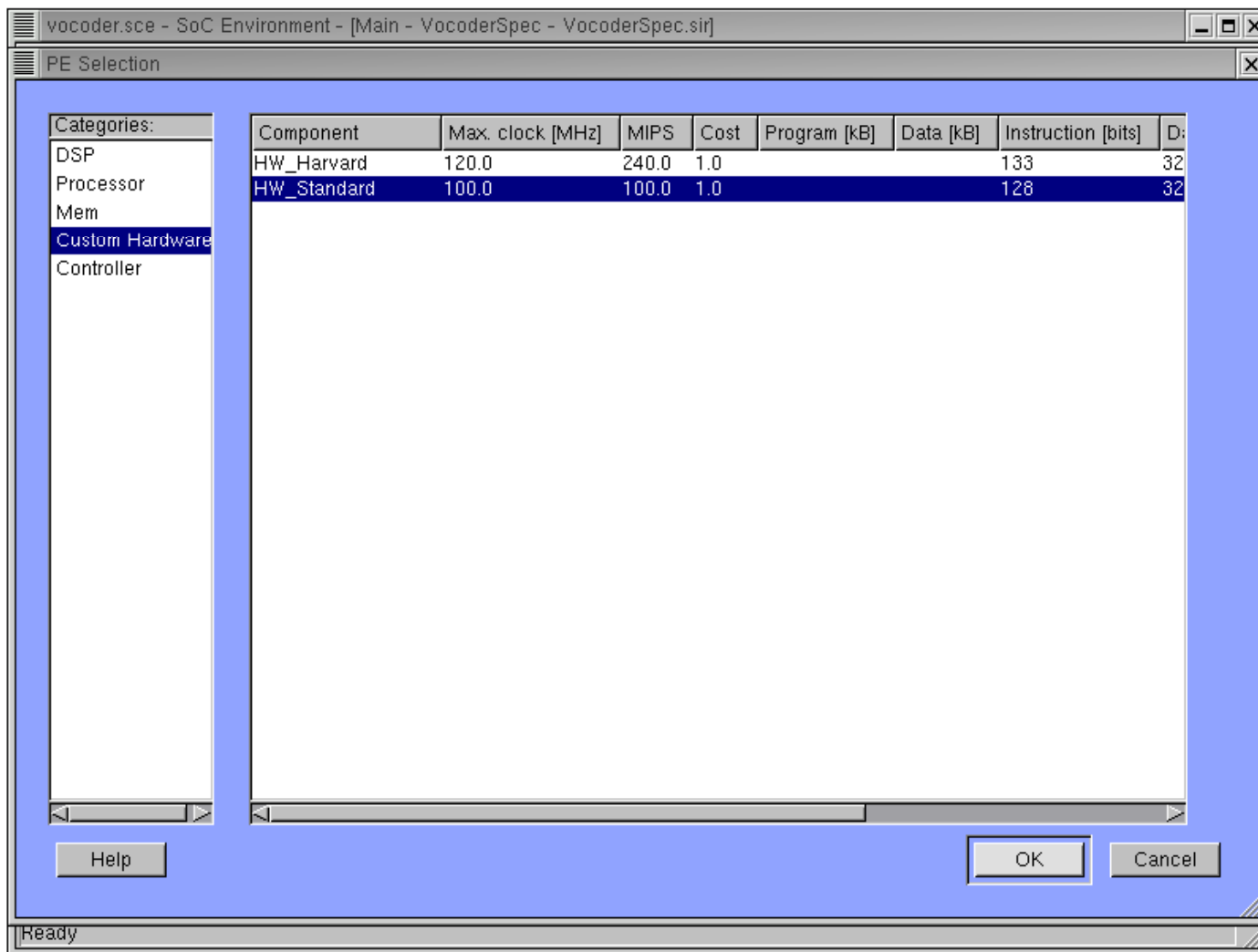
```

Start: retargetable profiling
Generating internal data structure for profiling
Deriving raw statistics from SIR file
Computing weighted statistics
Annotating weighted statistics to SIR file
End: retargetable profiling
    
```

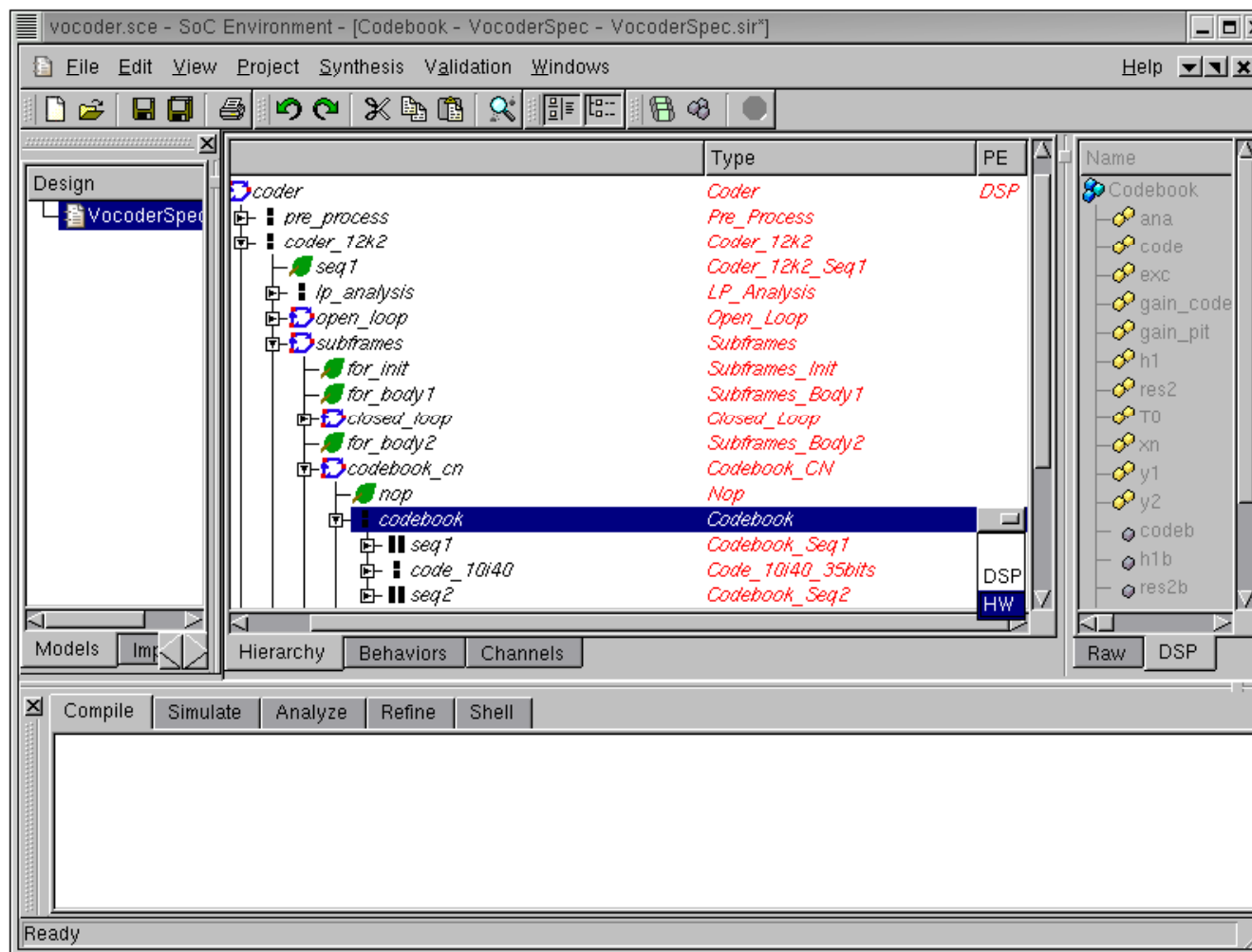
The status bar at the bottom of the window indicates 'Ready'.



Selecting additional custom HW including datapath and controller



Binding “codebook” to the custom datapath



Execution time of the DSP

The screenshot displays the 'vocoder.sce - SoC Environment' window. The main area shows a table of components and their execution statistics. The 'Behavior_DSP' component is highlighted with a red circle, indicating its computation time.

| Name | Type | N | Code | Computation |
|---------------------------|-------------------|---|------|--------------|
| Behavior_DSP | | 1 | | 2669435.0 us |
| AR_CC_cb_ana_HW_to_DSP | _IR_short_int_10_ | | | |
| AR_CC_code_HW_to_DSP | _IR_short_int_40_ | | | |
| AR_CC_exc_i_DSP_to_HW | _IS_short_int_40_ | | | |
| AR_CC_gain_code_HW_to_DSP | _IR_short_int | | | |
| AR_CC_gain_pit_DSP_to_HW | _IS_short_int | | | |
| AR_CC_h1_DSP_to_HW | _IS_short_int_40_ | | | |
| AR_CC_res2_DSP_to_HW | _IS_short_int_40_ | | | |
| AR_CC_T0_DSP_to_HW | _IS_short_int | | | |
| AR_CC_xn_DSP_to_HW | _IS_short_int_40_ | | | |
| AR_CC_y1_DSP_to_HW | _IS_short_int_40_ | | | |
| AR_CC_y2_DSP_to_HW | _IS_short_int_40_ | | | |
| AR_CC_y2_HW_to_DSP | _IR_short_int_40_ | | | |
| dtx_mode | in bool | | | |
| new_frame | in event | | | |

The console window at the bottom shows the following output:

```

Compile Simulate Analyze Refine Shell
Computing statistics for traffic
Computing Instance to Instance traffic
Computing statistics for storage
Annotating statistics to SIR file
End: Profiling and retargetable profiling
Ready
    
```



Execution time for hardware

| Name | Type | N | Code | Computation | Data |
|--------------------------|-------------------|---|------|-------------|-------|
| book_CN_HW | | 1 | | 543703.4 us | 10872 |
| R_CC_cb_ana_HW_to_DSP | _IS_short_int_10_ | | | | |
| R_CC_code_HW_to_DSP | _IS_short_int_40_ | | | | |
| R_CC_exc_i_DSP_to_HW | _IR_short_int_40_ | | | | |
| R_CC_gain_code_HW_to_DSP | _IS_short_int | | | | |
| R_CC_gain_pit_DSP_to_HW | _IR_short_int | | | | |
| R_CC_h1_DSP_to_HW | _IR_short_int_40_ | | | | |
| R_CC_res2_DSP_to_HW | _IR_short_int_40_ | | | | |
| R_CC_T0_DSP_to_HW | _IR_short_int | | | | |
| R_CC_xn_DSP_to_HW | _IR_short_int_40_ | | | | |
| R_CC_y1_DSP_to_HW | _IR_short_int_40_ | | | | |
| R_CC_y2_DSP_to_HW | _IR_short_int_40_ | | | | |
| R_CC_y2_HW_to_DSP | _IS_short_int_40_ | | | | |
| o_ana | short int [10] | | | | 4C |
| ode | short int [40] | | | | 16C |

Compile Simulate Analyze Refine Shell

Computing statistics for traffic
 Computing Instance to Instance traffic
 Computing statistics for storage
 Annotating statistics to SIR file
 End: Profiling and retargetable profiling

Ready

Assuming that 0.54+2.66 sec is acceptable



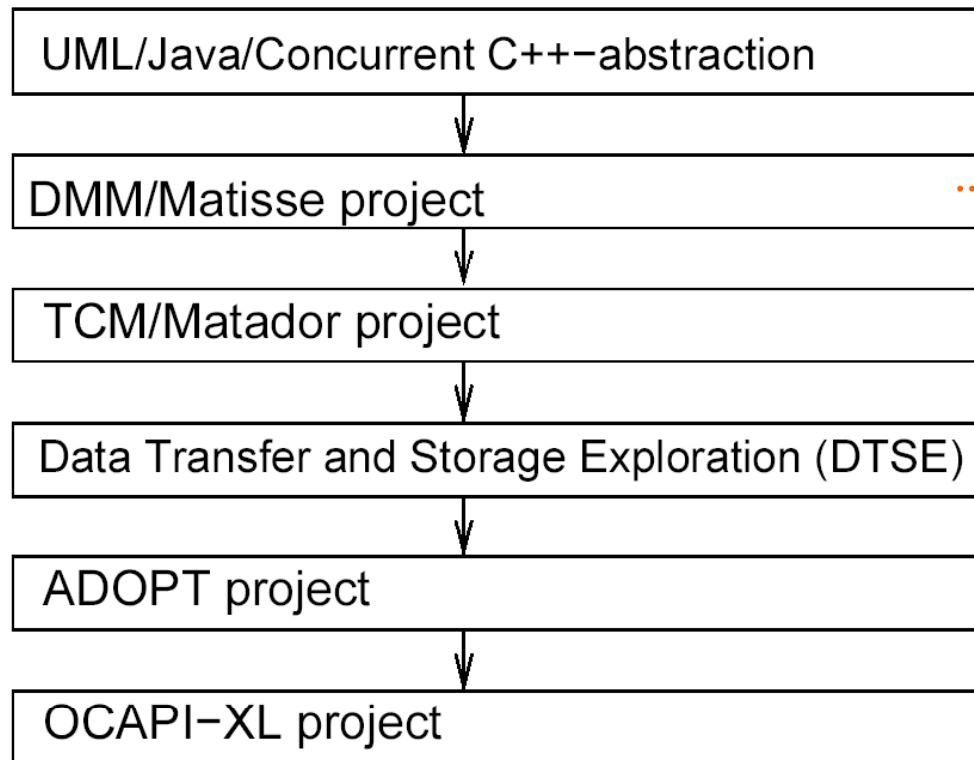
2. IMEC tool flow

IMEC = Interuniversitair Micro-Electronica Centrum,
Leuven, Belgium
(Large research facility)



Imec tool flow

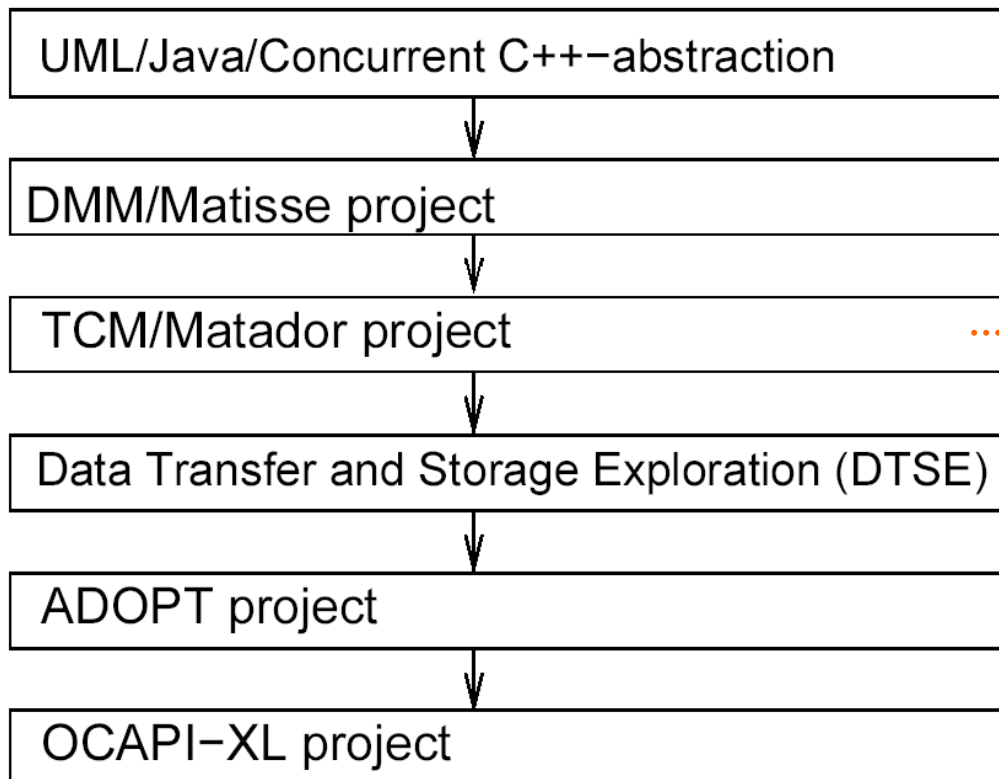
- Global view -



Considers the system at the concurrent process level as a set of concurrent and dynamic processes, whose specification consists of algorithms, abstract data types, communication primitives, and real-time requirements. Tools can perform source code transformations on the dynamic data types & provide also a memory pool organization in the virtual memory space.



Imec tool flow - Matador/TCM -



Again considering a system of concurrent processes. For these tools, the emphasis is on mapping tasks to processors. Different configurations of multi-processor systems are evaluated and curves of designs that are non-inferior to others are generated. These curves provide a view of the design space, and are the basis for final design decisions.



Matador/Task Concurrency Management (TCM)

Wong et al. [Wong et al., 2001]: configurations for a personal MPEG-4 player: combination of StrongArm processors and custom accelerators.

Found 4 configurations satisfying timing constraint of 30 ms.

| <i>Processor combination</i> | 1 | 2 | 3 | 4 |
|---------------------------------|---|---|---|----|
| Number of high speed processors | 6 | 5 | 4 | 3 |
| Number of low speed processors | 0 | 3 | 5 | 7 |
| Total number of processors | 6 | 8 | 9 | 10 |

For combinations 1 and 4, only one allocation of tasks to processors meets the timing constraints. For combinations 2 and 3, different time budgets lead to different task to processor mappings and different energy consumptions.



Pareto points

Multiobjective optimization:

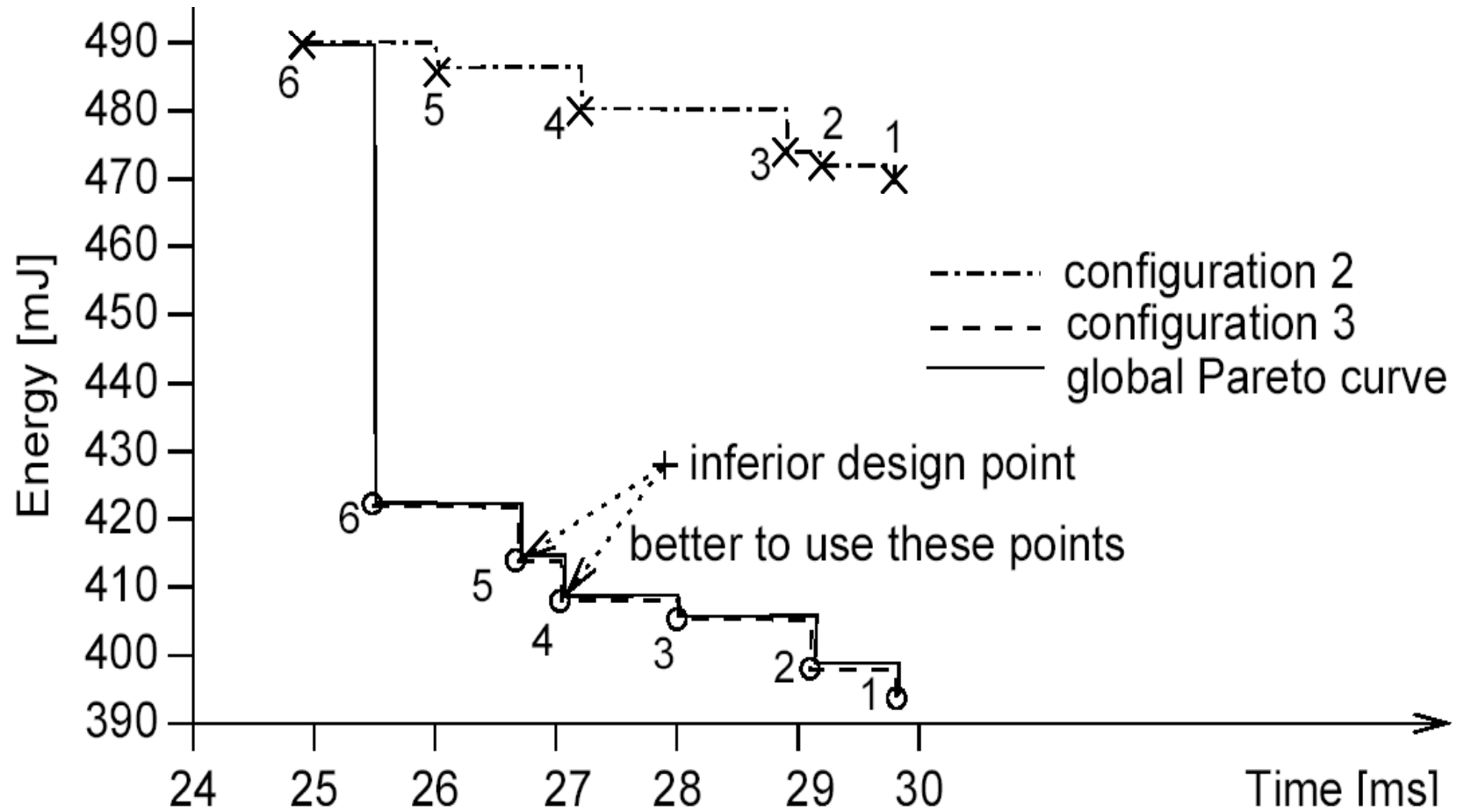
- several conflicting criteria
 - eg. performance vs. cost vs. power consumption

Definition: A (design) point J_i is **dominated** by point J_k if J_k is equal or better than J_i in each criterion ($J_i \leq J_k$).

Definition: A (design) point is **Pareto-optimal** or a **Pareto point**, if it is not dominated by any other point.



Pareto curves



Data Transfer and Storage Exploration (DTSE)

Reduction of the data transfers between processing components and at a reduction of the storage requirements.

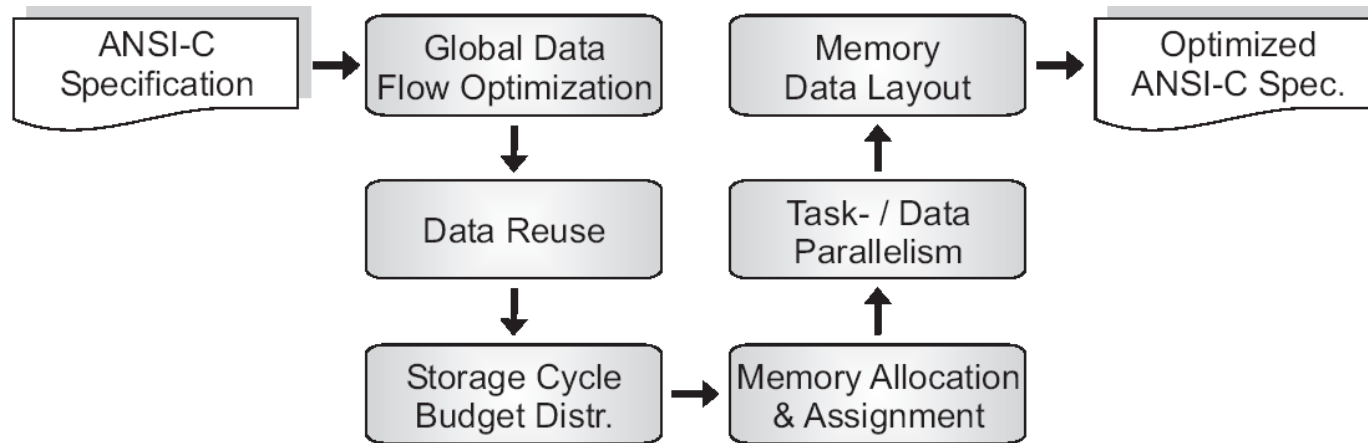
DTSE has proven to be highly effective in minimizing the energy cost related to data memory hierarchies.

For several typical embedded applications, reductions of main memory accesses, cache accesses and energy consumption between 50% and 80% have been reported.

© Falk, 2004



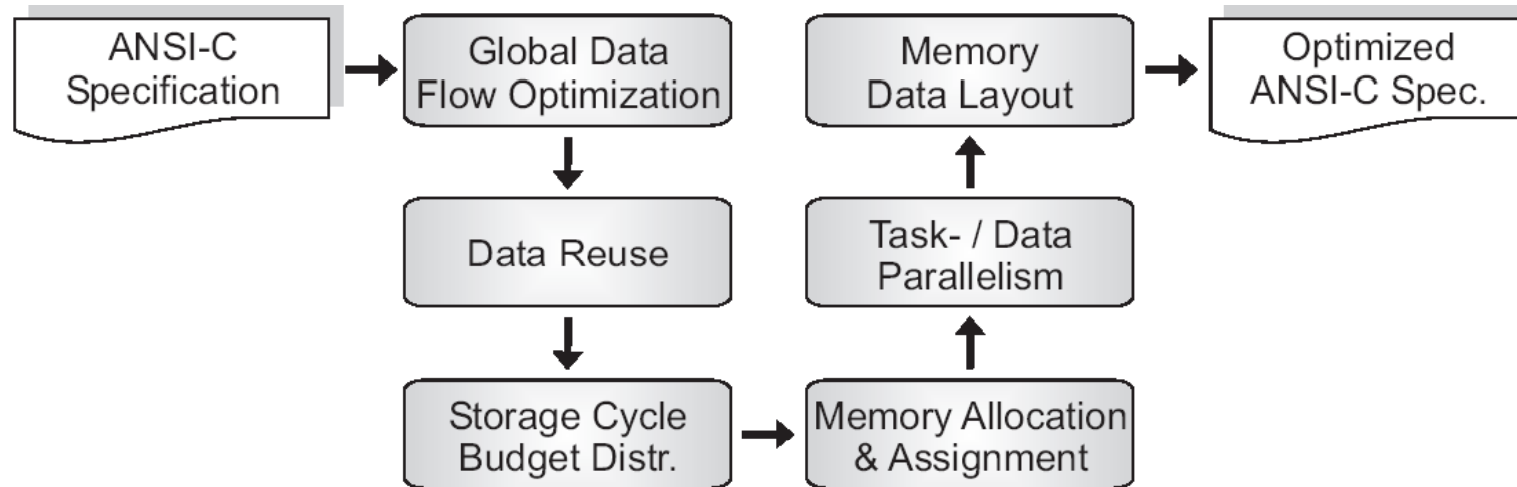
Data Transfer and Exploration (DTSE)



- A memory oriented data flow analysis is performed,
- followed by global data flow and loop transformations to reduce the amount of background memories;
- data reuse transformations exploit a distributed memories;
- storage cycle budget distribution determines the bandwidth requirements and the balancing of the available cycle budget over the different memory accesses.



Data Transfer and Exploration (DTSE)



- The memory hierarchy layer assignment produces a netlist of memory units & an assignment of variables to memories.
- For multi-processor systems, task- / data-parallelism exploitation minimizes communication & storage overhead caused by the parallel execution of subsystems;
- Data layout transformations map variables with non-overlapping lifetimes in the same physical memory location.



Address optimization (ADOPT)

Addressing is simplified in address optimization (ADOPT) tools. DTSE steps increase the addressing complexity of transformed applications, by introducing more complex index expressions, conditions ..

This overhead is neglected by the DTSE methodology.

The optimized memory system, will be power efficient but slow.

Parts of the additional complexity can be removed by source code optimizations for *address optimization* (ADOPT):

- algebraic cost minimization first minimizes operation instances. The goal is to find factorizations of addressing expressions in order to be able to reuse computations as much as possible. The reuse of expressions is based on common subexpression elimination techniques.



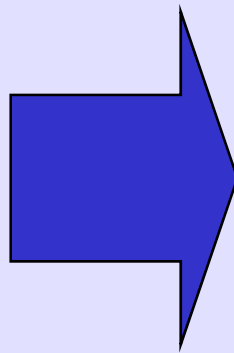
Address optimization (ADOPT)

High penalty by DTSE addressing using modulo (%) & /.
Optimization of this addressing is a major goal in ADOPT.

/ and % can be replaced by variables as follows:

```
for (j=0; j<n; j++)
```

```
  a[j/6][j%6] = ...;
```



```
  int jdiv6=0, jmod6=0;
```

```
  for (j=0; j<n; j++, jmod6++) {
```

```
    if (jmod6 >= 6) {
```

```
      jmod6 -= 6; jdiv6++; }
```

```
      a[jdiv6][jmod6] = ...; }
```

☞ addressing only consists of ++ and -- operators.

In many examples, ADOPT reduces runtimes by about 3.



Backend

- Compilers
- Mapping to configurable hardware using OCAPI-XL

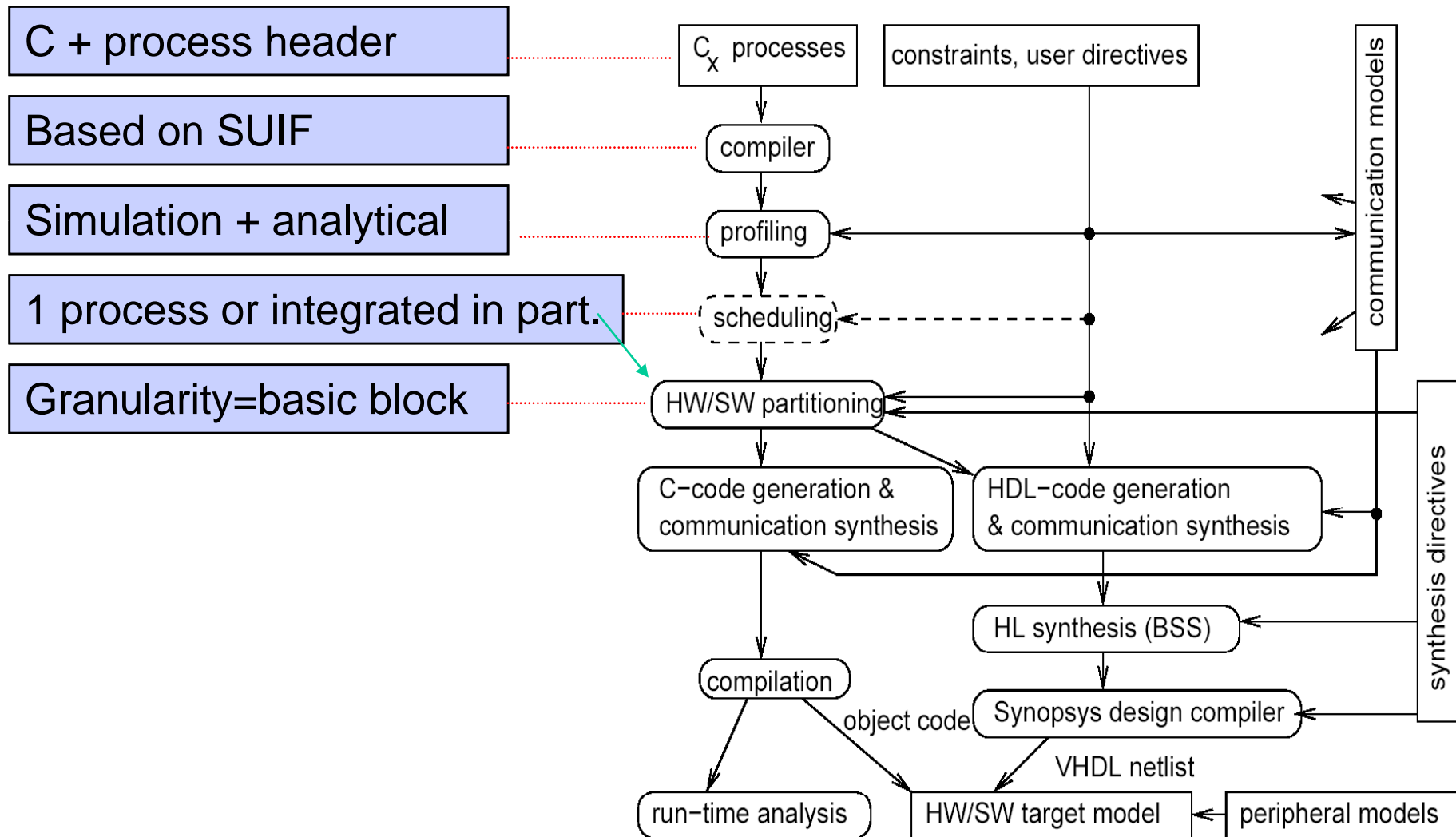


Cosynthesis for embedded micro-architectures (COSYMA)

Ernst et al., Univ. Braunschweig



Cosynthesis for embedded micro-architectures (COSYMA)



Ptolemy II

Ptolemy II supports specifications using different models of computation. In particular, it supports:

1. Communicating sequential processes (CSP).
2. Continuous time (CT): appropriate for ME, analog circuits. Supported by extensible differential equation solvers.
3. Discrete event model (DE): model used by many (e.g. VHDL) simulators.
4. Distributed discrete events (DDE).
5. Finite state machines (FSM).
6. Process networks (PN), using Kahn process networks
7. Synchronous dataflow (SDF)
8. Synchronous/reactive (SR) MoC. Discrete time, signals do not need to have a value at every clock tick. Esterel used.



Ptolemy II



Source ...

<http://ptolemy.eecs.berkeley.edu/ptolemyII/ptII3.0/ptII3.0.2/doc/index.htm>



Ptolemy II Documentation - Mozilla

http://ptolemy.eecs.berkeley.edu/ptolemyII/ptII3.0/ptII3.0.2/doc/index.htm

Ptolemy II heterogenous modeling and design

Go Pause Resume Stop

Model parameters:

stickiness:

Sticky Masses

This applet models two (sticky) masses on a flat frictionless table, as shown in the figure below. Each of the masses is attached to a spring with distinct spring constant. If the balls are not stuck together, they will independently swing back and forth. If they collide, they stick together, and swing together depending on their momentum when they collided. The stickiness is assumed to exponentially decay with rate

Ptolemy Project - UC Berkeley - EECS Comments to: ptII@eecs.berkeley.edu, Copyright © 1995-2003

execution finished.



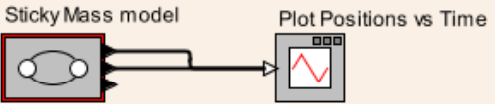
Ptolemy II Documentation - Mozilla

http://ptolemy.eecs.berkeley.edu/ptolemyII/ptII3.0/ptII3.0.2/doc/index.htm

Ptolemy II

heterogenous modeling and design

each state refines to a continuous-time model of the dynamics in that mode.



This top-level model contains only two components. The source, labeled "Hybrid System", outputs the positions of the two point masses. The sink plots the positions vs. time. The source is itself a modal model, implemented using the finite state machine (FSM) shown below:

V1

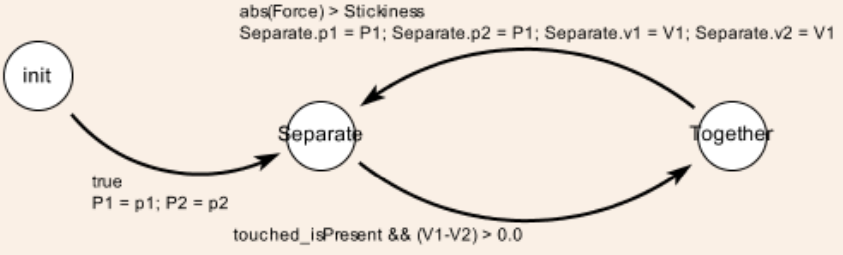
V2

P2

P1

touched

The sticky masses system has two modes of operation, "Separate" and "Together," corresponding to whether the point masses are stuck together. The "init" state has a transition that is used to initialize the "Separate" model (double click on that transition to see its actions).



The refinement for the Separate state models the situation where the point masses are separate, and is shown below:

Ptolemy Project - UC Berkeley - EECS [Comments to: ptll@eecs.berkeley.edu](mailto:ptll@eecs.berkeley.edu), Copyright © 1995-2003

Applet Multimode started



Ptolemy II Documentation - Mozilla

http://ptolemy.eecs.berkeley.edu/ptolemyII/ptII3.0/ptII3.0.2/doc/index.htm

Ptolemy II

heterogenous modeling and design

CT

The refinement for the Separate state models the situation where the point masses are separate, and is shown below:

Refinement Solver

This model gives two separate ordinary differential equations, one for each point mass attached to a spring. The ZeroCrossingDetector actor detects the collision of the point masses and emits the "touched" event.

V1 and V2 are velocities, and P1 and P2 are positions of the two masses.

The refinement for the Together state models the situation where the point masses are together, and is shown below:

Ptolemy Project - UC Berkeley - EECS Comments to: ptll@eecs.berkeley.edu, Copyright © 1995-2003

Applet Multimode started



Octopus

Addressing the poor match between the focus of **object-oriented design techniques** on the software object structure and the need to allocate operations to tasks.

This poor match was the main concern that was addressed in the design of OCTOPUS.

Totally software-oriented flow used at Nokia



Octopus

1. In the **systems requirement phase**, behavior is described by use case diagrams and use cases. The structure of the environment is described by a so-called context diagram.
- 2 In the **system architecture phase**, the structure is broken down into subsystems. Major interfaces between the subsystems are identified, but their behavior is not.
- 3 The **subsystem analysis phase** is done \forall subsystems. Class diagrams for the subsystems are generated. Behaviors of subsystems can be defined in various ways, including StateCharts, so-called event lists and event sheets.
- 4 The **subsystem design phase** generates outlines for processes/threads, classes and interprocess messages.
- 5 The **subsystem implementation phase** generates actual code of the selected programming language.

