[KGV83] S. Kirkpatrick, C.D. Gelatt, and M. P. Vecchi. \Optimization by simulated annealing,". *Science*, 220(4598):671{ 680, 1983.

[KL70] B.W. Kernighan and S. Lin. \An efficient heuristic procedure for partitioning graphs,". *Bell System Technical Journal*, February 1970.

[LT91] E.D. Lagnese and D.E. Thomas. \Architectural partitioning for system level synthesis of integrated circuits,". *IEEE Transactions on Computer-Aided Design*, July 1991.

[MK90] M.C. McFarland and T.J. Kowalski. \Incorporating bottom-up design into hardware synthesis,". *IEEE Transactions on Computer-Aided Design*, September 1990.

[NG92] S. Narayan and D.D. Gajski. \System clock estimation based on clock slack minimization,". In *Proceedings of the European Design Automation Conference (EuroDAC)*, 1992.

[NVG92] S. Narayan, F. Vahid, and D.D. Gajski. \System specification with the speccharts language,". In *IEEE Design & Test of Computers*, Dec. 1992.

[PK89] P.G. Paulin and J.P. Knight. \Algorithms for high-level synthesis,". In *IEEE Design & Test of Computers*, Dec. 1989.

[PPM86] A.C. Parker, T. Pizzaro, and M. Mlinar. \Maha : A program for datapath synthesis,". In *Proceedings of the Design Automation Conference*, 1986.

[TM91] D.E. Thomas and P. Moorby. *The Verilog Hardware Description Language*. Kluwer Academic Publishers, 1991.

[VG92] F. Vahid and D.D. Gajski. \Specification partitioning for system design,". In *Proceedings of the Design Automation Conference*, 1992.

[VGG93] F. Vahid, J. Gong, and D.D. Gajski. \A hardware-software partitioning algorithm for minimizing hardware,". UC Irvine, Dept. of ICS, Technical Report 93-38,1993.

References

[BHS91] F. Belina, D. Hogrefe, and A. Sarma. *SDL with Applications from Protocol Specifications*. Prentice Hall, 1991.

[BK87] G. Borriello and R.H. Katz. \Synthesis and optimization of interface transducer logic,". In *Proceedings of the International Conference on Computer-Aided Design*, 1987.

[EHB94] R. Ernst, J. Henkel, and T. Benner. \Hardware-software cosynthesis for microcontrollers,". In *IEEE Design & Test of Computers*, pages 64{ 75, December 1994.

[FM82] C.M. Fiduccia and R.M. Mattheyses. \A linear-time heuristic for improving network partitions,". In *Proceedings of the Design Automation Conference*, 1982.

[GD90] R. Gupta and G. DeMicheli. \Partitioning of functional models of synchronous digital systems,". In *Proceedings of the International Conference on Computer-Aided Design*, pages 216{ 219, 1990.

[GD92] R.K. Gupta and G. DeMicheli. \System-level synthesis using re-programmable components,". In *Proceedings of the European Conference on Design Automation (EDAC)*, pages 2{ 7, 1992.

[GD93] R. Gupta and G. DeMicheli. \Hardware-software cosynthesis for digital systems,". In *IEEE Design & Test of Computers*, pages 29{ 41, October 1993.

[GVN94] D.D. Gajski, F. Vahid, and S. Narayan. \A system-design methodology: Executable-specification refinement,". In *Proceedings of the European Conference on Design Automation (EDAC)*, 1994.

[Hal93] Nicolas Halbwachs. *Synchronous Programming of Reactive Systems*. Kluwer Academic Publishers, 1993.

[Hoa78] C.A.R. Hoare. \Communicating sequential processes,". *Communications of the ACM*, 21(8):666{ 677, 1978.

[IEE88] IEEE Inc., N.Y. *IEEE Standard VHDL Language Reference Manual*, 1988.

[JMP88] R. Jain, M. Mlinar, and A. Parker. \Area-time model for synthesis of non-pipelined designs,". In *Proceedings of the International Conference on Computer-Aided Design*, 1988.

[Joh67] S.C. Johnson. \Hierarchical clustering schemes,". *Psychometrika*, pages 241{ 254, September 1967.

[KC91] Y.C. Kirkpatrick and C.K. Cheng. \Ratio cut partitioning for hierarchical designs,". *IEEE Transactions on Computer-Aided Design*, 10(7):911{ 921, 1991.

SYSTEM-LEVEL METHODOLOGY and TECHNOLOGY

by

Daniel D. Gajski
Frank Vahid
Sanjiv Narayan

University of California at Irvine
Department of Computer Science
Irvine, CA 92715-3425

Telephone: 1-714-856-4155, 1-714-856-8059
Fax: 1-714-856-3429, 1-714-856-4056
Email: gajski@uci.edu, vahid@ics.uci.edu

Outline

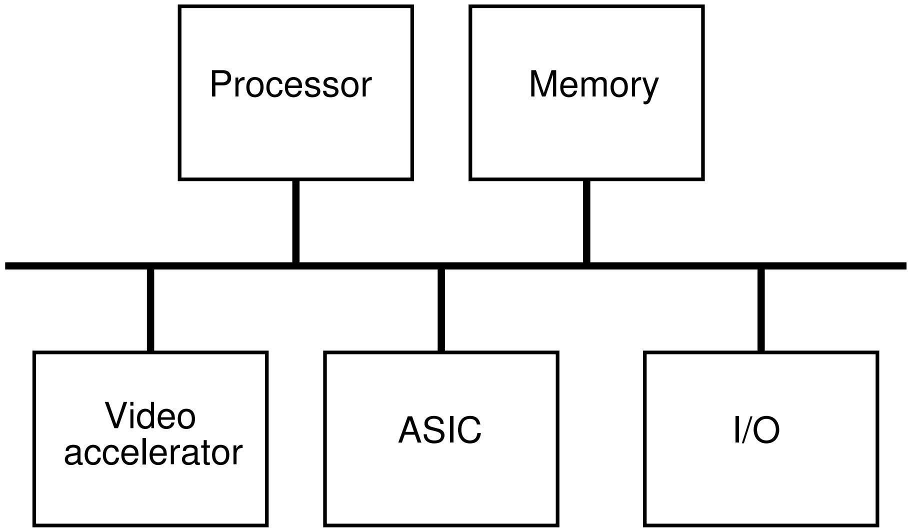
- Motivation
- Design models and architectures
- System-design languages
- An example
- Partitioning
- Estimation
- Refinement
- Methodology and environments
- Bibliography

Motivation

Executable specification

```
if  (x = 0) then
    y = a * b / 2
```

System implementation



Models

Languages

Partitioning

Estimation

Refinement

Architecture

UC Irvine

Design models and architectures

- Model: a conceptual view of a system

- Architecture: a topology of system components

Conceptual views of an elevator controller

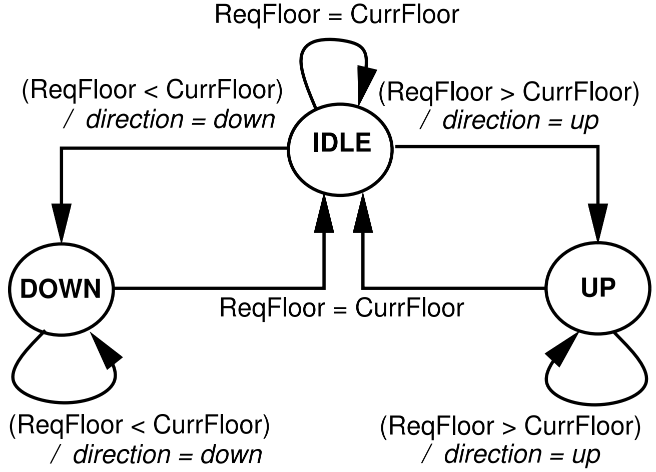
"if elevator is stationary and floor
requested is less than current floor,
then descend elevator to floor requested

if elevator is stationary and floor
requested is greater than current floor,
then ascend elevator to floor requested "

(a) English description

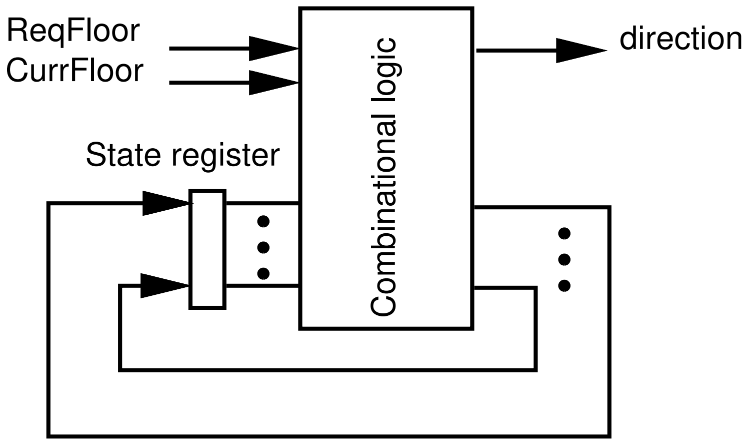
```
loop {
        if (ReqFloor == CurrFloor)
          direction = idle;
        else if (ReqFloor < CurrFloor)
          direction = down;
        else if  (ReqFloor > CurrFloor)
          direction = up;
 }
```

(b) Algorithmic view

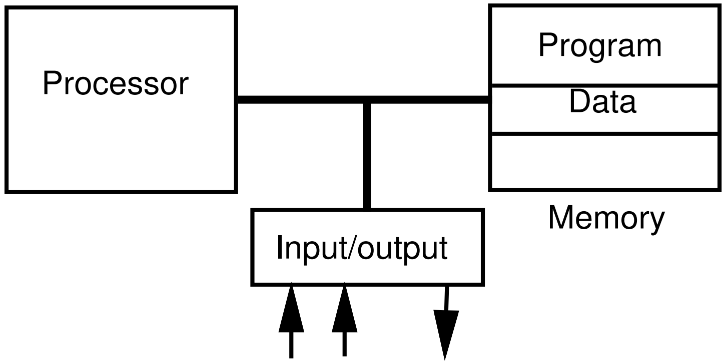


(c) State−machine view

Architectures for implementing the elevator controller



(a) Hardware implementation



(b) Software implementation

Design models

- State-oriented model
  Finite-state machine (FSM), Petri net, Hierarchical concurrent FSM

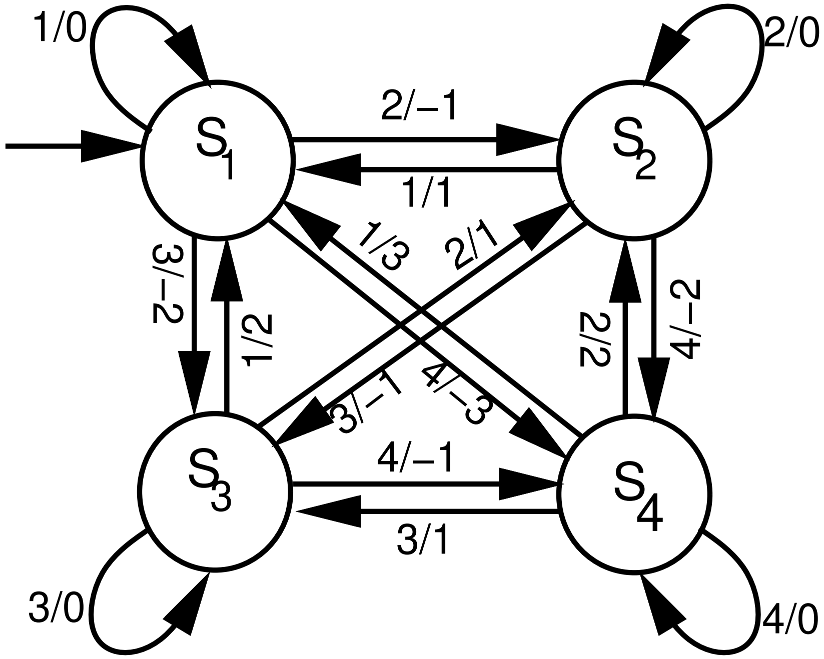
- Activity-oriented model
  Data-flow graph, Flowchart

- Structure-oriented model
  Block diagram, RT netlist, Gate netlist

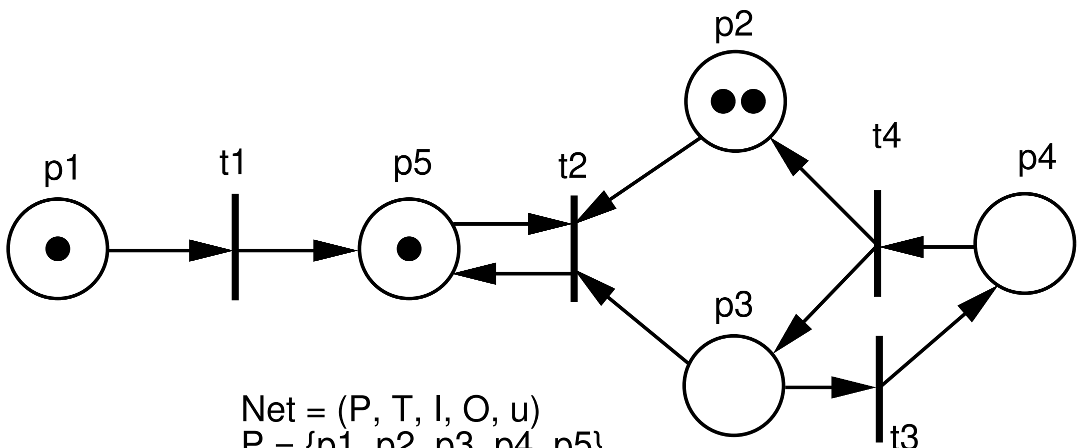
- Data-oriented model
  Entity-relationship diagram, Jackson's diagram

- Heterogeneous model
  Control-data flow graph, Structure chart, Programming language paradigm,
  Communicating sequential processes, Object-oriented paradigm,
  Program-state machine, Queueing model

Finite-state machine (FSM)



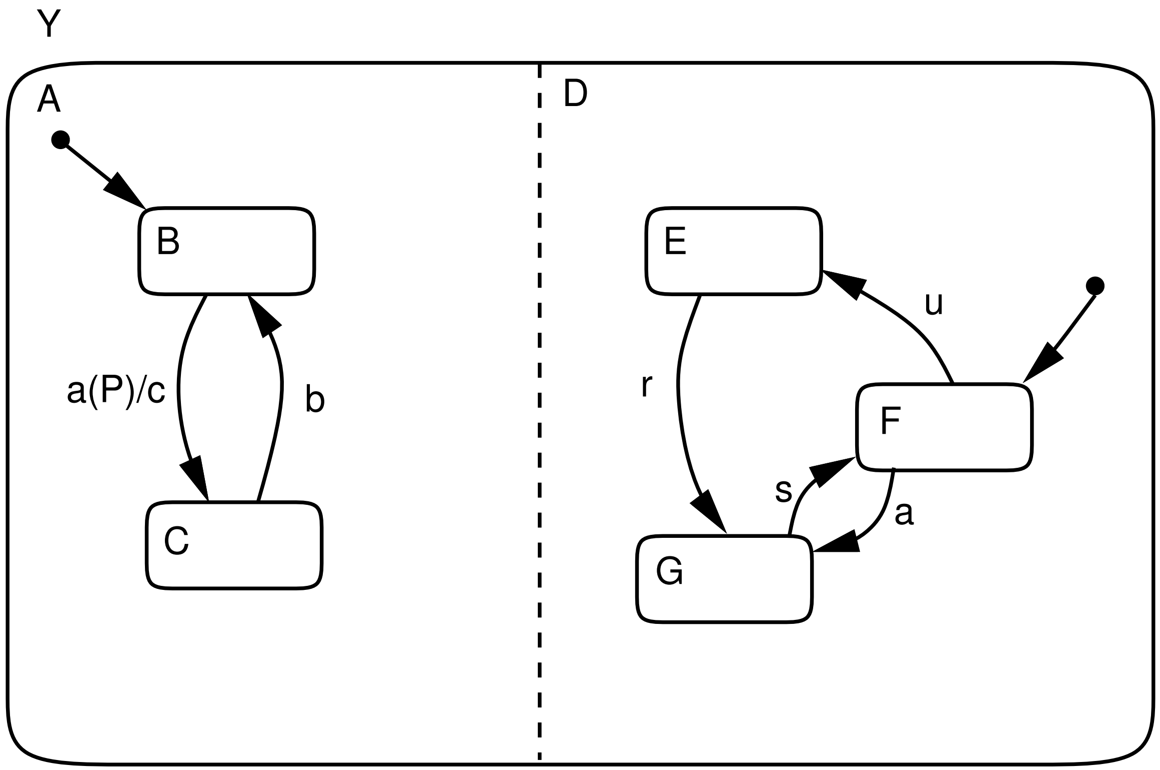
Petri net



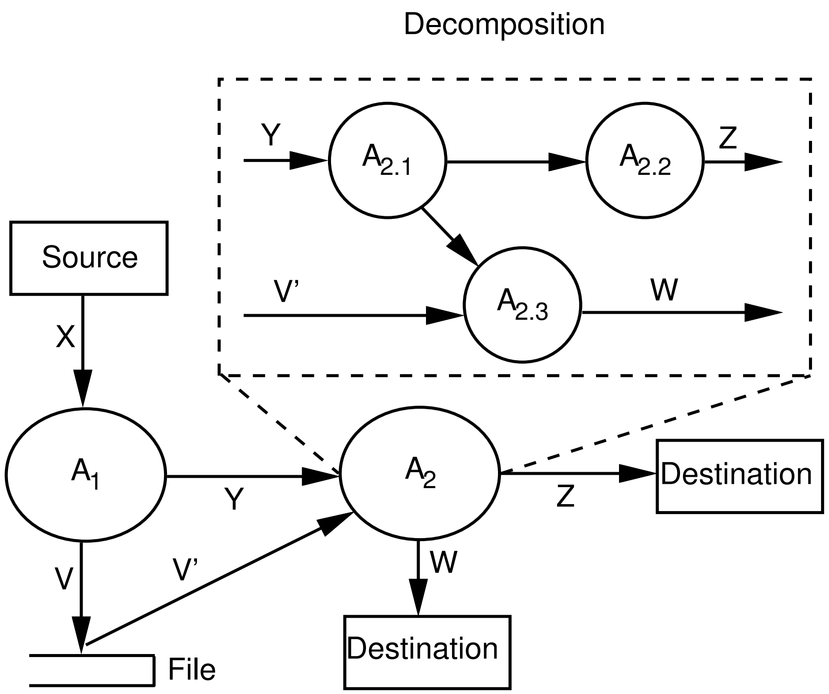
Net = (P, T, I, O, u)
P = {p1, p2, p3, p4, p5}
T = {t1, t2, t3, t4}
u = (1,2,0,0,1)

I(t1) = {p1}	O(t1) = {p5}
I(t2) = {p2,p3,p5}	O(t2) = {p5}
I(t3) = {p3}	O(t3) = {p4}
I(t4) = {p4}	O(t4) = {p2,p3}

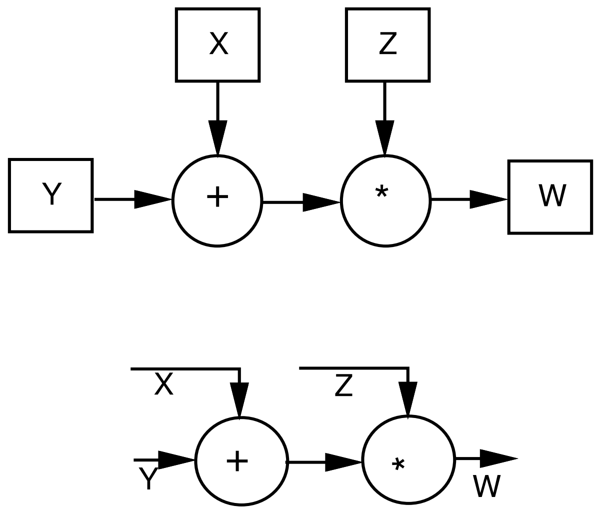
Hierarchical concurrent FSM



Data-flow graph

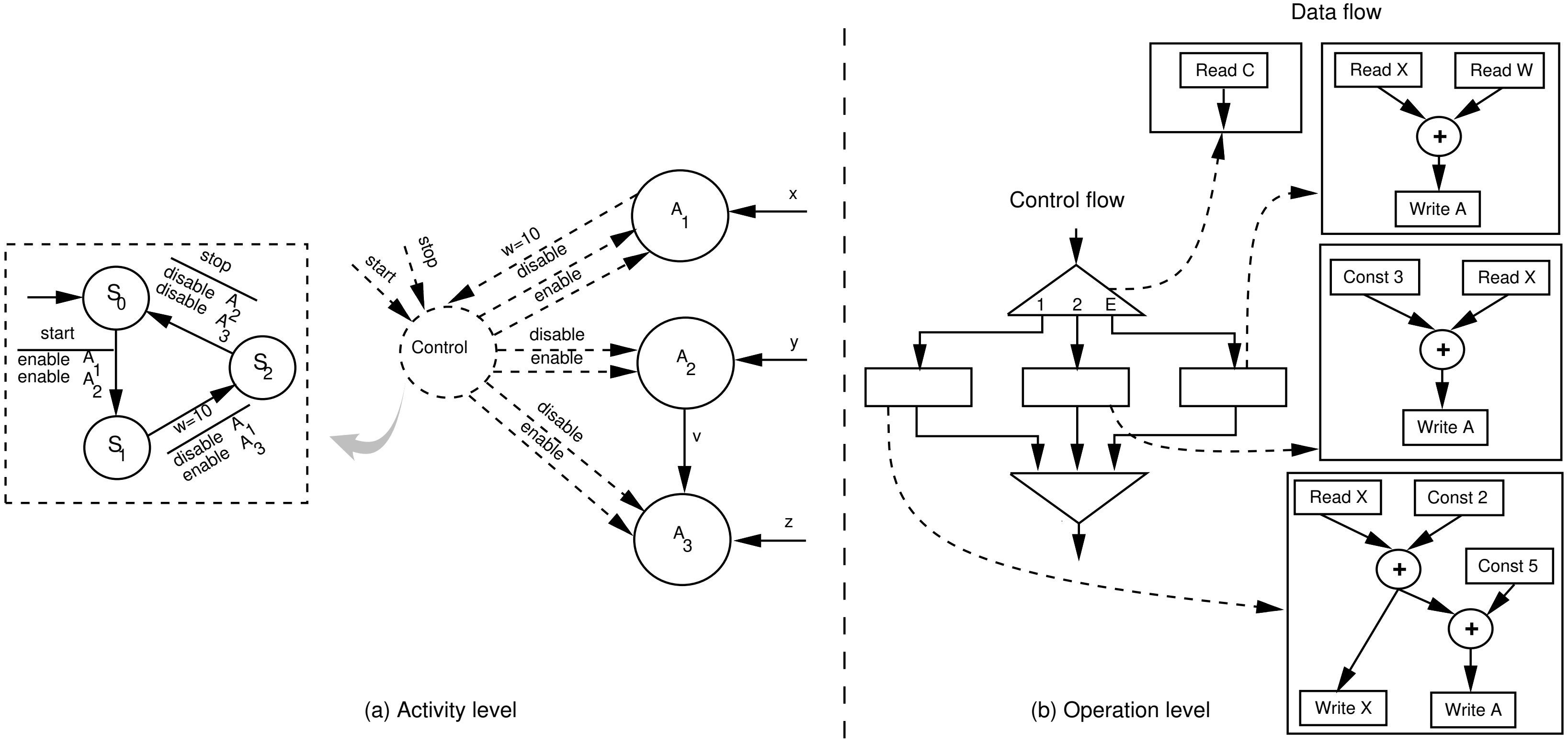


(a) Activity level



(b) Operation level

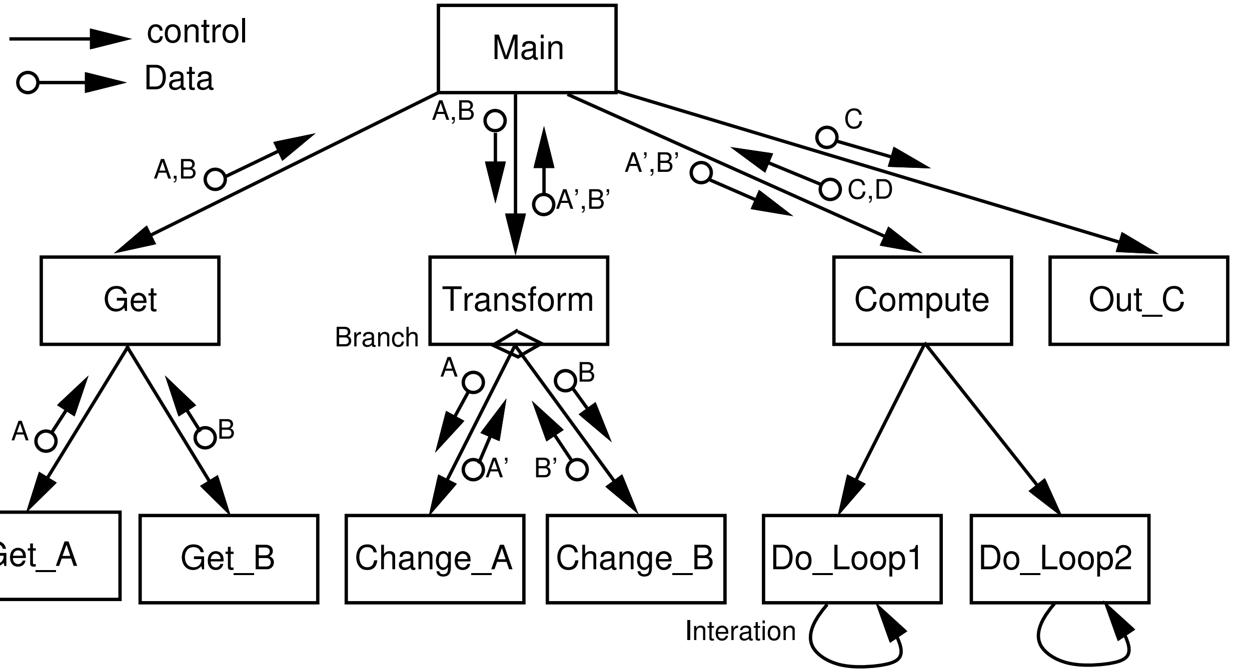
Control-data flow graph



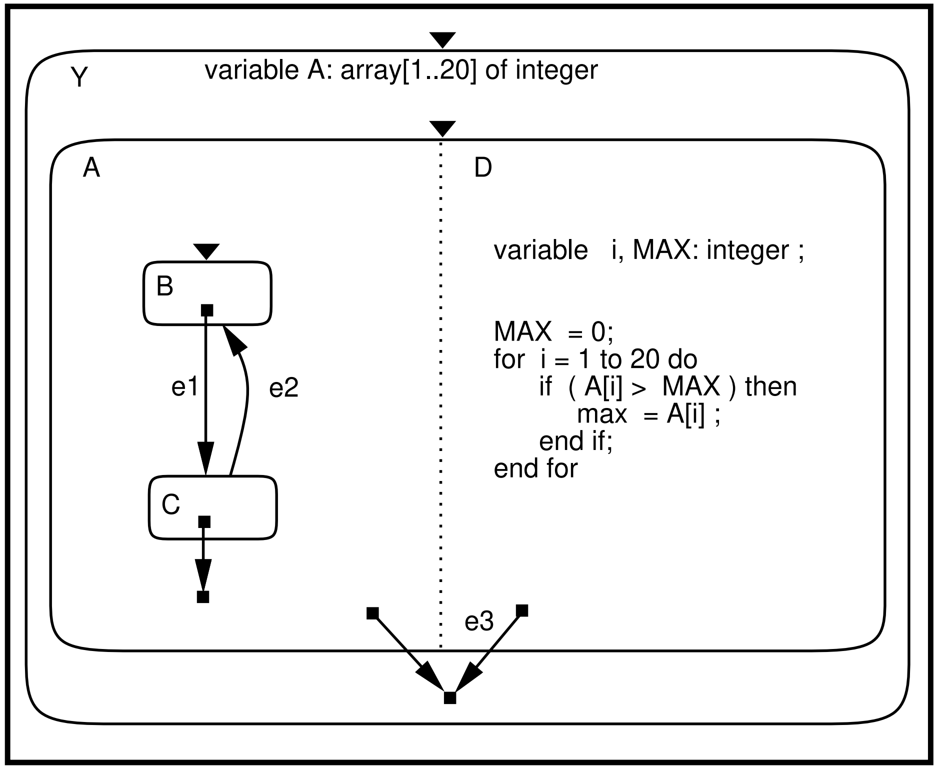
(a) Activity level

(b) Operation level

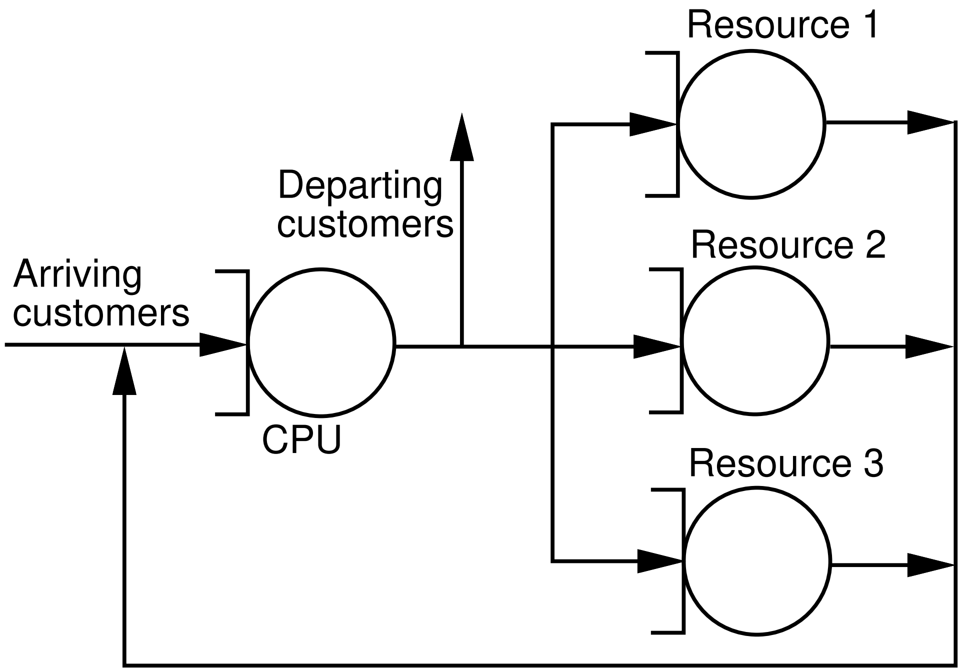
Structure chart



Program-state machine



Queueing model

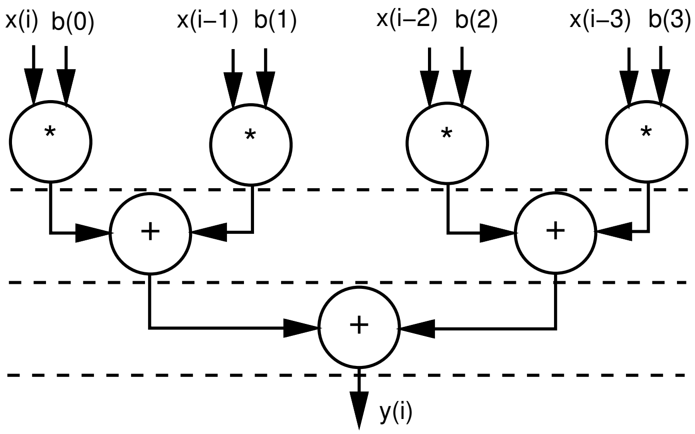


–Customer arriving rate
–Average service demand
–Average service time
–Utilization of the server

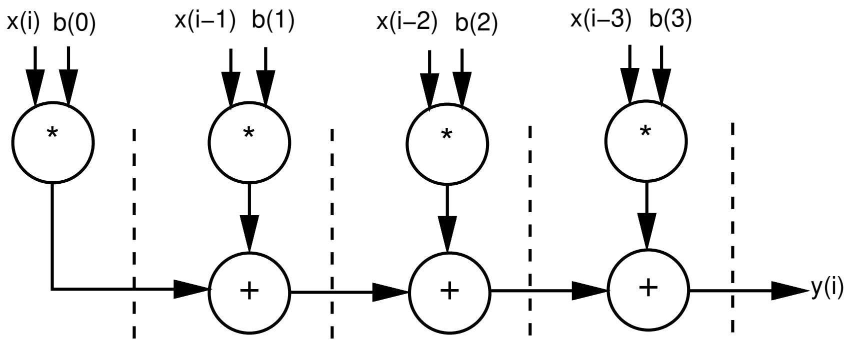
Architectures

- Finite-state machine controller (FSM)
- Digital signal processing datapath (DSP)
- Finite-state machine with datapath (FSMD)
- Complex instruction set computer (CISC)
- Reduced instruction set computer (RISC)
- Vector machine
- Very long instruction word computer (VLIW)
- Parallel processors

DSP

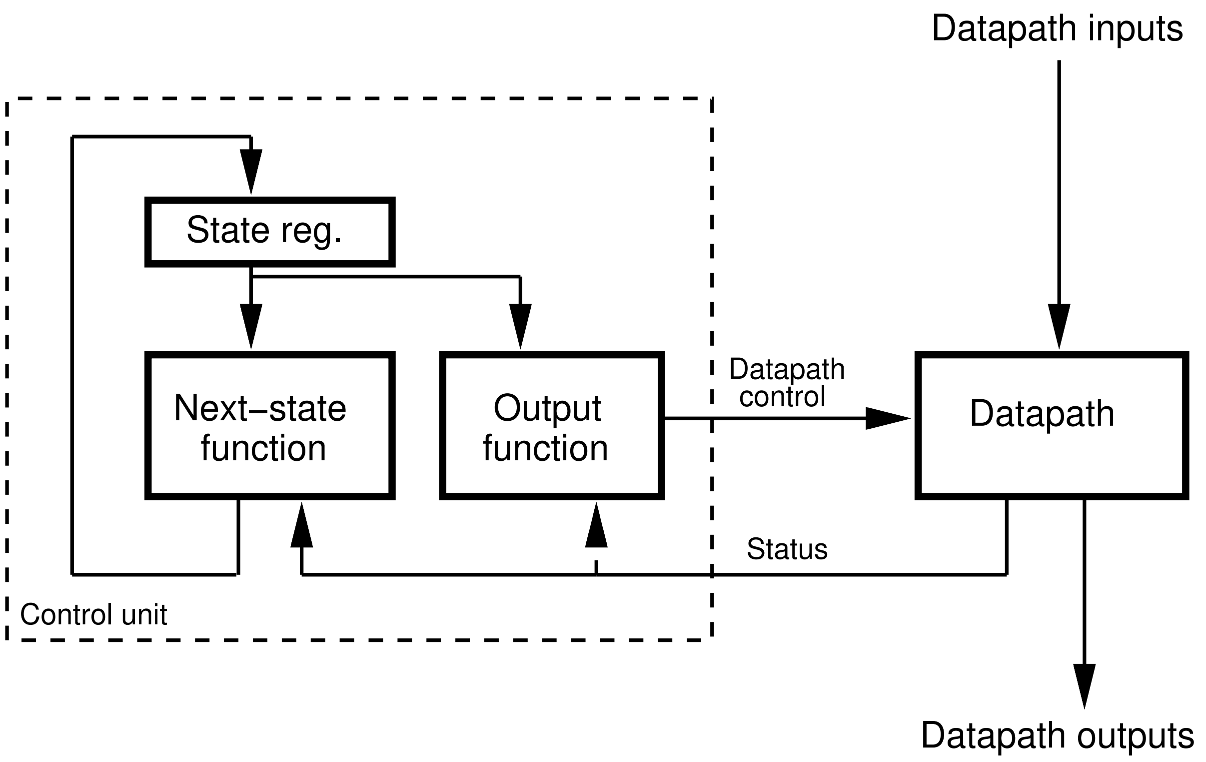


(a) Three stage pipeline



(b) Four stage pipeline

FSMD


Datapath inputs
State reg.
Next–state function
Output function
Datapath control
Datapath
Status
Control unit
Datapath outputs

System specification

- For every design, there exists a **conceptual view**
- Conceptual view depends on application
  - Computation : conceptualized as a program
  - Controller : conceptualized as a state-machine
- **Goal** of specification language
  - Capture conceptual view with minimum designer effort
- Ideal language

  1-to-1 mapping between conceptual model & language constructs

Outline

- Characteristics of commonly used conceptual models:
  Concurrency, hierarchy, synchronization
- Requirements for embedded system specification
- Evaluate HDLs with respect to embedded systems
  VHDL, Verilog, Esterel, CSP, Statecharts, SDL, SpecCharts

Concurrency

- **Behavior:** a chunk of system functionality
  - e.g. process, procedure, state-machine

- System often conceptualized as set of concurrent behaviors

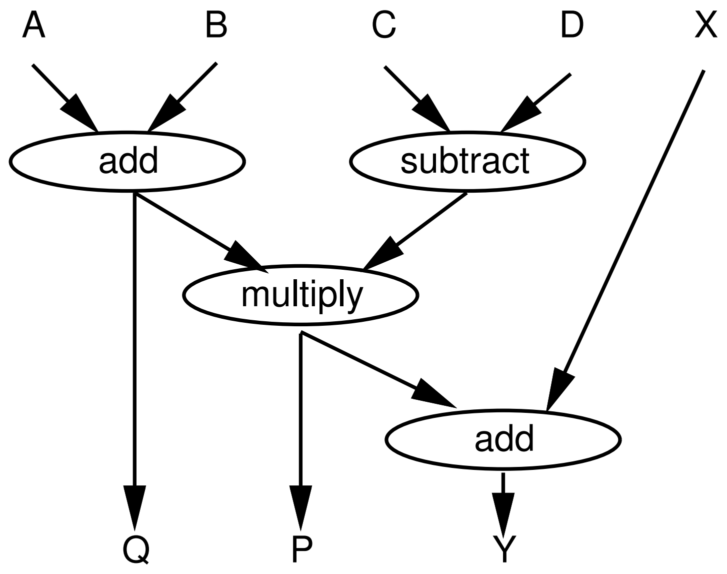
- Concurrency can exist at different abstraction levels:
  - Job-level
  - Task-level
  - Statement-level
  - Operation-level
  - Bit-level

- Two types of concurrency within a behavior
  - Data-driven, Control-driven

Data-driven concurrency

- Operations execute when input data is available
- Execution order determined by data dependencies

1: Q = A + B
2: Y = X + P
3: P = (C – D) * Q

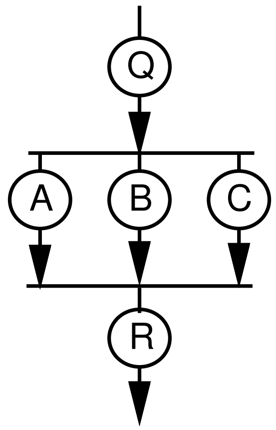


Control-driven concurrency

- Control thread : set of operations executed sequentially
- Concurrency represented by multiple control threads

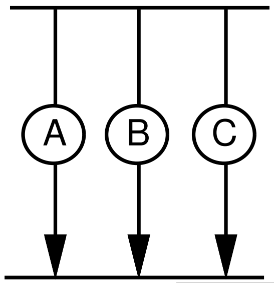
Fork-join statement

```
sequential  behavior X
begin
    Q();
    fork   A();  B(); C();  join;
    R();
end behavior X;
```



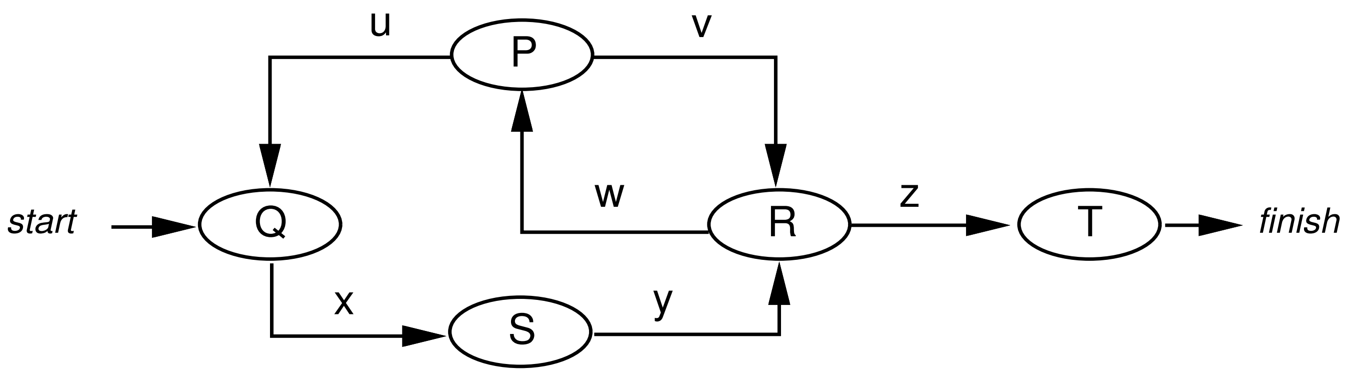
Process statement

```
concurrent  behavior X
begin
    process A();
    process B();
    process C();
end behavior X;
```



State-transitions

- Systems often are state-based, e.g. controllers
- State may represent
  - *mode* or stage of being
  - computation
- Difficult to capture using programming constructs

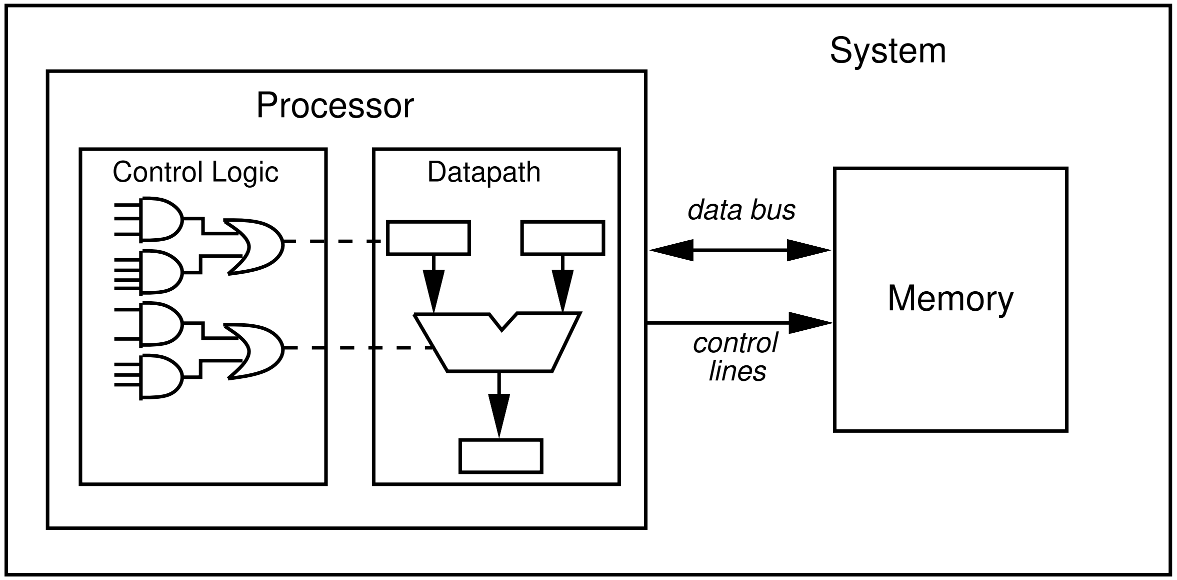


Hierarchy

- Required for managing system complexity
  - Allows system modeler to focus on one subsystem at a time
  - Enhances comprehension of system functionality
  - Scoping mechanism for objects like types and variables
- Two types of hierarchy
  - Structural hierarchy
  - Behavioral hierarchy

Structural hierarchy

- System represented as set of interconnected components
- Interconnections between components represent wires
- Several levels: systems, chips, RT-components, gates

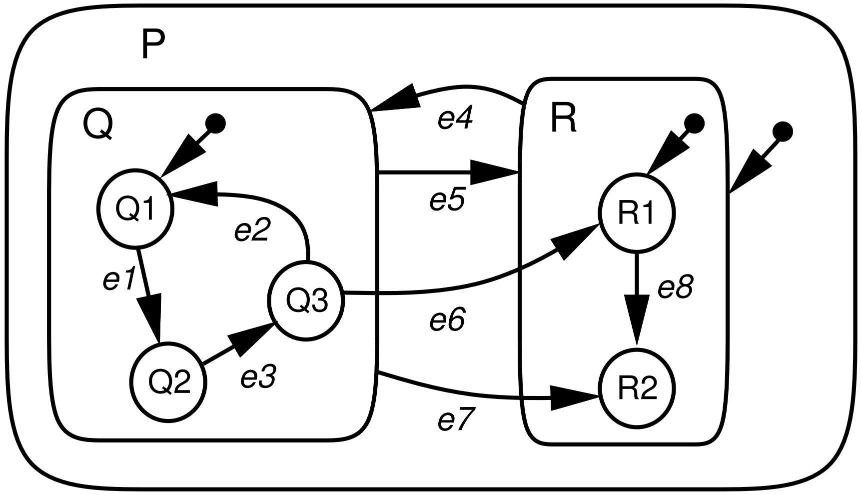


Behavioral hierarchy

- Ability to successively decompose behavior into sub-behaviors

```
behavior  P
    variable x, y;
begin
    Q(x) ;
    R(y) ;
end  behavior  P;
```

- Concurrent decomposition
  - Fork-join
  - Process
- Sequential decomposition
  - Procedure
  - State-machine



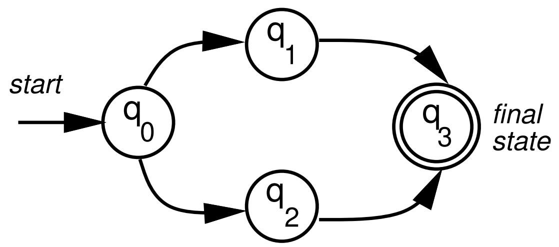
Programming constructs

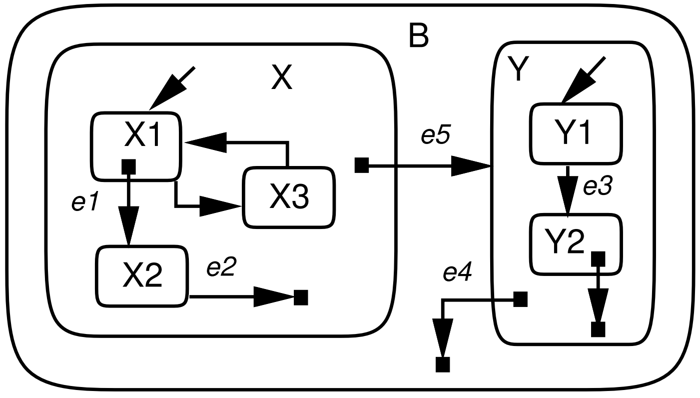
- Some behaviors easily conceptualized as sequential algorithms
- Wide variety of constructs available
  Assignment, branching, iteration, subprograms, recursion, complex data types (records, lists)

```
type      buffer_type  is array (1 to 10) of integer;
variable  buf : buffer_type;
variable  i, j : integer;

for i = 1 to 10
    for j = i to i
        if (buf(i) > buf(j)) then
            SWAP(buf(i), buf(j));
        end if;
    end for;
end for;
```

Behavioral completion

- Behavior *completes* when all computations performed
- Advantages
  - Behavior can be viewed without inter-level transitions
  - Allows natural decomposition into sequential subbehaviors

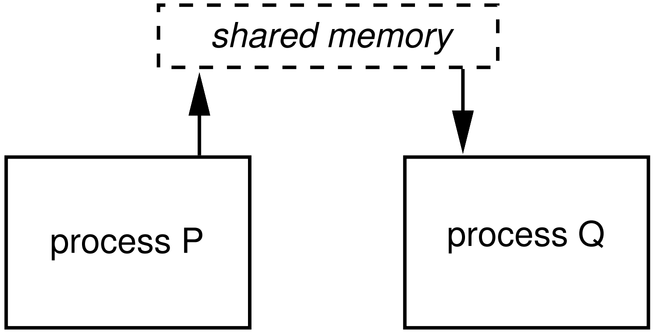




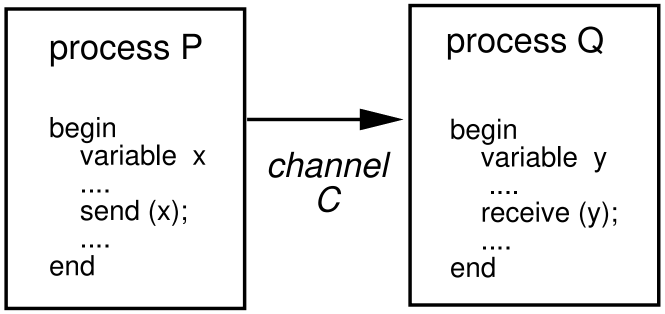
UC Irvine

Communication

- Concurrent behaviors exchange data
- Shared-memory model
  - Sender updates common medium
  - Persistent, Non-persistent



- Message-passing model
  - Data sent over abstract *channels*
  - Unidirectional / bidirectional
  - Point-to-point / multiway
  - Blocking / non-blocking



Synchronization

- Concurrent behaviors execute at different speeds

- Synchronization required when
  - Data exchanged between behaviors
  - Different activities must be performed simultaneously

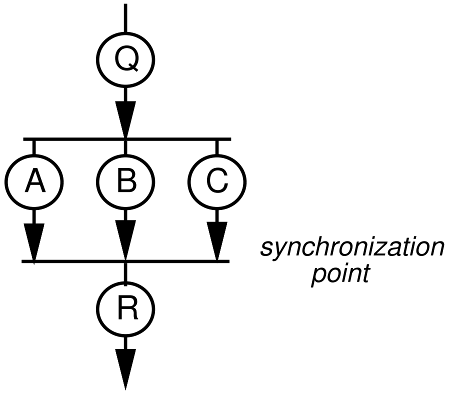
- Two types of synchronization mechanisms
  - Control-dependent
  - Data-dependent

Control-dependent synchronization

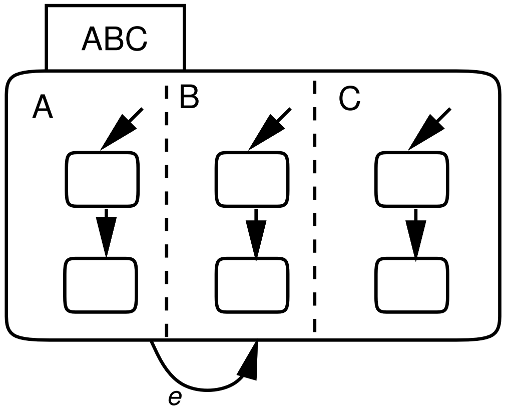
- Synchronization based on control structure of behavior

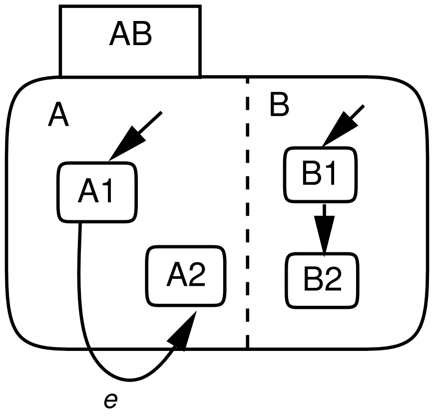
Fork-join

```
behavior X
begin
    Q();
    fork   A();  B(); C();  join;
    R();
end behavior X;
```



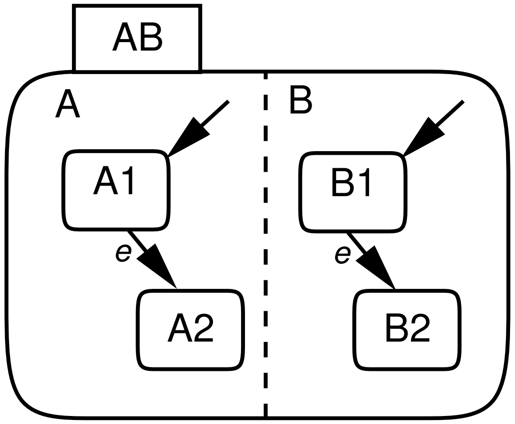
Reset



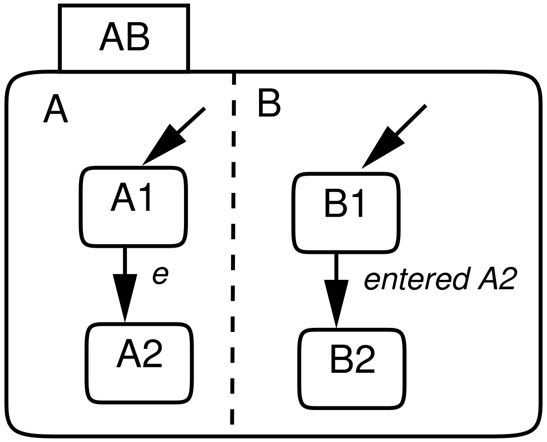


Data-dependent synchronization

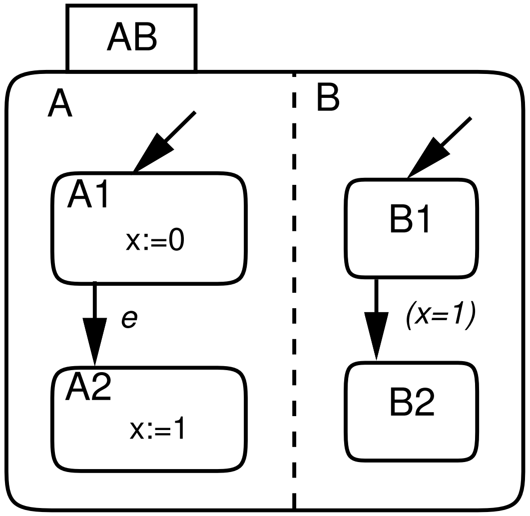
- Synchronization based on communication of data between behaviors



Synchronization by common event



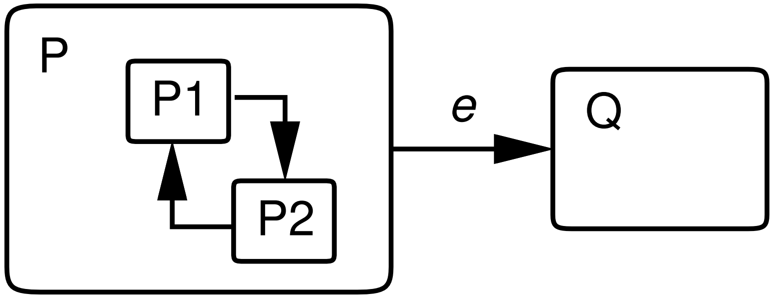
Synchronization by status detection



Synchronization by common variable

Exception handling

- Occurrence of event terminates current computation
- Control transferred to appropriate next mode
- Example of exceptions: interrupts, resets

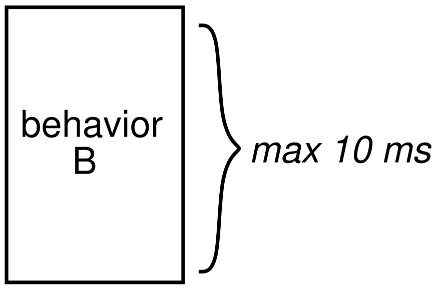


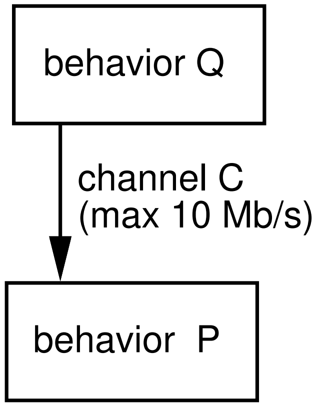
Timing

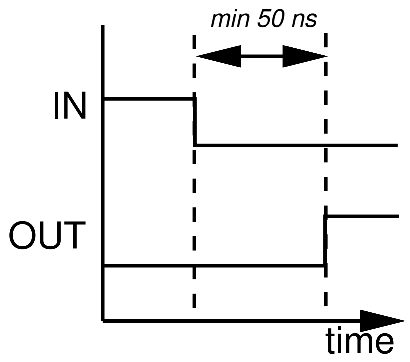
- Required to represent real world implementations
- **Functional timing**: affects simulation of system specification

  wait for 200 ns;

  A <= A + 1 after 100 ns;
- **Timing constraints**: guide synthesis and verification tools





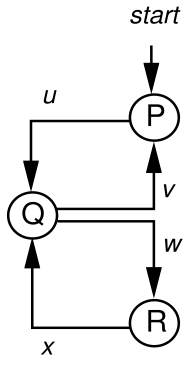


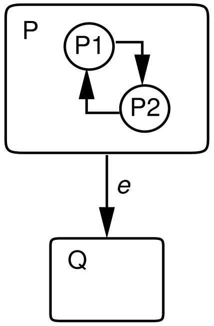
UC Irvine

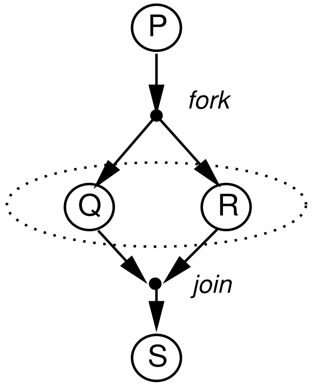
Embedded system specification

- Embedded system: behavior defined by interaction with environment
- Essential characteristics

State-transitions	Exceptions
Behavioral hierarchy	Concurrency
Programming constructs	Behavioral completion







VHDL

- IEEE standard, intended for documentation and exchange of designs [IEE88]

- **Characteristics supported**
  - Behavioral hierarchy : single level of processes
  - Structural hierarchy : nested blocks and component instantiations
  - Concurrency : task-level (process), statement-level (signal assignment)
  - Programming constructs
  - Communication : shared-memory using global signals
  - Synchronization : *wait on* and *wait until* statements
  - Timing : *wait for* statement, *after* clause in assignments

- **Characteristics not supported**
  - Exceptions : partially supported by guarded signal assignments
  - State transitions

Verilog and Esterel

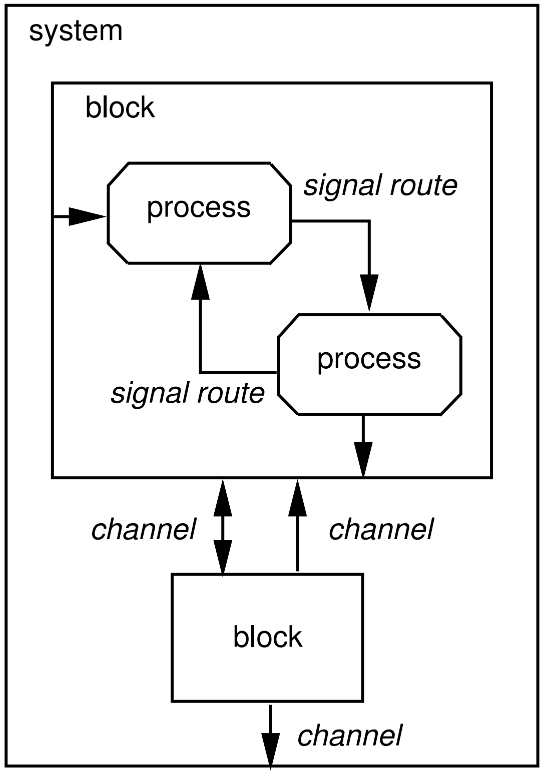
- Verilog [TM91] developed as proprietary language for specification, simulation
- Esterel [Hal93] developed for specification of reactive systems
- **Characteristics supported:**
  - Behavioral hierarchy : fork-join
  - Structural hierarchy : hierarchy of interconnected *modules*
  - Programming constructs
  - Communication : shared registers (Verilog) and broadcasting (Esterel)
  - Synchronization : *wait* for an event on a signal
  - Timing : modeling of gate, net, assignment delays in Verilog
  - Exceptions : *disable* (Verilog), *watching, do-upto, trap* statements (Esterel)
- **Characteristics not supported:** State transitions

SDL (Specification and Description language)

- CCITT standard in telecommunication for protocol specification [BHS91]

- **Characteristics supported**
  - Behavioral hierarchy : nested dataflow
  - Structural hierarchy : nested blocks
  - State transitions : state machine in processes
  - Communication : message passing
  - Timing : *timeouts* generated by timer object

- **Characteristics not supported**
  - Exceptions
  - Programming constructs



CSP (Communicating Sequential Processes)

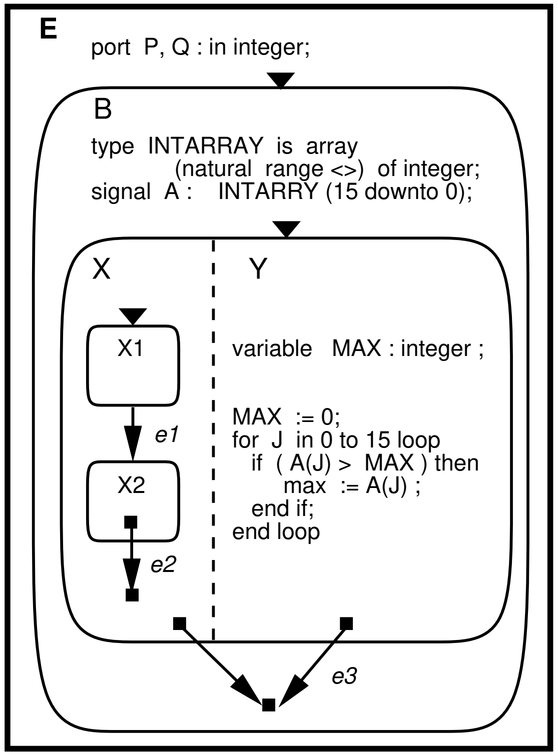
- Intended to specify programs running on multiprocessor machines [Hoa78]

- **Characteristics supported**
  - Behavioral hierarchy : fork-join using *parallel* command
  - Programming constructs
  - Communication : message passing using *input, output* commands
  - Synchronization : blocking message passing

- **Characteristics not supported**
  - Exceptions
  - State transitions
  - Structural hierarchy
  - Timing

SpecCharts

- Developed for embedded system specification [NVG92]
- PSM (program-state machine) model + VHDL
- **Characteristics supported**
  - Behavioral hierarchy : sequential/concurrent behaviors
  - State transitions: TOC (transition on completion) arcs
  - Communication : shared memory, message passing
  - Exceptions : TI (transition immediately) arcs
- **Characteristics similar to VHDL**
  - Programming constructs
  - Structural hierarchy
  - Synchronization and Timing



Summary

Language	Embedded System Features					
	State Transitions	Behavioral Hierarchy	Concurrency	Program Constructs	Exceptions	Behavioral Completion
VHDL	○	◐	●	●	○	●
Verilog	○	●	●	●	●	●
Esterel	○	●	●	●	●	●
SDL	●	◐	●	○	○	●
CSP	○	●	●	●	○	●
Statecharts	●	●	●	○	●	○
SpecCharts	●	●	●	●	●	●

● Feature fully supported ◐ Feature partially supported ○ Feature not supported

Specification example

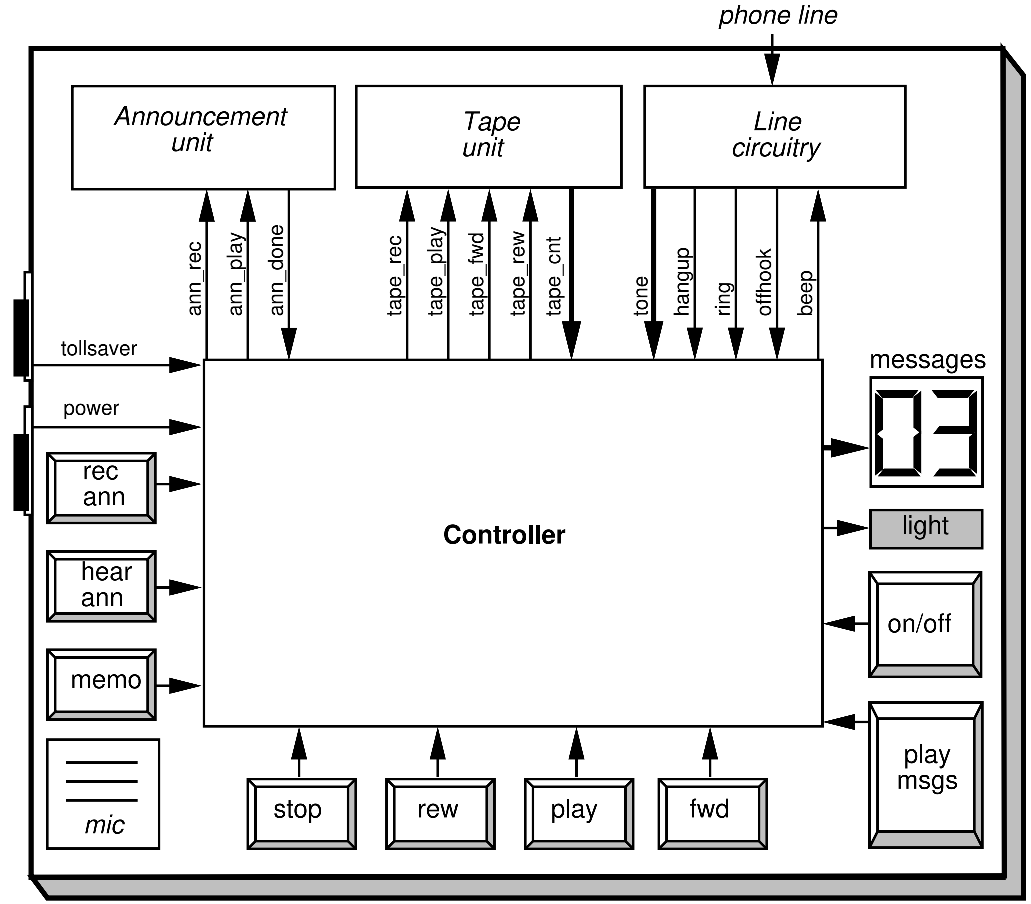
- An executable specification-language enables:
  - Early verification
  - Precision
  - Automation
  - Documentation

- A good language/model match reduces:
  - Capture time
  - Comprehension time
  - Functional errors

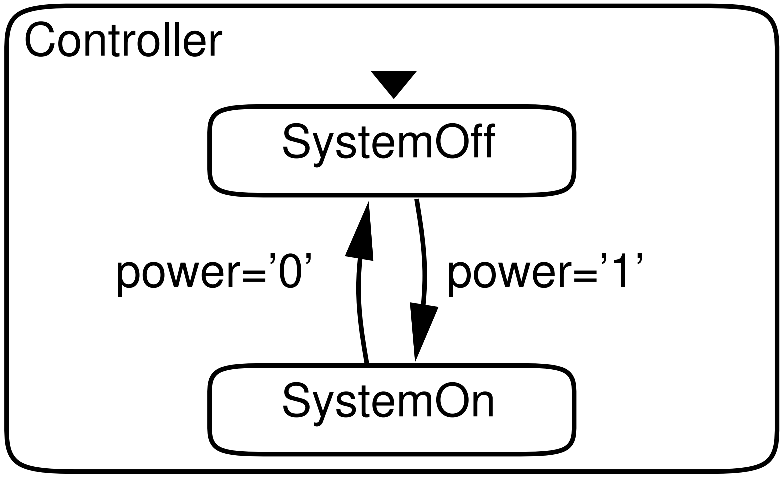
Outline

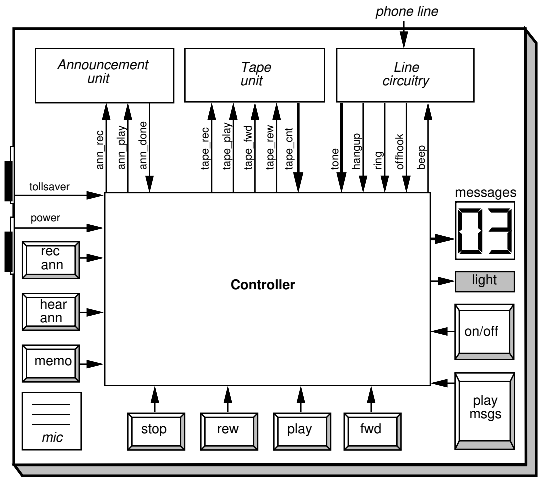
- Capture an example's model in a particular language
  - PSM model in the SpecCharts language
- Point out the benefits of a good language/model match
- Highlight experiments that demonstrate those benefits

Answering machine controller's environment



Highest-level view of the controller

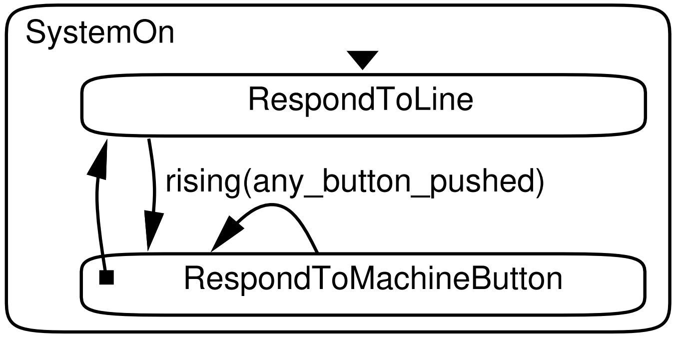


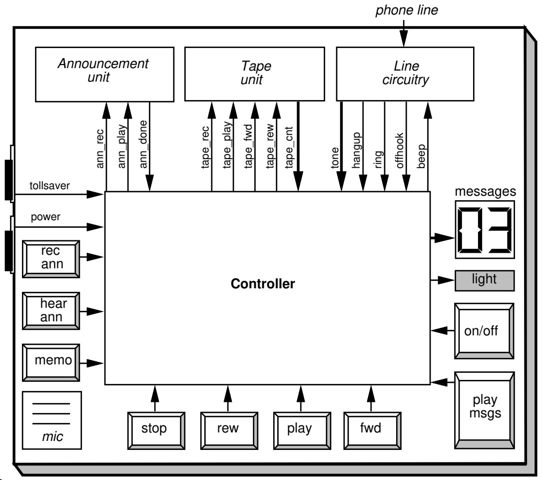


  UC Irvine

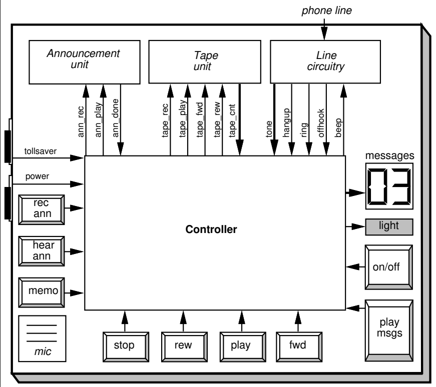
The SystemOn behavior

- System usually responds to the line
- Pressing any machine button gets immediate response



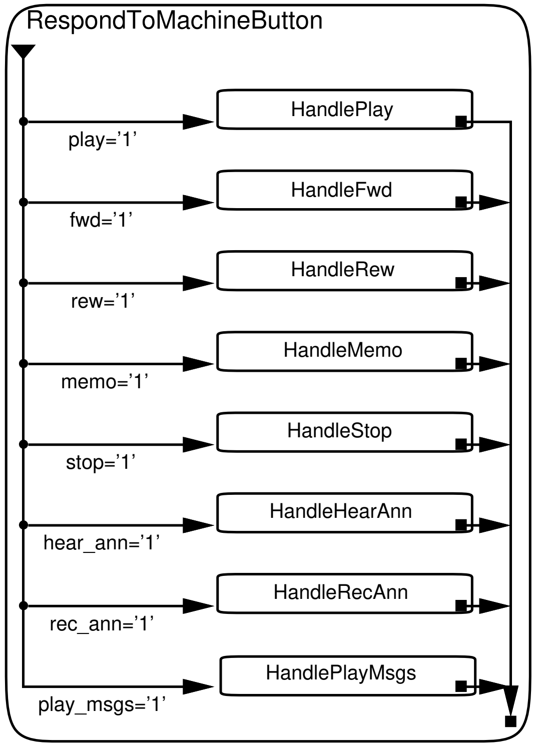


The RespondToMachineButton behavior




```
behavior  RespondToMachineButton
   type code is
begin
    if (play='1') then
        HandlePlay;
    elsif (fwd='1') then
        HandleFwd;
    elsif (rew='1') then
        HandleRew;
    elsif (memo='1') then
        HandleMemo;
    elsif (stop='1') then
        HandleStop;
    elsif (hear_ann='1')  then
        HandleHearAnn;
    elsif (rec_ann='1') then
        HandleRecAnn;
    elsif (play_msgs='1') then
        HandlePlayMsgs;
    end if;
end;
```

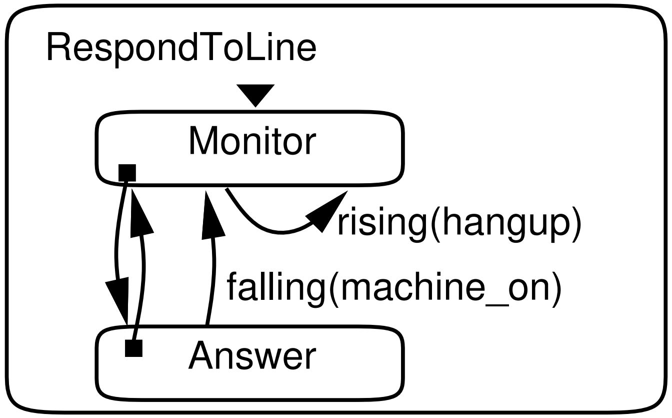
(a)

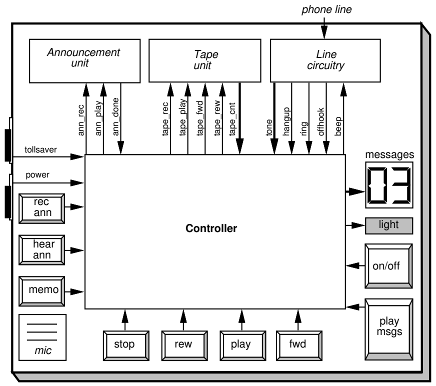


(b)

The RespondToLine behavior

- Monitors line for rings
- Answers line
- Responds to exceptions
  - Hangup
  - Machine turned off

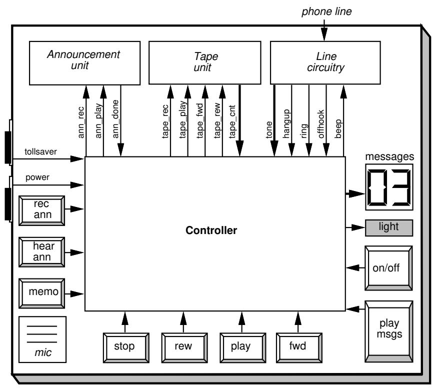




UC Irvine

The Monitor behavior

- Counts for required rings
- Requirements may change



Monitor

```
signal  rings_to_wait : integer range 1 to 20 := 4;
function  DetermineRingsToWait return integer is begin
     if ((num_msgs > 0) and (tollsaver='1')  and (machine_on='1')) then
        return(2);
     elsif (machine_on='1') then
        return(4);
     else
        return(15);
     end if;
end;
```

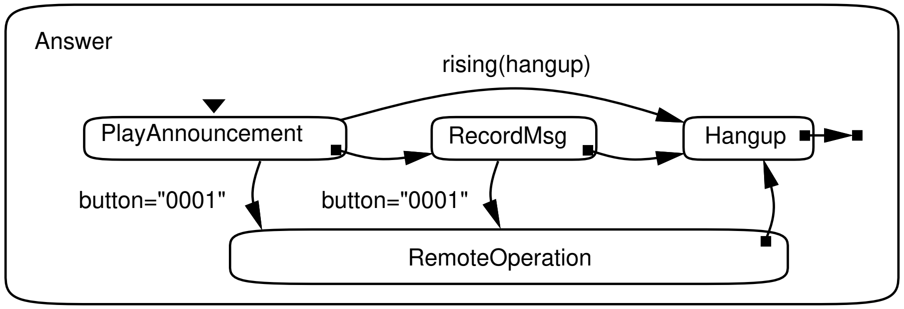
MaintainRingsToWait

```
loop
  rings_to_wait <= DetermineRingsToWait;
  wait on tollsaver, machine_on;
end loop;
```

CountRings

```
variable I : integer range 0 to 20;

i := 0;
while (i < rings_to_wait) loop
   wait on rings_to_wait, ring;
   if (rising(ring)) then
      i := i + 1;
   end if;
end loop;
```

The Answer behavior



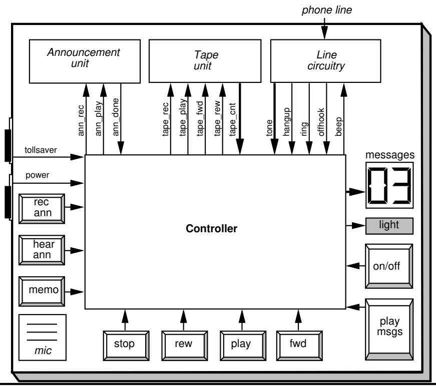
(a)

```
behavior  PlayAnnouncement  type code is
begin
    ann_play <= '1';
    wait until ann_done = '1';
    ann_play <= '0';
end;
```

(b)

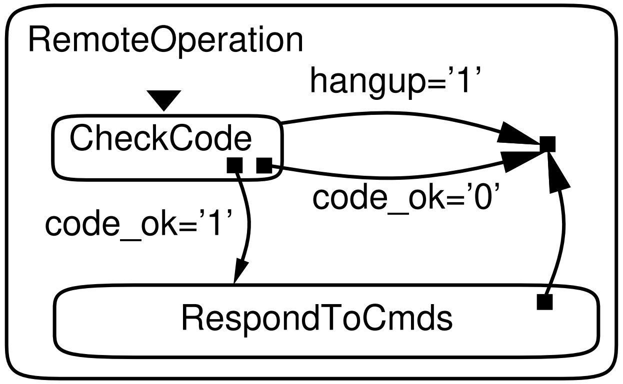
```
behavior  RecordMsg  type code is
begin
    ProduceBeep(1 s);
    if  (hangup = '0') then
        tape_rec <= '1';
        wait until hangup='1'  for 100 s;
        ProduceBeep(1 s);
        num_msgs <= num_msgs + 1;
        tape_rec <= '0';
    end if;
end;
```

(c)



The RemoteOperation behavior

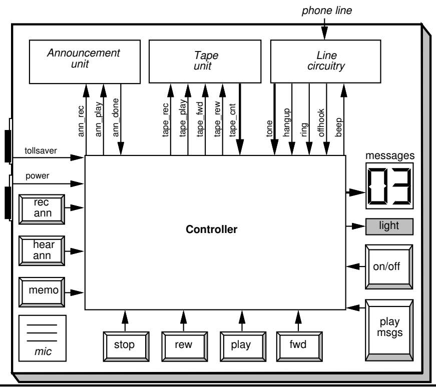
- Owner can operate machine remotely by phone
- Owner identifies himself by four button ID



(a)

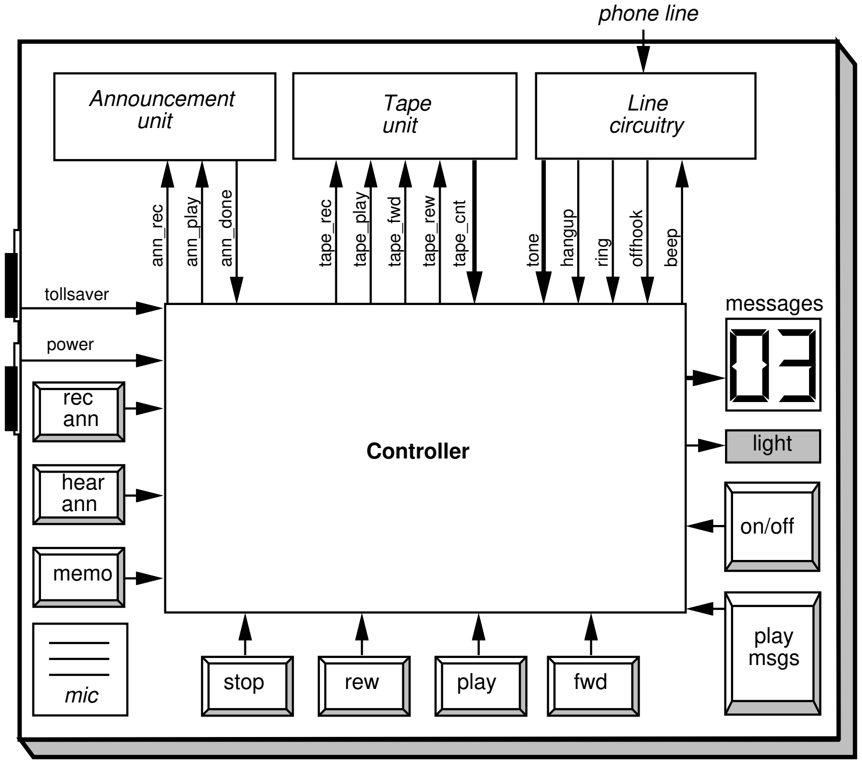
```
behavior  CheckUserCode  type code is
begin
   code_ok <= true;
   for (i in 1 to 4) loop
      wait until tone /= "1111" and tone'event;
      if (tone /= user_code(i)) then
          code_ok <= false;
      end if;
   end loop;
end;
```

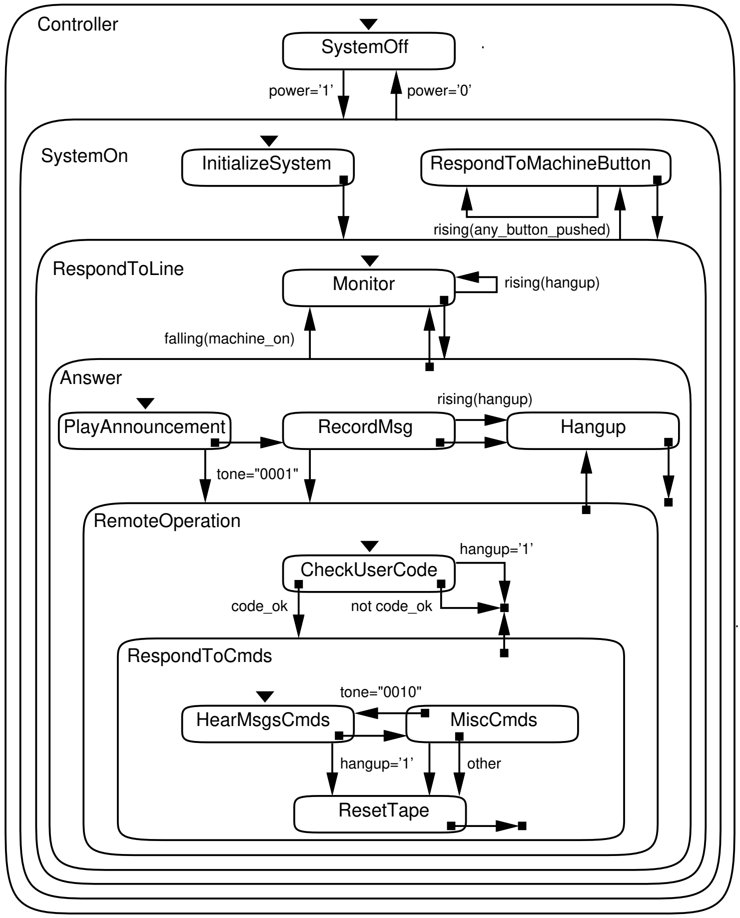
(b)



UC Irvine

The answering machine controller specification





Executable specification use

- Precision
  - Readability/precision compete in a natural language
  - Executable specification encourages precision
  - Designer asks questions, specification answers them

- Language/model match (SpecCharts/PSM):
  - Hierarchy
  - State-transitions
  - Programming constructs
  - Concurrency
  - Exceptions
  - Completion
  - Equivalence of states and programs

Specification capture experiment

	VHDL	SpecCharts
Average specification–time in minutes	40	16
Number of modelers	3	3
Number of incorrect specifications first time	2	0
Number of incorrect specifications second time	1	0

- VHDL modelers required 2.5 times longer
- Two VHDL specifications possessed control errors
- SpecCharts were effective for state-transitions and exceptions

Comparison of SpecCharts, VHDL and Statecharts

Answering machine example

		Conceptual model	SpecCharts	VHDL (hierarch.)	VHDL (flat)	Statecharts
Specification attributes	Program−states	42	42	42	32	80
	Arcs	40	40	40	152	135
	Control signals	––	0	84	1	0
	Lines/leaf	––	7	27	29	––
	Lines	––	446	1592	963	––
	Words	––	1733	6740	8088	––
Shortcomings	No sequential program constructs					X
	No hierarchy			X	X	
	No exception constructs			X	X	
	No hierarchical events				X	
	No state−transition constructs			X	X	

Design quality experiment

Design attribute	Designed from English	Designed from SpecCharts
Control transistors	3130	2630
Datapath transistors	2277	2251
Total transistors	5407	4881
Total pins	38	38

- No loss in design quality with an executable language

Summary

- Executable languages encourage precision and automation
- The language should support an appropriate model
  - Makes specification easy
- Strongly parallels programming languages
  - Structured vs. assembly languages
  - Object-oriented model and C++

System partitioning

- System functionality is implemented on system components
  - ASICs, processors, memories, buses
- Two design tasks:
  - **Allocate** system components or ASIC constraints
  - **Partition** functionality among components
- Constraints
  - Cost, performance, size, power
- Partitioning is a central system design task

Outline

- Structural vs. functional partitioning
- Natural vs. executable language specifications
- Basic partitioning issues and algorithms
- Functional partitioning techniques for hardware
- Hardware/software partitioning
- Functional partitioning techniques for software
- Exploring tradeoffs with functional partitioning

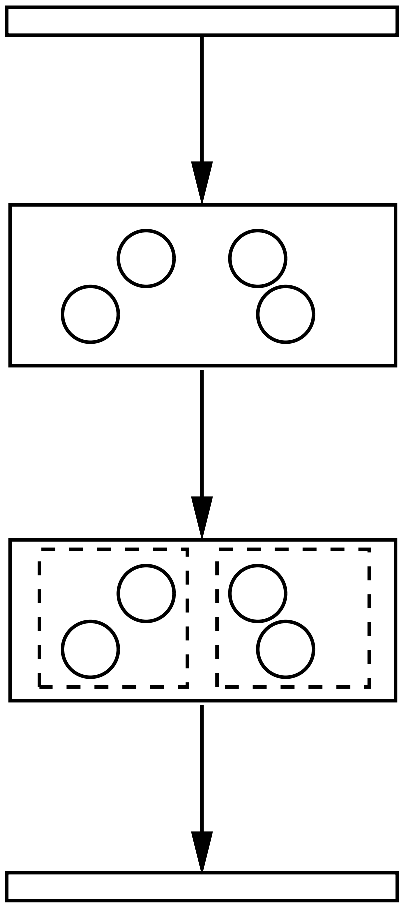
Structural vs. functional partitioning

- Structural: Implement structure, then partition
- Functional: Partition function, then implement
  - Enables better size/performance tradeoffs
  - Uses fewer objects, better for algorithms/humans
  - Permits hardware/software solutions
  - But, it’s harder than graph partitioning

Natural vs. executable language specifications

- Alternative methods for specifying functionality
- Natural languages common in practice
- Executable languages becoming popular
  - Automated estimation/partitioning explores solutions
  - Early verification reduces costly late changes
  - Precision eases integration

Basic partitioning issues



Specification abstraction−level

Granularity

Metrics and estimations

Partitioning algorithms

Objective and closeness functions

System−component allocation

Output

Basic partitioning issues (cont.)

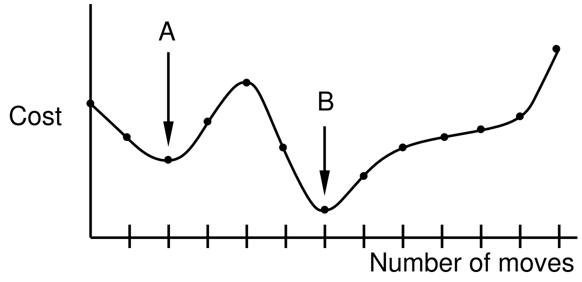
- Specification-abstraction level: input definition
  - Just indicating the language is insufficient
  - Abstraction-level indicates amount of design already done
  - e.g. task DFG, tasks, CDFG, FSMD
- Granularity: specification size in each object
  - Fine granularity yields more possible designs
  - Coarse granularity better for computation, designer interaction
  - e.g. tasks, procedures, statement blocks, statements
- Component allocation: types and numbers
  - e.g. ASICs, processors, memories, buses
- Output: format and uses
  - e.g. new specification, hints to synthesis tool

Basic partitioning issues (cont.)

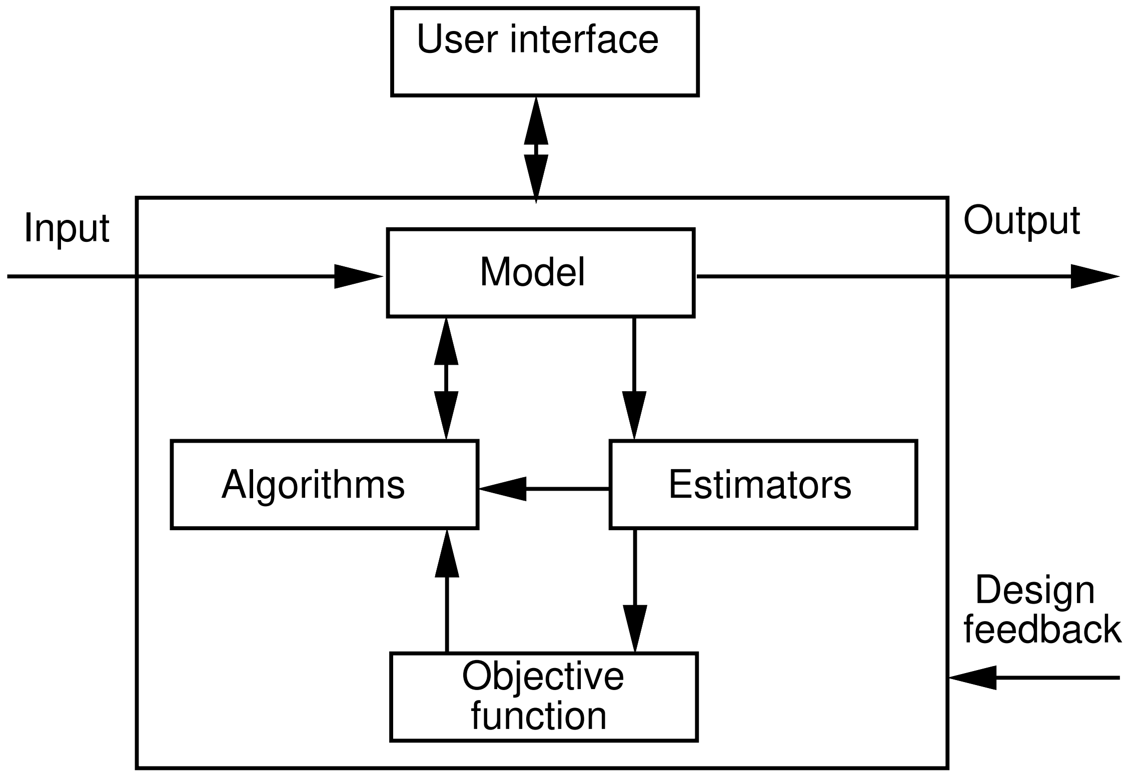
- Metrics and estimations: "good" partition attributes
  - e.g. cost, speed, power, size, pins, testability, reliability
  - Estimates derived from quick, rough implementation
  - Speed and accuracy are competing goals of estimation
- Objective and closeness functions
  - Combines multiple metric values
  - Closeness used for grouping before complete partition
  - Weighted sum common
  - e.g. $k_1F(area, c) + k_2F(delay, c) + k_3F(power, c)$

Basic partitioning issues (cont.)

- Algorithms: control strategies seeking best partition
  - Constructive creates partition
  - Iterative improves partition
  - Key is to escape local minimum



Typical partitioning-system configuration



Basic partitioning algorithms

- Clustering and multi-stage clustering [Joh67, LT91]
- Group migration (a.k.a. min-cut or Kernighan/Lin) [KL70, FM82]
- Ratio cut [KC91]
- Simulated annealing [KGV83]
- Genetic evolution
- Integer linear programming

Hardware/software partitioning

- Combined hardware/software systems are common
- Software is cheap, modifiable, and quick to design
- Hardware is fast
- Special algorithms are needed to favor software
- Proposed algorithms
  - Greedy [GD92]
  - Hill climbing [EHB94]
  - Binary-constraint search with hill climbing [VGG93]

Functional partitioning for systems: Vulcan, Cosyma

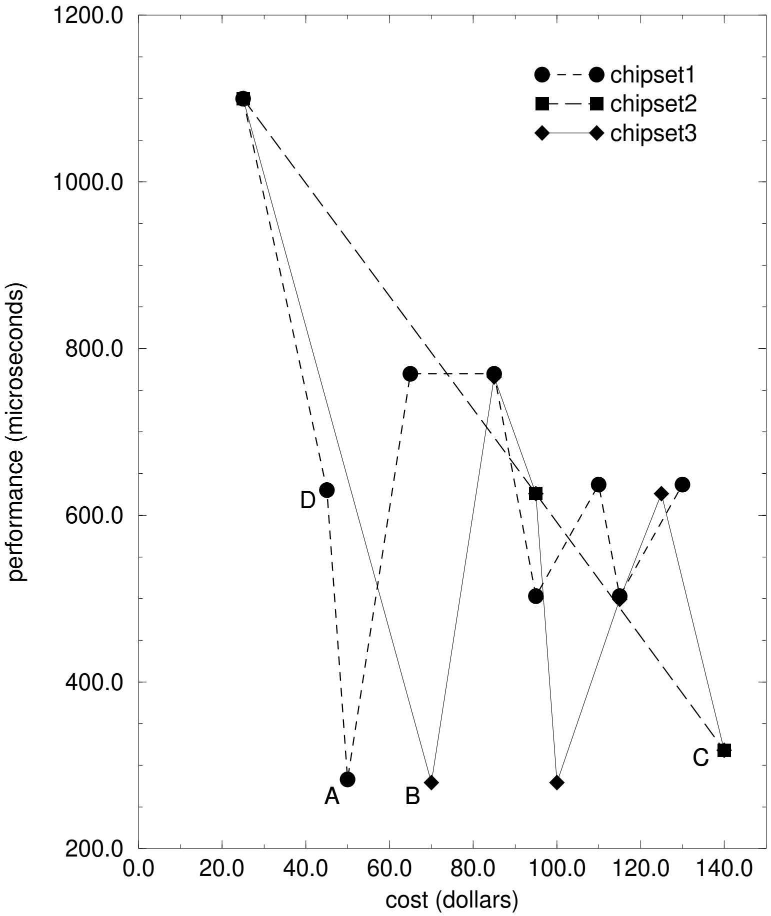
- Vulcan [GD90]I
  - Partitions CDFG operations among hardware only
  - Group migration and simulated annealing algorithms
- Vulcan II [GD93]
  - Partitions operations among hardware/software
  - Architecture: processor, hardware, memory, bus
  - All communication through memory
  - Uses greedy algorithm, extracts behaviors from hardware
- Cosyma [EHB94]
  - Partitions statement blocks among hardware/software
  - Architecture: processor, hardware, memory, bus
  - Simulated annealing, extracts behaviors from software

Functional partitioning for systems: SpecSyn

- Solves three partitioning problems
  - Behaviors to processors/ASICs
  - Variables to memories
  - Communication channels to buses
- Uses fast incremental-update estimators
- Covers both hardware and hardware/software partitioning [GVN94, VG92]

Exploring tradeoffs with functional partitioning

- Each line represents a different vendor's chip set
- Each point represents an allocation and partition
- Many designs quickly examined



Summary

- Partitioning heavily influences design quality
- Functional partitioning is necessary
- Executable specification enables:
  - Automation
  - Exploration
  - Documentation
- Variety of algorithms exist
- Variety of techniques exist for different applications

Future directions

- Metrics from real design to guide partitioning
- Comparison of functional partitioning algorithms
- Impact of metric selections and orderings
- Impact of of granularity on partition quality
- Exploitation of regularity in partitioning

Estimation

- Estimates allow
  - Evaluation of design quality
  - Design space exploration

- Design model
  - Represents degree of design detail computed
  - Simple vs. complex models

- Issues for estimation
  - Accuracy
  - Speed
  - Fidelity

Outline

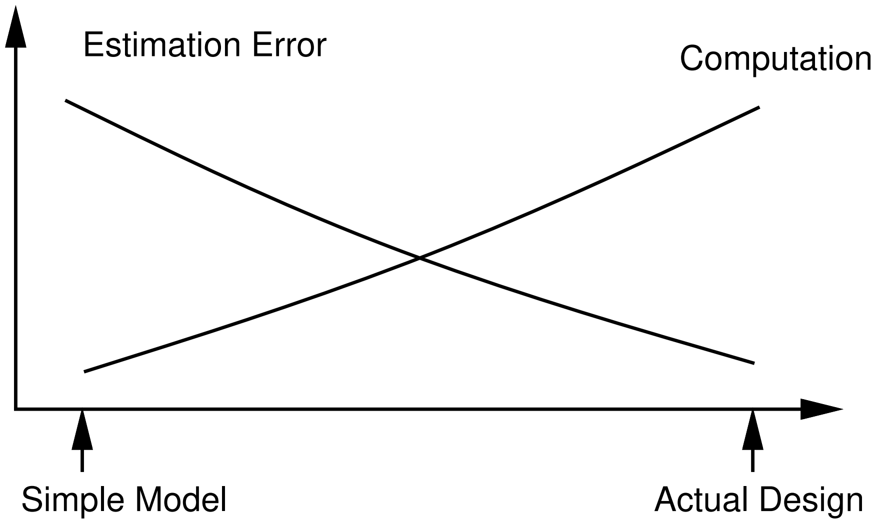
- Accuracy versus speed
- Fidelity
- Quality metrics
  - Performance metrics
  - Hardware and software cost metrics

Accuracy vs. Speed

- Accuracy: difference between estimated and actual value

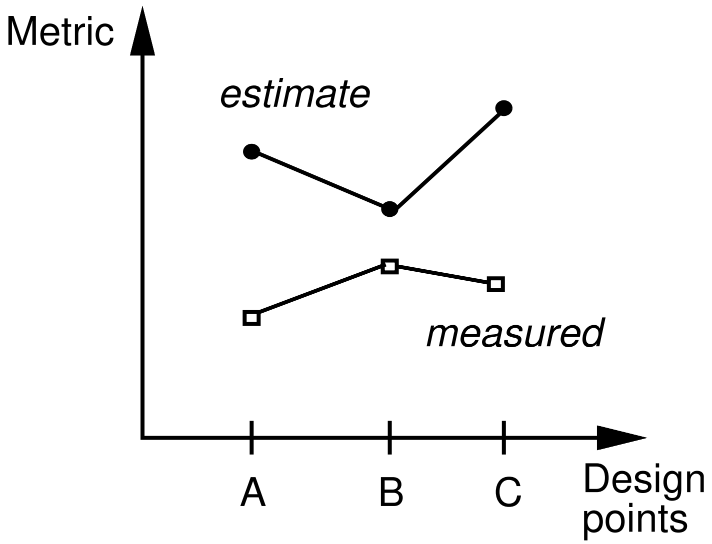
$$\mathcal{A} = 1 - \frac{\mid E(D) - M(D) \mid}{M(D)}$$

- Speed: computation time for obtaining estimate



Fidelity

- Estimates must predict quality metrics for different design alternatives
- Fidelity: % of correct predictions for pairs of design implementations
- Higher fidelity $\Longrightarrow$ correct decisions based on estimates



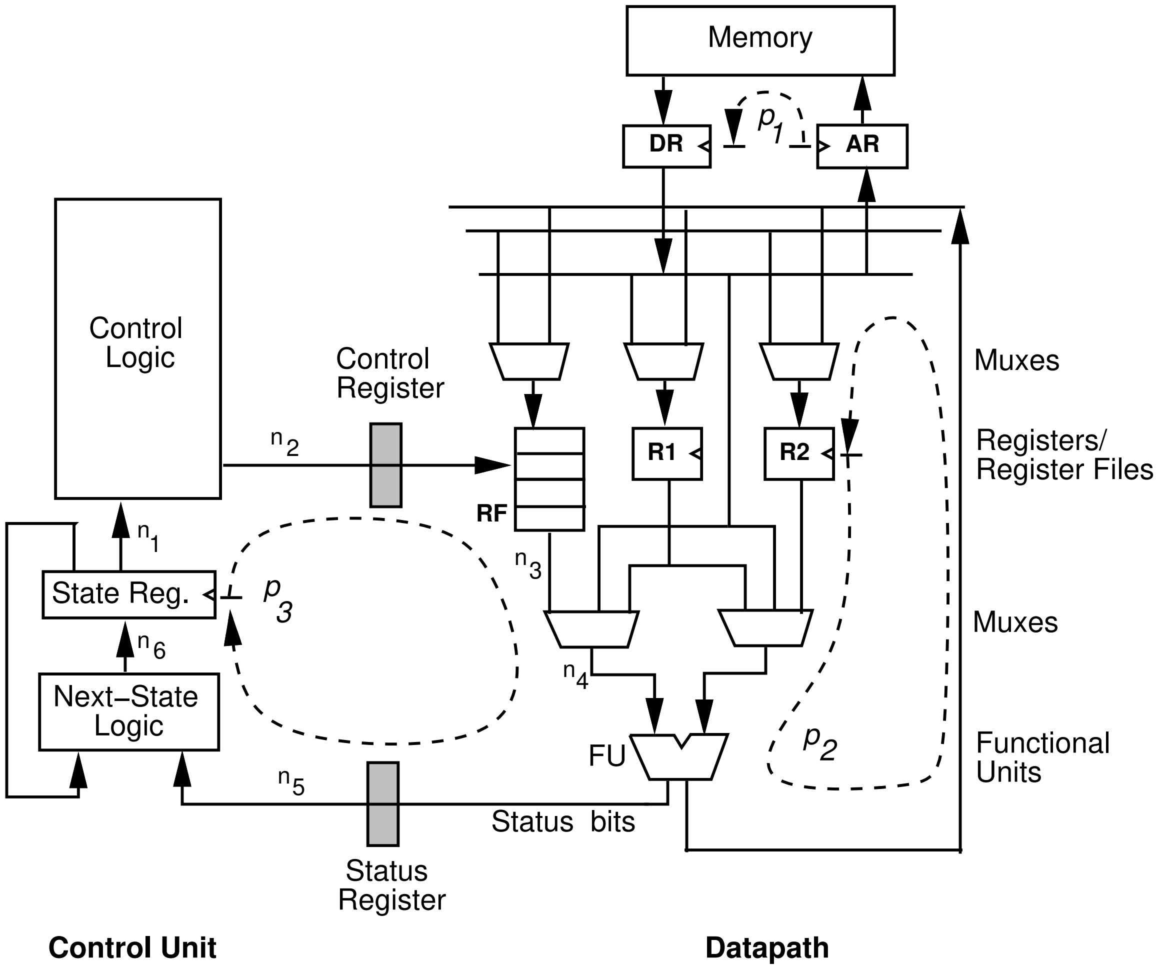
(A, B) =	$E(A) > E(B)$, $M(A) < M(B)$	✗
(B, C) =	$E(B) < E(C)$, $M(B) > M(C)$	✗
(A, C) =	$E(A) < E(C)$, $M(A) < M(C)$	✓

Fidelity = 33 %

Quality metrics

- Performance Metrics
  - Clock cycle, control steps, execution time, communication rates
- Cost Metrics
  - **Hardware**: manufacturing cost (area), packaging cost(pin)
  - **Software**: program size, data memory size
- Other metrics
  - Power, testability, design time, time to market

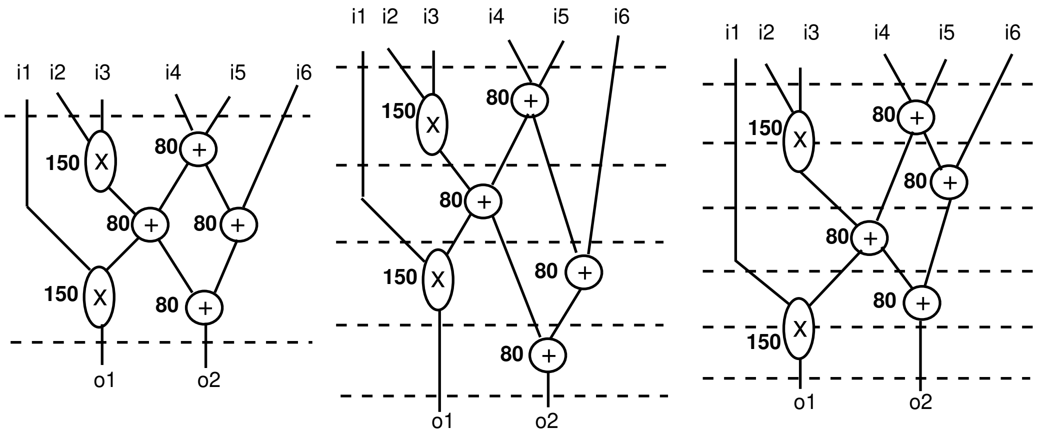
Hardware design model



UC Irvine

Clock cycle estimation

- Clock cycle determines:
  - Resources, execution time
- Determining clock cycle
  - Designer specified [PK89, MK90]
  - Maximum delay of any functional unit [PPM86, JMP88]
  - Clock utilization [NG92]



Clock Cycle : 380 ns
Exec. Time : 380 ns
Resources : 2 x, 4 +

Clock Cycle : 150 ns
Exec. Time : 600 ns
Resources : 1 x, 1 +

Clock Cycle : 80 ns
Exec. Time : 400 ns
Resources : 1 x, 1 +

Control steps estimation

- Operations in the specification assigned to control step

- Number of **control steps** determines:
  - Execution time of design
  - Complexity of control unit

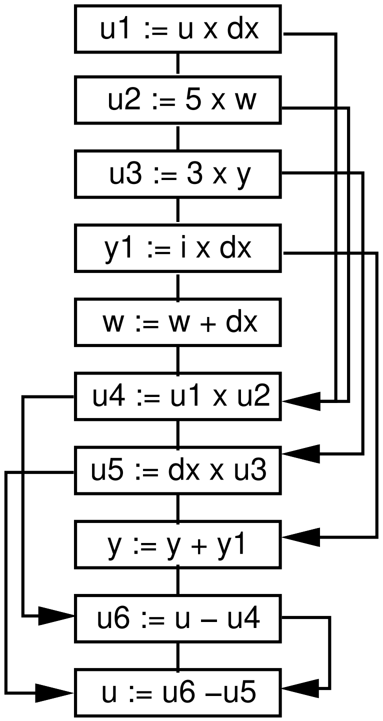
- Scheduling
  - Granularity is operations in a dataflow graph
  - Computationally expensive

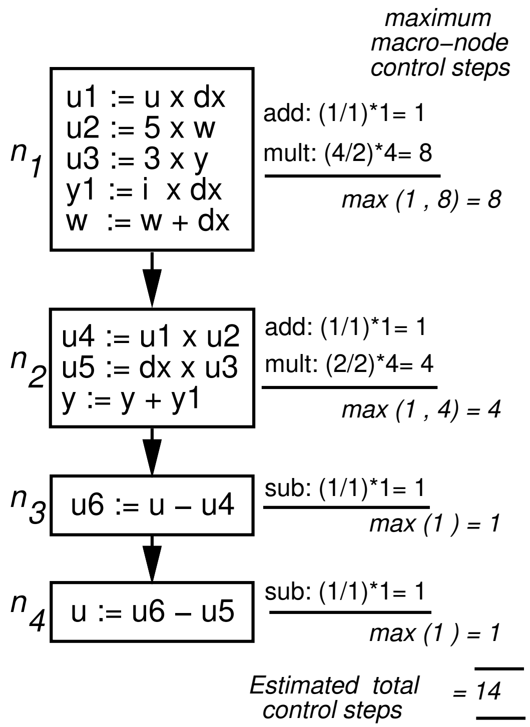
Operator-use method

- Granularity is statements in specification
- Faster than scheduling, average error 13%

t_i	$num(t_i)$	$clocks(t_i)$
add	1	1
mult	2	4
sub	1	1

```
u1 := u  x dx ;
u2 := 5  x w ;
u3 := 3  x y ;
y1 :=  i  x dx ;
w  := w + dx ;
u4 := u1 x u2 ;
u5 := dx x u3 ;
y   := y + y1 ;
u6 := u – u4 ;
u   := u6 – u5 ;
```

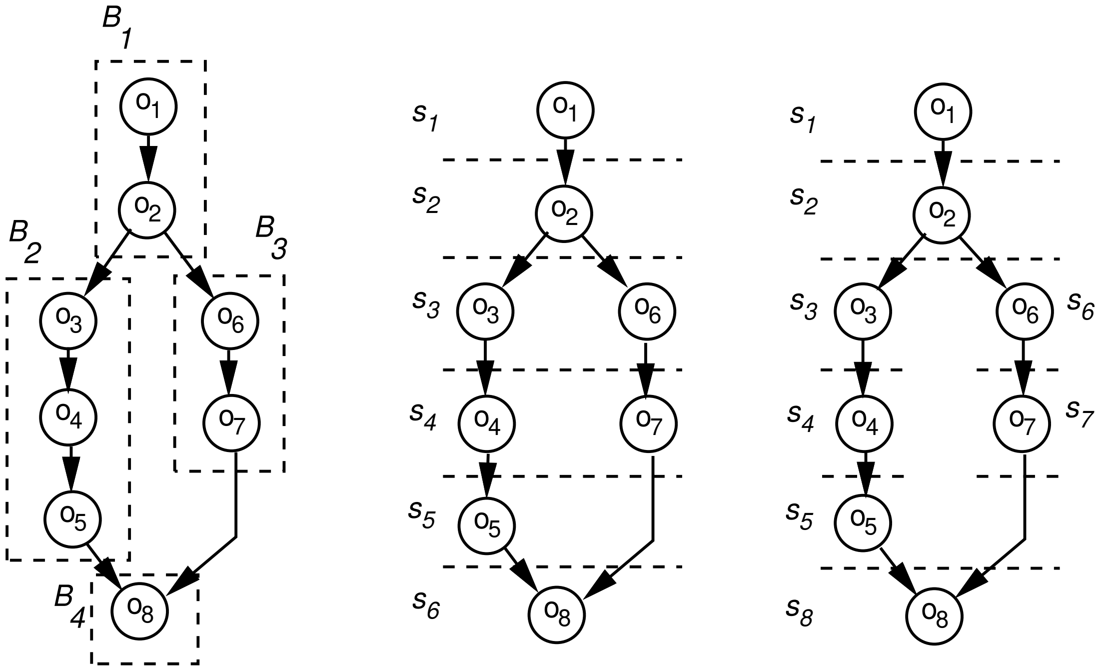




UC Irvine

Branching in behaviors

- Control steps maybe shared across exclusive branches
  - sharing schedule: fewer states, status register
  - non-sharing schedule: more states, no status registers



(a) (b) (c)

Execution time estimation

- Average start to finish time of behavior

- **Straight-line code behaviors**

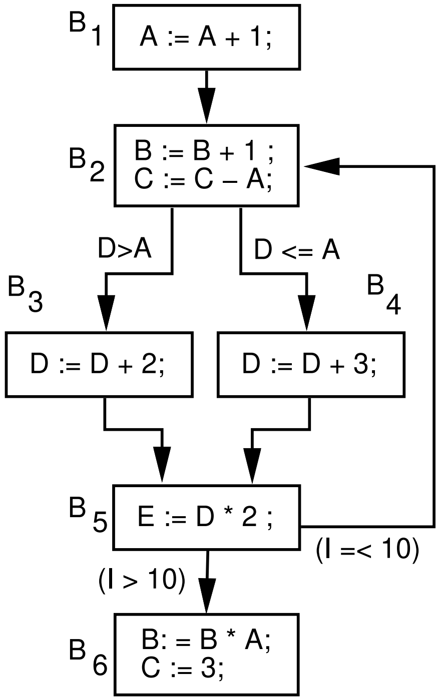
$$exectime(B) \ = \ csteps(B) \ \times \ clk$$

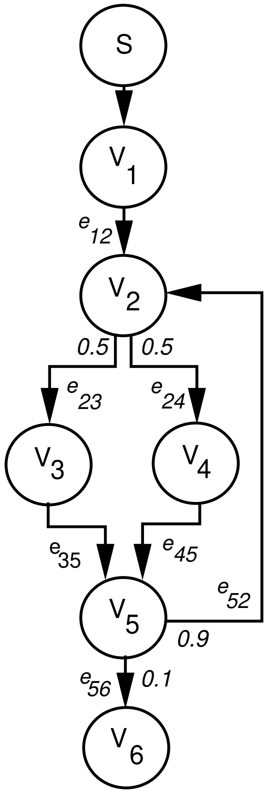
- **Behavior with branching**
  - Estimate execution time for each basic block
  - Create control flow graph from basic blocks
  - Determine branching probabilities
  - Formulate equations for node frequencies
  - Solve set of equations

$$exectime(B) = \sum_{b_i \in B} exectime(b_i) \times freq(b_i)$$

Probability-based flow analysis

```
A := A + 1;

for  I in 1 to 10  loop
     B := B + 1;
     C := C – A;

     if   (D  > A ) then
          D := D + 2;
     else
          D := D + 3;
     end if

     E := D * 2;
end loop;

B := B * A;
C := 3
```





Probability-based flow analysis

- Flow equations:

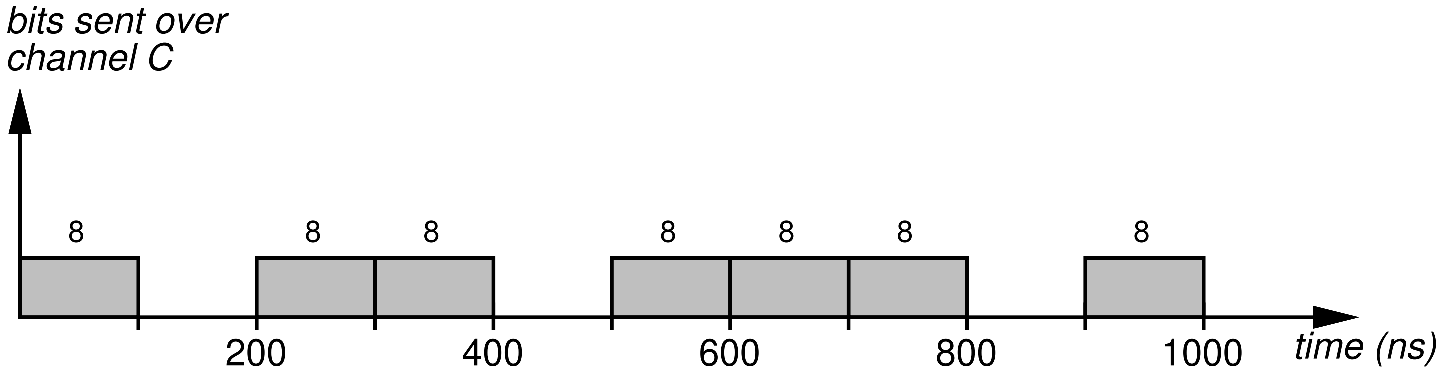
$$
\begin{aligned}
freq(S) &= 1.0 \\
freq(v_1) &= 1.0 \times freq(S) \\
freq(v_2) &= 1.0 \times freq(v_1) + 0.9 \times freq(v_5) \\
freq(v_3) &= 0.5 \times freq(v_2) \\
freq(v_4) &= 0.5 \times freq(v_2) \\
freq(v_5) &= 1.0 \times freq(v_3) + 1.0 \times freq(v_4) \\
freq(v_6) &= 0.1 \times freq(v_5)
\end{aligned}
$$

- Node execution frequencies:

$$
\begin{aligned}
freq(v_1) &= 1.0 & freq(v_2) &= 10.0 \\
freq(v_3) &= 5.0 & freq(v_4) &= 5.0 \\
freq(v_5) &= 10.0 & freq(v_6) &= 1.0
\end{aligned}
$$

- Can be used to estimate number of accesses to variables, channels or procedures

Communication rates



- **Average channel rate**

  rate of data transfer over lifetime of behavior

$$averate(C) = \frac{56\ bits}{1000\ ns} = 56\ Mb/s$$

- **Peak channel rate**

  rate of data transfer of single message

$$peakrate(C) = \frac{8\ bits}{100\ ns} = 80\ Mb/s$$

Communication rate estimation

- Total behavior execution time consists of
  - Computation time, $comptime(P)$, obtained from flow-analysis
  - Communication time, $commtime(P, C) \ = \ access(P, C) \times delay(C)$

- Total bits transferred by the channel,

$$total\_bits(P, C) \ = \ access(P, C) \ \times \ bits(C)$$

- Channel average rate

$$averate(C) \ = \ \frac{total\_bits(B, C)}{comptime(B) + commtime(B, C)}$$

- Channel peak rate

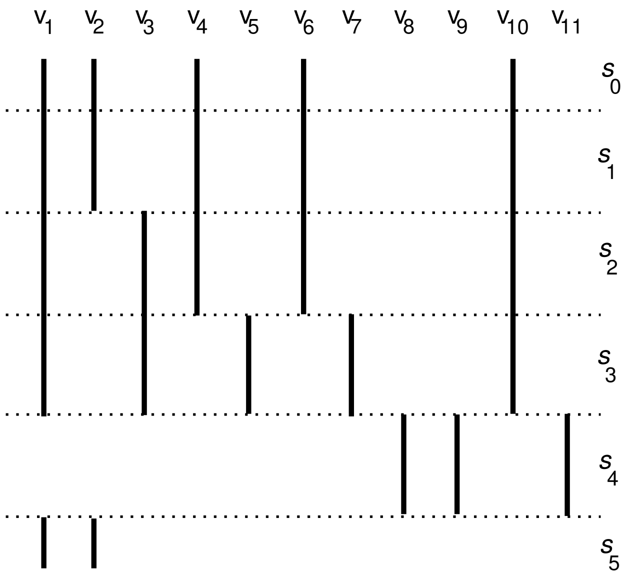
$$peakrate(C) = \frac{bits(C)}{protocol\_delay(C)}$$

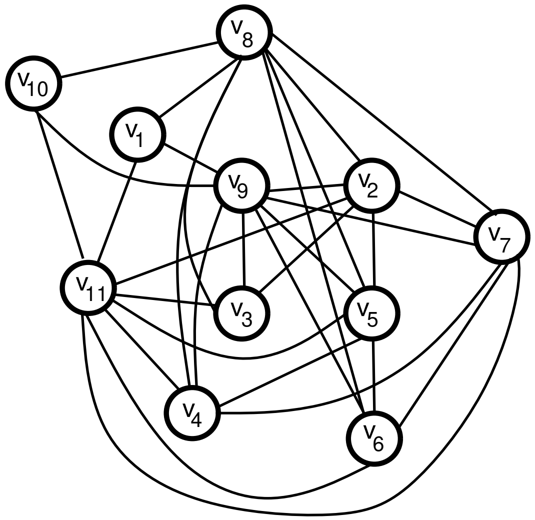
Area estimation

- Two tasks:
  - Determining number and type of components required
  - Estimating component size for a specific technology (FSMD, gate arrays etc.)
- Behavior implemented as a FSMD (finite state machine with datapath)
  - Datapath components: registers, functional units, multiplexers/buses
  - Control unit: state register, control logic, next-state logic
- We will discuss
  - Datapath component estimation
  - Control unit estimation
  - Layout area for a custom implementation

Storage-unit estimation

- Variables not used concurrently maybe mapped same storage-unit
- To use clique-partitioning, construct a graph where
  - Each variable represented by a vertex
  - Variables with non-overlapping lifetimes have an edge between] their vertices



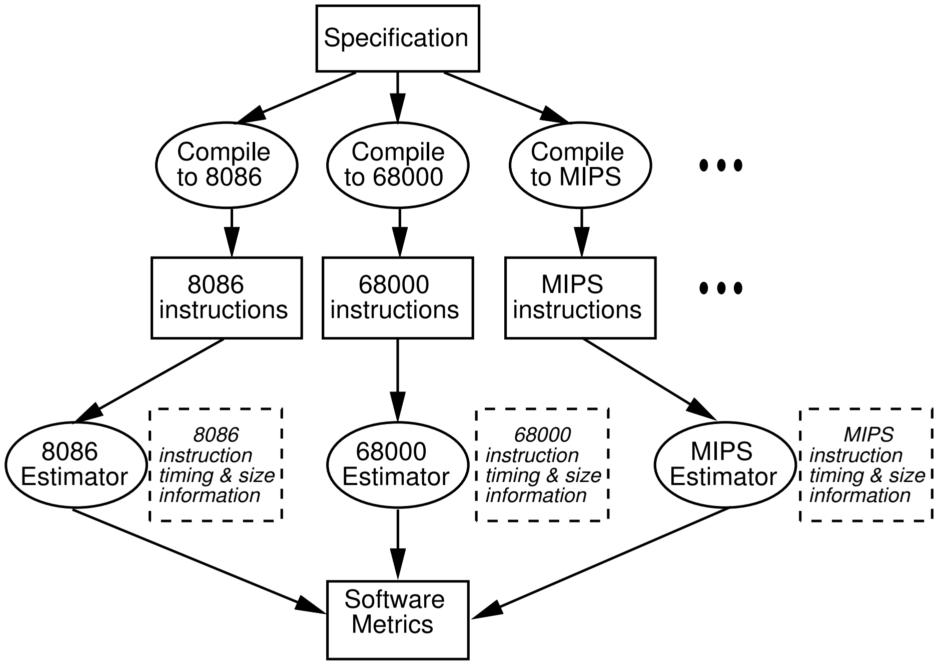


Cliques		Storage unit
$\{v_2, v_3\}$	=	R_1
$\{v_6, v_7, v_9\}$	=	R_2
$\{v_4, v_5, v_8\}$	=	R_3
$\{v_{10}, v_{11}\}$	=	R_4
$\{v_1\}$	=	R_5

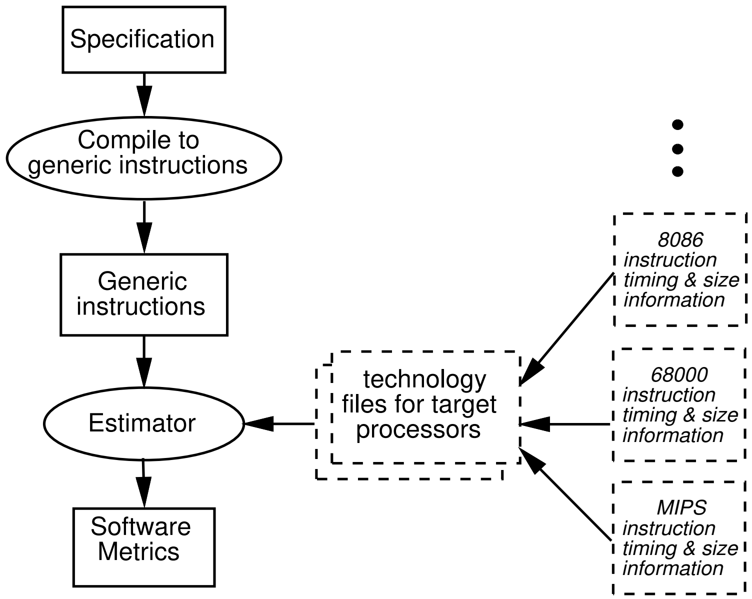
Functional-unit and interconnect-unit estimation

- Clique-partitioning can be applied
- For determining the number of FU's required, construct a graph where
  - Each operation in behavior represented by a vertex
  - Edge connects two vertices if
    - Corresponding operations assigned different control steps
    - There exists an FU that can implement both operations
- For determining the number of interconnect units, construct a graph where
  - Each connection between two units is represented by a vertex
  - Edge connects two vertices if corresponding connections not used in same control step

Software estimation models



Processor specific model



Generic model

Deriving processor technology files

Generic instruction

dmem3 = dmem1 + dmem2

8086 instructions

instruction	*clocks*	*bytes*
mov ax, word ptr[bp+offset1]	(10)	3
add ax, word ptr[bp+offset2]	(9 + EA1)	4
mov word ptr[bp+offset3], ax	(10)	3

68020 instructions

instruction	*clocks*	*bytes*
mov a6@(offset1), d0	(7)	2
add a6@(offset2), d0	(2 + EA2)	2
mov d0, a6@(offset3)	(5)	2

technology file for 8086

generic instruction	*execution time*	*size*
...		
dmem3 = dmem1 + dmem2	35 clocks	10 bytes
...		

technology file for 68020

generic instruction	*execution time*	*size*
...		
dmem3 = dmem1 + dmem2	22 clocks	6 bytes
...		

Software estimation

- **Program execution time**
  - Create basic blocks and compile into generic instructions
  - Estimate execution time of basic blocks
  - Perform probability-based flow analysis
  - Compute execution time of the entire behavior:

$$exectime(B) = \delta \times \left(\sum_{b_i \in B} exectime(b_i) \times freq(b_i) \right)$$

δ accounts for compiler optimizations

- **Program memory size**

$$progsize(B) = \sum_{g \in G} instr\_size(g)$$

- **Data memory size**

$$datasize(B) = \sum_{d \in D} datasize(d)$$

Summary and future directions

- We described methods for estimating:
  - Performance metrics: clock, control steps, execution time, communication rates
  - Cost metrics: design area, pins, program and data memory size
- Future directions:
  - Incorporating synthesis/compilation optimizations
  - New metrics for testability, power, integration cost, etc.
  - New architectural features for the estimation model

Refinement

- Functional objects are grouped and mapped to system components
  - Functional objects: variables, behaviors, and channels
  - System components: memories, chips or processors, and buses
- **Refinement** is update of specification to reflect mapping
- Need for refinement
  - Makes specification consistent
  - Enables simulation of specification
  - Generate input for synthesis, compilation and verification tools

Outline

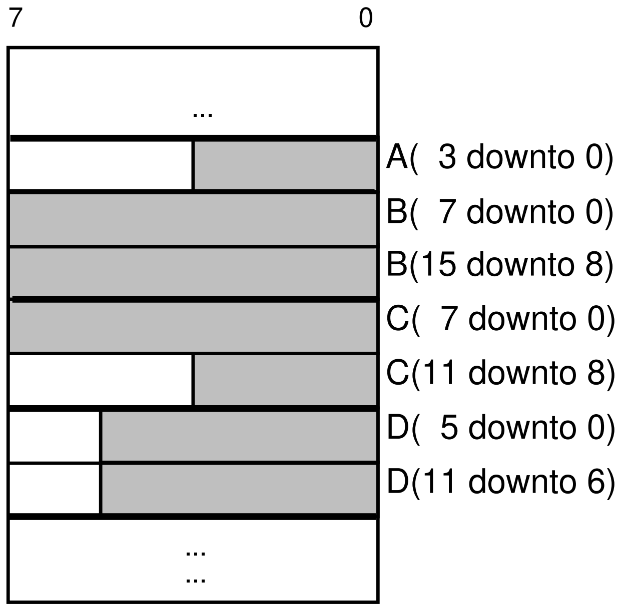
- Refining variable groups
- Channel refinement
- Resolving access conflicts
- Refining incompatible interfaces

Refining variable groups

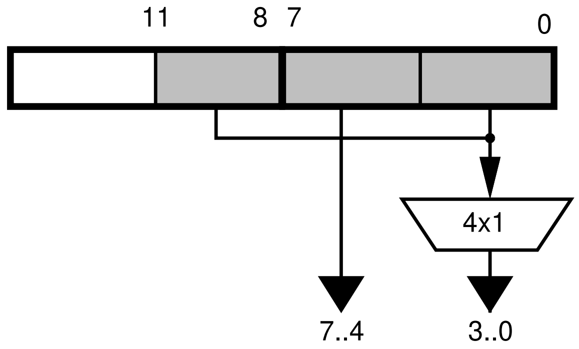
- Group of variables mapped to a memory
- Variable folding:
  - Implementing each variable in a memory with a fixed word size
- Memory address translation
  - Assignment of addresses to each variable in group
  - Update references to variable by accesses to memory

Variable folding

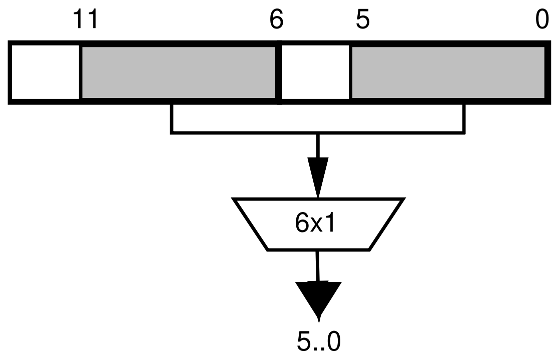
```
variable  A :  bit_vector( 3 downto 0) ;
variable  B :  bit_vector(15 downto 0) ;
variable  C :  bit_vector(11 downto 0) ;
variable  D :  bit_vector(11 downto 0) ;
```



8−bit Memory



to variable C in memory



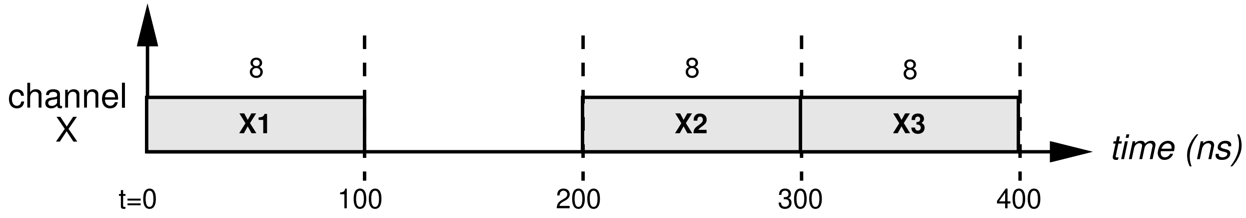
to variable D in memory

Refining channel groups

- Channels are virtual entities over which messages are transferred
- Bus is a physical medium that implements groups of channels
- Bus consists of:
  - wires representing data and control lines
  - protocol defining sequence of assignments to data and control lines
- Two refinement tasks
  - *Bus generation*: determining buswidth i.e. number of data lines
  - *Protocol generation*: specifying mechanism of transfer over bus

Characterizing communication channels

- For a given behavior P that sends data over channel C,
  - **Message size**, $bits(C)$: number of bits in each message
  - **Accesses**, $accesses(P, C)$: number of times P transfers data over C
  - **Average rate**, $averate(C)$: rate of data transfer of C over lifetime of behavior
  - **Peak rate**, $peakrate(C)$: rate of transfer of single message



$$bits(C) = 8 \text{ bits}$$

$$averate(C) = \frac{24\ bits}{400\ ns} = 60\ Mbits/s$$

$$peakrate(C) = \frac{8\ bits}{100\ ns} = 80\ Mbits/s$$

Characterizing buses

- For a given bus B,
  - **Buswidth** , $buswidth(B)$: number of data lines in B
  - **Protocol delay**, $protdelay(B)$: delay for single message transfer over bus
  - **Average rate**, $averate(B)$: rate of data transfer over B over lifetime of system
  - **Peak rate**, $peakrate(B)$: maximum rate of transfer of data on bus

$$peakrate(C) = \frac{buswidth(B)}{protdelay(B)}$$

Determining bus rates

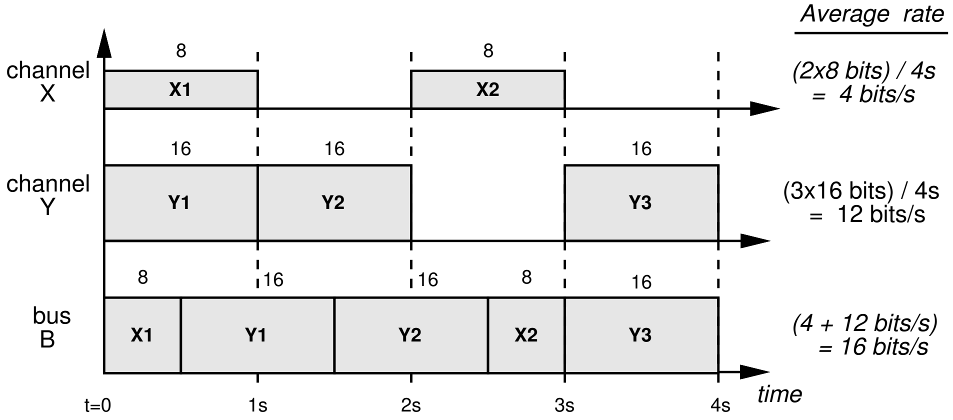
- Idle slots of a channel used for messages of other channels

- To ensure that channel average rates are unaffected by bus

$$averate(B) \geq \sum_{C \in B} averate(C)$$

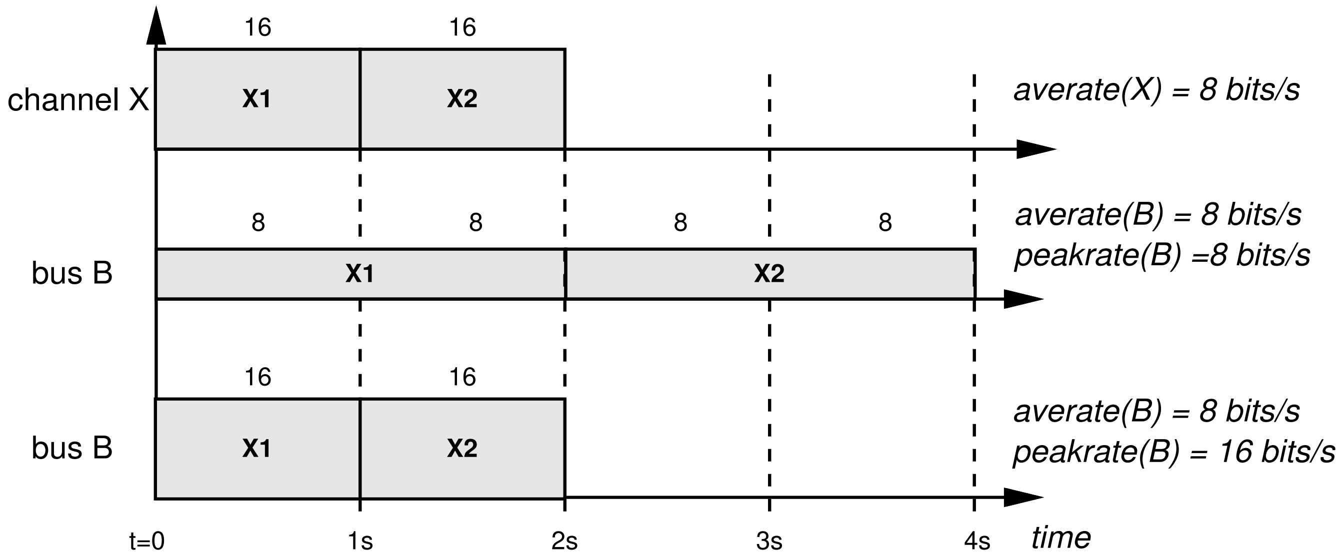
- Goal: to synthesize a bus that *constantly* transfers data i.e.

$$peakrate(B) = averate(C)$$



Constraints for bus generation

- **Buswidth**: affects number of pins on chip boundaries
- **Channel average rates**: affects execution time of behaviors
- **Channel peak rates**: affects time required for single message transfer



Bus generation algorithm

- **Compute buswidth range**: $minwidth = 1,\ maxwidth = \text{Max}(bits(C))$

- **For** $minwidth \leq currwidth \leq maxwidth$ **loop**
  - **Compute bus peak rate**:

    $$peakrate(B) = currwidth \div protdelay(B)$$

  - **Compute channel average rates**

    $$commtime(P) = access(P,C) \times [\ \lceil \frac{bits(C)}{currwidth} \rceil \times protdelay(B)\ ]$$

    $$averate(C) = \frac{access(P,C) \times bits(C)}{comptime(P) + commtime(P)}$$

  - **if** $peakrate(B) \geq \sum_{C \in B} averate(C)$ **then**

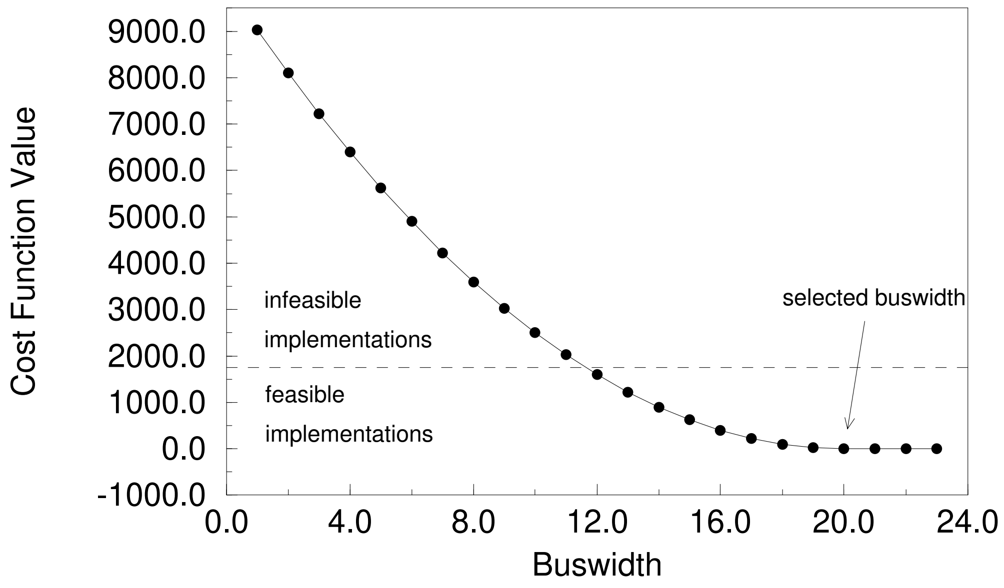
    if $bestcost > ComputeCost(currwidth)$ **then**

    $bestcost = ComputeCost(currwidth)$

    $bestwidth = currwidth$

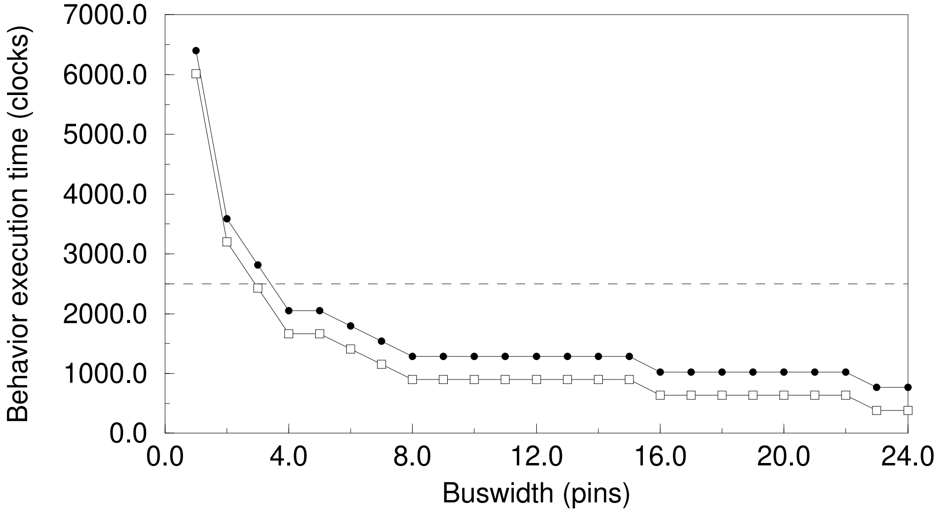
Bus generation example

- 2 behavior accessing 16 bit data over two channels
- Constraints specified for channel peak rates



Performance vs. buswidth tradeoffs

- Allows a buswidth to be selected, given performance constraints
  e.g. behavior $P1$ has performance constraint of 2500 clocks.
  buswidths of 4 or greater must be selected



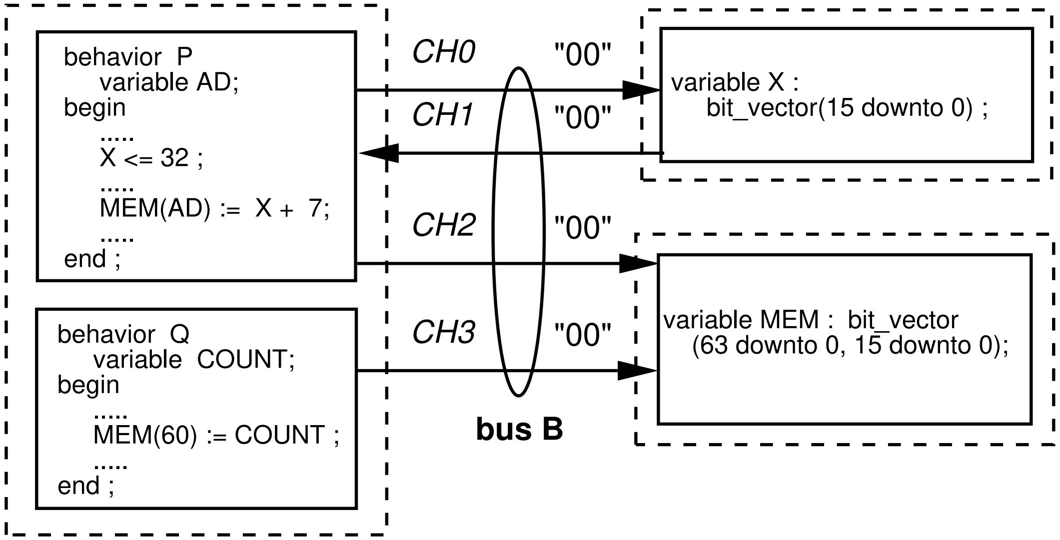
UC Irvine

Protocol generation

- Bus consists of several sets of wires:
  - **Data lines**, used for transferring message bits
  - **Control lines**, used for synchronization between behaviors
  - **ID lines**, used for identifying the channel active on the bus
- All channels mapped to bus share these lines
- Number of data lines determined by bus generation algorithm
- Protocol generation consists of six steps

Protocol generation

1. **Protocol selection**: full handshake, half-handshake etc.
2. **ID assignment**: N channels require $log_2(N)$ ID lines



UC Irvine

Protocol generation

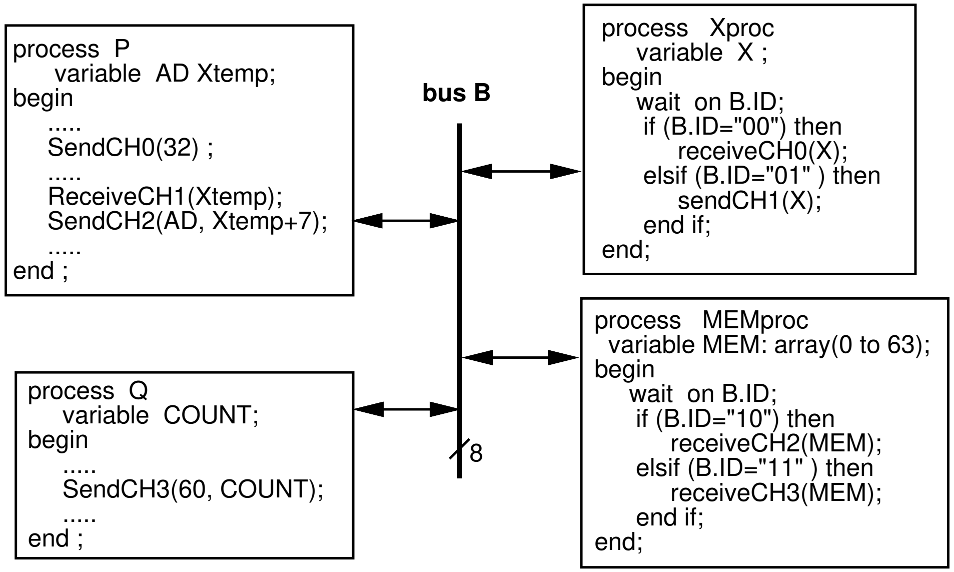
3. **Bus structure definition**

4. **Bus protocol definition**

```
type  HandShakeBus is record
        START, DONE : bit ;
        ID : bit_vector(1 downto 0) ;
        DATA  : bit_vector(7 downto 0) ;
end record ;

signal  B : HandShakeBus ;

procedure  ReceiveCH0( rxdata : out bit_vector)  is
begin
    for J in 1 to 2 loop
        wait  until  (B.START = '1') and (B.ID = "00") ;
        rxdata (8*J–1  downto 8*(J–1)) <= B.DATA ;
        B.DONE <= '1' ;
        wait until  (B.START = '0') ;
        B.DONE <= '0' ;
    end loop;
end ReceiveCH0;

procedure  SendCH0( txdata : in bit_vector)  is
begin
    bus B.ID <= "00" ;
    for J in 1 to 2 loop
        B.data <=  txdata(8*J–1  downto  8*(J–1)) ;
        B.START <= '1' ;
        wait until  (B.DONE = '1') ;
        B.START <= '0' ;
        wait until  (B.DONE = '0') ;
    end loop;
end SendCH0;
```

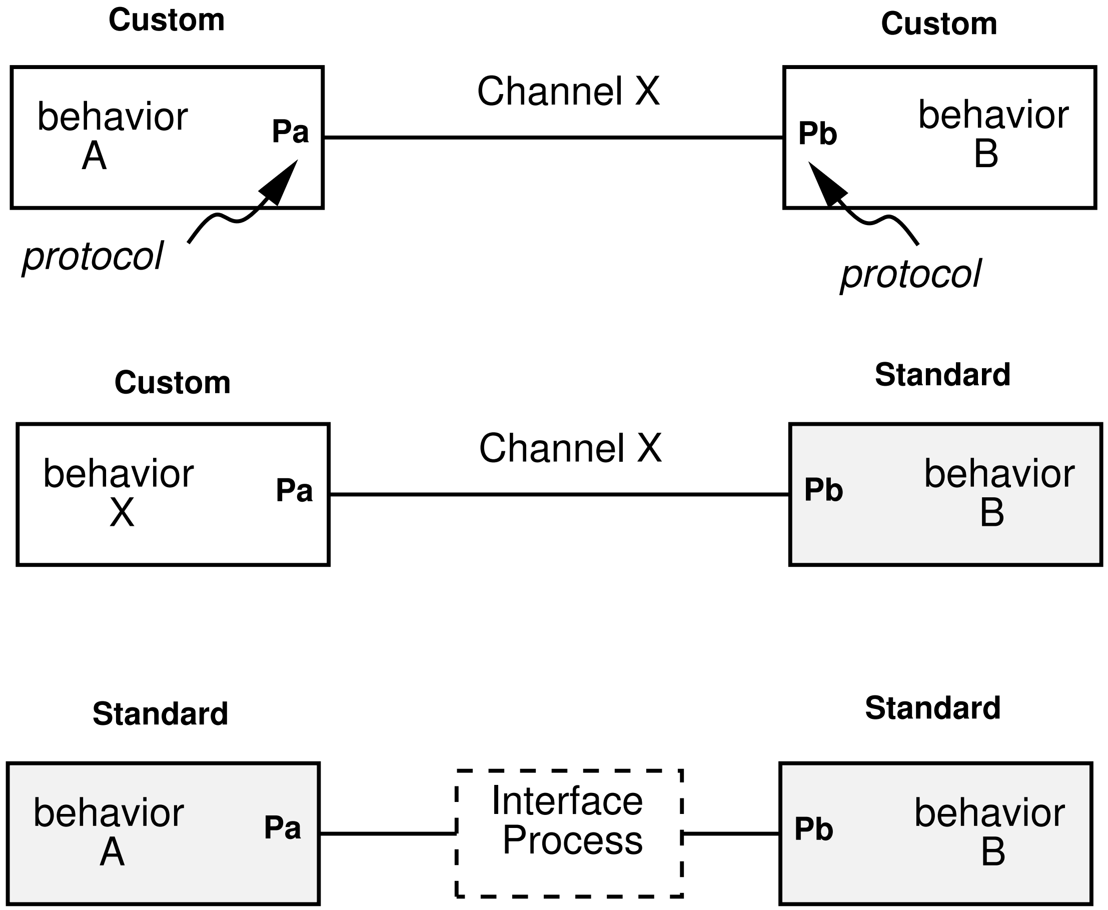
Protocol generation

5. **Update variable references**
6. **Generate behaviors for variables**



UC Irvine

Effect of binding on interfaces



Protocol operations

- Protocols usually consist of five *atomic* operations
  - waiting for an event on input control line
  - assigning value to output control line
  - reading value from input data port
  - assigning value to output data port
  - waiting for fixed time interval
- Protocol operations may be specified in one of three ways
  - Finite state machines (FSMs)
  - Timing diagrams
  - Hardware description languages (HDLs)

Interface process generation

- Input: HDL description of two fixed, but incompatible protocols
- Output: HDL process that translates one protocol to the other
  - i.e. responds to their control signals and sequence their data transfers
- Four steps required for generating interface process (IP):
  - Creating relations
  - Partitioning relations into groups
  - Generating interface process statements
  - interconnect optimization

IP generation: creating relations

- Protocol represented as an ordered set of relations
- Relations are sequences of events/actions

Protocol Pa

```
ADDRp <=  AddrVar(7 downto 0);
ARDYp <= '1';
wait until (ARCVp = '1' );
ADDRp <= AddrVar(15 downto 8);
DREQp <= '1';
wait  until (DRDYp = '1');
DataVar <=  DATAp;
```

→

Relations

```
A1  [ (true)   :
          ADDRp <=  AddrVar(7 downto 0)
          ARDYp <= '1'  ]
A2  [ (ARCVp = '1')  :
          ADDRp <=  AddrVar(15 downto 8)
          DREQp <= '1'  ]
A3  [ (DRDYp = '1')  :
          DataVar <=  DATAp  ]
```

IP generation: partitioning relations

- Partition the set of relations from both protocols into groups.
- Group represents a unit of data transfer

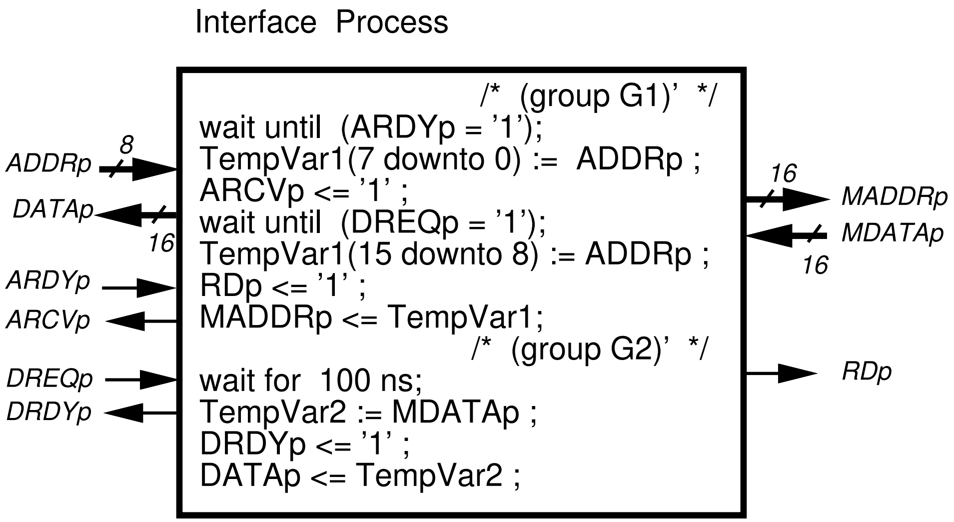
Protocol Pa		Protocol Pb	
A1 (8 bits out) **A2** (8 bits out)	→	**B1** (16 bits in)	*G1*
A3 (16 bits in)	←	**B2** (16 bits out)	*G2*

$$G_1 = (A_1 \; A_2 \; B_1) \qquad G_2 = (B_1 \; A_3)$$

IP generation: inverting protocol operations

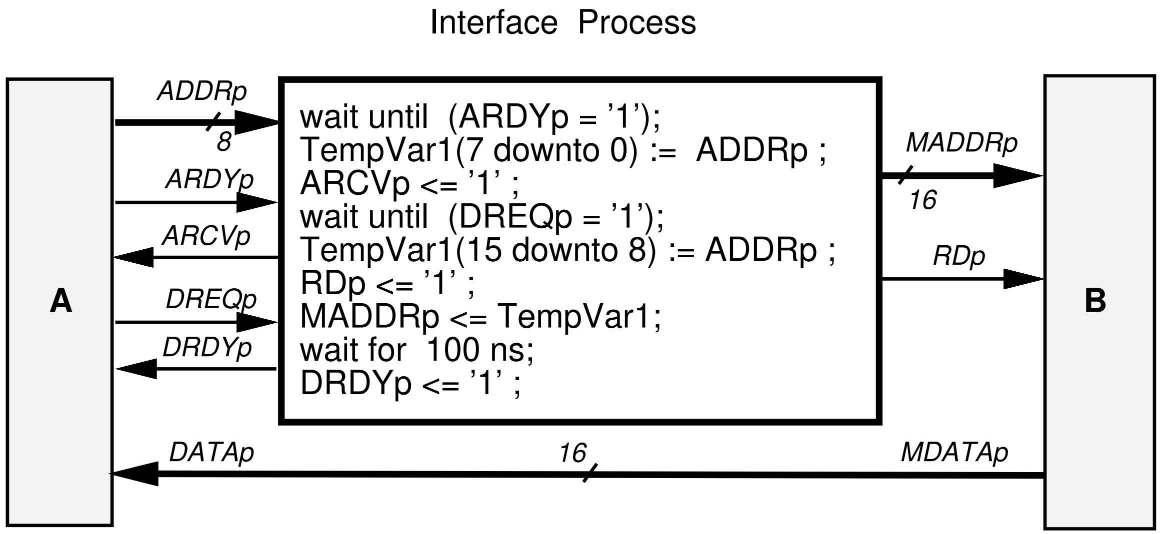
- For each operation in a group, add its *dual* to interface process
- Dual of an operation represents the complementary operation
- Temporary variable may be required to hold data values

Atomic operation	*Dual operation*
wait until (Cp = '1')	Cp <= '1'
Cp <= '1'	wait until (Cp = '1')
var <= Dp	Dp <= TempVar
Dp <= var	TempVar := Dp
wait for 100 ns	wait for 100 ns



IP generation: interconnect optimization

- Certain ports of both protocols may be directly connected
- Advantages:
  - Bypassing interface process reduces interconnect cost
  - Operations related to these ports can be eliminated from interface process

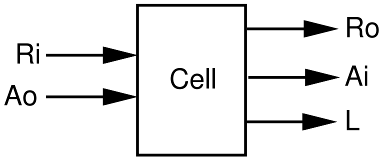


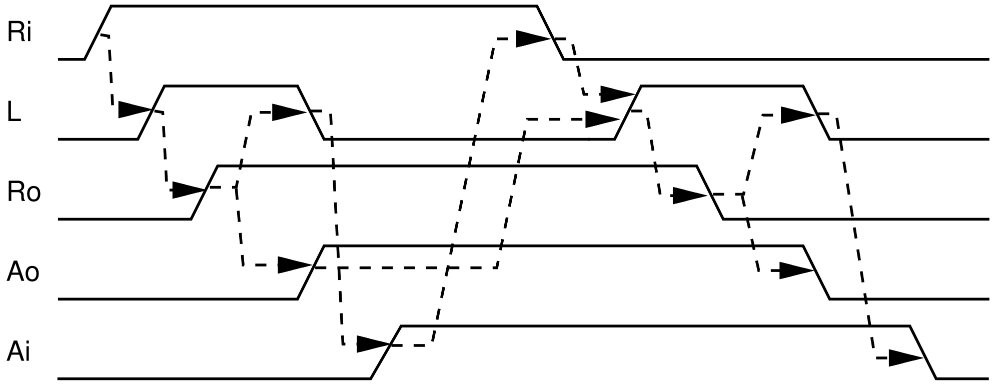
Transducer synthesis [BK87]

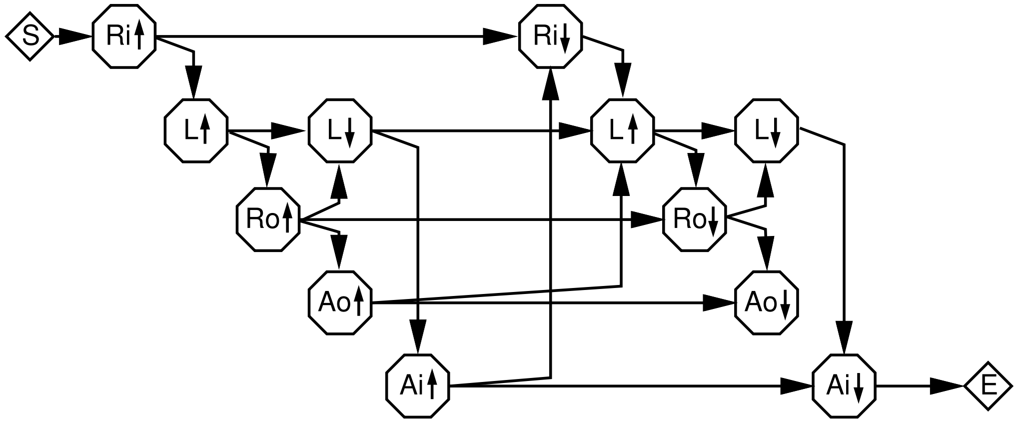
- Input: Timing diagram description of two fixed protocols
- Output: Logic circuit description of transducer
- Steps for generating logic circuit from timing diagrams:
  - Create event graphs for both protocols
  - Connect graphs based on data dependencies or explicitly specified ordering
  - Add templates for each output node in combined graph
  - Merge and connect templates
  - Satisfy min/max timing constraints
  - Optimize skeletal circuit

Generating event graphs from timing diagrams

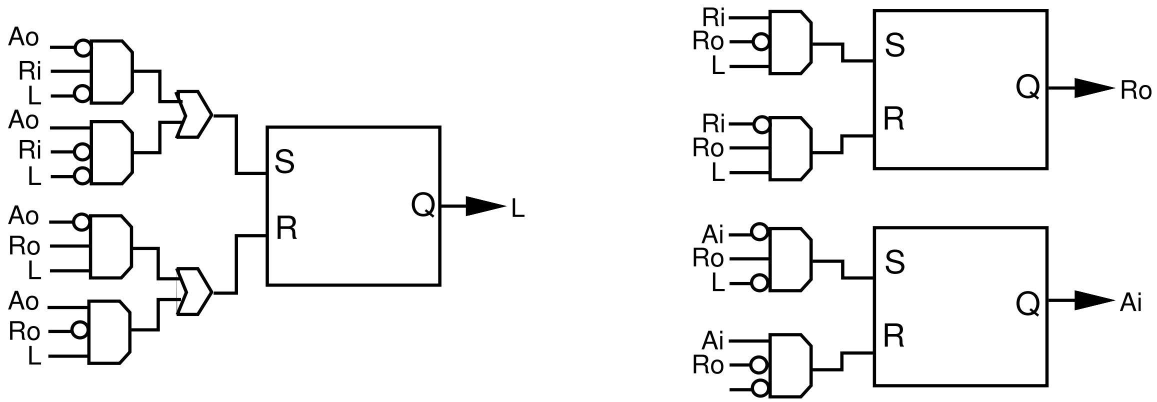
e.g. FIFO stack control cell







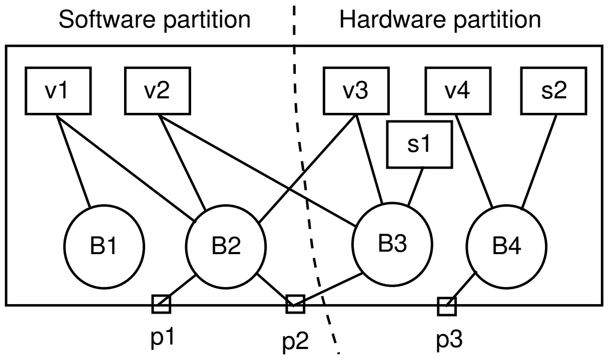
Deriving skeletal circuit from event graph



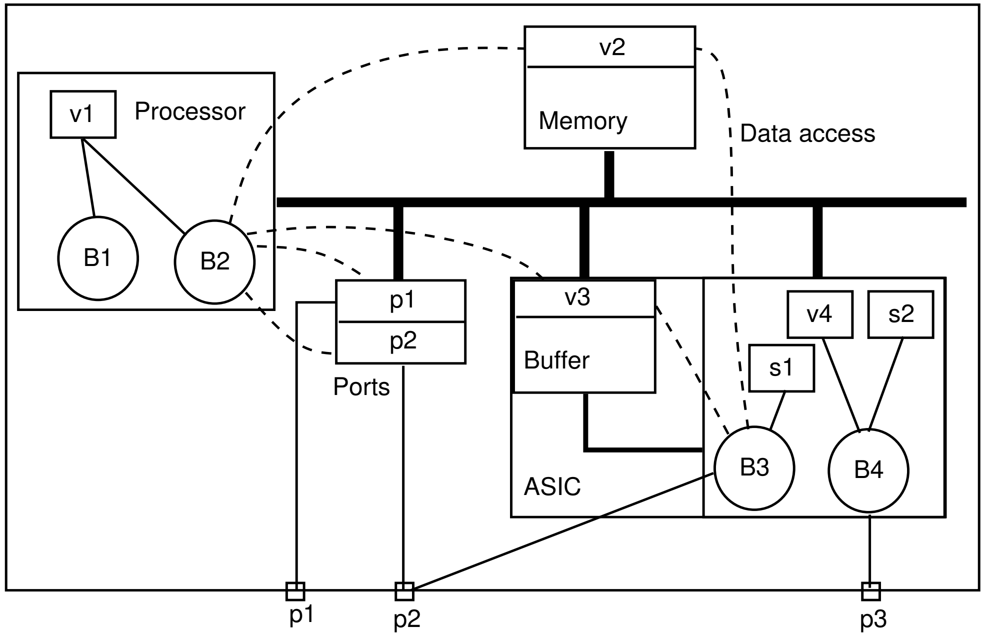
- Advantages:
  - Synthesizes logic for transducer circuit directly
  - Accounts for min/max timing constraints between events

- Disadvantages:
  - Cannot interface protocols with different data port sizes
  - Transducer not simulatable with timing diagram description of protocols

Hardware/Software interface refinement



(a) Partitioned specification



(b) Mapping to architecture

Tasks of hardware/software interfacing

- Data access (e.g., behavior accessing variable) refinement
- Control access (e.g., behavior starting behavior) refinement
- Select bus to satisfy data transfer rate and reduce interfacing cost
- Interface software/hardware components to standard buses
- Schedule software behaviors to satisfy data input/output rate
- Distribute variables to reduce ASIC cost and satisfy performance

Summary and future directions

- In this section, we described:
  - Refinement of variable groups: variable folding, address translation
  - Refinement of channel groups: bus and protocol generation
  - Resolution of access conflicts: arbiter generation
  - Refinement of incompatible interfaces: IP generation, transducer synthesis

- Future work should address the following issues:
  - Effects of bus arbitration delays on performance of a behavior
  - Developing metrics to guide selection of protocols and arbitration schemes
  - Efficient synthesis of arbiter and interface processes

Methodology

- Past design effort focused on lower levels
- Higher levels lack well-defined methodology and tools
- Paradigm shift to higher levels can increase productivity
- Need methodology and tools for system level

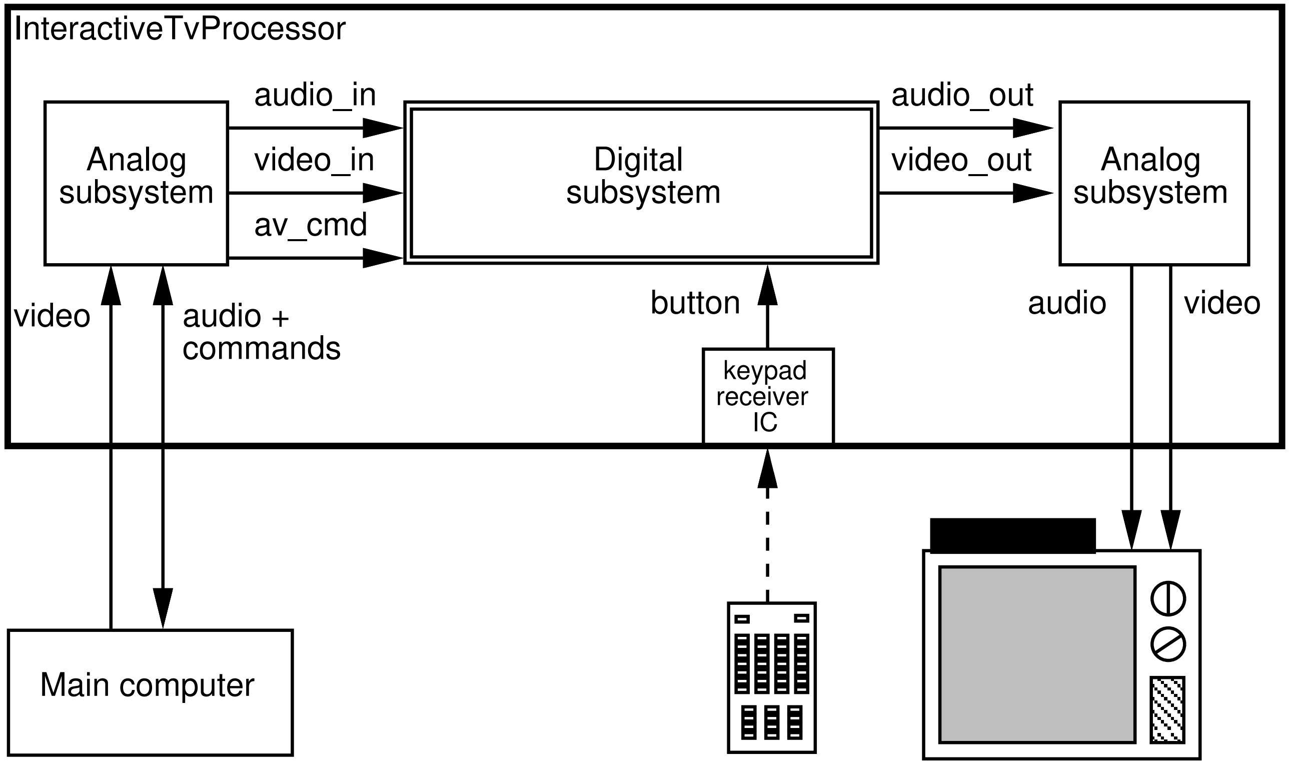
Outline

- Basic concepts in design methodology
- Example
- A design methodology
- A generic synthesis system
- Conceptualization environment

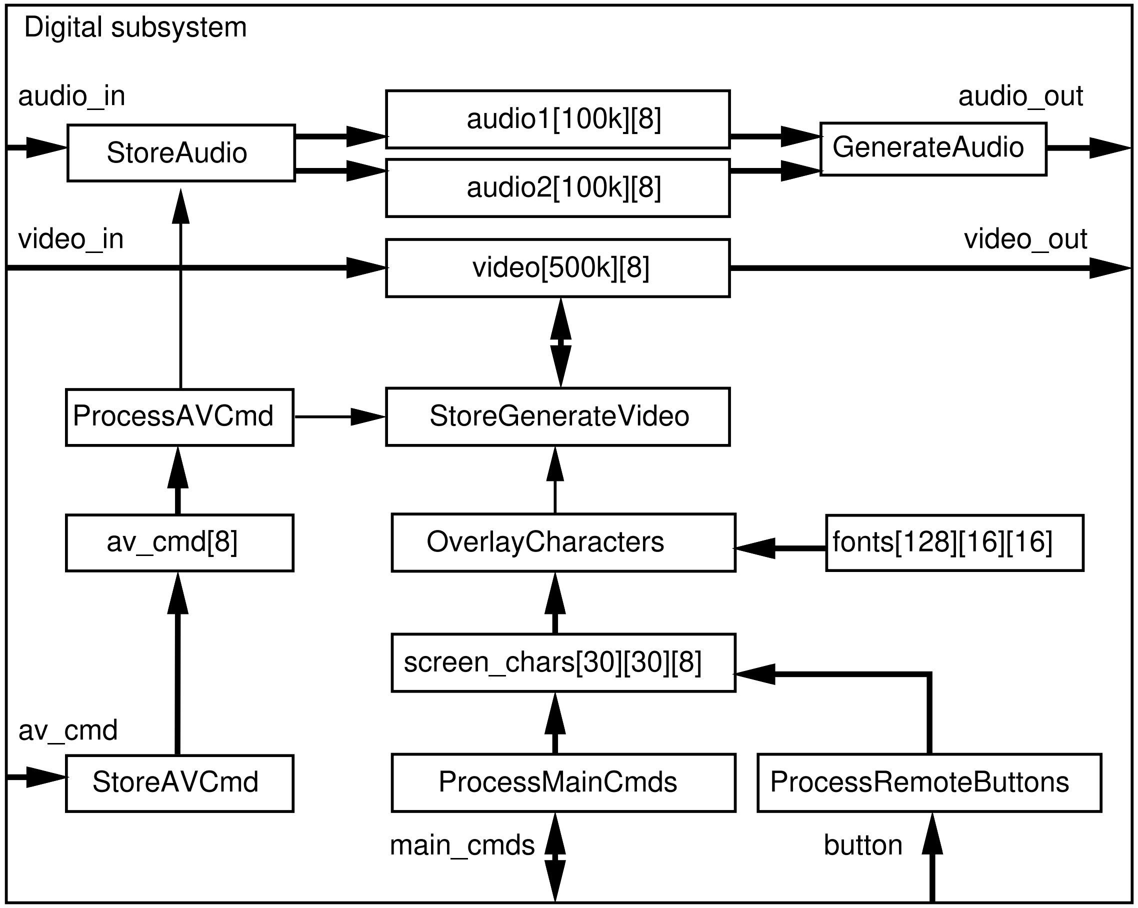
Items a design methodology must specify

- Syntax and semantics of input and output
- Algorithms for transforming input to output
- Components to be used in the design implementation
- Definition and ranges of constraints
- Mechanism for selection of architectural styles
- Control strategies (scenarios or scripts)

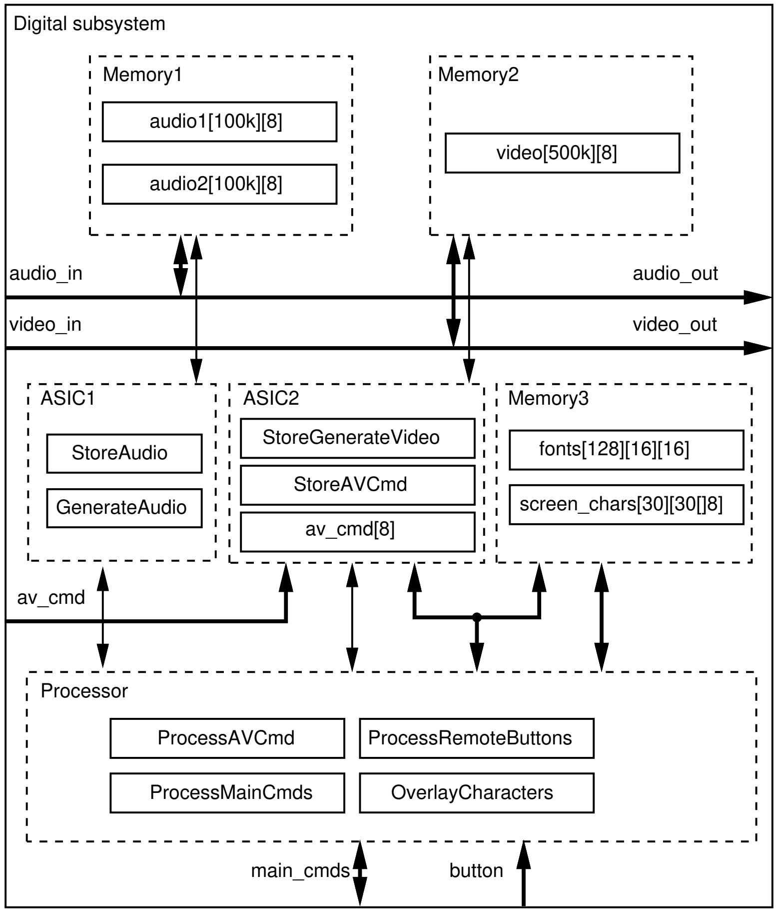
Example: Interactive TV processor



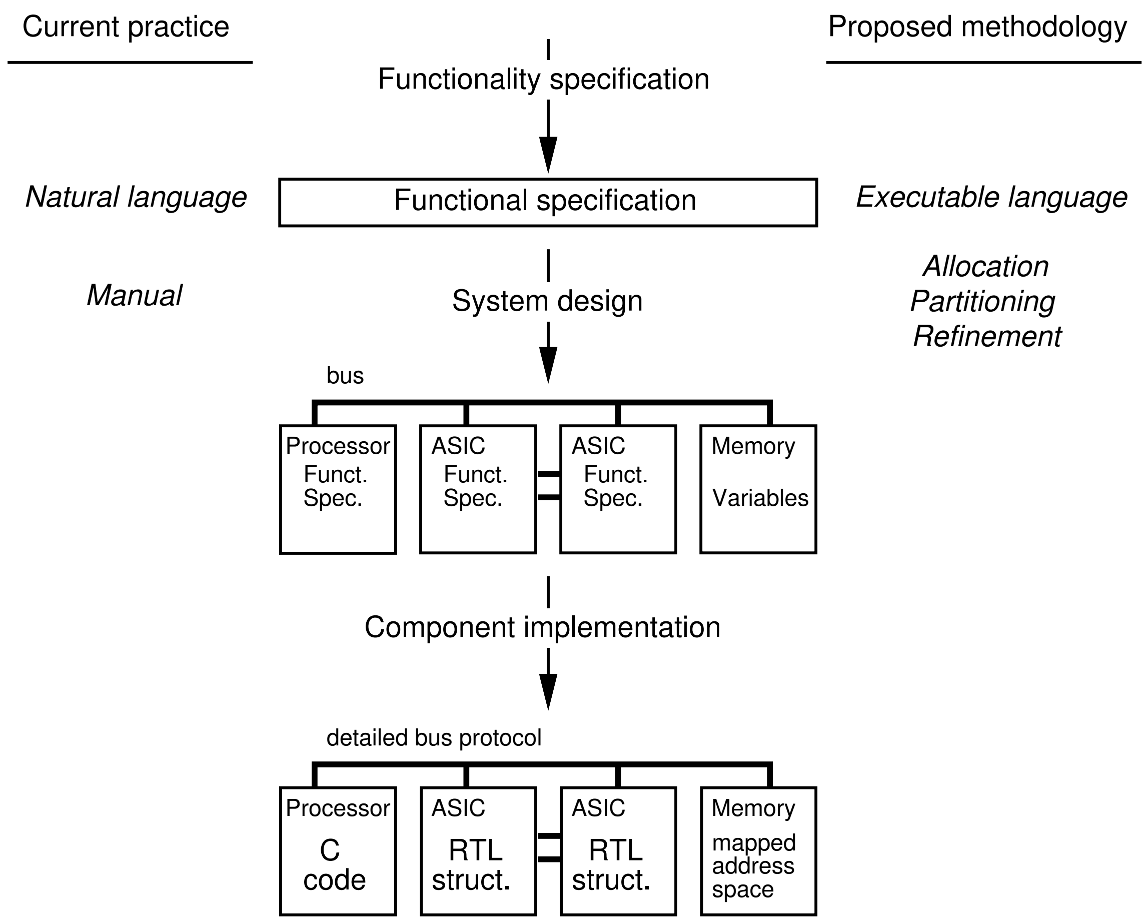
Example's dataflow behavior



Example's implementation after system design



An example design methodology

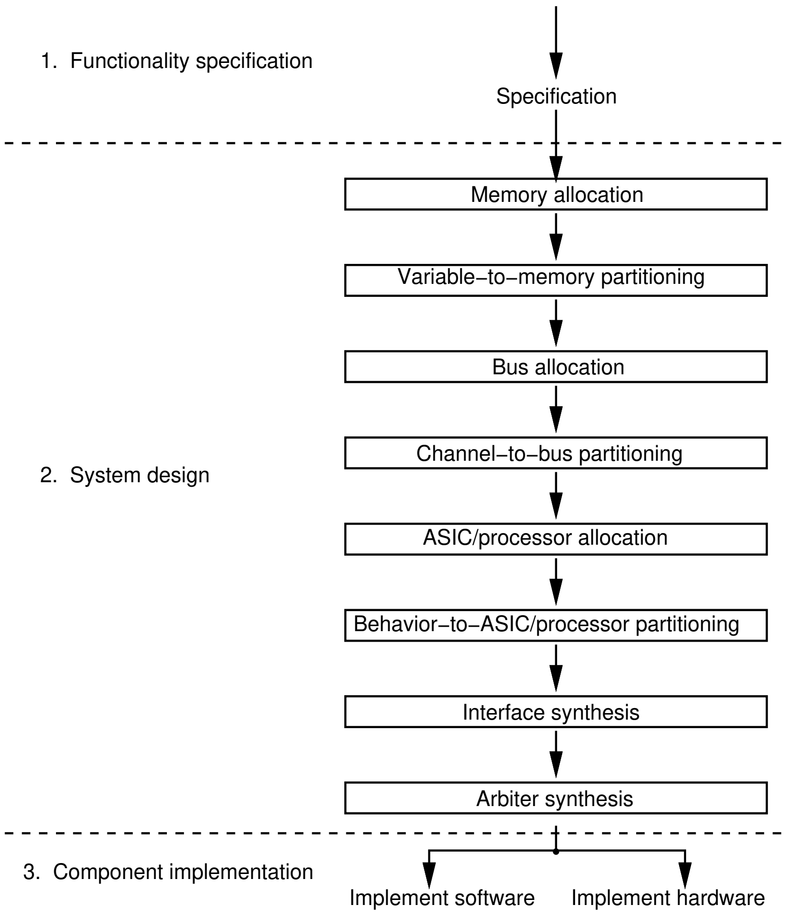


System-design tasks

System–design tasks

Functional objects	Allocation	Partitioning	Refinement
Variables	Memories	Variables to memories	Address assignment
Behaviors	Processors	Behaviors to processors	Interfacing
Channels	Buses	Channels to buses	Arbitration/protocols

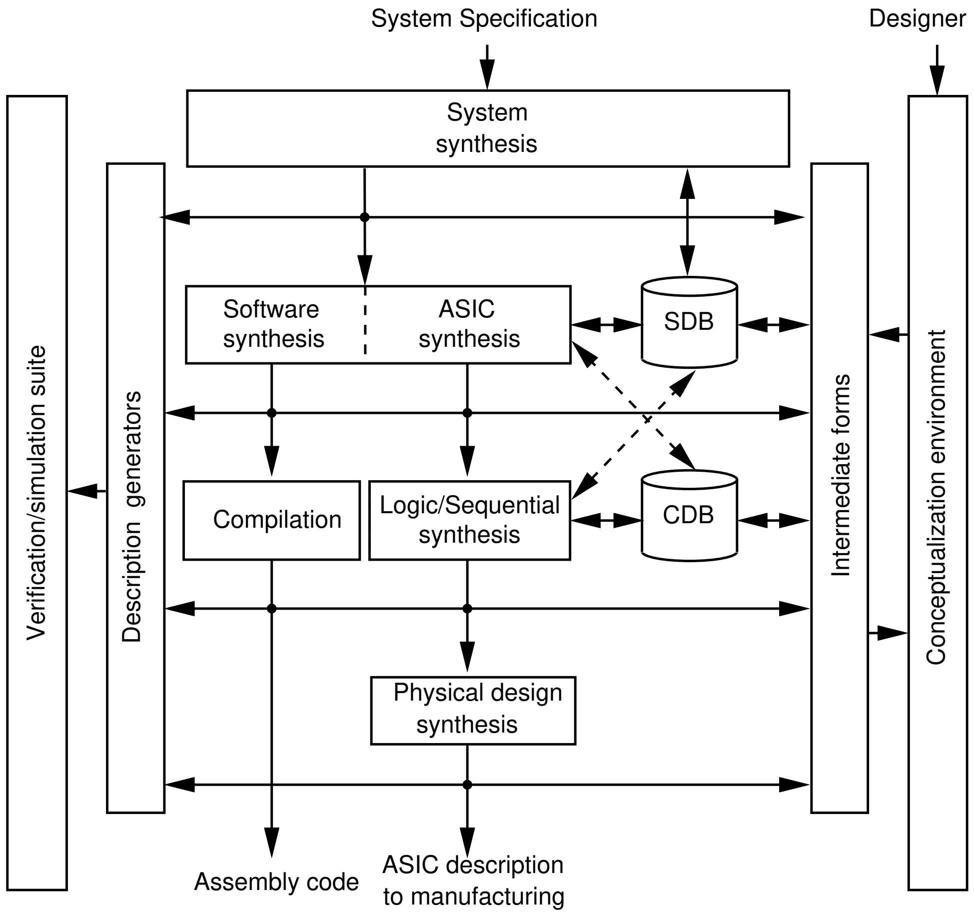
One possible ordering of tasks



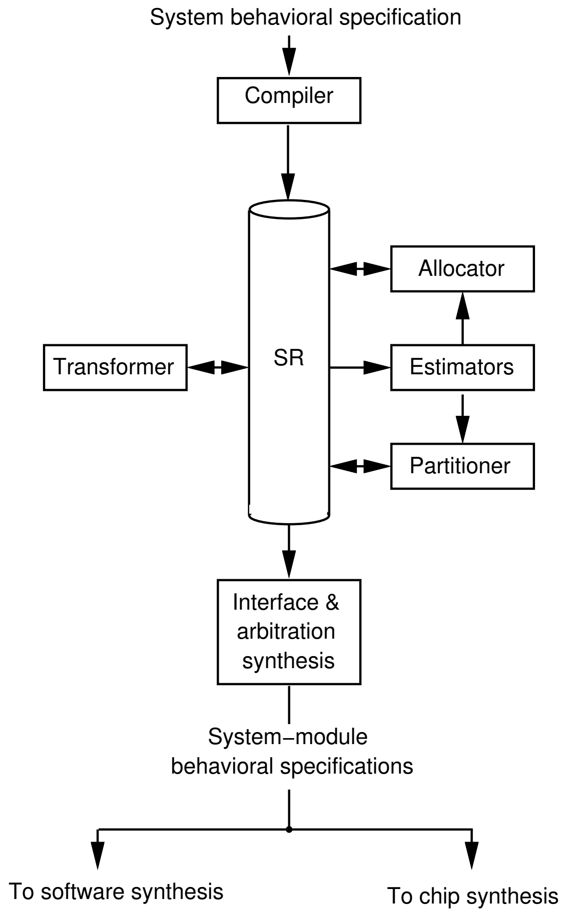
Generic synthesis system requirements

- Completeness
  - All levels of design, all implementation styles
- Extensibility
  - Allow addition of new algorithms and tools
- Controllability
  - User control of tools, design-quality feedback
- Interactivity
  - Partial design, design modification
- Upgradability
  - Evolve to describe-and-synthesize method

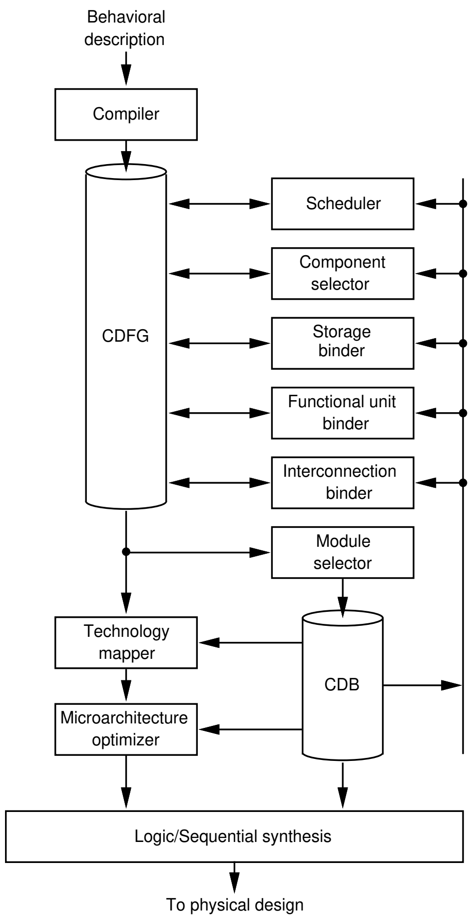
A generic synthesis system



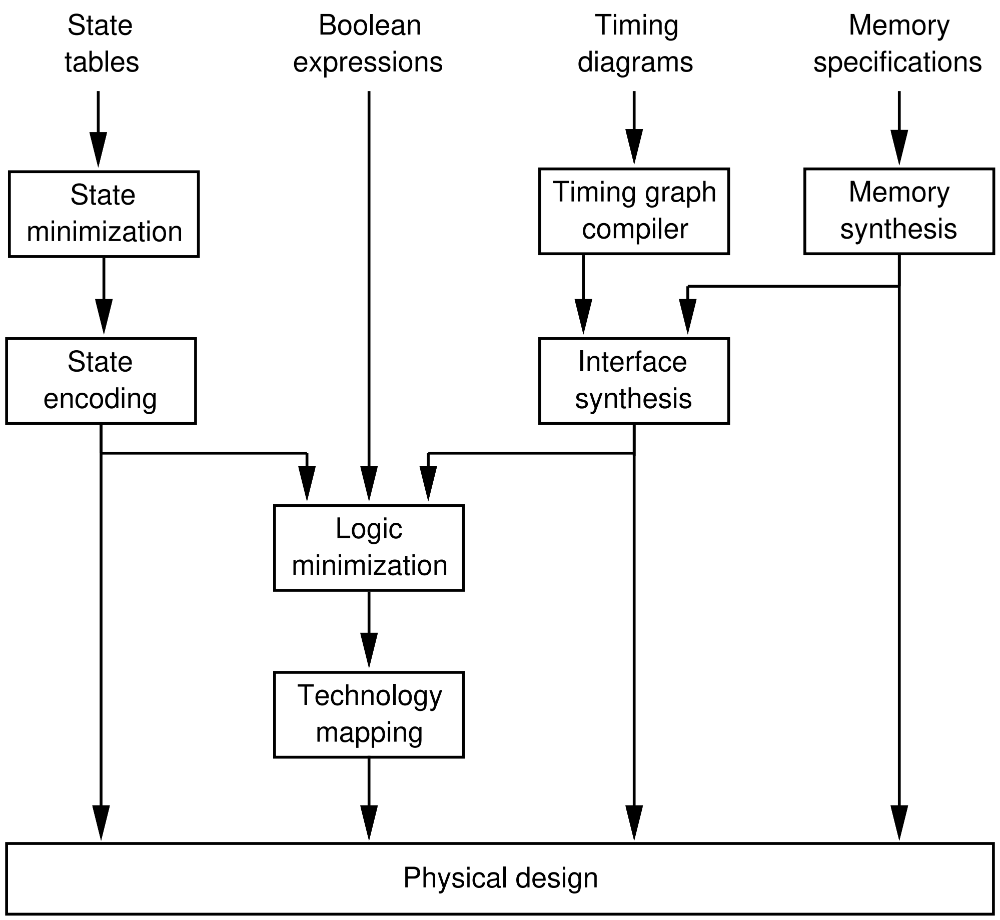
A generic system-synthesis tool



A generic chip-synthesis tool



A generic logic-synthesis tool



UC Irvine

Conceptualization environment

- Tool is only effective if the designer can use it
  - Understandable display of data
  - Highlight design parts that need attention
- Must support many design avenues

A system-synthesis tool interface

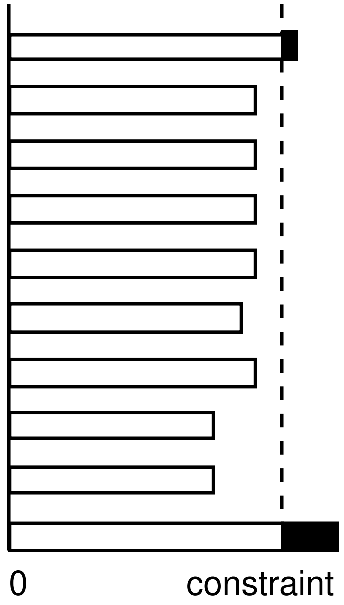
- Allocation
- Partition
- Estimates
- Constraints

Mappings	Module type	$	Execution time	Area	Pins	Instr
System		105 /100*				
ASIC1	X100	30		16000 /20000	46/60	
CaptureAudio			100/110			
GenerateAudio			100/110			
ASIC2	X100	30		18000 /20000	48/60	
CaptureGenerateVideo			100/110			
CaptureAVCmd			100/110			
Memory1	V1000	10				
audio_array1						
audio_array2						
Memory2	V1000	10				
video_array						
Processor1	Y900	25				6000 /5000*
ProcessRemoteButtons						
ProcessMiscCmds						

Cost: 5.43   View options   Partition/Allocate   Refine

An optional design view

Quality metric	Estimate/ Constraint	Violation?
$(System)	105/100	
Execution–time(CaptureAudio)	100/110	
Execution–time(GenerateAudio)	100/110	
Execution–time(CaptureGenerateVideo)	100/110	
Execution–time(CaptureAVCmd)	100/110	
Area(ASIC1)	16000/20000	
Area(ASIC2)	18000/20000	
Pins(ASIC1)	56/60	
Pins(ASIC2)	58/60	
Instr(Processor1)	6000/5000	



Summary

- Three-step design methodology
  - Functionality specification
  - System design
  - Component implementation
- Major tasks in system design
  - Allocation
  - Partitioning
  - Refinement
- Generic synthesis tool
- Conceptualization environment
  - Crucial to practical use

Future directions

- Advanced estimation methods
- Formal verification
- Testability
- Frameworks and databases
- Regularity exploiting
- System-level transformations
- Feedback incorporation