TOGAPS: A Testability Oriented Genetic Algorithm for Pipeline Synthesis

C. P. RAVIKUMAR^a and V. SAXENA^b

^aDepartment of Electrical Engineering, Indian Institute of Technology, New Delhi, India, ^bDepartment of Electrical Engineering, University of Illinois Urbana, USA

In this paper, we describe *TOGAPS*, a Testability-Oriented Genetic Algorithm for Pipeline Synthesis. The input to *TOGAPS* is an unscheduled data flow graph along with a specification of the desired pipeline latency. *TOGAPS* generates a register-level description of a datapath which is near-optimal in terms of area, meets the latency requirement, and is highly testable. Genetic search is employed to explore a 3-D search space, the three dimensions being the chip area, average latency, and the testability of the datapath. Testability of a design is evaluated by counting the number of self-loops in the structure graph of the data path. Each design is characterized by a four-tuple consisting of (i) the latency and schedule information, (ii) the module allocation, (iii) operation-to-module binding, and (iv) value-to-register binding. Accordingly, we maintain the population of designs in a hierarchical manner. The topmost level of this hierarchy consists of the latency and schedule information, which together characterize the timing performance of the design. The middle level of the hierarchy consists of a number of allocations for a given latency/schedule duplet. The lowest level of the hierarchy consists of a number of bindings for a specific latency/schedule/allocation. An initial population of designs is constructed from the given data flow graph using different latency cycles whose average latency is in the specified range. Multiple scheduling heuristics are used to generate schedules for the DFG. For each of the resulting scheduled data flow graphs, we decide on an allocation of modules and registers based on a lower bound estimated using the schedule and latency information. The operation-to-module binding and the value-to-register binding are then carried out. A fitness measure is evaluated for each of the resulting data paths; this fitness measure includes one component for each of the three search dimensions. Crossover and mutation operators are used to generate new designs from the current set of *parent* designs. The crossover operator attempts to combine the properties of two designs. The mutation operators include addition and deletion of pure delays before scheduling, as well as changes in the register and module allocation prior to binding. The genetic algorithm applies the rule of the *survival of the fittest* to obtain near-optimal solution to the otherwise intractable problem of data path synthesis. We have implemented *TOGAPS* on a Sun/SPARC 10 and studied its performance on a number of benchmark examples. Results indicate that *TOGAPS* finds area-optimal datapaths for the specified latency cycle, while reducing the number of self-loops in the data path.

Keywords: Data path synthesis, pipeline synthesis, genetic algorithm, evolutionary programming, testability oriented synthesis, BILBO architecture

1. INTRODUCTION

In this paper we present *TOGAPS* (A Testability Oriented Genetic Algorithm for Pipeline Synthesis), a high-level synthesis tool which generates a register-transfer-level description of a pipelined datapath designed to solve the problem specified by the input Data-Flow-Graph. The synthesis problem has been studied extensively in the recent past (see [2, 8, 15]). Our approach is different from the existing techniques in the following ways. We evaluate designs not only on their area and time performance, but also on their structural testability properties. We use a genetic algorithm for exploring the 3-dimensional design space characterized by the area, time, and testability properties.

Earlier efforts in high-level pipeline datapath synthesis have concentrated chiefly on minimization of chip area subject to timing constraints [5, 6, 12, 13, 15]. However, in the recent past, the testability of the generated datapath has also assumed importance [1, 9, 10]. Papachristou et al. discuss the concept of testable logic blocks (TLB) in the construction of testable datapaths [10]. A TLB essentially consists of a combinational logic block which is fed from a register R1 and which feeds another register R2. In the test mode, the register R1 can be configured as a test pattern generator and R2 can be configured as a signature analyzer. Simple rules are defined in [10] to interconnect TLBs without forming self loops in the structure graph of the datapath. The structure graph of a datapath is a graph whose nodes correspond to registers in the datapath. A directed edge is drawn from a node i to node j if the input of register j can be reached from the output of register i through a purely combinational path. A self-loop involving a node i in the structure graph would imply poor testability, since it would require that the register i be configured as a test pattern generator and a signature analyzer at the same time, an impossibility unless a concurrent BILBO is used [1, 10]. A concurrent BILBO is much more expensive in terms of area than a regular BILBO. It is easy to see that the binding phase in the synthesis process is most crucial in minimizing the number of self-loops in the structure graph. Avra [1]

generates register bindings to minimize the number of self-loops without altering the module binding. Mujumdar et al. use a network flow algorithm to carry out module and register bindings that minimize the number of self-loops [9]; they also proposed loop-breaking algorithms to further improve the testability of the resulting datapath. Thus the existing approaches for testability-oriented synthesis address the problem of improving testability at the binding level; performance optimization are first carried out using conventional scheduling and allocation heuristics. In *TOGAPS*, we address the issues of testability and area-time optimization in an integrated fashion at all the three levels of synthesis, namely, scheduling, allocation, and binding. To consider all the three design parameters simultaneously, we use a genetic algorithm, since it is a powerful optimization tool for exploring a large design space.

The paper is organized as follows. Section 2 describes the genetic algorithm for pipeline synthesis. In Section 3, we discuss the genetic operators (crossover and mutation) used in *TOGAPS*. The cost functions used in *TOGAPS*, and the techniques used to prune the large design space are discussed in Section 4. Results of experimentation are described in 5. Conclusions are presented in Section 6.

2. GENETIC ALGORITHM FOR PIPELINE SYNTHESIS

Genetic algorithms have been used extensively in solving difficult combinatorial optimization problems [3, 14]. The idea behind a genetic algorithm is the biological principle of survival of the fittest. A population of individuals (designs in our problem) is maintained at any point of time. A new generation of individuals is created by applying genetic operators such as *crossover* and *mutation*. The crossover operator ensures that the resulting offspring (new designs) have a mixture of parental properties; the mutation operator introduces new characteristics not present in the parent population. The population pool consisting of parents and the offspring are graded (according to

the cost function of the optimization problem) and the fittest individuals are retained as the next generation. Thus, over a sequence of generations, the overall properties of the population improves. The essential components of a genetic algorithm are the selection of the initial population, genetic operators, and the cost function used to grade individuals. In the subsequent sections, we describe these phases of the genetic algorithm, as applicable to the pipeline datapath synthesis problem.

2.1. Initial Population

In the pipeline synthesis problem, a point in the design space can be represented as a four-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{B}_M, \mathcal{B}_V)$, where

- $\mathcal{S}$ represents the latency cycle and scheduling information
- $\mathcal{A}$ represents the module and register allocation information
- $\mathcal{B}_M$ represents the operation-to-module binding
- $\mathcal{B}_V$ represents the value-to-register binding

We view the design space as a three-level hierarchy; at the top level of this hierarchy is the latency/schedule information. Recall that one of the inputs to *TOGAPS* is the range of acceptable pipeline average latencies. There exist a number of latency cycles for a specified average latency α . Any of these latency cycles, can be selected and used along with any of the possible schedules for the DFG. The duple consisting of a latency cycle L and a schedule S forms the root of the design hierarchy. Specifying (S, L) would indirectly specify a design subspace; all designs within this subspace have similar timing performance. For a given (S, L) , there is a lower bound on the number of resources (modules and registers). At the second level of the design hierarchy are a number of allocations that conform to this lower bound; we purposely consider designs which have more resources than the lower bound so as to perform area/testability tradeoff. Thus, specifying the allocation A of modules and registers further refines the design subspace defined by (S, L) . At the lowest level of design hierarchy are the

bindings $\mathcal{B}_M$ and $\mathcal{B}_V$. The module and the register bindings affect the structural testability properties such as the number of self loops. Specifying the binding information along with $(\mathcal{S}, \mathcal{A})$ completely specifies the design point. Figure 1 illustrates the design hierarchy.

2.2. Generation of Initial Schedules and Latency Cycles

As an aid towards limiting the size of the search space, *TOGAPS* initially asks for a lower and an upper bound on the desired average latency. We maintain a database of possible latency cycles for each value of the average latency. From this database, we select γ latency cycles randomly. For each of the γ latency cycles L_i , we form three schedule/latency pairs of the form (S_i, L_i) , where the schedule S_i is one of the following—a Force Directed Schedule (FDS) [13], an ASAP (As Soon As Possible) schedule and an ALAP (As Late As Possible) schedule [2]. Among the three scheduling algorithms, Force Directed Scheduling has the capability to consider the latency cycle in optimizing the number of resources [13]. We have selected the three scheduling algorithms in forming the schedule/latency pairs, with the view of providing diversity in the initial population.

2.3. Allocation

For a given schedule/latency pair (S_i, L_i) , a lower bound on the number of registers and modules is determined using what we call the *Concurrent Operation Enumeration Table* (COET). Let the latency cy-

FIGURE 1 Hierarchical Structure of the Population

cle in question be $(l_1, l_2, \dots, l_k)$. The COET is an $F \times P$ matrix, where F is the number of operations in the DFG, and P is given by $\sum_{i=1}^k l_i$. The entry $COET(i, j)$ is set to 1 if the operation i is active during time step j . Clearly, a lower bound on the number of functional modules of a type q is the maximum number of operations of type q that are concurrent at any particular time-step j . Define $\Delta_{i,q}$ to be 1 if operation i is of type q ; otherwise $\Delta_{i,q} = 0$. The lower bound LB_q for the number of functional modules of type q is given by

$$LB_q = \max_{j=1}^P \left(\sum_{i=1}^F COET(i, j) \cdot \Delta_{i,q} \right)$$

A lower bound on the number of registers is computed by performing a life time analysis on the values generated by each operation in the DFG. We extend the concepts introduced in [11] for latency cycles containing more than one term. The procedure to generate the initial allocation is shown in Figure 2. The allocation procedure uses a binding algorithm $bind()$, which is explained in the next subsection. If the allocation procedure cannot find any valid binding without violating the upper bound on the number of modules, it returns failure, and the corresponding schedule/latency pair (S_i, L_i) is discarded. For any (S_i, L_i) , the *TOGAPS* system experiments with a number of resource allocations. Let A_q be the number of modules of type q at the end of the *InitialAllocation* procedure. Then *TOGAPS* uses $A_q + \delta_q$ modules of type q for generating more allocations, where $0 < \delta_q < n_q$

```

procedure InitialAllocation()
(*  $\beta$  is the number of initial bindings required to be generated.
   An upper bound on module of type  $q$  is the number of
   operations of type  $q$ . *)
begin
  for each operation type  $q$  do
     $resource_q = LB_q$ ;
     $nobindings = TRUE$ ;
    while resources are within upper bound and  $nobindings$  do
      if  $bind() = FAILURE$  then
        add scarce resource
      else
         $nobindings = FALSE$ ;
    if  $nobindings = TRUE$  return (FAILURE)
     $b = 1$ ;
    while  $b < \beta$  do
      if  $bind() = SUCCESS$  then  $b = b + 1$ ;
    return (SUCCESS);
end

```

FIGURE 2 Algorithm for generating Initial Allocations

– A_q and n_q is the number of operations of type q . The δ_q are selected randomly in their respective range.

2.4. Binding

For a given operation i , we define the Exclusion List E_i as the list of operations which are concurrent with i during any of the time steps j , $1 \leq j \leq P$. Clearly, an operation k in E_i cannot be bound to the module to which i has been bound. The module binding procedure is the process of forming the reservation table R which has as many rows as there are functional modules and as many columns as the length of the critical path in the scheduled DFG. In *TOGAPS*, the operation-to-module binding is generated randomly, while ensuring that the resulting reservation table R has no conflicts.

LEMMA *The procedure $bind()$ shown in Figure 3 uses every allocated module and register at least once.*

```

procedure bind()
begin
  for each operation  $j$  do
    begin
       $A_j = \phi$ ;
      if all modules of type  $j$  have not been used at least once
        begin
           $A_j =$  any unused module of type  $j$ ;
          go to SKIP
        end
      for each module  $i$  whose type matches with  $j$  do
        if  $j$  is not in the exclusion list of  $i$  then
           $A_j = A_j \cup \{i\}$ ;
        if  $A_j = \phi$  then return (FAILURE);
    end
  SKIP:
    Randomly select module  $k$  from  $A_j$ ;
    Assign operation  $j$  to module  $k$ ;
    Add all elements of  $j$  to the exclusion list of  $k$ ;
    for the value  $v$  produced by  $j$  do
      begin
         $A_v = \phi$ ;
        for each register  $r$  do
          if  $v \notin$  the exclusion list of  $r$ 
            then
               $A_v = A_v \cup \{r\}$ ;
        if  $A_v = \phi$  then return (FAILURE);
        Randomly select a register  $s$  from  $A_v$ ;
        Assign value  $v$  to register  $s$ ;
        Add all elements of  $v$  to
          exclusion list of  $s$ ;
      end
    end
  return (SUCCESS);
end

```

FIGURE 3 Module and Register Binding

Proof: Note that the procedure *bind()* generates the availability list A_j of modules for each operation. If all the allocated modules of a particular type have not been used, the available-list consists of just one of the unused modules, which is hence selected by default. This ensures that all the allocated resources are used at least once.

After generating the bindings we determine the actual connections that exist between the various modules and registers. This gives us the number of self-loops that exist in the circuit and the number of multiplexors that will be required. We assume a multiplexor-tree structure in case there are more than two inputs that must be multiplexed at the input pin of a module or a register. We assume that there is an independent latch control for each register. This permits us to optimize the use of registers and avoid register-chains, as far as possible.

3. GENERATION OF NEW DESIGNS

In this section we present genetic operators used in *TOGAPS* to produce new members from the existing population of designs. The purpose of these operators is to create new designs which retain some traits of their parent designs, while introducing some new features. This increases the diversity in the population and enables us to explore the design space better. A good genetic operator should introduce a reasonable amount of change, be fast to implement and produce valid designs. Two kinds of genetic operators are used in *TOGAPS*:

1. **Crossover:** A crossover operator merges the characteristics of two parent designs to form new design(s).
2. **Mutation:** After a crossover operator has been used to generate a new design, a mutation operator may be applied to introduce random changes in the new design. The purpose of the mutation operator is to allow hill climbing and hence avoid a local optimal solution; if mutation operators are not employed, the final solution generated by the

genetic algorithm can become highly dependent on the initial population. If the mutation probability is p_μ , then, on an average, $N \cdot p_\mu$ of the N offspring are subjected to mutation.

We have explained in the previous section that the design population is maintained hierarchically. Accordingly, the genetic operators in *TOGAPS* also work at all the three levels of design hierarchy i.e. at the schedule and latency level, at the allocation level, and at the binding level.

3.1. Crossover Operators

TOGAPS uses three crossover operators, two of which work at the topmost level of design hierarchy, and a third one which works at the lowest level of design hierarchy.

3.1.1. Operators at the Schedule/Latency Level

The crossover operators γ_{MS} and γ_{RS} change the schedule, utilizing the information obtained by the topological sorting of the nodes of the DFG. An ordered list T of DFG nodes is said to be topologically sorted (t -sorted for brief) if $i < j$ and $T(i) < T(j)$ imply that $T(j)$ does not depend on $T(i)$. A t -sorted list corresponding to the DFG in Figure 4 [13], is $T = \{H, I, J, K, A, B, D, E, C, F, G\}$. Our crossover operators use the concept of topological sorting.

1. γ_{MS} (**Modify Schedule**): In this operator, we merge two designs (*Father* and *Mother*) and produce two new designs (*Son* and *Daughter*). Let n

FIGURE 4 Example of a Data Flow Graph

be the number of nodes in the DFG. Let T be a topologically sorted list of the n nodes in the DFG. We select a random number, distributed uniformly in the range 2 to $n - 1$ as the “pivot”. The procedure of Figure 5 is used to create a schedule S_{Son} from the parent designs. The latency information for *Son* is taken from *Father*. We repeat the above procedure for creating *Daughter*, with the roles of the *Father* and the *Mother* interchanged.

2. γ_{RS} (**Randomized Schedule**): This operator is similar to γ_{RS} , but for the fact the latencies of the children designs are determined randomly, rather than taking them from one of the parents.

3.1.2. Operator at the Binding level

This operator, which we call γ_{MB} (**Merge Bindings**), takes two bindings (*father* and *mother*) and generates two new bindings (*son* and *daughter*). For the *son*, we take the module bindings from the *father* and the register bindings from the *mother*, while for *daughter*, we take the module bindings from the *mother* and the register bindings from the *father*. Thereafter we determine the connections and the number of self-loops and multiplexors required by the two new bindings.

3.2. Mutation Operators

TOGAPS uses five different mutation operators, two of which work at the highest level of schedule &

```

procedure ScheduleSon( $T, i, Father, Mother$ )
(* Father and Mother are parent designs.  $T$  is a
   topologically sorted list of nodes in the DFG. *)
begin
  for  $j = 1$  to  $i$  do
    schedule  $T(j)$  as in parent design Father;
  for  $j = i + 1$  to  $n$  do
    begin
      Let  $\eta_j$  be the time step in which
        the schedule of Mother executes  $T(j)$ ;
      Let  $\xi_j$  be the earliest time step in which
         $T(j)$  can be scheduled without violating
        dependences due to  $T(1) \dots T(i)$ ;
      schedule  $T(j)$  in time step  $\max(\eta_j, \xi_j)$ ;
    end
  end

```

FIGURE 5 Modify-Schedule Crossover

latency, while the other three work at the next level of allocation. We give a brief description of each of the five operators.

3.2.1. Operators at the Highest Level

1. μ_{IPD} (**Insert Pure Delay**): It is well known that addition of a pure delay in the pipeline reservation table can often lead to superior performance [6]. As the name suggests, the operator μ_{IPD} adds a pure delay in the design schedule. This increases the total execution time required for the processing of one input instance. The selection of the time steps between which we insert the delay is random. For the new schedule/latency pair, we generate the allocations and bindings as described in the previous section.
2. μ_{IDF} (**Insert Delay for FDS Schedule**): Since the control step for an operation in the FDS schedule is intermediate between its ALAP and ASAP values, adding a pure delay gives the FDS algorithm greater flexibility in scheduling the operations. This is done with the hope of achieving a more uniform distribution of operations over various control steps and hence reduce the resource requirement.

In this operator we take the parent schedule and add a pure time delay in both the ASAP and ALAP schedules corresponding to the parent schedule. The delay is inserted between the same two time steps in both the ASAP and ALAP schedules. Thereafter we generate the FDS schedule for the new ALAP and ASAP schedules with the latency taken from the parent. We take the new FDS schedule and the latency from the parent to form a new duple at the highest level.

3.2.2. Operators at the Allocation Level

1. μ_{AR} (**Add register**): Addition of a register may improve the structural testability of the datapath, by giving more flexibility to the binding operator in minimizing the number of self-loops. The μ_{AR}

operator increases the number of registers in the allocation of the offspring, while retaining the number functional modules.

2. μ_{DR} (**Delete register**): The reverse of the μ_{AR} mutation operator, μ_{DR} operator deletes a randomly selected register and reassigns the values to the remaining registers. The deletion of a register reduces chip area, but may affect the testability of the data path.
3. μ_{MA} (**Modify Allocation**): This mutation is a generalization of μ_{AR} and μ_{DR} . When μ_{MA} is applied, the allocation of both functional modules and registers are modified. For instance, μ_{MA} may increase or decrease the number of adders; this manipulation is done while respecting the lower bound and upper bound on the number of functional modules. The reallocation is followed by the generation of bindings to form new designs.

4. PRUNING OF THE POPULATION SPACE

Once we have generated a number of designs from the current generation of designs, we need to prune the population space to eliminate poor designs. Genetic Algorithms (GA) make use of cost functions to select designs that shall survive until the next generation. In *TOGAPS*, the process of pruning of the population is carried out hierarchically. We start from the lowest level of bindings and proceed upwards towards the schedule/latency level, eliminating poor designs at each level. A separate cost-function formulation is used at each stage.

4.1. Pruning of Bindings

Recall that during the synthesis process, we create a number of bindings for each allocation. The first stage of pruning is carried out at the binding level. The cost function at this level considers the number of self-loops in the data path. For each allocation A_i , we calculate the average number of loops per binding l_i^{av} . We then visit each of the binding B_j , for the cur-

rent allocation A_i , and discard the binding if the number of self-loops in B_j is greater than l_i^{av} . This procedure is repeated for each of the allocations, for all the schedule/latency duples.

4.2. Pruning at the Allocation Level

In the next stage of the pruning process, we reduce the number of allocations for each of the schedule/latency pairs. The cost function at the allocation level considers the area requirement and the average number of self loops. For each schedule/latency pair (S_i, L_i) , we calculate the average functional area requirement α_i using an area estimator (Section 4.4) and the average number of self-loops λ_i . To calculate λ_i , we first visit all the bindings B_k for an allocation A_j and compute the average number of self-loops per binding, λ_j' , for A_j . Then we visit each allocation for the duple (S_i, L_i) and determine λ_i as the average of λ_j' . Thus, for each allocation A_j of duple (S_i, L_i) , we have

$$cost(A_j) = W_{area} * (\alpha(A_j)/\alpha_i) + W_{loops} * (\lambda_j'/\lambda_i) \quad (1)$$

where W_{area} and W_{loops} are positive fractional weights which are user-specified and indicate the relative importance of area and structural testability in the final solution. For a schedule/latency duple (S_i, L_i) , we discard all allocations A_j which have a cost $cost(A_j)$ greater than the average cost per allocation for that (S_i, L_i) .

4.3. Pruning at the Highest Level

The cost function at this level has two terms, a first term which measures the area versus time performance of the design, and a second term which measures the structural testability of the design. For each schedule/latency pair (S_i, L_i) ,

$$Cost(i) = W_{AreaLatency} * (Latency_i * Area_i / \delta) + W_{SelfLoops} * (SelfLoops_i / \psi)$$

where $Latency_i$ is the average latency for the duple i , $Area_i$ is the average area per allocation calculated over all allocations of the duple i , and $SelfLoops_i$ is the average number of self-loops per binding for all the designs originating from duple i . δ and ψ are both normalizing factors and are calculated over all the duples at the highest level of the design hierarchy. $W_{AreaLatency}$ and $W_{SelfLoops}$ are the weights attached to the two parts of the cost function; both these weights are positive real.

We remove those schedule/latency pairs whose cost is greater than the average cost per unit. The above cost function may be justified as follows. In a pipeline, as the average latency decreases, the number of resources increases, since the circuit will be required to perform a larger number of operations during a shorter time-period. The first term in the cost function reflects this trade-off. Comparison of two designs purely on the basis of their areas, without considering the average latency of the designs, would be inappropriate. The second part of the cost function reflects the average structural testability of the family of designs that correspond to the duple (S_i, L_i) . Thus the overall cost function reflects a tradeoff between area, delay, and structural testability properties of the designs.

4.4. Area Estimate

Since *TOGAPS* needs to estimate the area of a large number of designs, the area-estimator used should be fast, apart from being reliable. We use the following linear function which estimates the chip area in terms of the number of functional modules, registers and multiplexors used. Two types of functional modules are assumed, adders and multipliers.

$$Area = \alpha_{Adder} * n_{Add} + \alpha_{Mult} * n_{Mult} + \alpha_{Reg} * n_{Reg} + \alpha_{Mux} * n_{Mux} \quad (2)$$

where n_{Add} , n_{Mult} , n_{Reg} , and n_{Mux} are respectively the number of adders, multipliers, registers, and multiplexors required. α_{Adder} , α_{Mult} , α_{Reg} , α_{Mux} are respectively the area of a single adder, multiplier, register, and multiplexer. The module areas are library-specific; in *TOGAPS*, we used data from a XILINX 4000 FPGA library and counted the number of Configurable Logic Blocks (CLB) required for a module as an estimate of the module's area. When a k -input multiplexer, $k > 2$, is required we assumed that the same is implemented as a multiplexer tree using $k - 1$ two-input multiplexers. Since we are interested in comparing the resource requirement of various designs, we are not concerned by the absolute values generated during the estimate, as long as the values are proportional to the final area requirement for each design.

5. EXPERIMENTAL RESULTS

We ran *TOGAPS* on a number of benchmark circuits. To limit the size of the design space, we did not consider module chaining. In case of multi-step operations, we assumed that the concerned modules were in turn pipelined, ensuring that they are available to process a new input instance at every time step. In all the examples we were able to produce designs which were near-optimal in terms of the ALU resource requirement, while keeping the number of self loops small.

Tables I through IV show the results obtained us-

TABLE I Results for AR Filter (12 additions and 16 multiplications)

Avg. Lat.	lb A	lb M	A	M	INT	Area (CLB)	Self Loops	Max Loops	Min Loops	# Designs Searched	CPU (sec)
2	6	8	6	8	30	632	6	14	6	3452	51.24
3	4	6	6	8	31	633	6	14	5	3248	42.78
4	3	4	4	4	32	379	5	14	5	3750	49.20
5	3	4	3	4	32	370	7	14	5	3872	54.90

TABLE II Results for Bandpass Filter (17 additions and 12 multiplications)

Avg. Lat.	<i>lb</i> A	<i>lb</i> M	<i>A</i>	<i>M</i>	<i>INT</i>	Area (CLB)	Self Loops	Max Loops	Min Loops	# Designs Searched	CPU (sec)
2.5	7	5	9	6	31	450	5	15	5	3098	43.93
3.5	6	4	6	5	35	462	5	15	4	3326	49.20
4.0	5	3	6	3	38	353	6	16	5	3872	43.83
4.5	4	3	7	4	29	391	5	15	5	3234	47.64

ing *TOGAPS* on four different benchmark DFGs, namely, the AR-filter, Bandpass filter, FIR filter and Biquad filter [4]. The number of additions and multiplications in each DFG are indicated in the table captions. In all our experiments, we used 10 generations, 5 to 10 crossovers and mutations per generation. The initial number of schedule/latency pairs varied in accordance to the size of the desired range of latencies. For each schedule/latency pair, the number of initial allocations was chosen to be 5. For every allocation, the number of initial bindings was fixed at 10. *A*, *M*, and *INT* are the number of adders, multipliers, and registers+multiplexers used in each of the designs indicated. Area is the number of CLBs required. In each table, we show different designs obtained for different average-latencies. For a given average-latency, *lb(A)* and *lb(M)* are the lower bounds on the number of adders and number of multipliers, respectively. $lb(A)(lb(M)) = \lceil n/al \rceil$, where n = No. of Additions(Multiplications) in the DFG and al is the average latency. The number of self loops in the resulting data path, the functional area estimate, the number of designs searched, and the CPU requirement (SUN SPARC10) of *TOGAPS* are also indicated in the tables. During each experiment, we also kept track of the maximum number of self-loops and the minimum number of self-loops over the space of all the data paths examined by *TOGAPS* (*Max Loops* & *Min Loops* in the tables). It can be observed from the

tables that the majority of the designs generated using *TOGAPS* are indeed equal or close to the lower bounds on the number of functional modules, while keeping the number of self-loops in the data path at a low value. The number of self-loops in the final designs generated through *TOGAPS* were, as a rule, close to the minimum number of self-loops encountered.

6. CONCLUSIONS

Assuming BIST as the testing technique, we have presented a high-level synthesis tool named *TOGAPS* for automatic generation of testable data paths for pipelined designs. In addition to testability, the objective of the synthesis tool is to achieve designs that meet the timing specification and optimize the number of resources required. The optimization technique employed in *TOGAPS* is based on genetic search. The advantage of a genetic search algorithm is that it can handle multiple objective functions and constraints (such as testability, number of resources, and timing information) through an appropriately designed cost function. The relative importance of each objective function can be controlled through user-specified weights. A novel hierarchical representation of a design was used in our implementation to achieve efficient search, storage, and pruning of the design space. This concept of hierarchy can be extended in useful

TABLE III Results for FIR Filter (15 additions and 8 multiplications)

Avg. Lat.	<i>lb</i> A	<i>lb</i> M	<i>A</i>	<i>M</i>	<i>INT</i>	Area (CLB)	Self Loops	Max Loops	Min Loops	# Designs Searched	CPU (sec)
3.0	5	3	5	3	30	320	4	15	4	5750	62.22
3.5	5	3	6	3	24	309	4	14	4	5310	58.40
4.0	4	2	4	2	32	256	5	15	3	4984	61.58
4.0	4	2	6	2	28	262	4	13	3	3066	34.37

TABLE IV Results for Biquad Filter (15 additions and 7 multiplications)

Avg. Lat.	lb A	lb M	A	M	INT	Area (CLB)	Self Loops	Max Loops	Min Loops	# Designs Searched	CPU (sec)
2.0	8	4	8	4	27	399	5	12	5	4034	43.58
3.0	5	3	5	3	36	343	6	12	5	1170	12.00
4.0	4	2	5	3	26	308	5	13	4	5186	53.21
5.5	3	2	4	3	28	307	5	13	5	5020	53.75

ways; for instance, we are now exploring the possibility of introducing a further level of hierarchy at the top which allows us to explore *behavioral level transformations* of the data flow graph.

The framework of *TOGAPS* offers flexibility in a number of ways. Although this paper has focussed on BIST architecture for data paths, it is not difficult to support other testing methodologies such as scan, partial scan, or multiple scan by suitably modifying the testability measure. The genetic algorithm allows the flexibility of adding other components to the cost function, such as power dissipation.

Acknowledgements

We thank Prof. Rajiv Jain and Mr. Ashutosh Mujumdar of the Department of Electrical and Computer Engineering, University of Wisconsin, for providing us with the benchmark circuits and for helpful discussions.

References

- [1] L. Avra. Allocation and assignment in high-level synthesis for self-testable data paths. In *Proceedings of International Test Conference*, pages 463–472, 1991.
- [2] D.D. Gajski and L. Ramachandran. Introduction to High-Level Synthesis. *IEEE Design and Test of Computers*, pages 44–54, Winter 1994. Tutorial.
- [3] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, Reading, Massachusetts, 1989.
- [4] R. Jain. High level synthesis benchmarks. Personal Communication.
- [5] K. Kucukcakar. *System-Level Synthesis Techniques With Emphasis on Partitioning and Design Planning*. PhD thesis, University of Southern California, October 1991.
- [6] P.M. Kogge. *The Architecture of Pipelined Computers*. Hemisphere Publishing Corporation, NY, 1981.

- [7] S.-P. Lin, C. Njinda, and M. Breuer. Generating a Family of Testable Designs using the BILBO Methodology. *Journal of Electronic Testing: Theory and Applications*, pages 71–89, 1993.
- [8] A.C. Parker, M. McFarland and R. Camposano. The high level synthesis of digital systems. *Proceedings of the IEEE*, 78, February 1990.
- [9] A. Mujumdar, K. Saluja, and R. Jain. Incorporating testability considerations in high-level synthesis. In *Proceedings of the Fault-Tolerant Computing Symposium*, pages 272–279, 1992.
- [10] C.A. Papachristou et al. A framework for high-level synthesis with self-testability. Technical report, Computer Engineering and Science Department, Case Institute of Technology, Case Western Reserve University, February 1991. CES-91-03.
- [11] K.K. Parhi. Calculation of minimum number of registers in arbitrary life time cycles. In *Proceedings of the VLSI Design 1994 Conference*, 1994. Calcutta, India.
- [12] N. Park and A.C. Parker. SEHWA a program for synthesis of pipelines. In *Proceedings of the 23rd IEEE/ACM Design Automation Conference*, pages 454–460, 1986.
- [13] P.G. Paulin and J.P. Knight. Force directed scheduling for the behavioral synthesis of ASICs. *IEEE Transactions of CAD*, 8(6):661–679, 1989.
- [14] C.P. Ravikumar and A.K. Gupta. A genetic algorithm for mapping tasks onto a reconfigurable parallel processor. *IEE Proceedings (E)*; 142:81–86, March 1995.
- [15] L. Stok. Data Path Synthesis. *Integration, the VLSI Journal*, pages 1–71, December 1994. Tutorial.

Authors' Biographies

C.P. Ravikumar obtained his Ph.D. in Computer Engineering from the Department of Electrical Engineering systems at the University of Southern California, Los Angeles, in 1991. During 1987–1991, he was a research and teaching assistant in the Computer Engineering Division of the EE-Systems Department at the University of Southern California. From 1991–1995 he was an Assistant Professor at the Department of Electrical Engineering, Indian Institute of Technology, Delhi. He is currently an Associate Professor in the Department of Electrical Engi-

neering at IIT Delhi. He is the Indian Editor of the "International Journal of VLSI Design" and serves on the editorial board of the journal "Computers and Informatics". His research interests are in the areas of VLSI Design Automation, High Level Synthesis of Digital Systems, Testing of digital systems, high performance computer architecture, and high speed networks.

Vikram Saxena received his B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, New Delhi in 1994. Since June 1994, he has been as a research assistant with the Analog and Digital Circuits group, in the Coordinated Science Laboratory at the University of Illinois at Urbana-

Champaign, where he is pursuing graduate studies in the Department of Electrical and Computer Engineering. During the summer of 1995 he worked in the SEG-CAD group at the Digital Equipment Corporation, Hudson, Massachusetts. Earlier, he has held summer positions at the Center for Development of Advanced Computing (CDAC), Pune (1993) and at the Center for Biomedical Engineering at IIT Delhi (1992). He is a recipient of the National Talent Search Scholarship. His research interests are in the area of CAD for VLSI circuits, including power and current estimation, synthesis for testability and low power, fault simulation, test generation for sequential circuits and parallel algorithms for CAD.