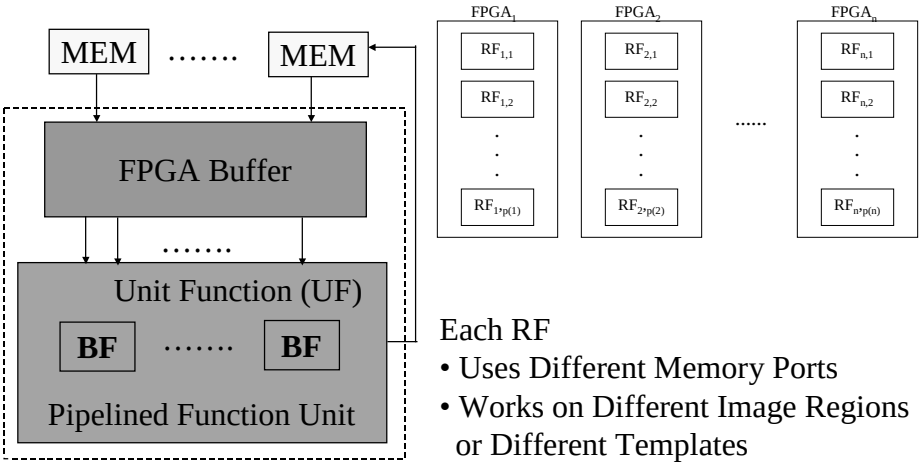
GTM Design Structure



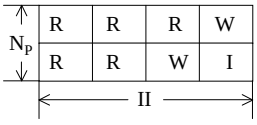
Region Function (RF)

Each RF
- Uses Different Memory Ports
- Works on Different Image Regions or Different Templates

Memory Access Pattern (MAP)

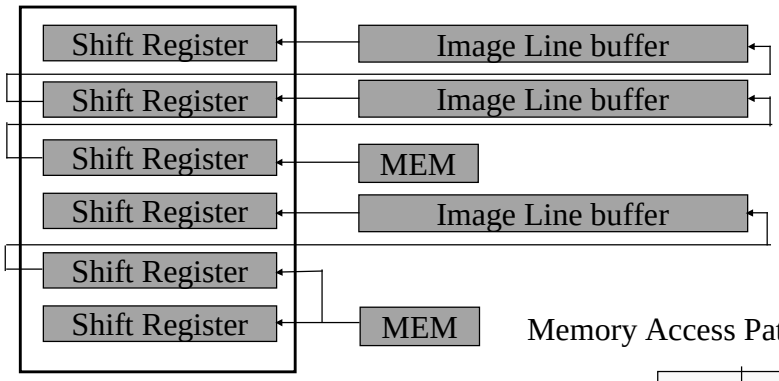
For a given pair of PF (Packing Factor) and II (Initiation Interval), MAP includes

- Number of Memory Port Used (N_P)
- Number of Memory Reads ($N_R(P)$) From Each Memory Port ($N_R(1)$, $N_R(2)$, …, $N_R(N_P)$)—Reading Pattern
- Number of Memory Writes ($N_W(P)$) From Each Memory Port ($N_W(1)$, $N_W(2)$, …, $N_W(N_P)$)—Writing Pattern



$$N_R = \sum_{p=1}^{N_P} N_R(p) \qquad N_W = \sum_{p=1}^{N_P} N_W(p)$$

FPGA Buffer Structure



Template Window

Memory Access Pattern

R	W
R	R

II=2

N_P=2

PF=1

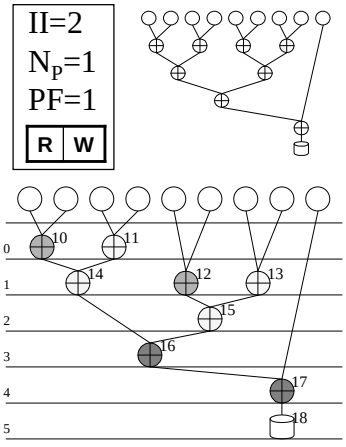
Unit Function Synthesis

II=2

N_P=1

PF=1

R	W



Cycle	Iterations		
	1	2	3
0	10,11		
1	12,13,14		
2	15	10,11	
3	16	12,13,14	
4	17	15	10,11
5	18	16	12,13,14
6		17	15
7		18	16
8			17
9			18

10,11
12,13,14
15,10,11
16,12,13,14
17,15,10,11
18,16,12,13,14
17,15
18,16
17
18

R					W

Memory Access Schedule

Performance: SL+II*(N-1)

Constraints of MAP

$$N_R = \sum_{p=1}^{N_P} N_R(p) \quad \text{and} \quad N_W = \sum_{p=1}^{N_P} N_W(p)$$

$$N_R(p) \geq 0 \text{ and } N_W(\text{p}) \geq 0, \quad p = 1,2,\ldots N_P$$

$$0 < N_R(p) + N_W(\text{p}) \leq \text{II}, \quad p = 1,2,\ldots N_P$$

$$1 \leq N_P \leq N_{MP} \quad \text{and} \quad N_W = N_{MW} \times PF$$

$$1 \leq N_R \leq R_{WIN} \quad \text{or} \quad N_R = N_{AP}$$

$$PF = 2^{\text{b}}, \; 0 \leq b \leq \log_2(W_{MP} / B_{DATA})$$

$$II_{MIN} \leq II \leq II_{UB}$$

$$II_{MIN} = \lceil (1 + N_{MW} \times PF) / N_{MP} \rceil$$

$$II_{UB} = N_{AP} + N_{MW} \times PF$$

Quality Measures of MAP

1. Number of Memory Port Used (N_P)
2. Number of of Line Buffers (N_{LB})
3. Initiation Interval (II)
4. Memory Size Requirement (S_M)

$$S_M = \max\{S_M(p) : 1 \leq p \leq N_P\}$$

$$S_M(p) = \begin{cases} 1 + N_W(p) \times \alpha_W & \text{when } N_R(p) \neq 0 \\ N_W(p) \times \alpha_W & \text{when } N_R(p) = 0 \end{cases}$$

Non-Dominated MAP

Definition: Given PF and II, the set of all MAPs is denoted by $S_{MAP}(pf,II)$. Given PF and all possible II, the set of all MAPs is denoted by $S_{MAP}(pf)$. For two MAPs A and B in $S_{MAP}(pf)$, A dominates B, if

A's $N_P \leq$ B's N_P, and
A's $N_{LB} \leq$ B's N_{LB}, and
A's II $\leq$ B's II, and
A's $S_M \leq$ B's S_M

If a MAP in $S_{MAP}(pf)$ is not dominated by any other MAP in $S_{MAP}(pf)$, it is considered non-dominated.

Algorithm: Compute Non-Dominated Patterns in $S_{MAP}(pf,II)$

```
S_NMAP(pf, II) = Φ;
While S_MAP(pf, II) ≠ Φ
{
   pick a ∈ S_MAP(pf, II);
   Temp_Set = S_MAP(pf, II) \ {a};
   do = true;
   for each e ∈ Temp_Set and do == true
  {
     if a dominates e then S_MAP(pf, II) = S_MAP(pf, II) \ {e};
     else if e dominate a then {S_MAP(pf, II) = S_MAP(pf, II) \ {a}; do = false;}
  }
  if do == true then S_NMAP(pf, II) = S_NMAP(pf, II) ∪ {a};
}
```

Naïve Enumeration of $S_{MAP}(pf,II)$

for $(N_R = 1 \text{ to } R_{WIN})$ or $(N_R = N_{AP})$

for $(N_P = 1 \text{ to } N_{MP})$

$$(3.1)\quad \begin{cases} \sum_{p=1}^{N_P} N_R(p) = N_R \\ II \geq N_R(1) \geq N_R(2) \geq \cdots \geq N_R(N_P) \geq 0 \end{cases}$$

$$(3.2)\quad \begin{cases} \sum_{p=1}^{p=N_P} N_W(p) = N_M \\ 0 \leq N_W(p) \text{ and } 1 \leq N_R(p) + N_W(p) \leq II,\ p = 1,2,\ldots N_P \\ N_R(p) = N_R(p+1) \Rightarrow N_W(p) \geq N_W(p+1), \\ p = 1,2,\ldots, N_P - 1 \end{cases}$$

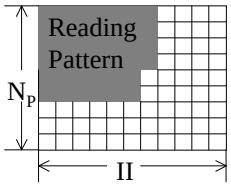
MAP Pruning

Reading Pattern Pruning

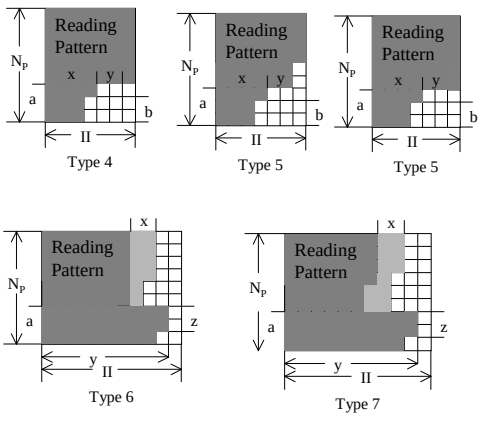
Definition: A minmax decomposition of an integer n with respect to another integer q is $(i_1, i_2, \ldots, i_q)$ that minimizes i_1 subject to

$$\sum_{j=1}^{q} i_j = n$$

$$i_1 \geq i_2 \geq \ldots \geq i_q \geq 1$$



Writing Pattern Pruning



Experimental Results (1)

Efficiency of Pruning

R_{WIN}	N_{AP}	N_{PMAP}	N_{NMAP}	Time (S)	$\frac{N_{NMAP}}{N_{PMAP}}$
3	9	96	50	0.06	52%
9	20	288	130	0.17	45%
25	45	816	351	0.66	38%
45	65	1492	632	2.14	42%
65	75	2169	915	4.89	42%
85	95	2845	1199	10.22	42%
105	115	3522	1481	19.11	42%
125	135	4199	1765	33.28	42%

R_{WIN}—Rows of Template Window
N_{AP}—Number of Active Points
N_{PMAP}—Number of Pruned MAPs
N_{NMAP}—Number of Non-Dominated MAPs

Experimental Results (2)

Efficiency of Pruning

W_{WIN}	N_{AP}	N_{PMAP}	Time
3	9	72	0.02
9	20	214	0.06
15	30	346	0.17

With Pruning

W_{WIN}	N_{AP}	N_{MAP}	Time
3	9	4032	1.48
9	20	61133	30.59
15	30	285966	252.38

Without Pruning

Experimental Results (3)

Optimal Mapping

N_{MP}=1
C_{IMG}=360
R_{WIN}=3
C_{WIN}=4
N_{AP}=9
W_{MP}=16
B_{DATA}=8
α_W=1.0

All Non-Dominated MAPs

PF	II	MAP	Buffer Type
2	3	RWW	Full/Packing
2	4	RRWW	Hybrid/Packing
1	2	RW	Full
2	5	RRRWW	Partial/Packing
1	3	RRW	Hybrid
1	4	RRRW	Partial
2	11	RRRRRRRRRWW	Internal
1	10	RRRRRRRRRW	No

Experimental Results (4)

Optimal Mapping

If FPGA has 560 CLBs
Then Partial Buffer with Packing can be selected.

Area Estimate

Buffer	UF	Total
648	244	892
458	232	690
530	100	630
264	240	504
339	112	451
114	116	260
32	184	216
0	92	92

Experimental Results (5)

Optimal Mapping

If FPGA has 560 CLBs
Then Partial Buffer
can be selected

MAPs that are Considered
By Human Designers

PF	II	MAPs	Buffer Type
1	2	RW	Full
1	4	RRRW	Partial
2	11	RRRRRRRRRWW	internal
1	10	RRRRRRRRRW	No

Area Estimate

Buffer	UF	Total
530	100	630
114	116	260
32	184	216
0	92	92