# Type-Directed Scheduling of Streaming Accelerators – Technical Appendix

David Durst, Matthew Feldman, Dillon Huff, David Akeley, Ross Daly, Gilbert Bernstein, Marco Patrignani, Kayvon Fatahalian, Pat Hanrahan

This appendix contains additional material from the original paper. Appendix A contains formulas for calculating area and delay for  $\mathbf{L}^{st}$  operators. Appendix B formalizes  $\mathbf{L}^{seq}$  and Appendix C formalizes  $\mathbf{L}^{st}$ .

## A Formulas for Lst Operator Properties

For simplicity, we provide only the formulas for the operators used in the code example in the paper.

### A.1 Areas of Lst operators (excerpts)

Area a is measure of FPGA resources required to implement an operator. Area is a vector of two components, storage and compute, to account for the fact that FPGAs have different resources for storing data and performing computation [2].

 $counter\_size(n)$  computes the area for a counter that counts up to n.

```
area(tuple) = \{compute = 0, storage = 0\}
        area(map_s f) = n * area(f)
where the input has type SSeq n t
        area(map_t f) = area(f)
       area(map2_s f) = n * area(f)
where the input has type SSeq n t
       area(map2_t f) = area(f)
     area(reduce_s f) = (n-1) * area(f)
where the input has type SSeq n t
     area(reduce_t f) = area(f) +
                        \{compute = 0,
                        storage = sizeof(t) +
                        counter_size(n)
where the input has type TSeq n i t
        area(shift_s) = \{compute = 0, storage = 0\}
        area(shift_t) = \{compute = 0, storage = sizeof(t)\}
where the input has type TSeq n i t
    area(select_1d_s) = \{compute = 0, storage = 0\}
    area(select_1d_t) = \{compute = 0, storage = 0\}
        area(reshape) = see[1]
          area(g . f) = area(g) + area(f)
```

## A.2 Delays of Lst operators (excerpts).

Delay is a measure of time (in clocks) between the first element of an input sequence arriving at a operator, and the first element emitted by the operator. A fully combinational adder has zero delay. Both the full parallel map\_s and fully sequential map\_t operators begin emitting output as soon as their contained function f does, so the delay of these higher order operators is the same as the delay of f.

```
\label{eq:delay} \begin{split} \operatorname{delay}\left(\operatorname{add}\right) &= 0 \\ \operatorname{delay}\left(\operatorname{tuple}\right) &= 0 \\ \operatorname{delay}\left(\operatorname{map\_s} \ f\right) &= \operatorname{delay}\left(f\right) \\ \operatorname{delay}\left(\operatorname{map\_t} \ f\right) &= \operatorname{delay}\left(f\right) \\ \operatorname{delay}\left(\operatorname{reduce\_s} \ f\right) &= 0 \\ \operatorname{delay}\left(\operatorname{reduce\_t} \ f\right) &= n-1 \\ \end{split} \\ \text{where the input has type TSeq n i t} \\ \operatorname{delay}\left(\operatorname{shift\_s}\right) &= 0 \\ \operatorname{delay}\left(\operatorname{shift\_t}\right) &= 0 \\ \operatorname{delay}\left(\operatorname{select\_1d\_s}\right) &= 0 \\ \operatorname{delay}\left(\operatorname{select\_1d\_t}\right) &= j \\ \end{split} \\ \text{where the selecting the jth element} \\ \operatorname{delay}\left(\operatorname{reshape}\right) &= \operatorname{see}\left[1\right] \\ \operatorname{delay}\left(f\right) &= \operatorname{delay}\left(f\right) + \operatorname{delay}\left(g\right) \end{split}
```

#### **B** L<sup>seq</sup> Formalisation

#### B.1 Terms

```
t::= undef | n \in \mathcal{N} | b \in \{\text{True}, \text{False}\}

| \lambda x : \tau . t | x | [t, \ldots, t] | \langle t, \ldots, t \rangle | t.i

| \text{tuple\_to\_seq } t | \text{seq\_to\_tuple } t

| \text{not } t | t == t | t \text{ op } t \text{ s.t. op } \in \{+, -, *, /\}

| t \text{ bop } t \text{ s.t. bop } \in \{\lor, \land\}

| \text{const\_gen } t

| \text{shift } t t | \text{up\_1d } t t | \text{select\_1d } t t

| \text{partition } t t t | \text{unpartition } t

| \text{map } t t | \text{map2 } t t t | \text{reduce } t t

B.2 Values

v ::= \text{undef } | \text{n} | \text{b} | \lambda x : \tau . t | [v_1, \ldots, v_n] | \langle v, \cdots, v \rangle
```

# B.3 Types

PLDI '20, June 15-20, 2020, London, UK 2020.

```
\tau ::= \mathbb{N} \mid \mathbb{B}
\sigma ::= \operatorname{seq} n \, \sigma \mid \sigma \times \sigma \mid \tau
f ::= \sigma \to \sigma \mid \sigma
```

#### **B.4** L<sup>seq</sup> Evaluation Contexts

```
\begin{split} E &::= [\cdot] \,|\, E \,t \,|\, (\lambda x : \tau. \,t) \,E \\ &|\, [E, \dots, t_n] \,|\, [v_1, \dots, E, \dots, t_n] \,|\, [v_1, \dots, v_{n-1}, E] \,|\, \langle E, \dots, t_n \rangle \\ &|\, \langle v_1, \dots, E, \dots, t_n \rangle \,|\, \langle v_1, \dots, v_{n-1}, E \rangle \\ &|\, \text{tuple\_to\_seq} \,\, E \,|\, \text{seq\_to\_tuple} \,\, E \\ &|\, \text{not} \,\, E \,|\, E \,=\, t \,|\, v \,=\, E \,|\, E \,\text{ op } \,t \,|\, n \,\text{ op } \,E \\ &|\, \text{for } E \,|\, E
```

#### **B.5** L<sup>seq</sup> Program Contexts

```
 \begin{array}{l} \mathfrak{c} ::= [\cdot] \mid \mathfrak{c} \mid t \mid t \mid t \mid \lambda x : \tau . \, \mathfrak{c} \\ \mid [\mathfrak{c}, \ldots, t_n] \mid [t_1, \ldots, \mathfrak{c}, \ldots, t_n] \mid [t_1, \ldots, t_{n-1}, \mathfrak{c}] \mid \langle \mathfrak{c}, \ldots, t_n \rangle \\ \mid \langle t_1, \ldots, \mathfrak{c}, \ldots, t_n \rangle \mid \langle t_1, \ldots, t_{n-1}, \mathfrak{c} \rangle \\ \mid \text{tuple\_to\_seq } \mathfrak{c} \mid \text{seq\_to\_tuple } \mathfrak{c} \\ \mid \text{not } \mathfrak{c} \mid \mathfrak{c} = t \mid t == \mathfrak{c} \mid \mathfrak{c} \text{ op } t \mid n \text{ op } \mathfrak{c} \\ \mid \mathfrak{c} \mid \text{bop } t \mid t \text{ bop } \mathfrak{c} \\ \mid \text{lut\_gen } \mathfrak{c} \mid t \mid \text{lut\_gen } t \mid \mathfrak{c} \mid \text{const\_gen } \mathfrak{c} \\ \mid \text{shift } n \mid \mathfrak{c} \mid \mathfrak{up\_1d} \mid n \mid \mathfrak{c} \mid \text{select\_1d} \mid n \mid \mathfrak{c} \\ \mid \text{partition } n \mid \mathfrak{n} \mid \mathfrak{c} \mid \text{unpartition } \mathfrak{c} \\ \mid \text{map } \mathfrak{c} \mid t \mid \text{map } t \mid \mathfrak{c} \mid \text{map } 2 \mid \mathfrak{c} \mid t \mid \text{map } 2 \mid t \mid \mathfrak{c} \mid \mathsf{c} \mid
```

# C Lst Formalisation

#### C.1 Terms

```
\begin{array}{l} t ::= \mathsf{undef} \mid n \in \mathcal{N} \mid b \in \{\mathsf{True}, \mathsf{False}\} \\ \mid \lambda x : \tau. \ t \mid x \mid [t, \dots, t] \mid \langle t, \dots, t \rangle \mid t.i \\ \mid \mathsf{not} \ t \mid t == t \mid t \ \mathsf{op} \ t \ \mathsf{s.t.} \ \mathsf{op} \in \{+, -, *, /\} \\ \mid \ t \ \mathsf{bop} \ t \ \mathsf{s.t.} \ \mathsf{bop} \in \{\lor, \land\} \\ \mid \ \mathsf{const\_gen} \ t \\ \mid \ \mathsf{shift\_s} \ t \ t \mid \mathsf{shift\_t} \ t \ t \mid \mathsf{up\_1d\_s} \ t \ t \mid \mathsf{up\_1d\_t} \ t \ t \\ \mid \ \mathsf{select\_1d\_s} \ t \ t \mid \mathsf{select\_1d\_t} \ t \ t \\ \mid \ \mathsf{map\_s} \ t \ t \mid \mathsf{map\_t} \ t \ t \mid \mathsf{map2\_s} \ t \ t \ t \mid \mathsf{map2\_t} \ t \ t \ t \\ \mid \ \mathsf{reduce\_s} \ t \ t \mid \mathsf{reduce\_t} \ t \ t \\ \mid \ \mathsf{reshape} \ \sigma \ \sigma \ t \\ \end{array}
```

#### C.2 Values

```
v := \text{undef} \mid \mathbf{n} \mid \mathbf{b} \mid \lambda x : \tau. \ t \mid [v_1, \dots, v_n]_s \mid | [v_1, \dots, v_n]_t \mid \langle v, \dots, v \rangle \mid \text{invalid}
```

#### C.3 Types

```
\tau ::= \mathbb{N} \mid \mathbb{B}
\sigma ::= \operatorname{sseq} n \, \sigma \mid \operatorname{tseq} n \, n \, \sigma \mid \sigma \times \sigma \mid \tau
f ::= \sigma \to \sigma
```

#### C.4 Lst Evaluation Contexts

```
\begin{split} E &:= [\cdot] \, | \, E \, t \, | \, (\lambda x : \tau. \, t) \, E \\ &| \, [E, \ldots, t_n]_s \, | \, [v_1, \ldots, E, \ldots, t_n]_s \, | \, [v_1, \ldots, v_{n-1}, E]_s \, | \, [E, \ldots, t_n]_t \\ &| \, [v_1, \ldots, E, \ldots, t_n]_t \, | \, [v_1, \ldots, v_{n-1}, E]_t \\ &| \, \langle E, \ldots, t_n \rangle \, | \, \langle v_1, \ldots, E, \ldots, t_n \rangle \, | \, \langle v_1, \ldots, v_{n-1}, E \rangle \\ &| \, \text{not} \, \, E \, | \, E \, = t \, | \, v \, = E \, | \, E \, \text{ op } \, t \, | \, n \, \text{ op } \, E \\ &| \, \, \text{not} \, \, E \, | \, E \, \text{op } \, t \, | \, n \, \text{ op } \, E \\ &| \, \, \text{const\_gen} \, E \\ &| \, \, \text{shift\_s} \, \, n \, E \, | \, \text{shift\_t} \, \, n \, E \, | \, \text{up\_1d\_s} \, \, n \, E \, | \, \text{up\_1d\_t} \, \, n \, E \\ &| \, \, \text{select\_1d\_s} \, \, n \, E \, | \, \text{select\_1d\_t} \, \, n \, E \\ &| \, \, \text{map\_s} \, \, E \, t \, | \, \text{map\_s} \, \, v \, E \, | \, \text{map\_t} \, \, t \, t \, | \, \text{map\_t} \, \, v \, E \\ &| \, \, \text{map2\_s} \, \, E \, t \, t \, | \, \text{map2\_s} \, \, v \, E \, t \, | \, \text{map2\_s} \, \, v \, v \, E \\ &| \, \, \text{reduce\_s} \, \, E \, t \, | \, \text{reduce\_s} \, \, v \, E \\ &| \, \, \text{reduce\_t} \, \, E \, t \, | \, \text{reduce\_t} \, \, v \, E \\ &| \, \, \text{reshape} \, \, \sigma \, \sigma \, E \end{split}
```

#### C.5 Lst Evaluation Contexts

```
 c := [\cdot] \mid c \ t \mid (\lambda x : \tau . \ t) \ c \\ \mid \ [c, \dots, t_n]_s \mid [t_1, \dots, c, \dots, t_n]_s \mid [t_1, \dots, t_{n-1}, c]_s \mid [c, \dots, t_n]_t \\ \mid \ [t_1, \dots, c, \dots, t_n]_t \mid [t_1, \dots, t_{n-1}, c]_t \\ \mid \ \langle c, \dots, t_n \rangle \mid \langle t_1, \dots, c, \dots, t_n \rangle \mid \langle t_1, \dots, t_{n-1}, c \rangle \\ \mid \ \text{not} \ c \mid c == t \mid t == c \mid c \ \text{op} \ t \mid n \ \text{op} \ c \\ \mid \ c \ \text{bop} \ t \mid b \ \text{bop} \ c \\ \mid \ c \ \text{bop} \ t \mid b \ \text{bop} \ c \\ \mid \ c \ \text{const\_gen} \ c \\ \mid \ select\_1d\_s \ n \ c \mid select\_1d\_t \ n \ c \\ \mid \ select\_1d\_s \ n \ c \mid select\_1d\_t \ n \ c \\ \mid \ map\_s \ c \ t \mid map\_s \ t \ c \mid map\_t \ t \ c \\ \mid \ map2\_s \ c \ t \mid map2\_s \ t \ c \mid map2\_s \ t \ t \ c \\ \mid \ reduce\_s \ c \ t \mid reduce\_s \ t \ c \\ \mid \ reduce\_t \ c \ t \mid reduce\_t \ t \ c \\ \mid \ reshape \ \sigma \ \sigma \ E
```

#### References

- [1] Thaddeus Koehn and Peter Athanas. 2016. Arbitrary streaming permutations with minimum memory and latency. In 2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD). IEEE, 1–6.
- [2] Xilinx, Inc. 2018. Zynq-7000 SoC Data Sheet: Overview. Xilinx, Inc.

Figure 1. L<sup>seq</sup> Typing Rules

Figure 2. L<sup>seq</sup> Primitive Reduction Rules

**Figure 3.** Lst typing rules. Duplicates from the Lseq are omitted.

$$\begin{split} & \text{shift\_s} \ n' \ [v_1, \dots, v_n] \leadsto^p \ [undef, \dots, v_1, \dots, v_{n-n'}]_t \\ & \text{shift\_t} \ n' \ [v_1, \dots, v_n, \dots, v_{n+i}]_t \leadsto^p \ [undef, \dots, v_1, \dots, v_{n-n'}, \dots, v_{n+i}]_t \\ & \text{up\_1d\_s} \ n \ [v]_s \leadsto^p \ [v, \dots, v]_s \\ & \text{up\_1d\_t} \ n \ [v, \dots, v_{n+i}]_t \leadsto^p \ [v, \dots, v_{n+i}]_t \\ & \text{select\_1d\_s} \ n' \ [v_1, \dots, v_n]_s \leadsto^p \ [v_{n'}]_s \\ & \text{select\_1d\_t} \ n' \ [v_1, \dots, v_n, \dots, v_{n+i}]_t \leadsto^p \ [v_{n'}, \dots, v_{n+i}]_t \\ & \text{map\_s} \ (\lambda x : \tau. t) \ [v_1, \dots, v_n]_s \leadsto^p \ [(\lambda x_1 : \tau. t) \ v_1, \dots, (\lambda x_n : \tau. t) \ v_n]_s \\ & \text{map\_t} \ (\lambda x : \tau. t) \ [v_1, \dots, v_n, \dots, v_{n+i}]_t \leadsto^p \ [(\lambda x_1 : \tau. t) \ v_1, \dots, (\lambda x_n : \tau. t) \ v_n, \dots, v_{n+i}]_t \\ & \text{map2\_s} \ (\lambda x : \tau. (\lambda x' : \tau'. t)) \ [v_1, \dots, v_n]_s \ [v'_1, \dots, v'_n]_s \leadsto^p \ [(\lambda x_1 : \tau. (\lambda x'_1 : \tau'. t)) \ v_1 \ v'_1, \dots, (\lambda x_n : \tau. (\lambda x'_n : \tau'. t)) \ v_n \ v'_n]_s \\ & \text{map2\_t} \ (\lambda x : \tau. (\lambda x' : \tau. t)) \ [v_1, \dots, v_n, \dots, v_{n+i}]_t \ [v'_1, \dots, v'_n, \dots, v'_{n+i}]_t \\ & \text{preduce\_s} \ (\lambda x : \tau \times \tau. t) \ [v_1, \dots, v_n]_s \leadsto^p \ [(\lambda x_1 : \tau. \tau. t) \ v_1 \ ((\lambda x_2 : \tau \times \tau. t) \ v_2 \ (\dots, ((\lambda x_n : \tau \times \tau. t) \ (v_1, v_2)))))]_s \\ & \text{reduce\_t} \ (\lambda x : \tau \times \tau. t) \ [v_1, \dots, v_n, \dots, v_{n+i}]_t \rightsquigarrow^p \ [(\lambda x_1 : \tau \times \tau. t) \ v_1 \ ((\lambda x_2 : \tau \times \tau. t) \ v_2 \ (\dots, ((\lambda x_n : \tau \times \tau. t) \ (v_1, v_2))))), \dots, v_{n+i}]_t \\ & \text{reshape} \ \sigma \ \sigma' \ v \leadsto^p \ v' \ \text{s.t. v'} \ \text{and v are equal when converted to a flat seq} \end{aligned}$$

Figure 4. Lst Primitive Reduction Rules. Duplicates from the Lseq are omitted.