# Type-Directed Scheduling Of Streaming Accelerators – Technical Appendix

David Durst<sup>1</sup>, Matthew Feldman<sup>1</sup>, Dillon Huff<sup>1</sup>, David Akeley<sup>2</sup>, Ross Daly<sup>1</sup>, Gilbert Bernstein<sup>3</sup>,
Marco Patrignani<sup>1,4</sup>, Kayvon Fatahalian<sup>1</sup>, Pat Hanrahan<sup>1</sup>
Stanford University, USA<sup>1</sup> University of California, Los Angeles, USA<sup>2</sup> University of California, Berkeley, USA<sup>3</sup>
CISPA Helmholtz Center for Information Security, Germany<sup>4</sup>

CCS Concepts: • Hardware → Hardware description languages and compilation; • Software and its engineering → Data types and structures; Data flow languages; • Computer systems organization → Data flow architectures

#### **ACM Reference Format:**

David Durst<sup>1</sup>, Matthew Feldman<sup>1</sup>, Dillon Huff<sup>1</sup>, David Akeley<sup>2</sup>, Ross Daly<sup>1</sup>, Gilbert Bernstein<sup>3</sup>, Marco Patrignani<sup>1,4</sup>, Kayvon Fatahalian<sup>1</sup>, Pat Hanrahan<sup>1</sup>. 2020. Type-Directed Scheduling Of Streaming Accelerators – Technical Appendix. In *Proceedings of the 41st ACM SIGPLAN International Conference on Programming Language Design and Implementation (PLDI '20)*, June 15–20, 2020, London, UK. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3385412. 3385983

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 formalises  $\mathbf{L}^{seq}$  and Appendix C formalises  $\mathbf{L}^{st}$ .

# 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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

```
PLDI '20, June 15–20, 2020, London, UK
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7613-6/20/06...$15.00
https://doi.org/10.1145/3385412.3385983
```

```
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 L<sup>st</sup> 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.

```
delay (add) = 0
  delay (tuple) = 0
  delay (map_s f) = delay (f)
  delay (map_t f) = delay (f)
  delay (reduce_s f) = 0
  delay (reduce_t f) = n - 1
where the input has type TSeq n i t
  delay (shift_s) = 0
  delay (shift_t) = 0
  delay (select_1d_s) = 0
  delay (select_1d_t) = j - 1
e selecting the ith element
```

where the selecting the jth element

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

#### **B.1** Terms

```
\begin{array}{ll} 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{tuple\_to\_seq} \ t \mid \mathsf{seq\_to\_tuple} \ t \\ \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 \{\vee, \wedge\} \\ \mid \ \mathsf{const\_gen} \ t \\ \mid \ \mathsf{shift} \ t \ t \mid \mathsf{up\_1d} \ t \ t \mid \mathsf{select\_1d} \ t \ t \\ \mid \ \mathsf{partition} \ t \ t \ t \mid \mathsf{unpartition} \ t \\ \mid \ \mathsf{map} \ t \ t \mid \mathsf{map2} \ t \ t \ | \ \mathsf{reduce} \ t \ t \\ \end{array}
```

#### **B.2** Values

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

#### **B.3** Types

```
\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 &::= \left[ \cdot \right] \mid E \; t \; | \; (\lambda x : \tau . \; t) \; E \\ &\mid \; [E, \ldots, t_n] \mid [v_1, \ldots, E, \ldots, t_n] \mid [v_1, \ldots, v_{n-1}, E] \mid \langle E, \ldots, t_n \rangle \\ &\mid \; \langle v_1, \ldots, E, \ldots, t_n \rangle \mid \langle v_1, \ldots, v_{n-1}, E \rangle \\ &\mid \; \text{tuple\_to\_seq} \; E \mid \text{seq\_to\_tuple} \; E \\ &\mid \; \text{not} \; E \mid E = t \mid v == E \mid E \; \text{op} \; t \mid n \; \text{op} \; E \\ &\mid \; \text{bop} \; t \mid b \; \text{bop} \; E \\ &\mid \; \text{lut\_gen} \; E \; t \mid \text{lut\_gen} \; v \; E \mid \text{const\_gen} \; E \\ &\mid \; \text{shift} \; n \; E \mid \text{up\_1d} \; n \; E \mid \text{select\_1d} \; n \; E \\ &\mid \; \text{partition} \; n \; n \; E \mid \text{unpartition} \; E \\ &\mid \; \text{map} \; E \; t \mid \text{map} \; v \; E \mid \text{map2} \; v \; E \; t \mid \text{map2} \; v \; E \; t \mid \text{map2} \; v \; E \\ &\mid \; \text{reduce} \; E \; t \mid \text{reduce} \; v \; E \end{split}
```

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

```
\begin{array}{l} \mathbf{c} ::= \left[ \cdot \right] \mid \& t \mid t \& \mid \lambda x : \tau . \& \\ \mid \left[ [\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 \mathsf{tuple\_to\_seq} \ \mathsf{c} \mid \mathsf{seq\_to\_tuple} \ \mathsf{c} \\ \mid \mathsf{not} \ \mathsf{c} \mid \mathsf{c} = t \mid t == \mathfrak{c} \mid \mathsf{c} \ \mathsf{op} \ t \mid n \ \mathsf{op} \ \mathsf{c} \\ \mid \mathsf{c} \ \mathsf{bop} \ t \mid t \ \mathsf{bop} \ \mathsf{c} \\ \mid \mathsf{lut\_gen} \ \mathsf{c} \ t \mid \mathsf{lut\_gen} \ t \ \mathsf{c} \mid \mathsf{const\_gen} \ \mathsf{c} \\ \mid \mathsf{shift} \ n \ \mathsf{c} \mid \mathsf{up\_1d} \ n \ \mathsf{c} \mid \mathsf{select\_1d} \ n \ \mathsf{c} \\ \mid \mathsf{partition} \ n \ n \ \mathsf{c} \mid \mathsf{unpartition} \ \mathsf{c} \\ \mid \mathsf{map} \ \mathsf{c} \ t \mid \mathsf{map2} \ t \ \mathsf{c} \ t \mid \mathsf{map2} \ t \ \mathsf{c} \ t \mid \mathsf{map2} \ t \ \mathsf{c} \ \mathsf{c} \\ \mid \mathsf{reduce} \ \mathsf{c} \ t \mid \mathsf{reduce} \ t \ \mathsf{c} \end{array}
```

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

**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.

# C.5 Lst Evaluation Contexts

```
 c := [\cdot] \mid c \ t \mid (\lambda x : \tau. \ t) \ c \\ \mid \ [c, \ldots, t_n]_s \mid [t_1, \ldots, c, \ldots, t_n]_s \mid [t_1, \ldots, t_{n-1}, c]_s \mid [c, \ldots, t_n]_t \\ \mid \ [t_1, \ldots, c, \ldots, t_n]_t \mid [t_1, \ldots, t_{n-1}, c]_t \\ \mid \ \langle c, \ldots, t_n \rangle \mid \langle t_1, \ldots, c, \ldots, t_n \rangle \mid \langle t_1, \ldots, 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{onst\_gen} \ c \\ \mid \ shift\_s \ n \ c \mid shift\_t \ n \ c \mid up\_1d\_s \ n \ c \mid up\_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 \ c \ t \mid map\_t \ t \ c
```

```
map2_s c t t | map2_s t c t | map2_s t t c map2_t c t t | map2_t t t t | map2_t t t c reduce_s t t | reduce_s t c reduce_t t t | reduce_t t t c 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.