GMU SHA Core Interface & Hash Function Performance Metrics
Interface
Why Interface Matters?

- Pin limit

Total number of i/o ports $\leq$ Total number of an FPGA i/o pins

- Support for the maximum throughput

Time to load the next message block $\leq$ Time to process current block
Interface: Two possible solutions

Length of the message communicated at the beginning

+ easy to implement passive source circuit

- area overhead for the counter of message bits

- more intelligent source circuit required

+ no need for internal message bit counter

msg_bitlen
message
zero_word

end_of_msg
SHA core
SHA Core: Interface & Typical Configuration

• SHA core is an active component; surrounding FIFOs are passive and widely available
• Input interface is separate from an output interface
• Processing a current block, reading the next block, and storing a result for the previous message can be all done in parallel
SHA Core Interface
SHA Core Interface + Surrounding FIFOs

Input FIFO

clk → input clk
rst → input rst

ext_idata → ext_idata
fifoin_full → fifoin_full
fifoin_empty → fifoin_empty
fifoin_write → fifoin_write
write → write
read → read

ich → input idata
fifoin_empty → fifoin_empty
fifoin_read → fifoin_read

Output FIFO

clk → output clk
rst → output rst

ext_odata → ext_odata
fifout_empty → fifout_empty
fifout_full → fifout_full
fifout_write → fifout_write
write → write
read → read
Communication Protocol for Unpadded Messages

a) $w \text{ bits}$

- \text{msg\_bitlen}

- \text{message}

- \text{zero\_word}$

b) $w \text{ bits}$

- \text{seg\_0\_bitlen}

- \text{seg\_0}

- \text{seg\_1\_bitlen}

- \text{seg\_1}

- \ldots

- \text{seg\_n-1\_bitlen}

- \text{seg\_n-1}

- \text{zero\_word}
SHA Core Interface with Additional Faster I/O Clock

Diagram of SHA core with interfaces:
- `io_clk`, `clk`, `rst`
- `w`
- `dout` and `w` inputs
- `src_ready` and `dst_ready` outputs
- `src_read` and `dst_write` inputs
SHA Core Interface with Two Clocks + Surrounding FIFOs

Input FIFO

- ext_idata
- w
- fifoin_full
- full
- write
- fifoin_write
- fifoin_empty
- empty
- read

SHA core

- iodata
- w
- src_ready
- src_read
- dst_ready
- dst_write

Output FIFO

- ext_odata
- w
- fifoout_full
- full
- write
- fifoout_write
- fifoout_empty
- empty
- read
Communication Protocol for Padded Messages Without Message Splitting

\[ \text{msg\_len\_ap} \mid \text{last} = 1 \]
\[ \text{msg\_len\_bp} \]

\[ w \text{ bits} \]

- \text{msg\_len\_ap} – message length after padding [bits]
- \text{msg\_len\_bp} – message length before padding [bits]
**Communication Protocol for Padded Messages With Message Splitting**

- **seg_0_len_ap** | **last=0**
  - **seg_0**

- **seg_1_len_ap** | **last=0**
  - **seg_1**

- **seg_n-1_len_ap** | **last=1**
  - **seg_n-1_len_bp**
  - **seg_n-1**

seg_i_len_ap – segment i length after padding*
[bits]

seg_i_len_bp – segment i length before padding
[bits]

* For all \( i < n-1 \) segment i length after padding is assumed to be a multiple of the message block size, \( b \) [characteristic to each function], and thus also the word size, \( w \).

The last segment cannot consist of only padding bits. It must include at least one message bit.
Performance Metrics
Performance Metrics - Speed

Throughput for Long Messages [Mbit/s]

Throughput for Short Messages [Mbit/s]

Execution Time for Short Messages [ns]

Allows for easy cross-comparison among implementations in software (microprocessors), FPGAs (various vendors), ASICs (various libraries)
Performance Metrics - Speed

Time to hash N blocks of message [cycles] = Htime(N)

The exact formula from analysis of a block diagram, confirmed by functional simulation.

Minimum Clock Period [ns] = T

From a place & route and/or static timing analysis report file.
Time to Hash \( N \) Blocks of the Message [clock cycles]

\[
\text{HTime}(N) = c_{\text{INIT}} + \left\lfloor \frac{c_{\text{IN}}}{r_{\text{IO}}} \right\rfloor + c_{\text{BLOCK}} \cdot N + c_{\text{FINAL}} + \left\lfloor \frac{c_{\text{OUT}}}{r_{\text{IO}}} \right\rfloor
\]

- \( c_{\text{INIT}} \) is the number of clock cycles necessary to establish communication with the source of data (typically, Input FIFO) and read the length of the message (in our formulas we assume that the length of the message is smaller than \( 2^w \)).
- \( c_{\text{IN}} \) is the number of clock cycles required to read the very first block of the message. \( c_{\text{IN}} = \text{Block\_size}/w \).
- \( c_{\text{BLOCK}} \) is the number of clock cycles required to process one block of the message.
- \( c_{\text{FINAL}} \) is the number of clock cycles required for the finalization. We assume that only one finalization is required per entire message (if the finalization needs to be repeated for every block of the message, its number of clock cycles is included in \( c_{\text{BLOCK}} \)).
- \( c_{\text{OUT}} \) is the number of clock cycles required to write hash value to the destination circuit (typically Output FIFO). \( c_{\text{OUT}} = \text{output\_size}/w \).
Performance Metrics - Speed

Minimum time to hash N blocks of message [ns] = $H_{time}(N) \cdot T$

Maximum Throughput (for long messages) = $\frac{block\_size}{T \cdot (H_{time}(N+1) - H_{time}(N))}$

Effective maximum throughput for short messages:

$$Throughput_{eff} = \frac{N \cdot block\_size}{(T \cdot H_{Time}(N))}.$$
Performance Metrics - Speed

Maximum Throughput (for long messages) = \frac{\text{block_size}}{T \times \text{block_processing_time}}

- From specification
- From place & route report and/or static timing analysis report
- From analysis of block diagram and/or functional simulation
Performance Metrics - Area

Resource Utilization\textsubscript{Spartan3} = (\#CLB slices, \#BRAMs, \#MULs)  
Resource Utilization\textsubscript{Cyclone III} = (\#LE, \#memory\_bits, \#MULs).

For the basic, folded, and unrolled architectures, we force these vectors to look as follows through the synthesis and implementation options:

\begin{align*}
\text{Resource Utilization}\textsubscript{Spartan3} &= (\#CLB slices, 0, 0) \\
\text{Resource Utilization}\textsubscript{Cyclone III} &= (\#LE, 0, 0) \\
\end{align*}
Choice of Optimization Target

Primary Optimization Target: **Throughput to Area Ratio**

Features:

- practical: good balance between speed and cost
- very reliable guide through the entire design process, facilitating the choice of
  - high-level architecture
  - implementation of basic components
  - choice of tool options
- leads to high-speed, close-to-maximum-throughput designs
Our Design Flow

- Specification
  - Datapath
    - Block diagram
  - VHDL Code
  - Formulas for Throughput & Hash time

- Interface
  - Controller
    - ASM Chart
  - Max. Clock Freq.
  - Resource Utilization

- Library of Basic Components
- Controller Template

Throughput, Area, Throughput/Area, Hash Time for Short Messages
How to compare hardware speed vs. software speed?

EBASH reports (http://bench.cr.yp.to/results-hash.html)

In graphs

Time(n) = Time in clock cycles vs. message size in bytes for n-byte messages, with n=0,1, 2, 3, … 2048, 4096

In tables

Performance in cycles/byte for n=8, 64, 576, 1536, 4096, long msg

Performance for long message = \frac{\text{Time}(4096) - \text{Time}(2048)}{2048}
How to compare hardware speed vs. software speed?

Throughput [Gbit/s] = \[ \frac{8 \text{ bits/byte} \cdot \text{clock frequency [GHz]}}{\text{Performance for long message [cycles/byte]}} \]