# Homework 3 – Number Format and Recursive Algorithm

## Objective

In this homework, you will design a hardware kernel for a simple recursive algorithm using an optimized number format. The function to be implemented is defined as $z = x^{100} + x^{99} + \cdots + x^2 + x$, which can be calculated by the following recursive algorithm:

$$
\begin{aligned}
&y = 1;\\
&z = 0;\\
&for\ (i = 1; i \le 100; i = i + 1)\ \{\\
&\quad y = y \times x;\\
&\quad z = z + y;\\
&\}.
\end{aligned}
$$

Note that both the operations insides the for loop have loop-carried data dependency.

## Tasks

1. Given that the input vector of this function is generated by the MATLAB function $x = (rand(n, 1) - 0.5) * 4$, the input samples has a range of $x \in (-2, 2)$. Please design a fixed-point data format ($W_{Tot}$, $T_{Fr}$) that can well represent the input vector with a quantization error of $|e| \le 0.005$. Explain how your fixed-point data format is derived.
2. Given the fixed-point data format at the input $x$ ($W_{Tot}$, $T_{Fr}$), what is the required fixed-point data format at the output $z$ ($W_{Tot}$, $T_{Fr}$) that can keep full precision of the computation results. Explain how your fixed-point data format is derived.

   Hint: Full precision means that the no quantization error is introduced by any part of the arithmetic units.
3. Due to the large dynamic range of the system, keeping full precision of the results is impractical. If we only care about keeping the full precision of 5 significant digits of the results (in decimal), what is the required floating-point data format ($W_{Man}$, $W_{Exp}$) that should be used by the entire system? Explain how your floating-point data format is derived.

   Hint: The word length of mantissa ($W_{Man}$) determines the number of significant digits of the results, while the word length of exponent ($W_{Exp}$) determines the range of the results. Assume the floating-point data format has a bias of 2^($W_{Exp}$-1)-1 by default. Assume $0.01 \le |x| < 2$.

Prof. Fengbo Ren

School of Computing, Informatics, and Decision Systems—Arizona State University

4. Design a highly recursive hardware architecture implementing the function with the customized floating-point data format that you derive in Task 3. To save area, the architecture should be maximally folded (unfolding factor = 1) using one multiplier and one adder only. Map this architecture into a Xilinx Virtex 7 XC7VX485T-2FFG1761 FPGA target. Show the synthesized system frequency and resource usage. Take a screen shot to show your proposed recursive architecture and label the latency of each arithmetic blocks. Explain what the loop latency and the number of iterations required for producing a valid result $z$ are in this architecture. To achieve the maximum data processing throughput, what is the interleaving factor that should be used? Based on this interleaving factor, what is the actual data processing throughput (measured at the output) of the FPGA implementation?

   Hint: MUXes and registers are allowed. You may have more than one loops on the data path, in which case the loop latencies must be balanced. The data processing throughput is measured as how many valid results of $z$ can be expected at the output per second. It is a function of the loop latency, the interleaving factor, the number of iteration required, the unfolding factor, and the system frequency.

5. Unfold and redesign the recursive architecture with an unfolding factor of 4. Map the unfolded (unfolding factor = 4) architecture into a Xilinx Virtex 7 XC7VX485T-2FFG1761 FPGA target. Show the synthesized system frequency and resource usage. Take a screen shot to show your proposed recursive architecture and label the latency of each arithmetic blocks. Explain what the loop latency and the number of iterations required for producing a valid result $z$ are in this architecture. To achieve the maximum data processing throughput, what is the interleaving factor that should be used? Based on this interleaving factor, what is the actual data processing throughput (measured at the output) of the FPGA implementation?
6. Based on the implementation results from Task 5 and 6, discuss the impact of architectural interleaving and unfolding on the system throughput and resource utilization of an FPGA implementation, respectively.

## Due Date

11:59pm on Mar. 29th, Wednesday.

## Instructions on Submission

- Submit your Simulink models in Task 4 and 5.
  - Include the MATLAB scripts for configuring the test bench if any. Your submission should allow the grader to verify your design by running the Simulink test bench.
- Submit your solution in PDF format.

- Put all of your files into a folder first and then compress the submission folder into a zip archive file named cen598-<lastname>-hw3.zip and submit it on Blackboard.