CSE 246: Computer Arithmetic Algorithms and Hardware Design

Course Project: Design of Fixed-Point Parallel-Form Digital Filters

Vikas Paliwal

Problem Statement

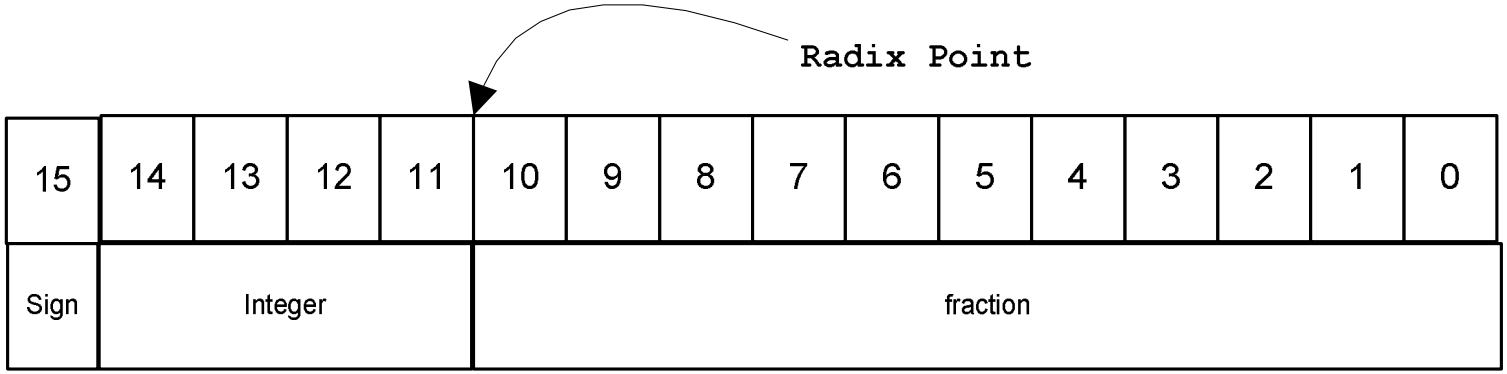
- To derive a technique for designing fixed-point IIR filters in parallel form, i.e. the radix location, bit widths for fractional and integral parts to achieve the target of error bounds and overflow avoidance.
- Study based on analysis and modeling, study of digital filter principles and existing research on direct-form IIR realizations

Agenda:

- Fixed-Point Arithmetic
  - DSPs versus FPGAs
- Digital Filters
  - Parallel Forms
  - Finite word challenges
- Parallel form digital filter design
  - Integer bit-length
  - Fraction bit-length

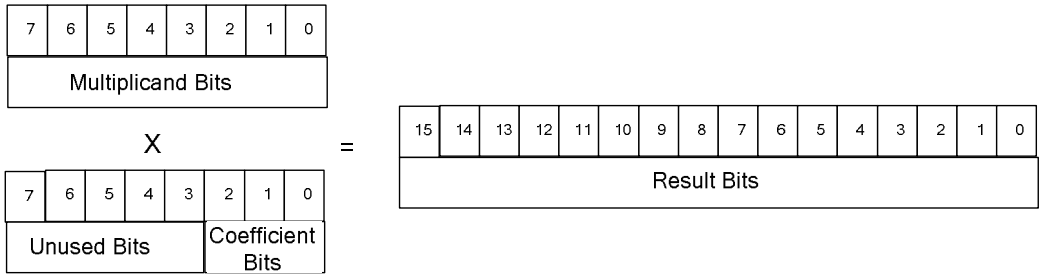
Fixed Point Arithmetic

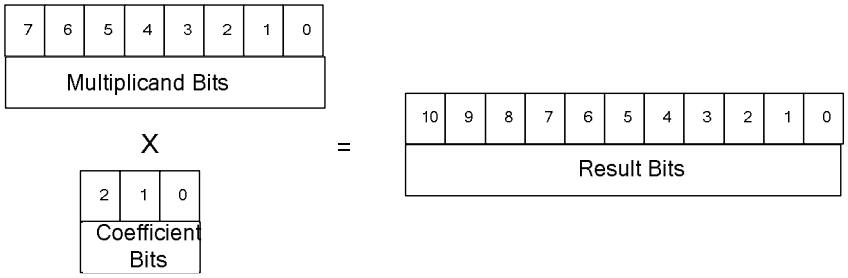
- Fixed point numbers have fixed number of decimal points before and after the radix point.
- The dynamic range of such numbers is limited due to fixed radix position as compared to floating point.



DSPs vs FPGAs Fixed Point design

- DSPs have fixed register lengths, thus a single fixed point used throughout
- FPGAs in contrast allow separate fixed point format for each signal





Digital Filters

- Do the filtering process for incoming digital samples by means of mathematical operations
- Performance of filters is closely tied to its pole locations
- Performance of filter sensitive to accuracy in mathematical operations
- Mathematical representation facilitates realization in several forms

IIR Filters

- Infinite impulse response (IIR) filters have non-zero response for an impulse over infinite time
- Can be represented in Auto-Regressive Moving Average (ARMA) equations of type:

$$y[n] = \sum_{m=1}^{N} a_m y[n-m] + \sum_{m=0}^{N} b_m x[n-m]$$

$$H[z] = \frac{b_0 + b_1 z - 1 + b_2 z^{-2} \ldots + b_M z^{-M}}{1 + a_1 z - 1 + a_2 z^{-2} \ldots + a_N z^{-N}}$$

Parallel-form IIR filters

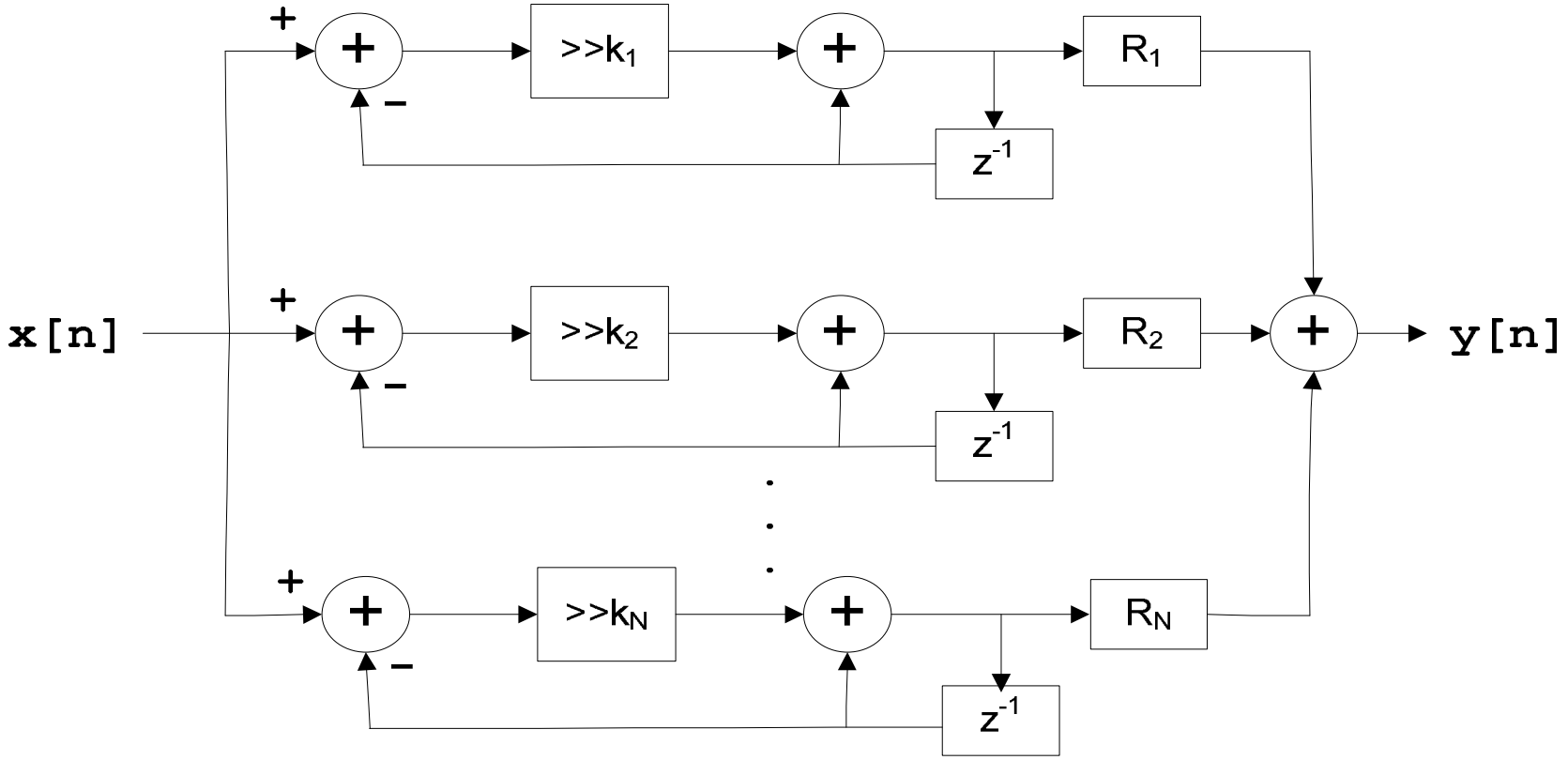
- Direct-form of filters is common but often parallel form is easier and simpler for particular applications

$$H[z] = \sum_{m=1}^{N} \frac{R_m}{1 - p_m z^{-1}}$$

- Discrete time filters with short sampling periods have all poles around 1. Binary poles of type $p_m = (1 - \frac{1}{2^{k_m}})$ are very common due to ease of calculations

Parallel form IIR filter

- Parallel form IIR filter with binary poles can be easily realized using simple bit shifts and multiplication



Finite word effects in IIR filters

- The finite word length used to represent the calculated variables results in loss of computational accuracy
- Example, in the parallel form IIR filter with binary coefficients, information is lost in each branch due to right bit shifts
- The truncation loss is carried over in IIR filter and possibly amplified as well

Residue Bit Width in IIR filters

- The integer part of m-th residue is simply $\lceil log_2 R_m \rceil + 1$
- The bits required for the fractional part of the residue can be derived from the desired level of accuracy for the zeroes with respect to the unit circle e.g. for a 2nd order filter, bits required for residue R1 assuming R2 to be accurate for δ perturbation in zero.

$$\lceil -log_2(\frac{\delta(R_1+R_2)}{R_2(p_2-p_1)}) \rceil + 1$$

Bit overflow avoidance

- It is desirable to have enough integer bits in the intermediate signals and final outputs so that overflow never occurs.
- Bit overflow in signal v[n] using the upper bound on input signal x[n] and the transfer function to v[n], h_{vx}.

$$|v_{max}| \leq \{\sum_{k=0}^{\infty} |h_{v,x}[k]|\} \cdot |x_{max}|$$

Bit overflow avoidance

- The overflow avoidance bound can be found using the inverse z-transform in each of the branches of parallel architecture.
- Example, for single pole IIR filter $H(z) = \frac{R_1}{1-p_1 z^{-1}}$, maximum value of output signal is $\frac{R_1 x_{max}}{1-p_1}$, so minimum number of bits to be used for integral part is

$$\lceil log_2(\frac{R_1 x_{max}}{1-p_1}) \rceil + 1$$

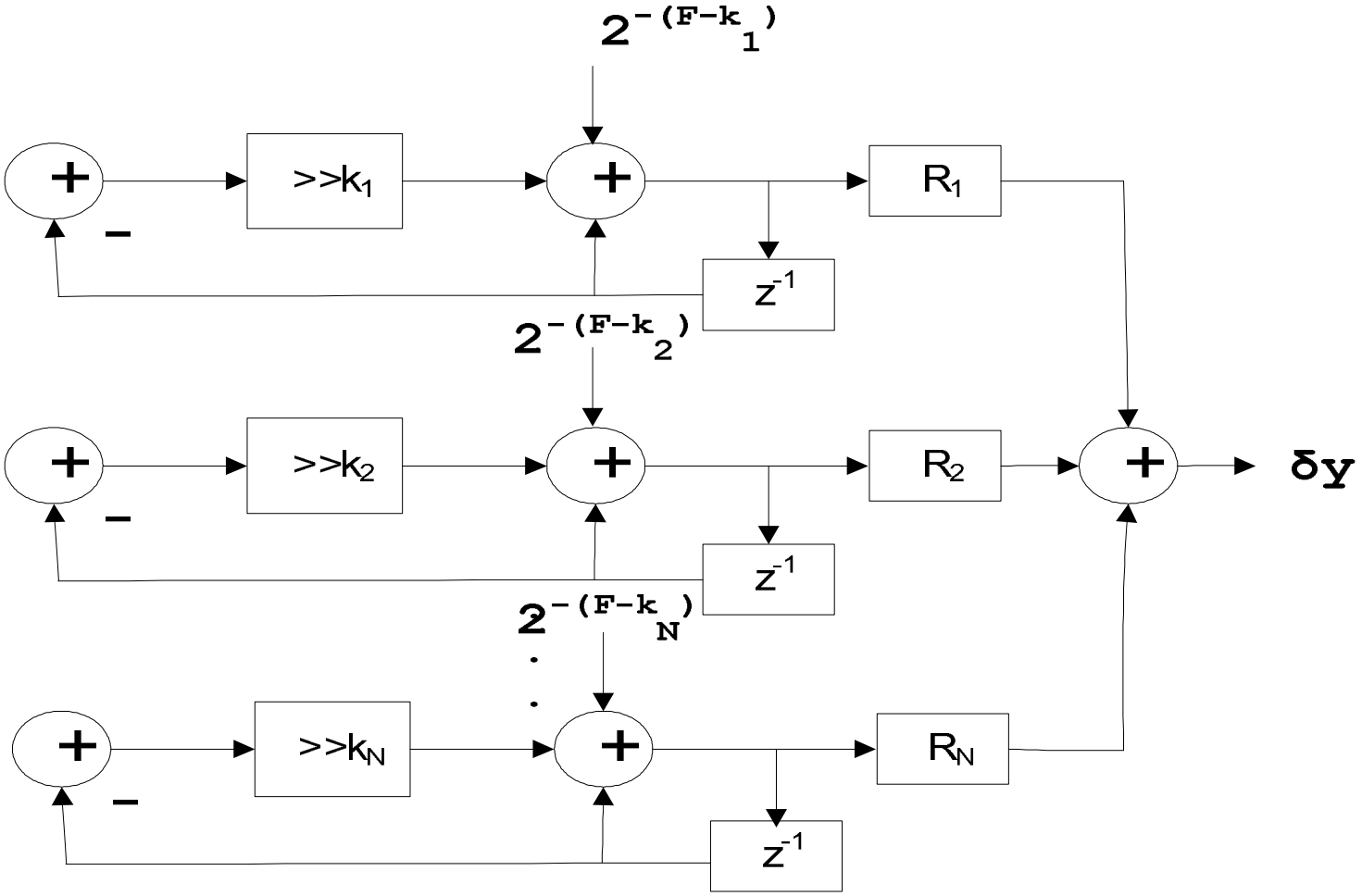
Bit width for fractional parts

- The fractional part bit width is driven the desired level of error bound in outputs
- It can be understood that the right bit shifts in filter realization result in truncation losses
- Truncation losses can get keep accumulating and produce large errors
- Thus quantifying the errors is needed

Output error control

- One approach to model the impact of truncation errors is keep feeding the maximum possible permissible error in the nearest summation branches and derive the bounds based on transfer functions from individual errors to output
- In parallel-form realization, truncations are done in each branch which needs to be modeled differently as opposed to direct-form where all error is lumped in a single adder.

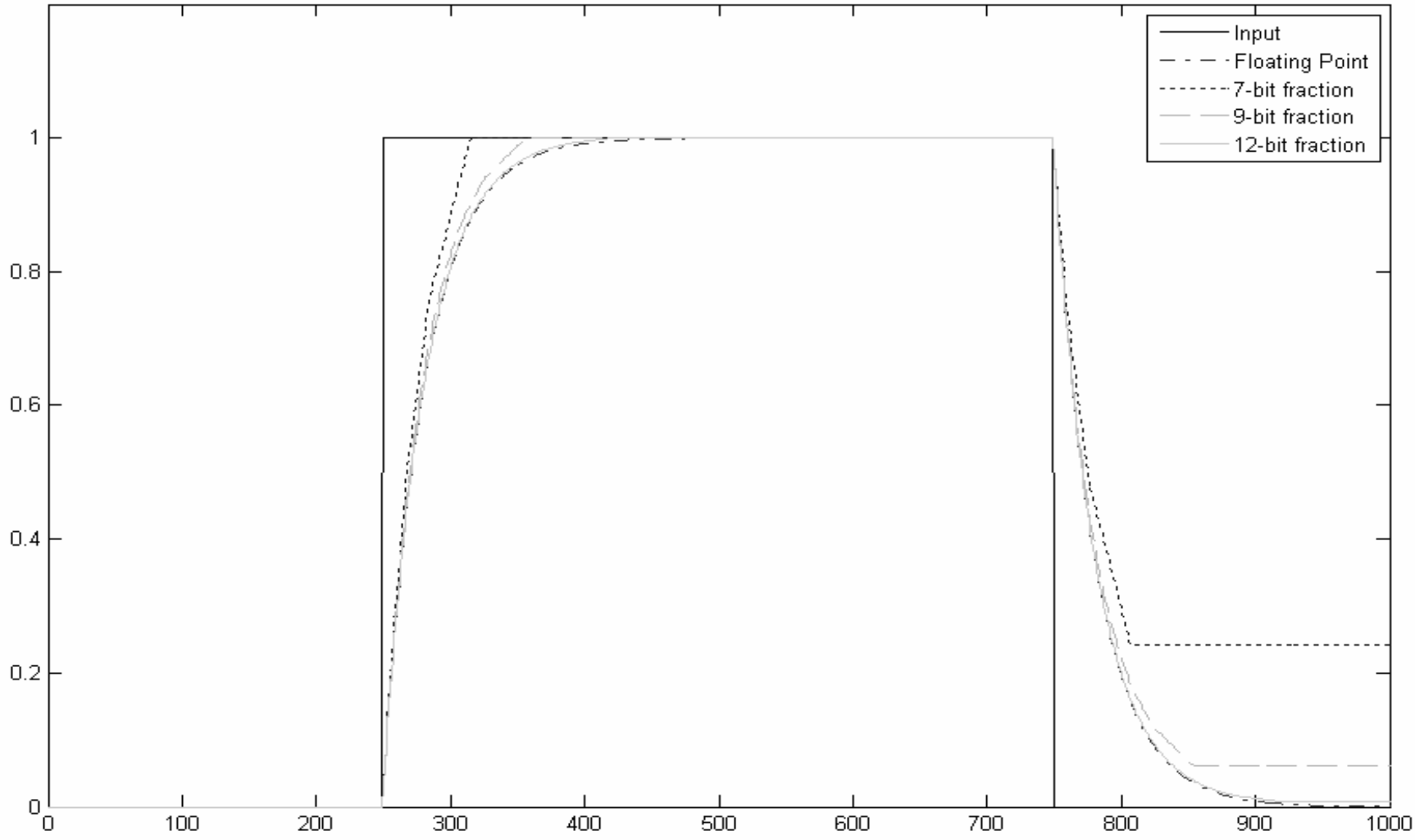
Truncation Error Modeling



Truncation Error Modeling

- Assuming that each right shift in m-th branch can produce a maximum truncation error of $2^{-(F-k_1)}$ where F is the number of fraction bits used for representing the intermediate filter value.
- This error can be fed to adder after the shift register, and a criterion for total permissible error can be derived using the transfer function $|\delta y_{max}| = \sum_{m=1}^{N} R_m 2^{-(F-2k_m)}$ final output,
- Required number of fractional bits in output are $log_2|\delta y_{max}|$

Truncation Error Results for single-pole IIR filter



Complete IIR filter design approach

- The parallel-form for the transfer function is realized
- The bit widths for the residue values is calculated based on integral part and permissible error in zero placements
- The integral part of intermediate variables and output is calculated using maximum bound theorem
- Truncation error is modeled in each branch and summed to derive the number of fraction bits.

References

- Randy Yates, Practical Considerations in Fixed-Point FIR Filter Implementations, Digital Signal Labs White paper, November 2006
- J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications,*Prentice-Hall, New Jersey, 1996.
- J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory,* Macmillan, New York, 1992.
- Joan Carletta Robert Veillette Frederick Krach Zhengwei Fang, *Determining Appropriate Precisions for Signals in Fixed-point IIR Filters,* DAC 2003, June 2-6,2003, Anaheim, Califomia, USA.
- Andrew G. Dempster, and Malcolm D. Macleod, Use of Minimum-Adder Multiplier Blocks in FIR Digital Filters, IEEE Transactions on Circuits and Systems-11; Analog and Digital Signal Processing, Vol. 42, No. 9, September 1995
- Nurhan Karaboğa1, Bahadır Çetinkaya1 , *Efficient Design of Fixed Point Digital FIR Filters by Using Differential Evolution Algorithm,* Volume 3512 Computational Intelligence and Bioinspired Systems.

Questions

- ????