Design and Implementation of a Variable Order Markov Modeling Unit for Data Compression

Second European SystemC Users Group Meeting
Munich, Germany

Claudia Feregrino Uribe
C.Feregrino-Uribe@lboro.ac.uk

www.lboro.ac.uk/departments/el/research/sys/index.html

June 30th, 2000

Loughborough University

Presentation Structure

- Introduction to Data Compression
- X-Match Data Compressor
- PPM and Shift Models
- Shift Model Architecture
- Design in SystemC
- Conclusions

Data Compression

- What is it?

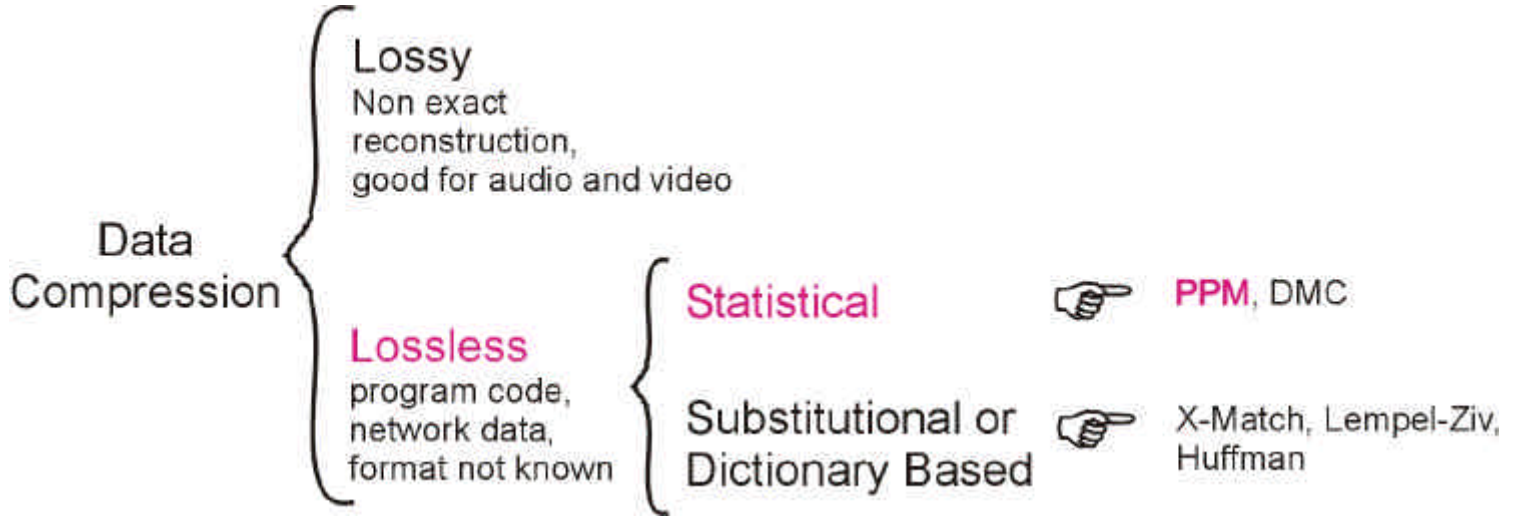
  Identification and removal of redundancy from a piece of data

- Decompression

  The reinsertion of some or all of that redundancy

- Compression Effects:
  - Reduces the bandwidth required to transmit information increasing the capacity of a data channel
    - faster, cheaper transmission
  - Reduces the physical storage requirements of a block of data
    - cheaper storage, lower power

Data Compression Classification



Current Compressors

Commercial Chips

IBM, AHA, Hi/fn and DCP

- Dictionary based
- Single symbol at a time
- Compression speeds 80 Mbytes/sec

X-Match class of data compressors

- CAM based, Selectively-Shiftable for model updating
- Partial Matching for improved compression
- Compress 4 bytes at a time - 176 Mbytes/sec

Commercial Chips

DEVELOPERS	IBM	AHA	STAC Electronics	DCP	Loughborough University	
CHIP	ALDC1-40S	AHA3231	Hi/fn 9610	DCP816	X-Match	
PROCESS	IBM CMOS 0.8 μ triple-level gate array/std cell	0.5 μ CMOS	0.35 μ gate array/std cell	1.0 μ CMOS gate array	ACTEL 0.25 μ FLASH CMOS FPGA	ALTERA 0.18 μ SRAM CMOS FPGA
COMPLEXITY	Not Stated	Not Stated	100 Kgates	15 Kgates	61 Kgates	61 Kgates
CLOCK SPEED	40 MHz	40 MHz	40 MHz	40 MHz	33 MHz	44 MHz
THROUGHPUT	40 MB/s	20 MB/s	80 MB/s	210 KB/s	132 MB/s	176 MB/s
COMPRESSION RATIO	0.44	0.51	0.43	0.47	0.53	0.53

Presentation Structure

- Introduction to Data Compression
- X-Match Data Compressor
- PPM and Shift Models
- Shift Model Architecture
- Design in SystemC
- Conclusions

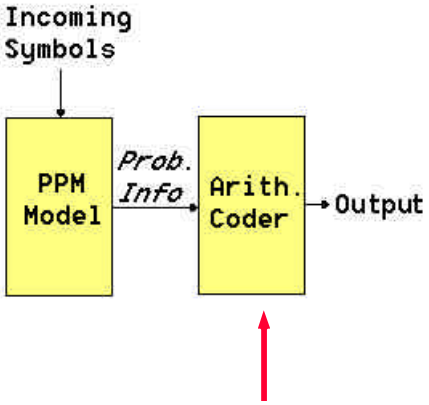
PPM Algorithm

MODEL

- Leading algorithm in lossless data compression
- Adaptive - based on a variable length context
  - *Example:* English alphabet
    u - 2.4% qu - 99.1%
- Variable order

	ABC	BC	C	C
Context order	2''	1''	0''	-1''

- Its order dictated by the longest context



$$\text{range} = \text{high} - \text{low};$$

$$\text{high} = \text{low} + \text{range} * \frac{CF_{S_{i-1}}}{CF_{S_0}} - 1;$$

$$\text{low} = \text{low} + \text{range} * \frac{CF_{S_i}}{CF_{S_0}};$$

CODER

Shift Model

- A fast hardware-friendly model

- Eliminates divide operation in arithmetic coding
  - Much faster operation
  - Keep CF_{S0} constant

$$range = high - low;$$

$$high = low + range * \frac{CF_{S_{i-1}}}{CF_{S_0}} - 1;$$

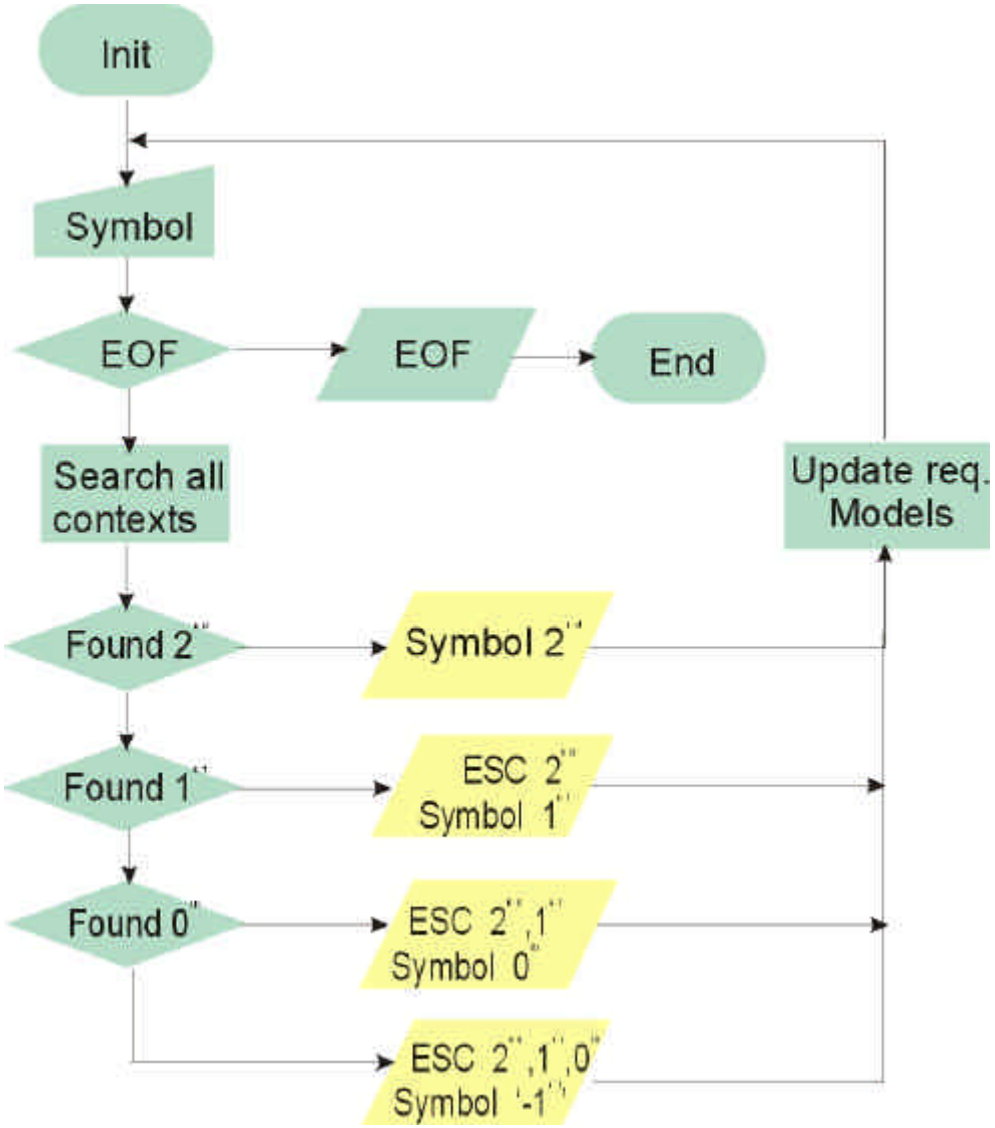
$$low = low + range * \frac{CF_{S_i}}{CF_{S_0}};$$

- Tokens: Number assigned to a symbol proportional to its occurrence in the input stream.
  - ➡ Total number of tokens CF_{S0} is fixed

- Adjusting symbol probabilities by reallocating tokens
  - Incoming symbol gets more tokens
  - Other symbols donate tokens to incoming one
  - CF_{S0} remains unchanged

- Token allocation must be: simple, fast and accurate

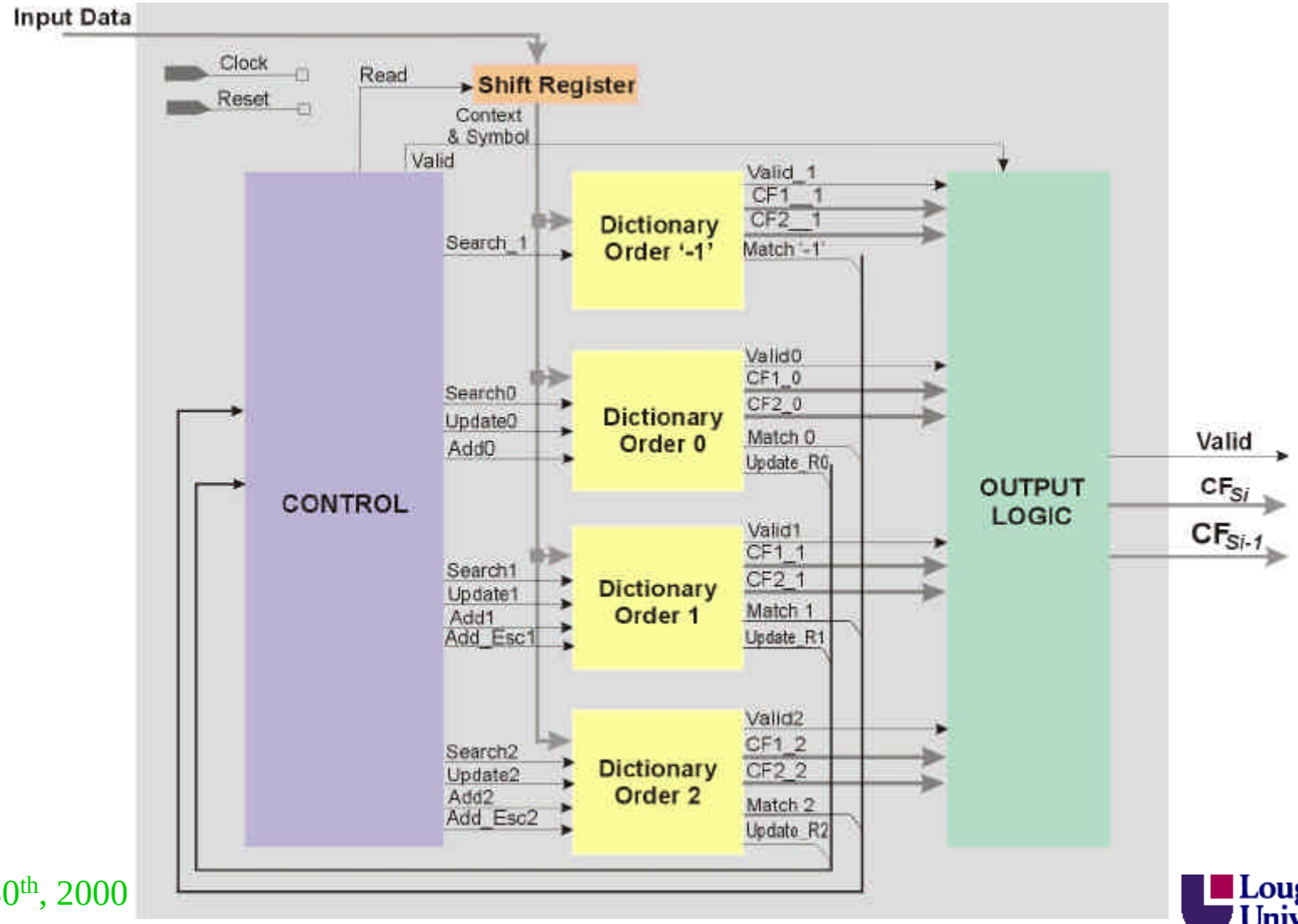
Shift Model Flow Diagram



Presentation Structure

- Introduction to Data Compression
- X-Match Data Compressor
- PPM and Shift Models
- Shift Model Architecture
- Design in SystemC
- Conclusions

Shift Model Architecture

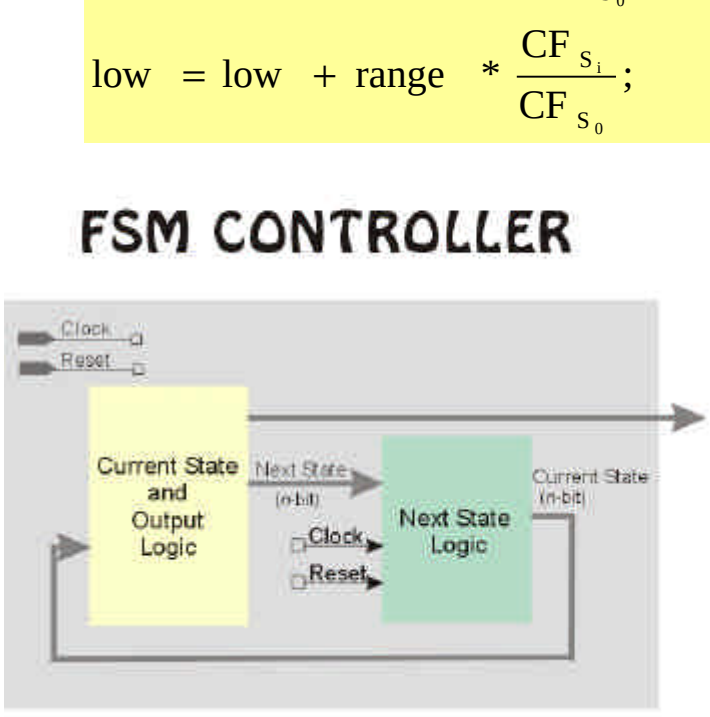


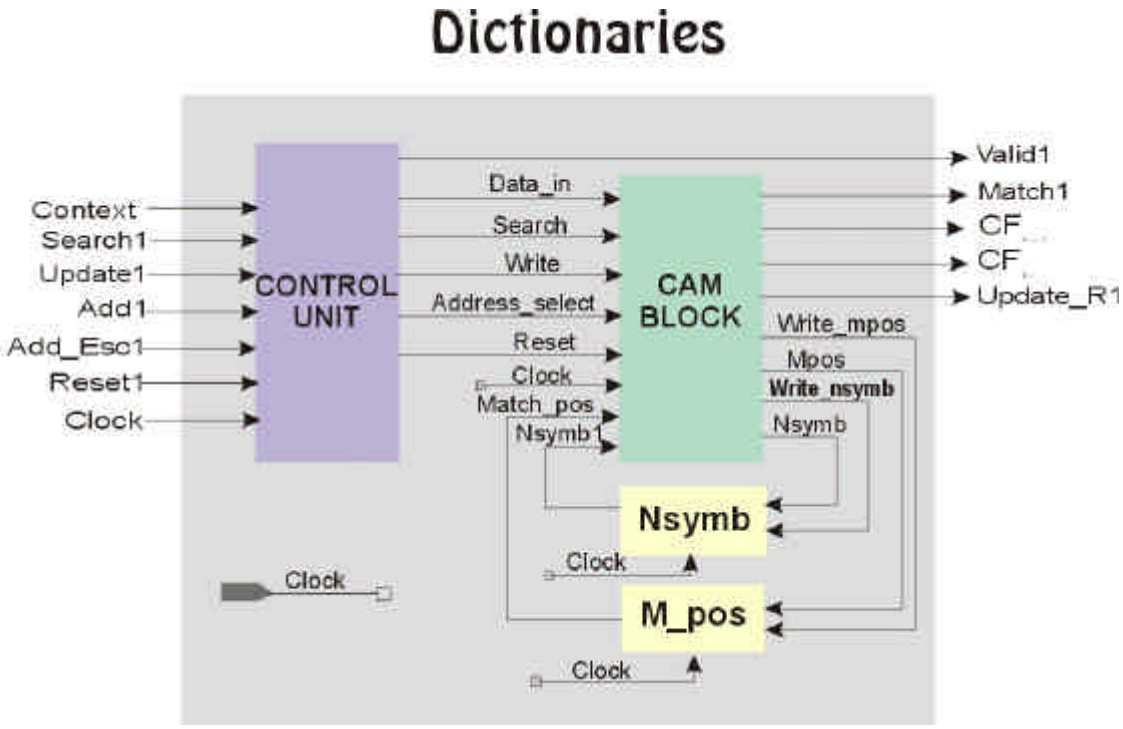
Shift Model Architecture (2)

$$range = high - low;$$

$$high = low + range * \frac{CF_{S_{i-1}}}{CF_{S_0}} - 1;$$

$$low = low + range * \frac{CF_{S_i}}{CF_{S_0}};$$





Presentation Structure

- Introduction to Data Compression
- X-Match Data Compressor
- PPM and Shift Models
- Shift Model Architecture
- Design in SystemC
- Conclusions

Shift Model design in SystemC

- Shift Model is complex - but not difficult to model
  - Previous software algorithm
  - Familiarity with C++ and VHDL languages
- High-level functional model
- Exploits concurrency of SystemC
  - Search and Update operations
- Functionality described in processes - Methods
- Large number of ports - Named Connections
- User Defined Data Types

SystemC Characteristics

- Structured hierarchical design methodology
- Model functionality - easy to implement
- Hardware-oriented with C++ flexibility
- Cycle-accurate modeling
- High-speed simulation

SystemC - for first time users

- Follow examples but don't trust them blindly
- Help from the SystemC forum
- Do not think sequentially as C or C++ - for non HDL users

Presentation Structure

- Introduction to Data Compression
- X-Match Data Compressor
- PPM and Shift Models
- Shift Model Architecture
- Design in SystemC
- Conclusions

Conclusions

GOOD POINTS

- Cycle-accurate model
- High speed simulation
- Hierarchical design
- Hw-oriented with C++ flexibility
- Help from the forum

BAD POINTS

- Errors in examples
- Complex waveform tracing of internal signals

Questions ?