Efficient Multiple Regular Expression Matching on FPGAs based on Extended SHIFT-AND Method

*Yusaku Kaneta, Shingo Yoshizawa, Shin-ichi Minato, Hiroki Arimura, and Yoshikazu Miyanaga
(Graduate School of IST, Hokkaido University, Japan)

- **Large-scale pattern matching problem on hardwares**
  - A large number (e.g. thousands) of complex (e.g. regular expressions) against high-speed data streams (e.g. of several Gbps)
  - Static compilation hardware vs Dynamic reconfigurable hardware

- **Our Static BP-NFA architecture**
  - Static compilation hardware for a subclass of regular expressions, called linear regular expressions, based on bit-parallel method
  - Construction of our Static BP-NFA:
    1. Linear regular expression: \( R \)
      \[
      R = A \ B# \ [^A] \ B# \ C
      \]
    2. NFA: \( N_R \)
    3. Circuit on FPGA: \( M_R \)

Pattern matching module for \( R \)

High-speed data stream

Bit-parallel method is one of the pattern matching techniques on a general CPU
Experimental results

- **Performance evaluation and resource usage**
  - Throughput: \( T = 1.6 \) Gbps
  - Number of PMMs: \( N = 1500 \) PMMs (19987 total chars)
  - Compilation time: \( C = 6000 \) sec for \( N = 1500 \)
    - Too expensive to frequently modify input patterns

- **Comparisons against other pattern matching hardwares**
  - Our Static BP-NFA for STR is faster than TNFA-based hardware and comparable with SHIFT-OR based hardware

<table>
<thead>
<tr>
<th>Design</th>
<th>Class</th>
<th>Device</th>
<th>Throughput</th>
<th>Logic Cells/char</th>
<th>#Char Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static BP-NFA for STR (SASIMI2010, ours)</td>
<td>STR</td>
<td>Xilinx Virtex-5 LX50</td>
<td>1.6 Gbps</td>
<td>2.30</td>
<td>19887</td>
</tr>
<tr>
<td>TNFA-based hardware (FCCM2001)</td>
<td>REG</td>
<td>Xilinx Virtex 100</td>
<td>0.5 Gbps</td>
<td>66</td>
<td>29</td>
</tr>
<tr>
<td>SHIFT-OR based hardware (FPT2006)</td>
<td>STR</td>
<td>Altera Stratix EP1S40</td>
<td>2.3 Gbps</td>
<td>0.96</td>
<td>5004</td>
</tr>
<tr>
<td>Dynamic BP-NFA for STR (FPT2010, to appear)</td>
<td>STR</td>
<td>Xilinx Virtex-5 LX300</td>
<td>2.9 Gbps</td>
<td>10.8</td>
<td>8192</td>
</tr>
<tr>
<td>Dynamic BP-NFA for EXT (FPT2010, to appear)</td>
<td>EXT</td>
<td>Xilinx Virtex-5 LX300</td>
<td>1.6 Gbps</td>
<td>24.6</td>
<td>4096</td>
</tr>
</tbody>
</table>

**STR**: string patterns, **EXT**: extended string patterns, **REG**: regular expressions