FUJITSU

shaping tomorrow with you

EXI (Efficient XML Interchange)

April 11, 2016

Takuki Kamiya

Fujitsu Laboratories of America

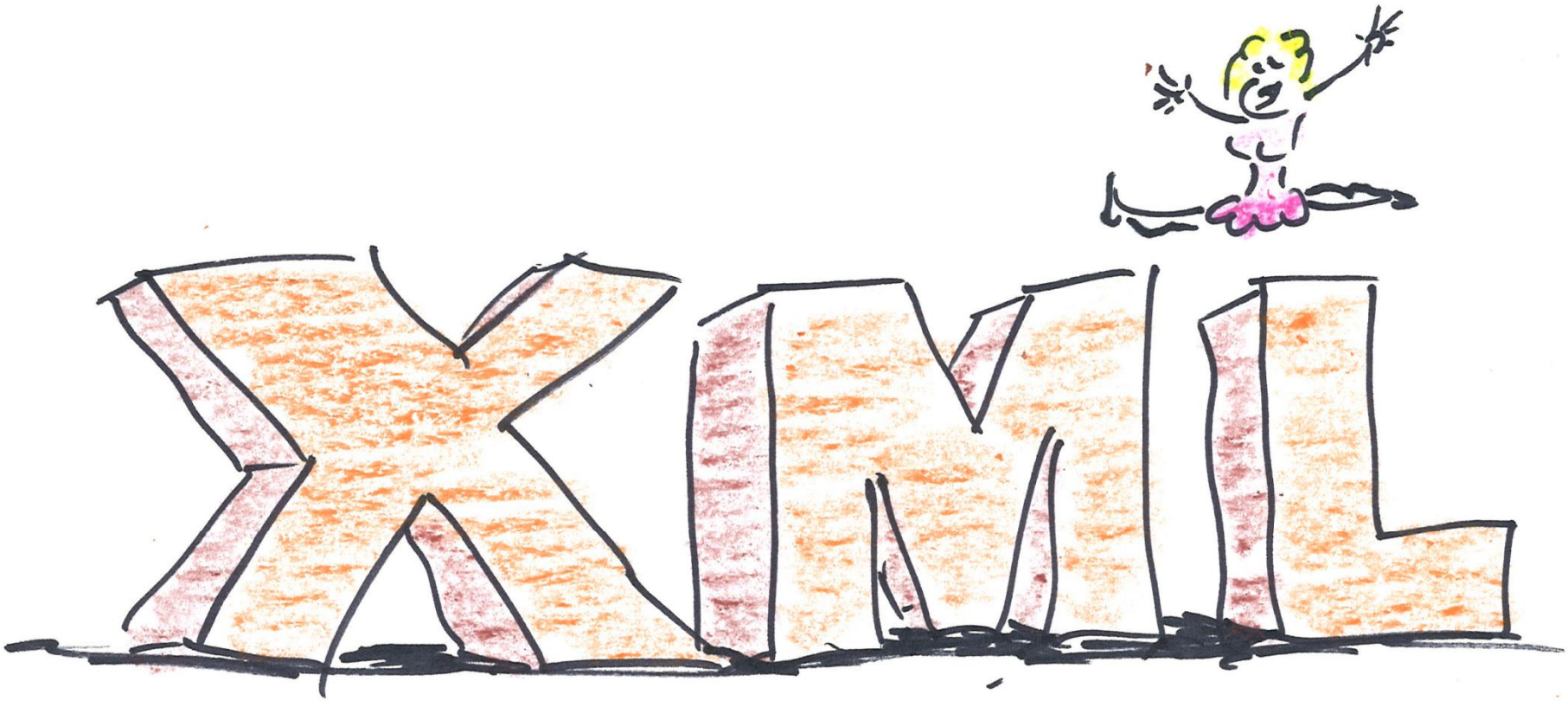
What is EXI?

FUJITSU

(From Wikipedia)

Efficient XML Interchange (EXI) is **a binary XML format** which was adopted as a **Recommendation by the World Wide Web Consortium (W3C)** on 10 March 2011.

FUJITSU



XML (Extensible Markup Language)

FUJITSU

- W3C (World Wide Consortium) created XML in 1998 as a format for data exchange on the Web.
- Now, XML is a commodity with robust technology stack on top of it.
  - XML Schema, XPath2/XQuery, XSLT, etc.
- By using XML:
  - Technologies and tools built around XML are all available for you
  - Easy to leverage thousands of Web Services already supporting XML, or easy to expose your function as a Web Service

Those Left Behind

FUJITSU

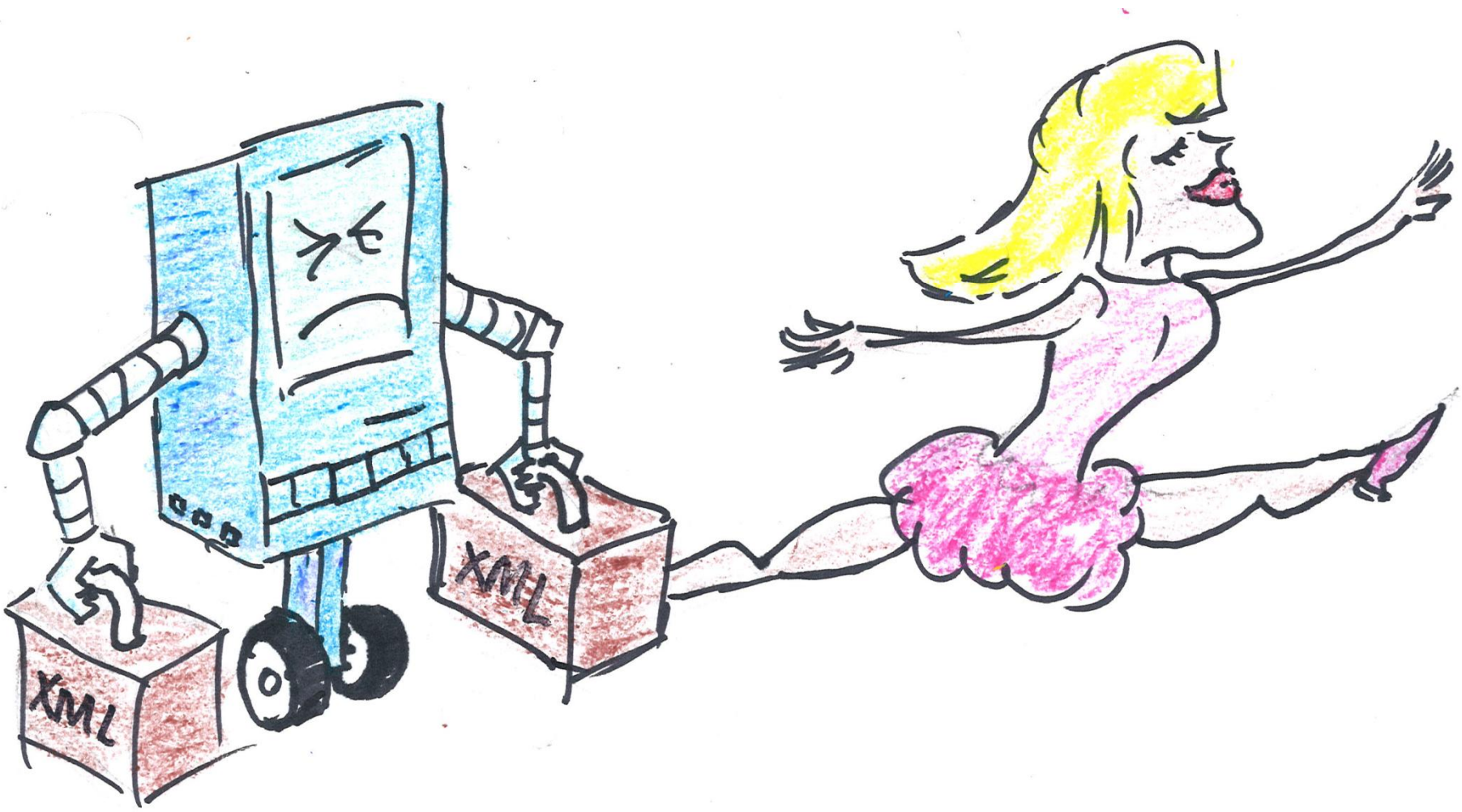
However....

There were use cases that wanted to use XML, but found it difficult, including use cases such as:

- Use cases that involve very large (Giga bytes) of data between servers. XML was too bloated. Traditionally, flat files or CSV files served those use cases.
- Use cases that involve sensors and microcontrollers. Text was not amenable to those devices. Traditionally, ASN.1 PER or hand-crafted format was used.

In Need of Help

FUJITSU



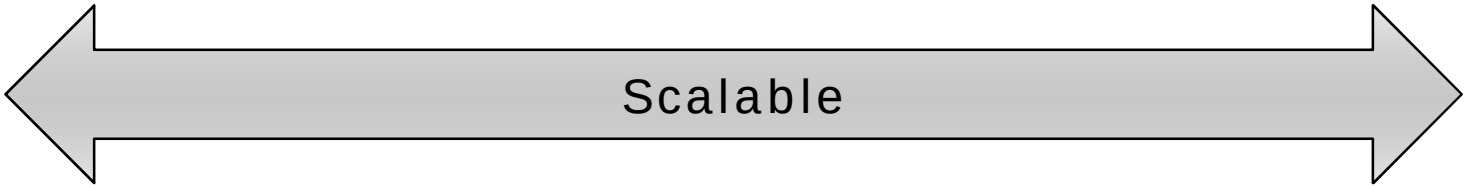
Requirements

FUJITSU

- Large Data
- Abundant Computing Resource (e.g. servers)

- XML works fine, but
  EXI provides efficiency

- Small Data
- Resource Constrained
  (e.g. Small Things)



Factors in Use Cases

FUJITSU

- Use of Schema
  - With Schema
    - Totally-defined schema
    - Partially-defined schema
  - Without Schema – Rapidly Changing Structure
- Use of Compression
  - With Compression – Extra CPU cycle available
  - Without Compression
    - Very Small Message
    - Can’t Afford the Cost of Compression

EXI

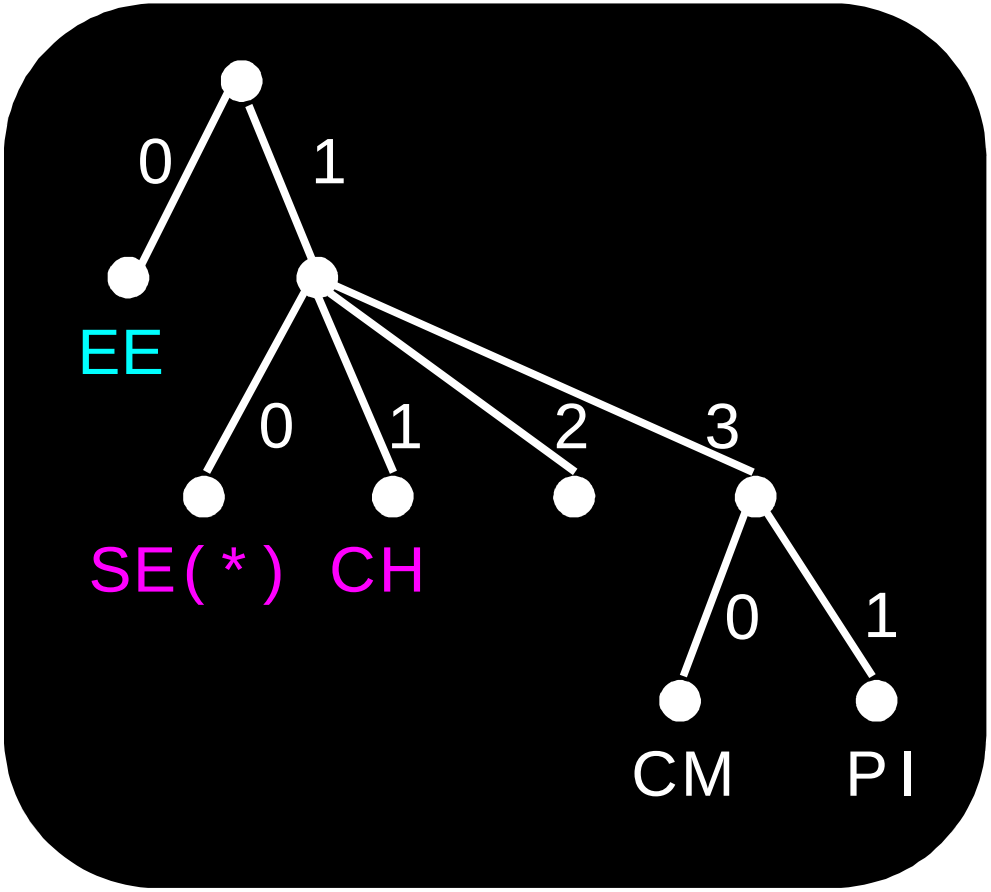
FUJITSU

- EXI grammar is able to be Schema-informed
  - Use of schemas is optional.
  - When there is a schema, EXI leverages the schema.
    - Partial schema is OK.
  - When no schema is available, EXI still provides compact encoding.
- Integrated Compression
  - Use of compression is optional.
  - Compression scheme (i.e. DEFLATE) integrated with grammar-based tokenization.

Baseline Schema-less Grammar

FUJITSU

- Informed by no schema, following baseline grammar is used



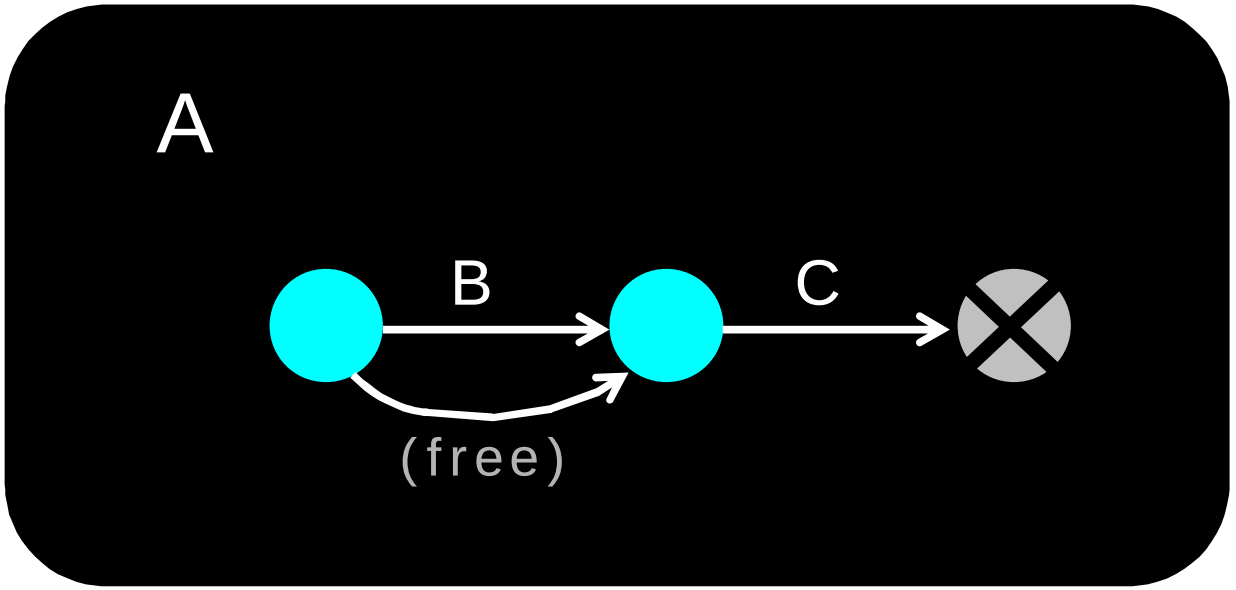
	Path	Encoding
EE	0	0
SE(*)	1.0	1 00
CH	1.1	1 01
CM	1.3.0	1 11 0
PI	1.3.1	1 11 1

Event Code Tree (for element-content)

What If We Have Schema?

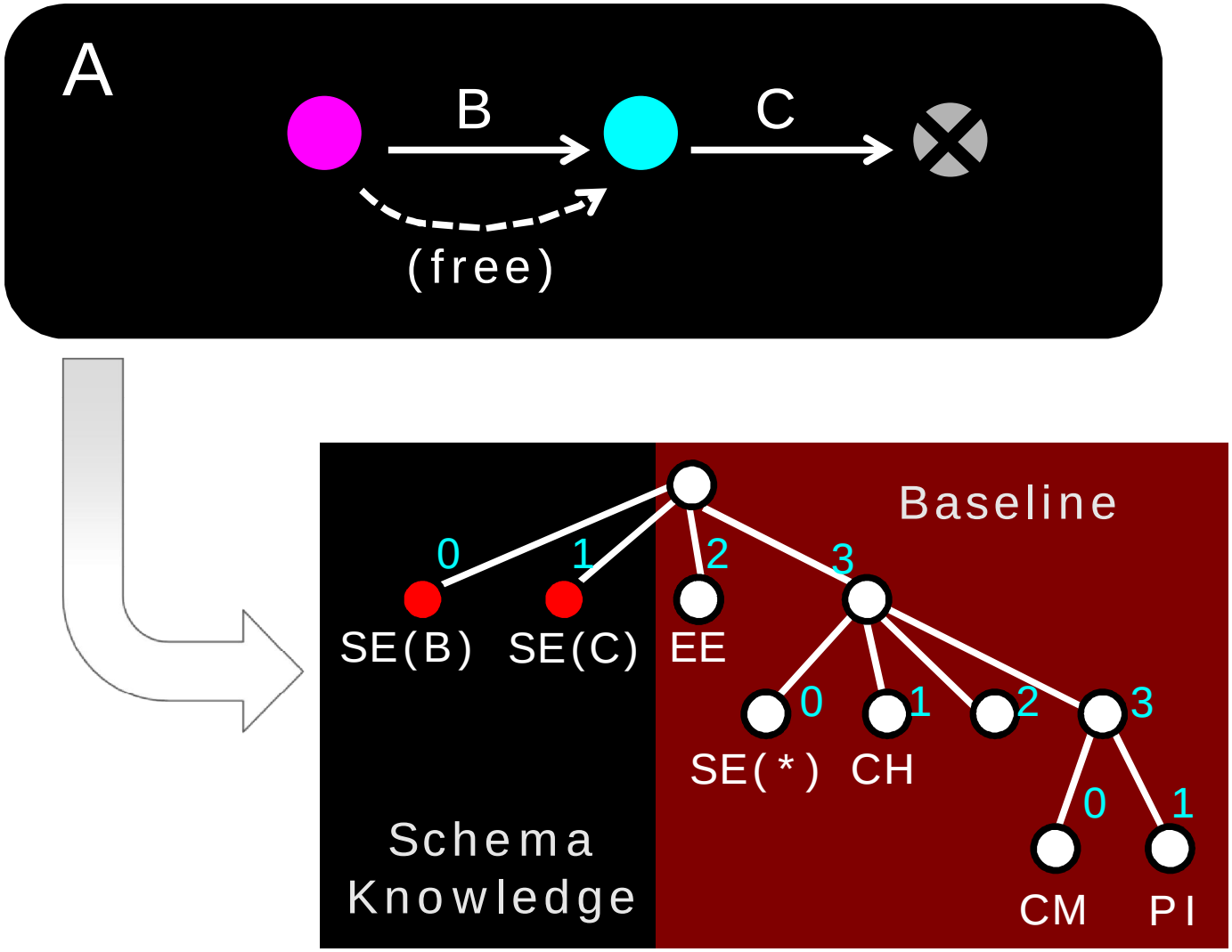
FUJITSU

A := sequence (B? , C)



Automaton

Schema-informed Grammar



Event Code Tree

Schema-informed Grammar

FUJITSU

	Path	Encoding	# of bits
SE(B)	0	00	2
SE(C)	1	01	2
EE	2	10	2
SE(*)	3.0	11 00	4
CH	3.1	11 01	4
CM	3.3.0	11 11 0	5
PI	3.3.1	11 11 1	5

EXI Encoding Result Examples

FUJITSU

JTLM Data (360 bytes)

```
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/" ↓
                  xmlns:xsd="http://www.w3.org/2001/XMLSchema" ↓
                  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">↓
  <soapenv:Body>↓
    <subscriberEventNotificationResponse xmlns="http://www.firesupport.com/schema/tlmsubs/operations"/>↓
  </soapenv:Body>↓
</soapenv:Envelope>↓
```

Location Data (103 bytes)

```
<dahut-sighting xmlns='http://berjon.com/ns/dahut-sighting' id='robin' lat='2.38323' long='48.86484'/>↓
```

Settings

- ✓ Schema-Informed
- ✓ No Compression

Format	Size (bytes)	Ratio
XML	360	100%
ASN.1	105	29%
EXI	**39**	**11%**

JTLM

Format	Size (bytes)	Ratio
XML	103	100%
ASN.1	27	26%
EXI	**17**	**17%**

Location

Processing Efficiency

FUJITSU

Location Data (103 bytes)

Encode

Format	TPS	Ratio
XML (Java)	15858	1
ASN.1 (C)	310862	x19.6
EXI (Java)	**185029**	**x11.7**

Decode

Format	TPS	Ratio
XML (Java)	9216	1
ASN.1 (C)	318419	x34.6
EXI (Java)	**277409**	**x30.0**

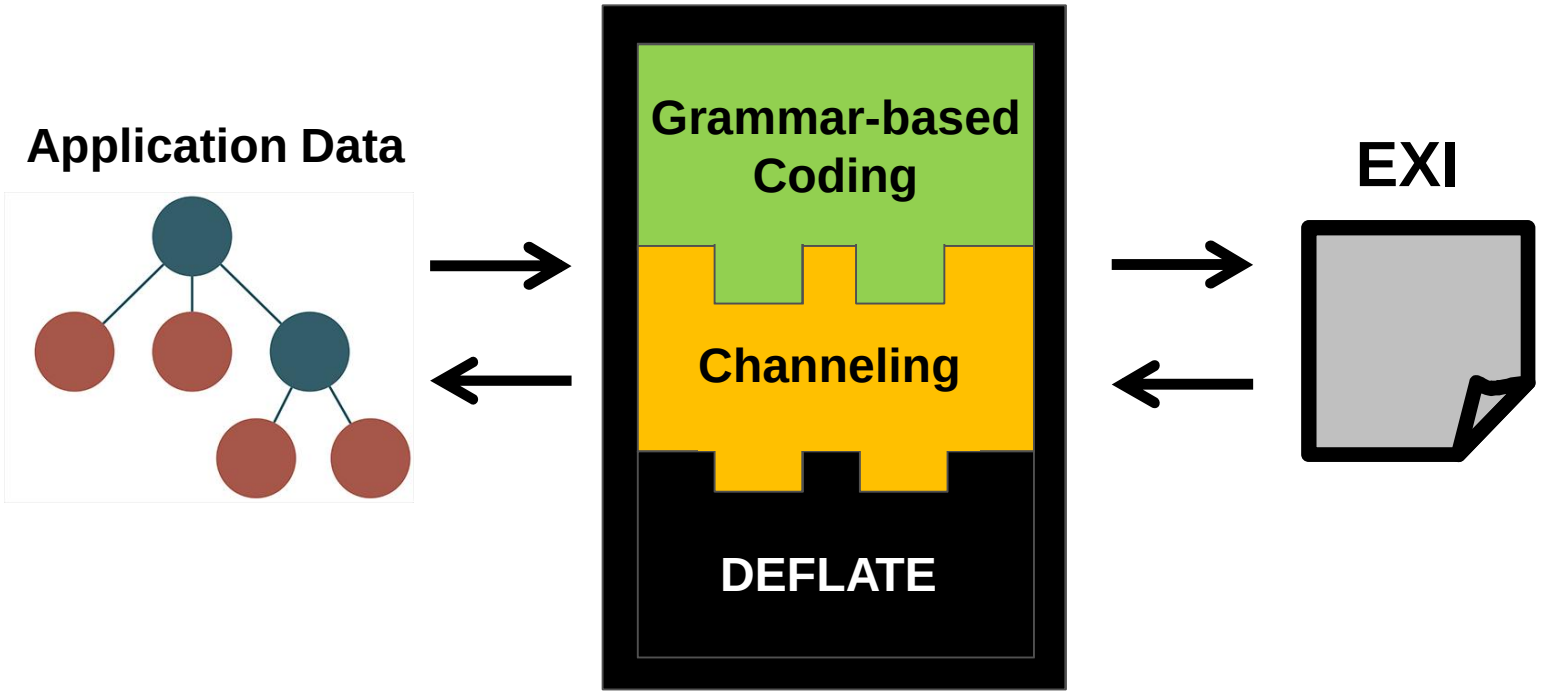
Settings

- ✓ Schema-Informed
- ✓ No Compression

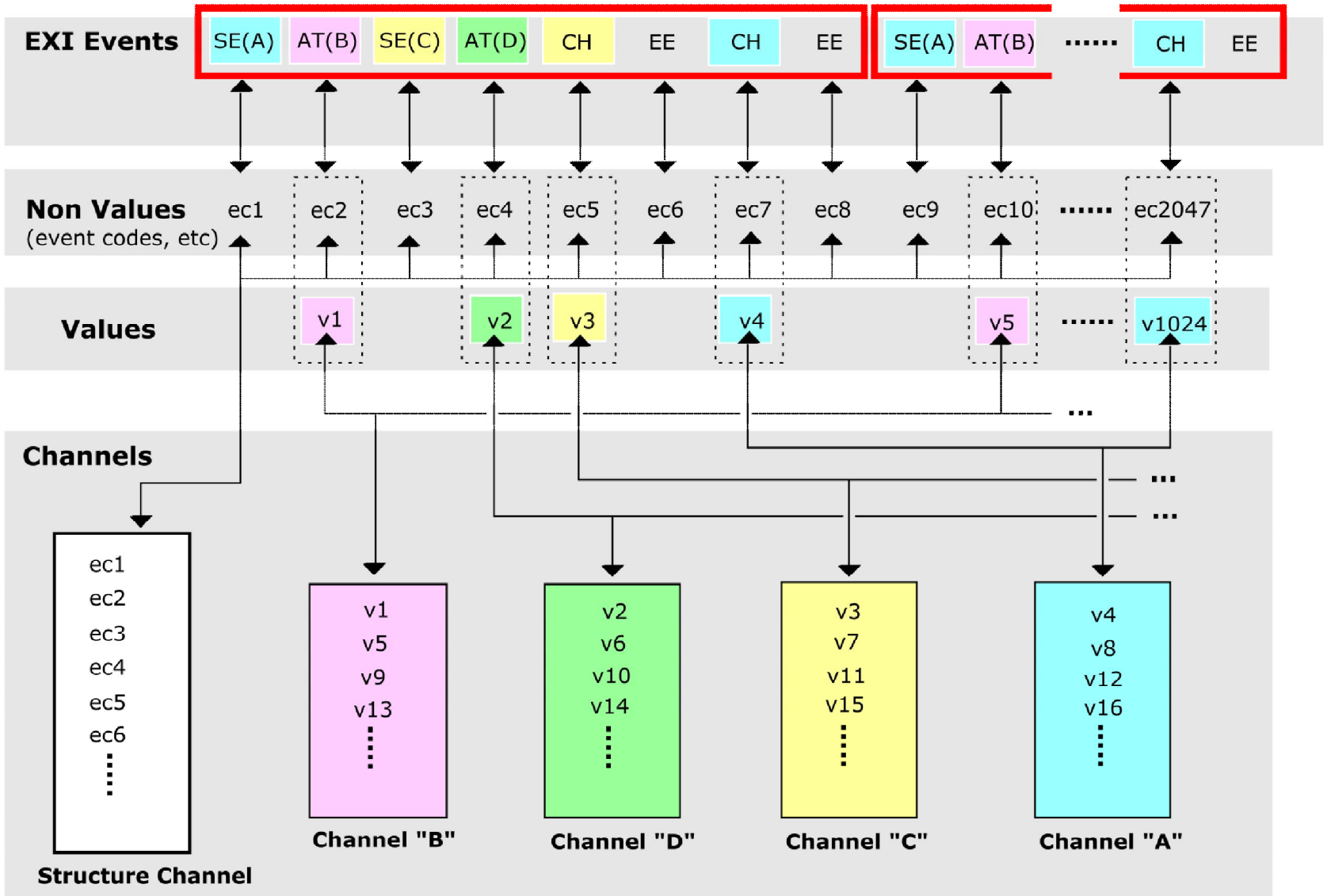
Integrated Compression

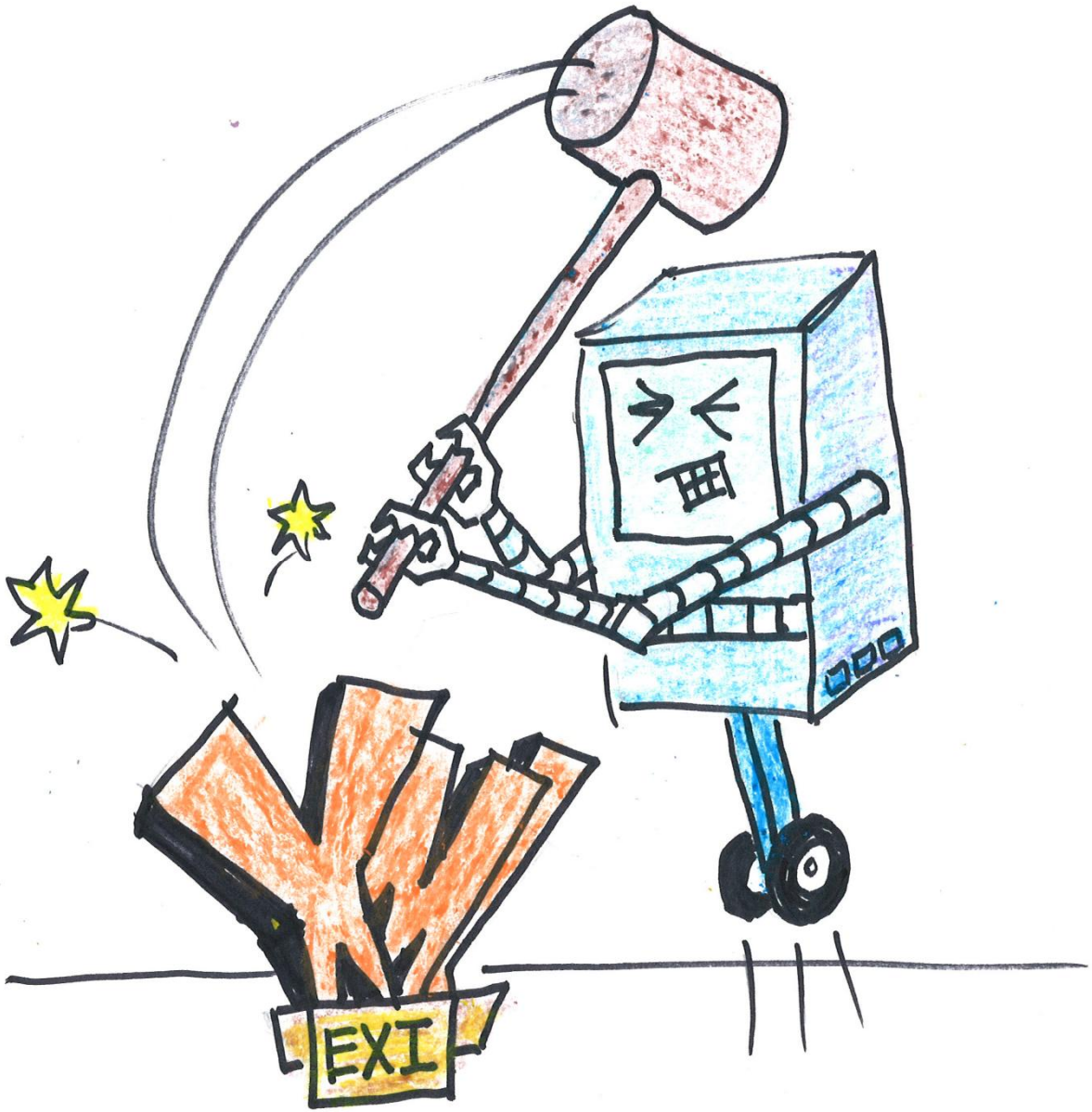
FUJITSU

- Compression Scheme (DEFLATE) is integrated into Grammar-based Coding.
- Channeling in the middle binds them together.



Channeling


EXI Events
SE(A)
AT(B)
SE(C)
AT(D)
CH
EE
CH
EE
SE(A)
AT(B)
.....
CH
EE
Non Values
(event codes, etc)
ec1
ec2
ec3
ec4
ec5
ec6
ec7
ec8
ec9
ec10
.....
ec2047
Values
v1
v2
v3
v4
v5
.....
v1024
...
Channels
...
...
ec1
ec2
ec3
ec4
ec5
ec6
Structure Channel
v1
v5
v9
v13
Channel "B"
v2
v6
v10
v14
Channel "D"
v3
v7
v11
v15
Channel "C"
v4
v8
v12
v16
Channel "A"



EXI Compression Result Example - XBRL

FUJITSU

- **JTLM Data ≈ 1 MB (915 KB)**

	JTLM 937,005 bytes	Ratio
XML+GZIP	111.23 KB	100%
EXI	**7.70 KB**	**7%**

Settings

✓ Schema-Informed

✓ With Compression

- **XBRL Data 315 MB**

	Size		Encode		Decode	
	MB	vs. GZIP	Time	vs. GZIP	Time	vs. GZIP
XML+GZIP	3.70 MB	100%	6.18 s	100%	4.25 s	100%
EXI	**1.64 MB**	**44%**	**3.67 s**	**59%**	**2.64 s**	**62%**
XML+LZMA	1.01 MB	27%	37.47 s	606%	3.34 s	79%

Implementations

FUJITSU

Name	Platform/Language	Organization	License
EXIficient	Java, JavaScript, C/C++	Siemens	MIT License
EXIP	C	EISLAB	BSD 3-Clause License
OpenEXI	Java, .Net	Fujitsu	Apache License
Efficient XML	Java, .Net, C/C++	AgileDelta	Commercial
OSS EXI Tools	.Net, C/C++	OSS Nokalva	
…	**…**	**…**	**…**