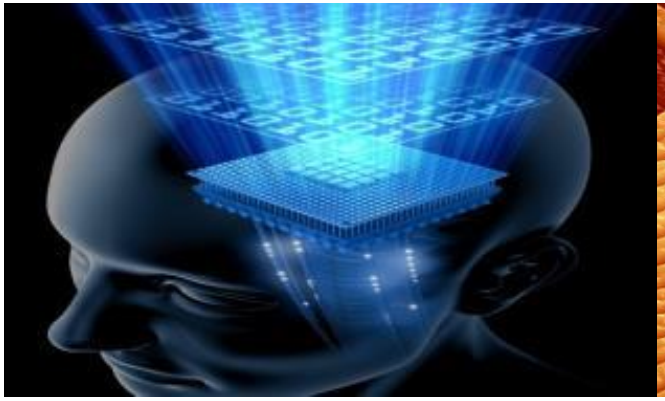
THE UNIVERSITY of ADELAIDE

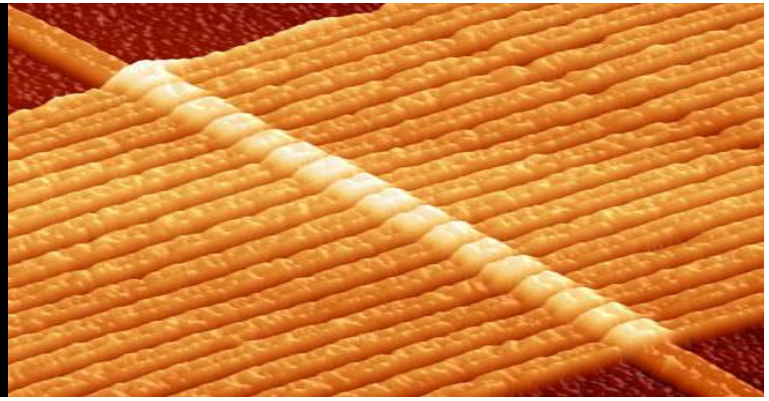
Street: A Hardware Production Engine for Soar

Dr. Braden Phillips

Mostafa Numan

Jesse Frost

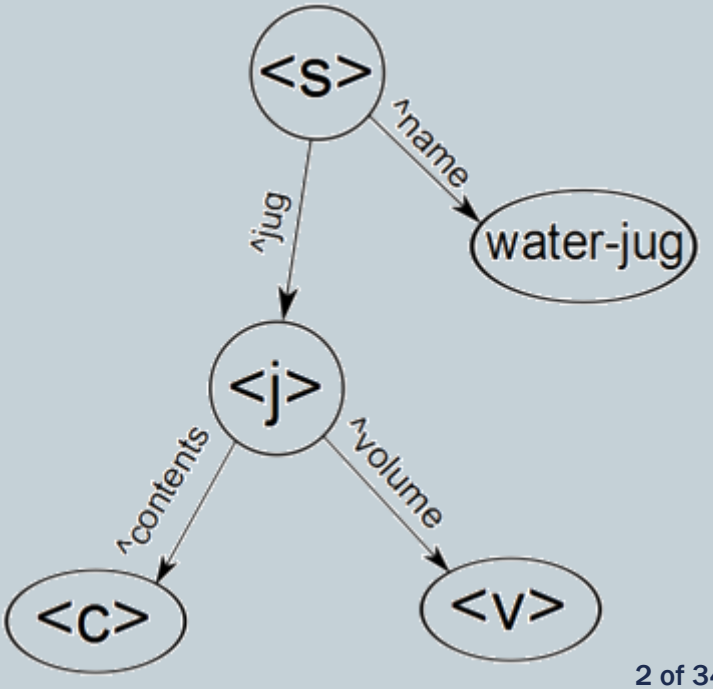




Soar, Production Systems

- Working Memory Encoded as WMEs – which can be represented in a graph
- Productions test for subsets of WM that match a specified pattern

```
sp {
        (state <s> ^name water-jug)
        (<s> ^jug <j>)
        (<j> ^contents <c>)
        (<j> ^volume <v>)
-->
        (<j> ^empty (- <v> <c>))
}
```

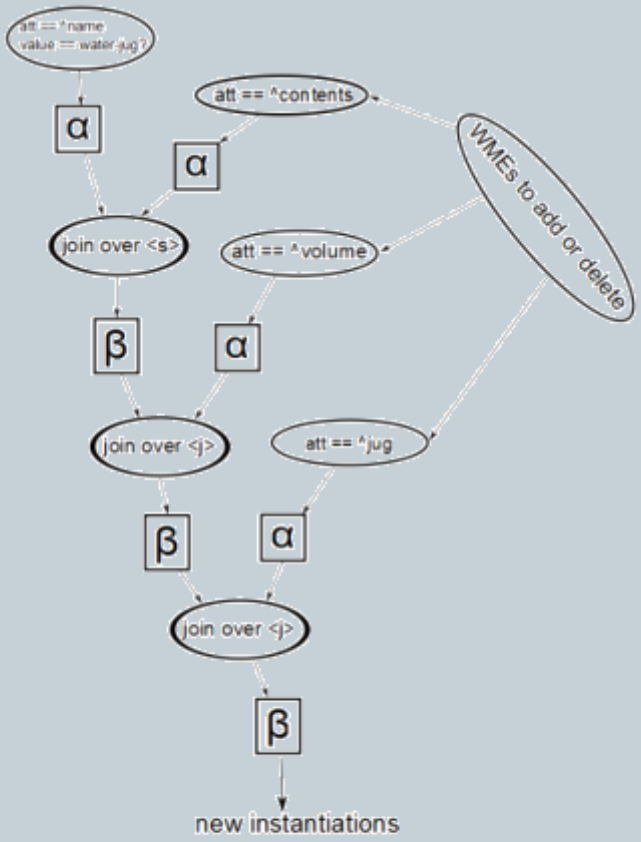


Soar, Production Systems

- How do you find instantiations in working memory?
- Brute force: search every subset all the time
- Rete: track changes to rule satisfaction

Rete

- Rete exploits temporal redundancy
- Stores candidate subsets of working memory
- Tracks changes to these subsets over time



Forgy, Charles L. *On the efficient implementation of production systems*. Carnegie-Mellon University, 1979.

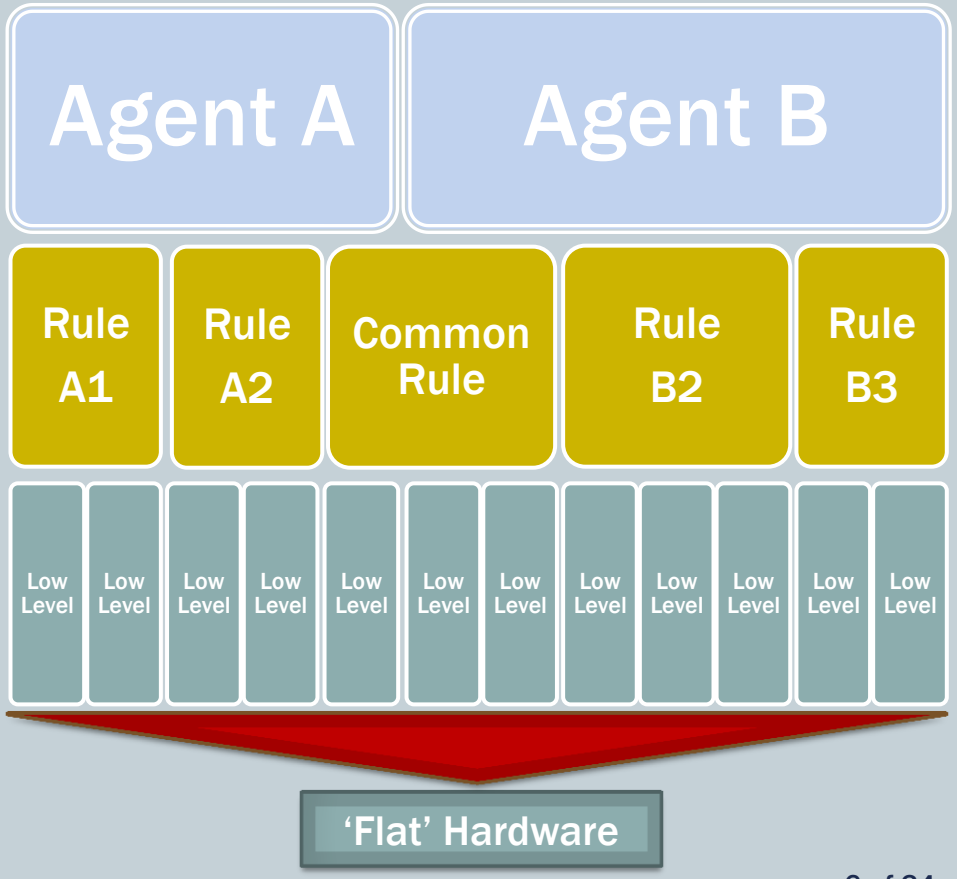
Treat

- Treat is a variation of Rete originally for parallel architectures
- Information in ‘conflict set’ of current instantiations and the beta memories is redundant
- Does not maintain partial joins, just recalculates the information as its needed

Miranker, Daniel P. *TREAT: A Better Match Algorithm for AI Production Systems*. University of Texas at Austin, 1987.

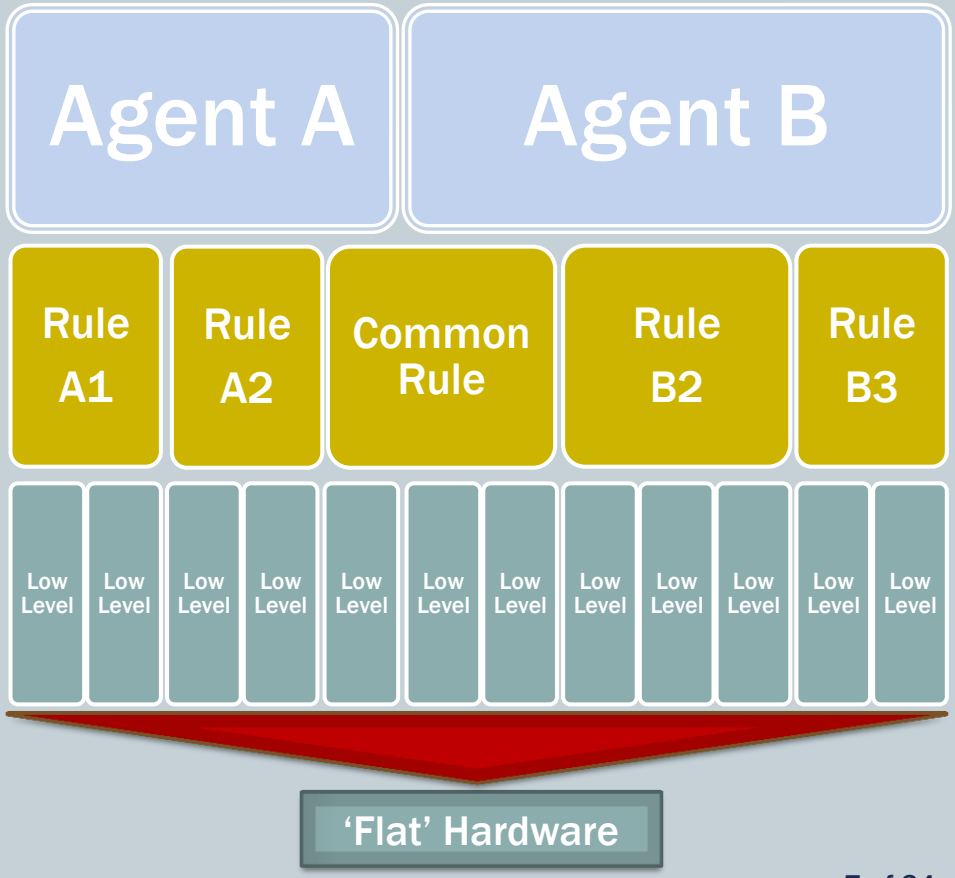
Parallelism

- We want a system that allows parallelism at multiple levels
- General Purpose Architectures have bottlenecks



Parallelism

- A hardware platform should scalable at all levels – need a flat uniform architecture
- Modern Technology makes custom architectures accessible

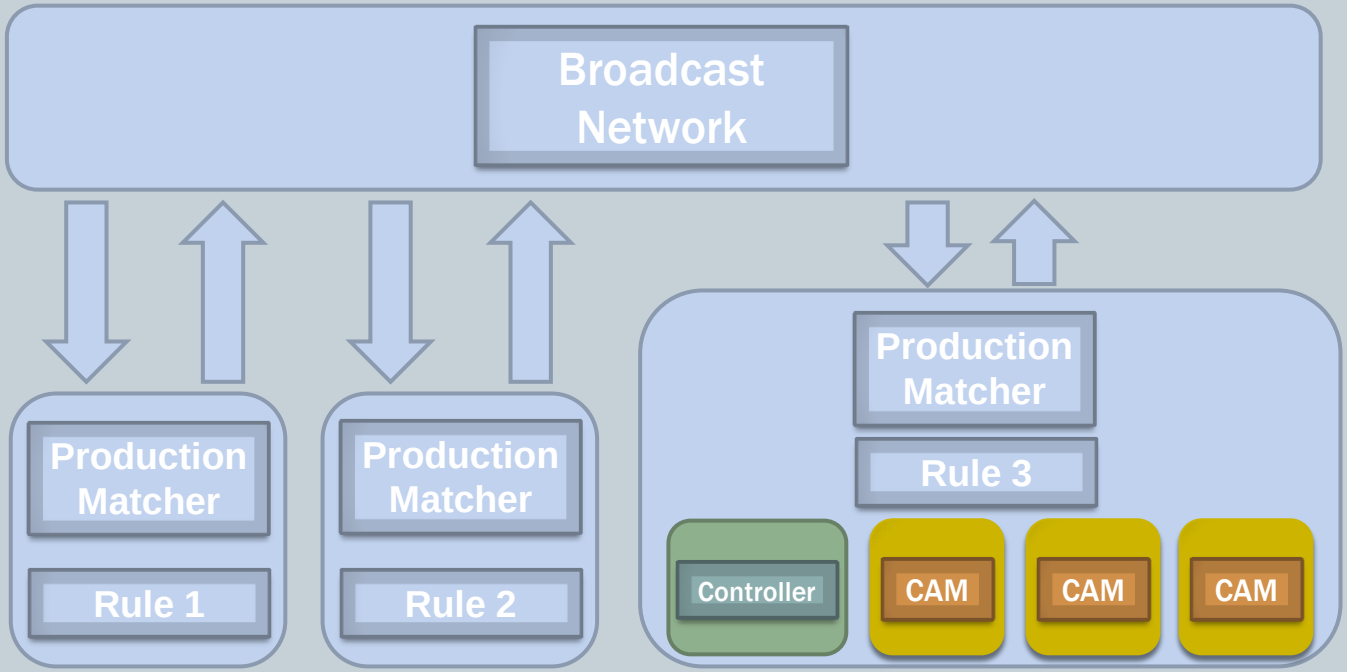


Street

- Street – the name might be from ‘Soar-Treat’
- Not really either
- A hardware platform for executing a parallel production language directly
- Hardware blocks perform the match-act cycle directly (not general purpose processors running a production interpreter)

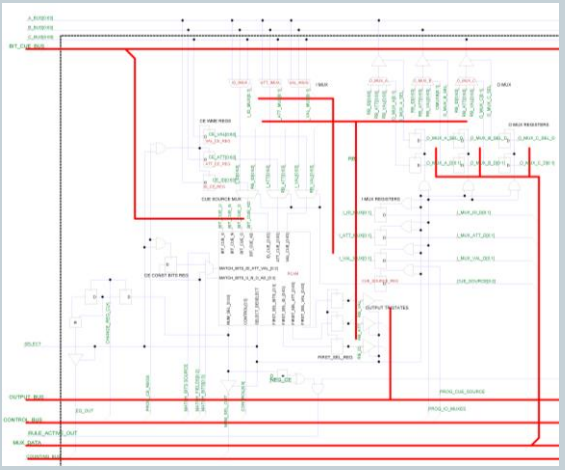
Street

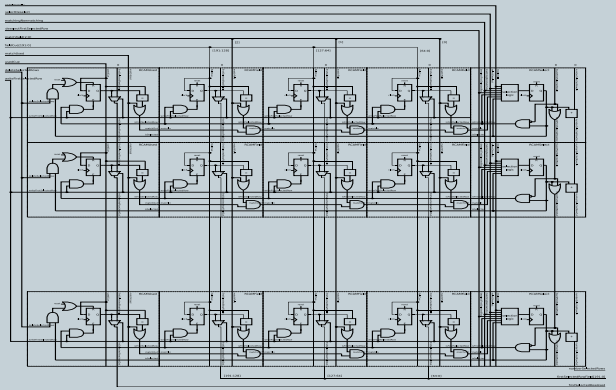
- Use ideas from Rete/Treat
- Subsets of working memory kept in specialised CAMs



Production Matcher

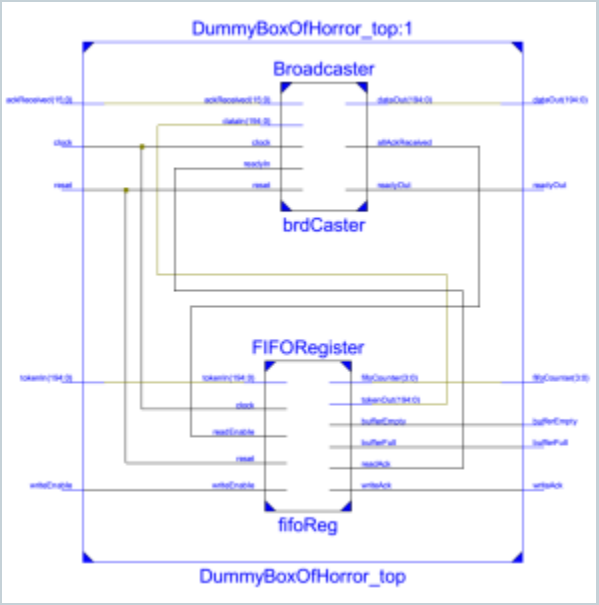
- Subsets of working memory stored in collections of CAMs
- CAMs designed for single cycle execution of the operations we need
- Each rule is assigned one CAM for each condition element





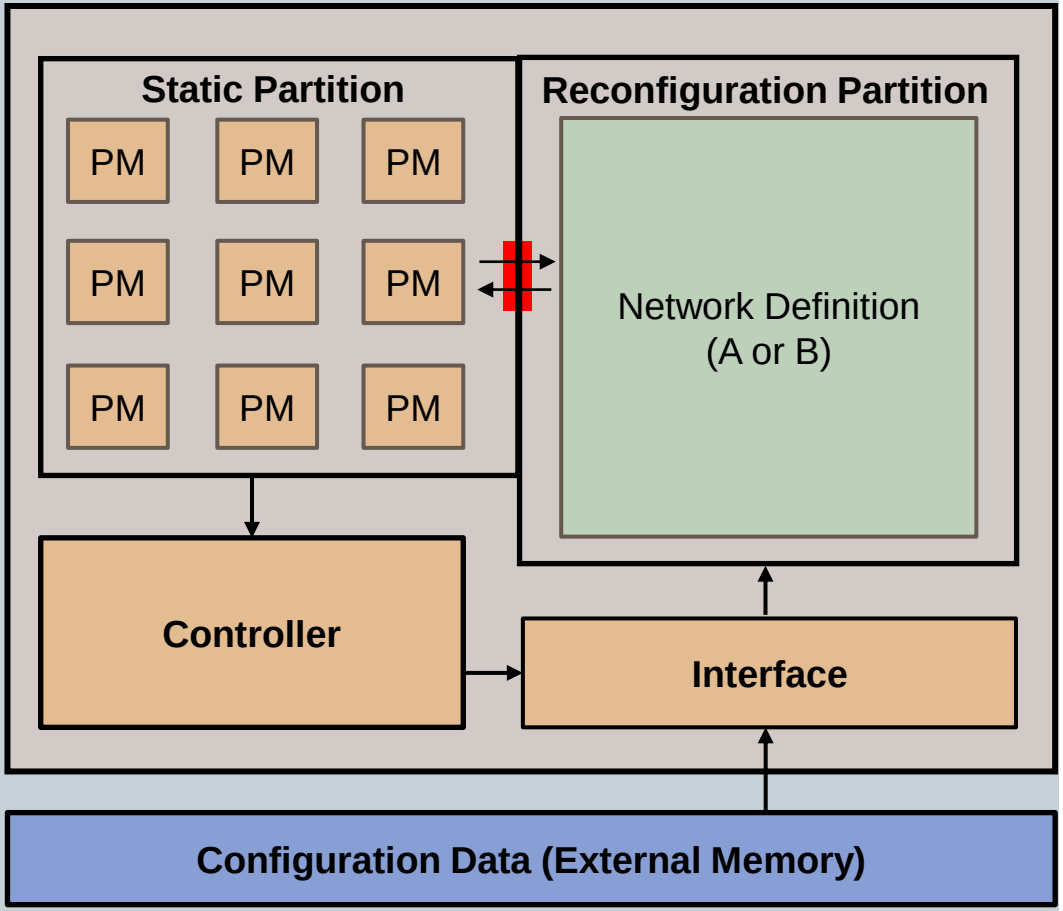
Broadcast Network

- Broadcasts WM changes and synchronises production matchers
- Small broadcasters can be built from FIFO registers
- Gives the architecture a two-section design, one containing matchers and one containing the broadcaster.

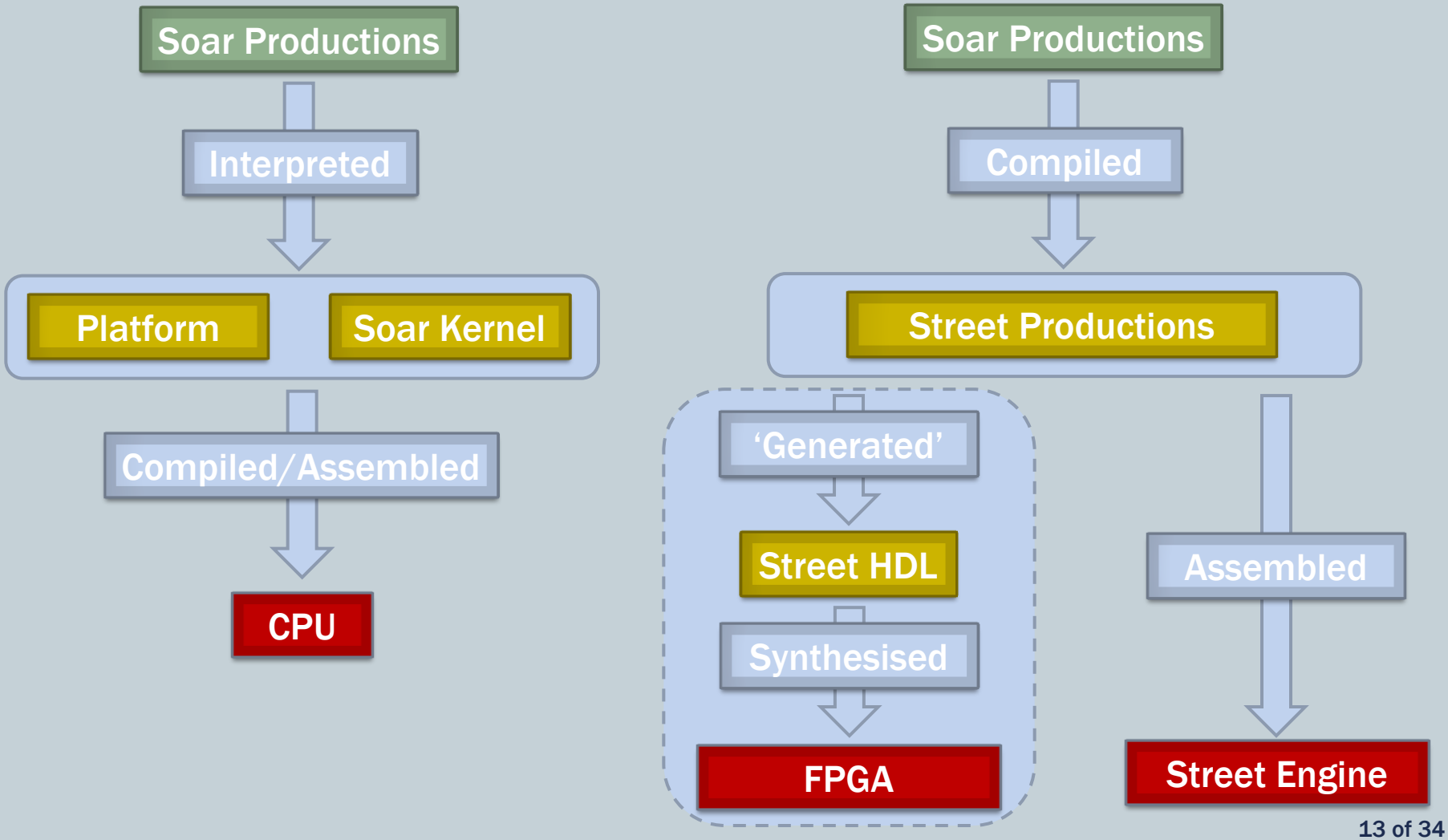


Network-on-Chip

- Network-on-Chip for large systems
- Network definition can be reconfigured dynamically
- Broadcast Network is a potential bottleneck



Street


Soar Productions
Interpreted
Platform
Soar Kernel
Compiled/Assembled
CPU
Soar Productions
Compiled
Street Productions
'Generated'
Street HDL
Synthesised
FPGA
Assembled
Street Engine

Street

- Work in progress
- Sorry, no ‘Soar-on-a-Chip’ just yet
- Preliminary results suggest we could fit several hundred productions on a large commercial FPGA
- Working on proof-of-concept
- Synchronisation and some more complex behaviours for Soar require more rules than the same agent in Soar
- Broadcast Network needs analysis

Vision

- Application Specific Device – Similar to existing FPGA devices but with CAM blocks rather that conventional LUTs
- Very fast cycle times – orders of magnitude faster than software -
- As many components of Soar implemented in simple production language as possible – RISC vs CISC
- A shift of the programming paradigm from serial rule firing to massively parallel rule firing

Questions

?