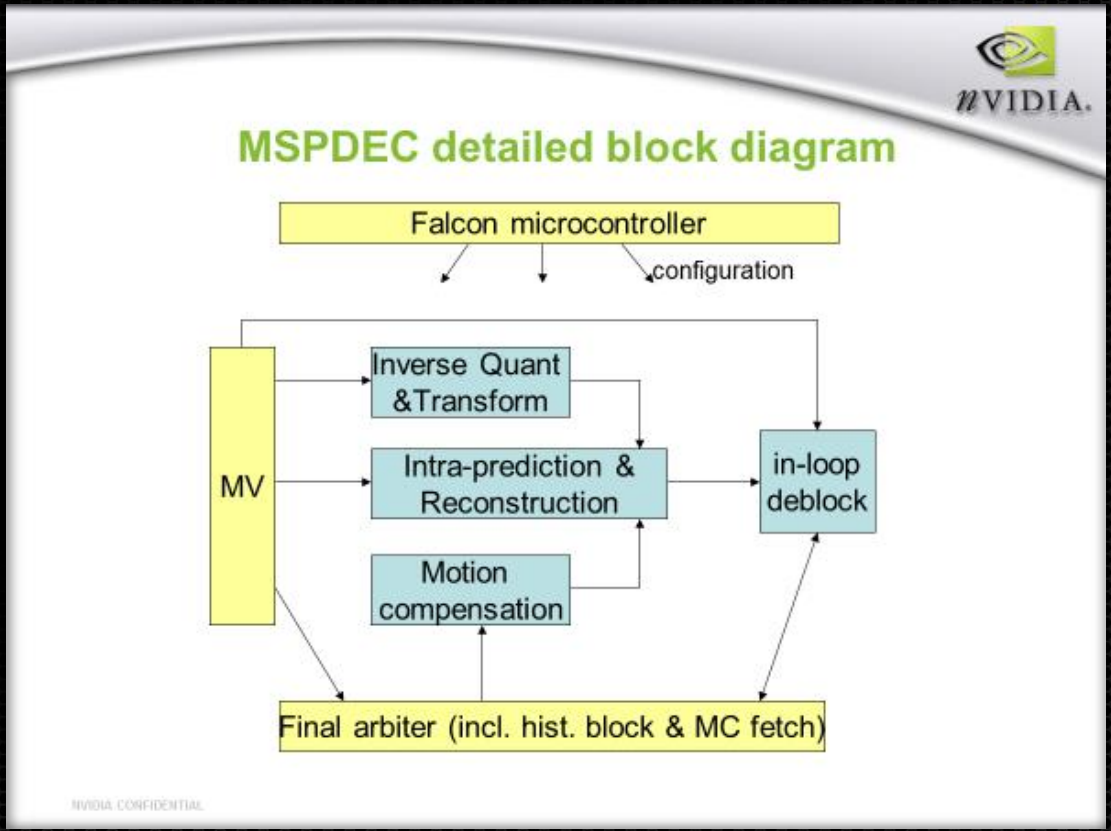

RISC-V in NVIDIA
6th RISC-V Workshop, Shanghai, May 2017
NVIDIA

Falcon: NVIDIA's proprietary RISC

- Falcon = FAst Logic CONtroller
- General purpose embedded processor
- Design started in ~2005; production ~2007



Falcon's history

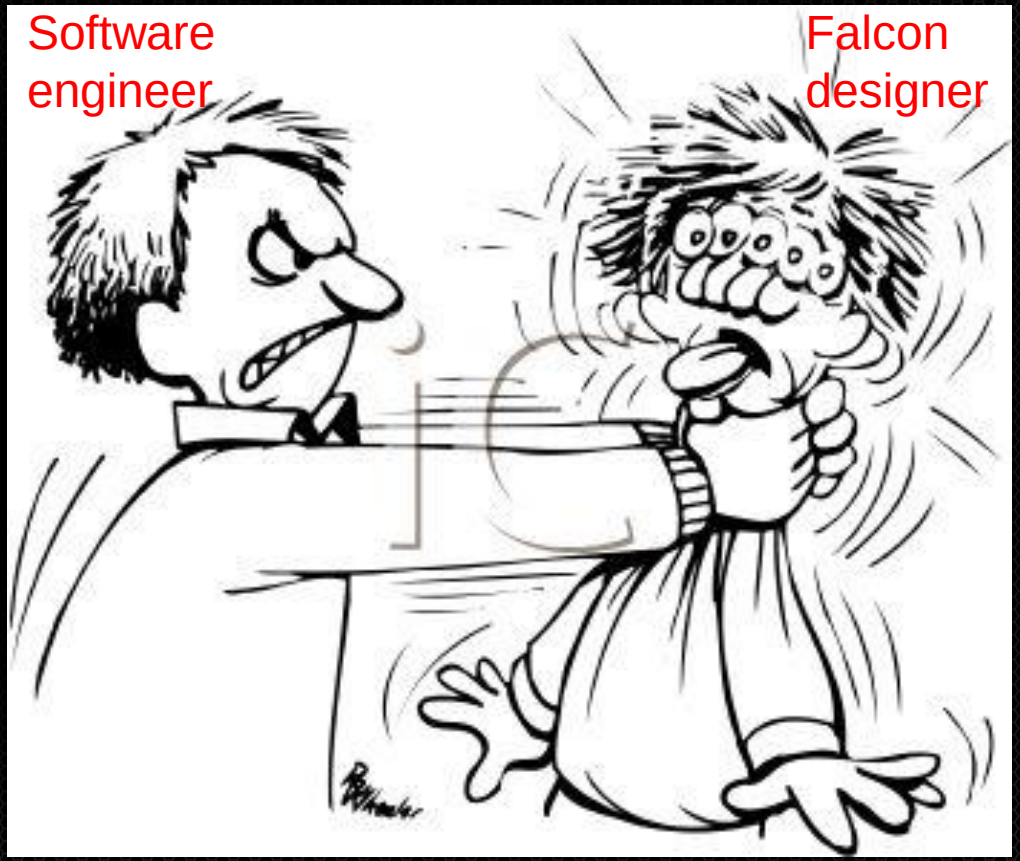
- Embedded in 15+ designs
- Taped out in ~50 chips
- Shipped ~3 billion times
- No stop-ship bugs

Falcons shipped estimate	
dGPU Volume /year	50M*
Years Falcon shipping	10
Avg. #Falcons / GPU	10
Avg. NVIDIA market share	60%
Total shipped	3 billion

http://www.anandtech.com/show/10864/discrete-desktop-gpu-market-trends-q3-2016

Why replace such a successful design?

- Use cases getting more complex
  - Large complex SW
  - External SW
  - Threaded SW
  - Large virtual memory space
- Falcon limitations
  - Low performance
  - No caches (Icache added lately)
  - No thread protection
  - 32-bit address range
  - One size fits all



Selecting the next architecture

Technical criteria

- >2x performance of Falcon
- <2x area cost of Falcon
- Support for caches as well tightly coupled memories
- 64-bit addresses
- Suitable for modern OS

Considered architectures

- ARM
- Imagination Technologies MIPS
- Synopsys ARC
- Cadence Tensilica
- RISC-V

Why RISC-V for Falcon Next

RISC-V is the only architecture that meets all our criteria

https://riscv.org/wp-content/uploads/2016/07/Tue1100_Nvidia_RISCV_Story_V2.pdf

Item	Requirement	ARM A53	ARM A9	ARM R5	RISC-V Rocket	NV RISC-V
Core perf	>2x falcon	Yes	Yes	Yes	Yes	Yes
Area (16ff)	<0.1mm^2	No	No	Yes	Yes	Yes
Security	Yes	TZ	TZ	No	Yes	Yes
TCM	Yes	Yes	No	Yes	No	Yes
L1 I/D $	Yes	Yes	Yes	Yes	Yes	Yes
Addressing	64bit	Yes	No	No	Yes	Yes
Extensible ISA	Yes	No	No	No	Yes	Yes
Safety (ECC/Parity)	Yes	Yes	Yes	Yes	Yes	Yes
Functional Simulation model	Yes	Yes	No	No	No	Yes

Scalability - beyond Falcon replacement

- Falcon ISA is monolithic
- RISC-V is flexible
  - 32,64, and 128-bit versions
  - Cost/performance and supervisor options
  - Custom extensions
- NV-RISCV uses RV64IM_Sdef ISA
- New opportunities
  - Address lower cost and higher perf problems
  - Backward compatibility allows opening up to 3rd party programmers
  - Mix and match internally and externally developed cores

Open source architecture

- Control
  - Match NVIDIA interfaces and tools
  - Original reason for Falcon
- Quality
  - Large community of contributors
  - E.g. memory model tuning
- Cost of ownership
  - No license, royalty fees
  - ISA, tools from community

	Licensed (ARM)	NVIDIA proprietary (Falcon)	Open source (RISC-V)
Control	-	+	+
Quality	0	0	+
Cost	-	-	+

Why contribute to RISC-V?

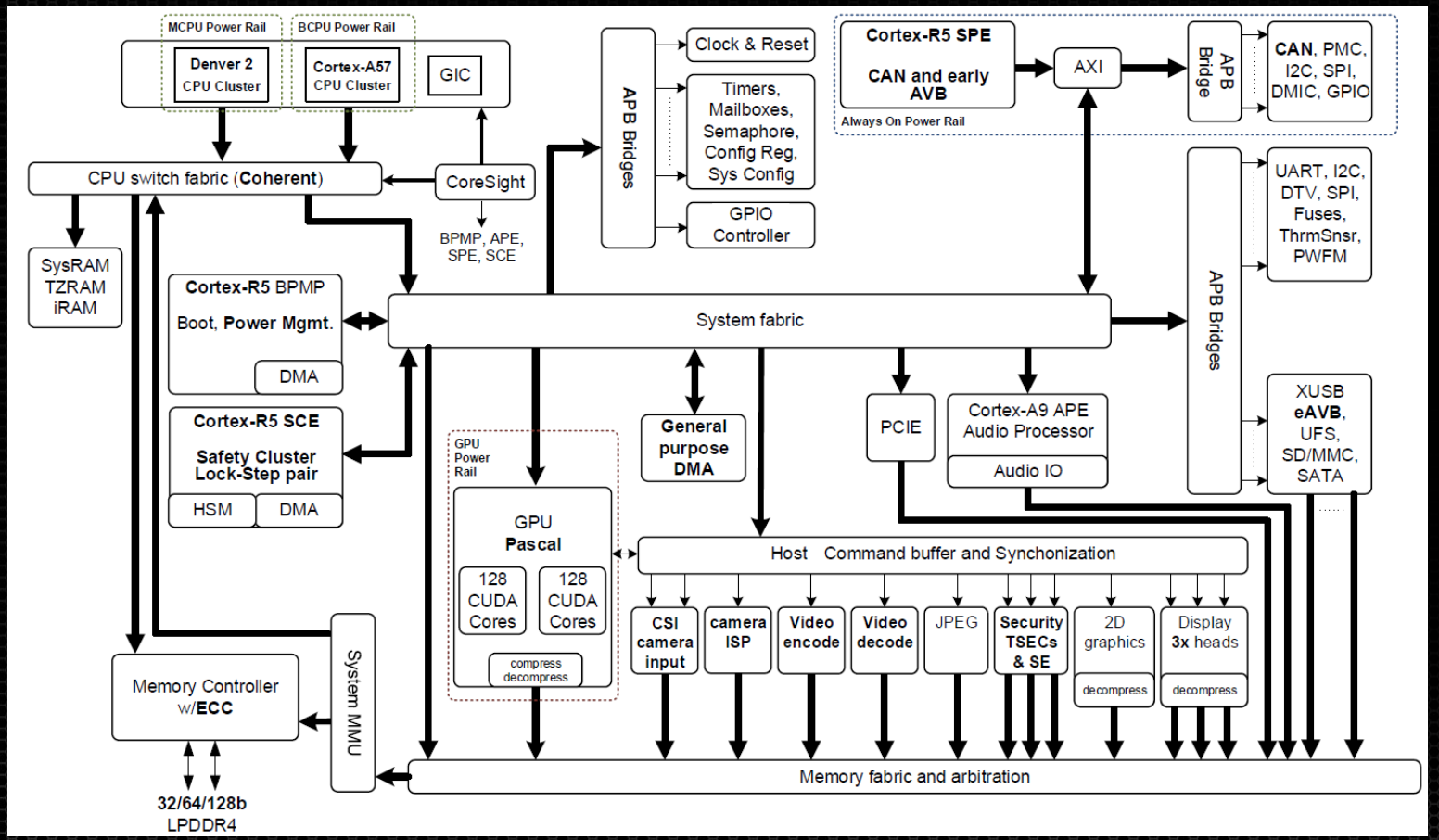
Polder model:
cooperation despite differences

- Benefit from thriving RISC-V community and architecture
- Influence the direction so RISC-V and (y)our interests align

Memory model workgroup

- Daniel Lustig chairing the workgroup
- Presentation on Wednesday
- Workgroup meeting on Thursday

Why is it important to NVIDIA?



Jetson TX2

- ~20 CPU cores
- 256 GPU cores
- Complex bus topology
- Coherent and non-coherent traffic

NVIDIA

What is the problem?

Core0	Core1
{[a] == 1}	{[b] == 1}
ld [a] → x	ld [b] → y
st [b], 0	st [a], 0
{x == 0}	{y == 0}

That obviously cannot happen!

What is the problem?

Core0

{[a] == 1}

ld [a] → x

st [b], y

{x == 0}

Core1

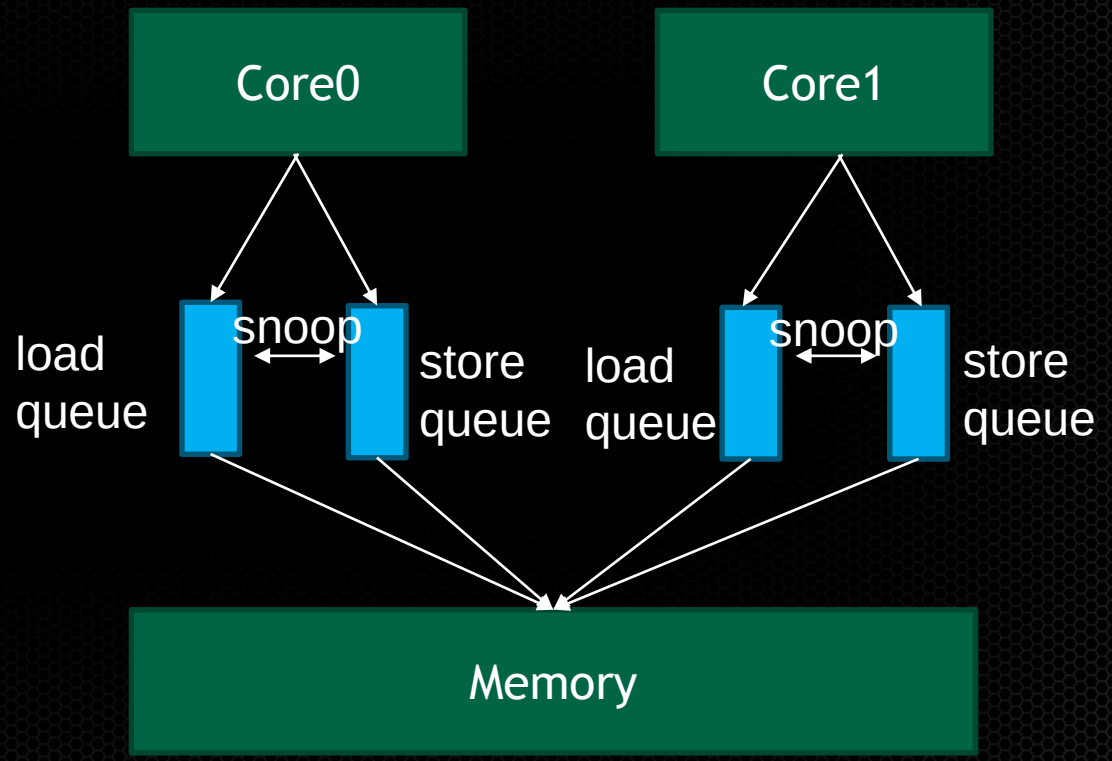
{[b] == 1}

ld [b] → y

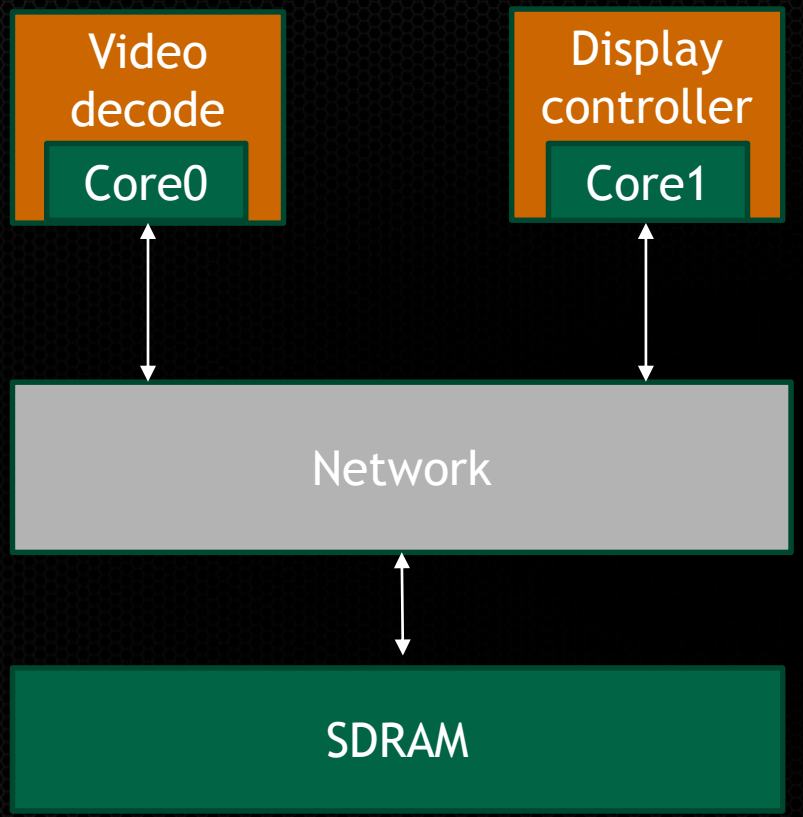
st [a], x

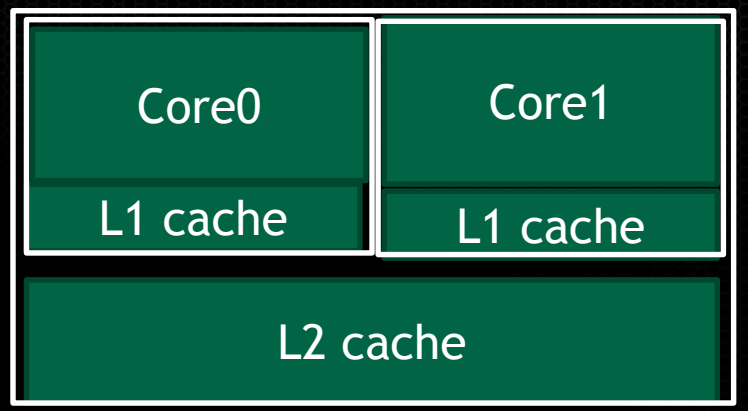
{y == 0}

Or can it?



Should HW or compiler prevent this?





Not here

Maybe here

Memory model final remarks

- Need to balance between performance and ease of use
  - Weak memory model allows for more HW optimizations
  - Strong memory model allows for simpler SW
- May need flexibility to switch /combine depending on use case
- Different choices and vagueness exists in established architectures

Security architecture workgroup

- Joe Xie chairing the workgroup
- Helped organize the workshop
- Workgroup meeting on Thursday

Why is it important to NVIDIA?

the total revenue lost to pirated games was approximately $74.1 billion in 2014

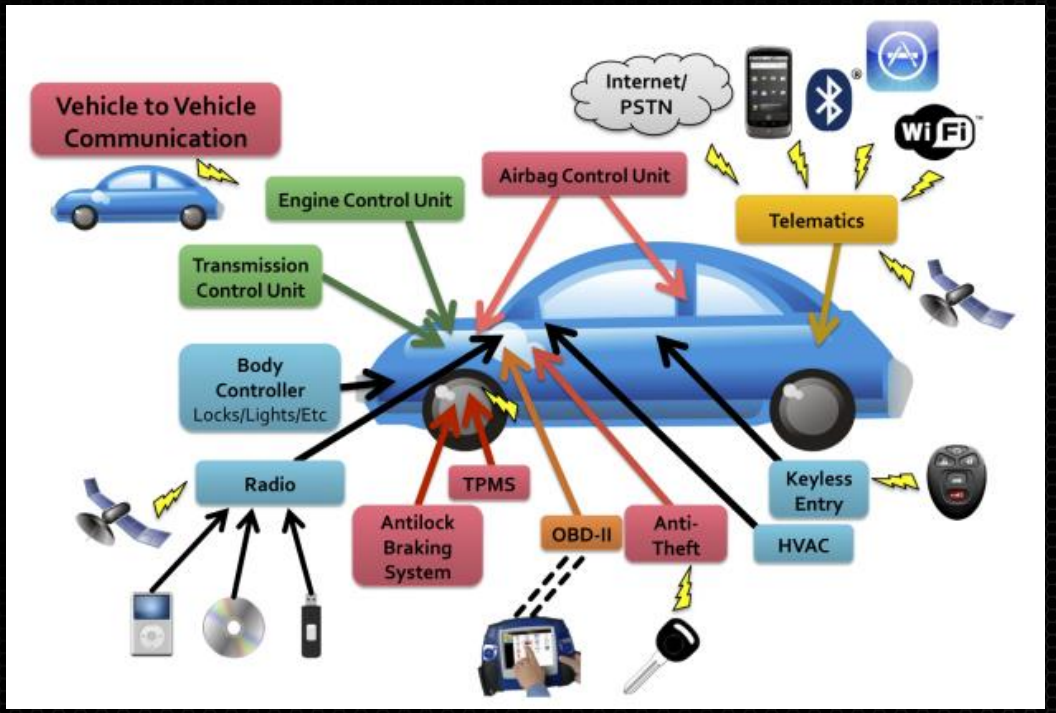
http://gearnuke.com/video-game-piracy-rise-will-cost-industry-much-makes/

Chinese company hacks Tesla car remotely

http://www.cnbc.com/2016/09/20/chinese-company-hacks-tesla-car-remotely.html

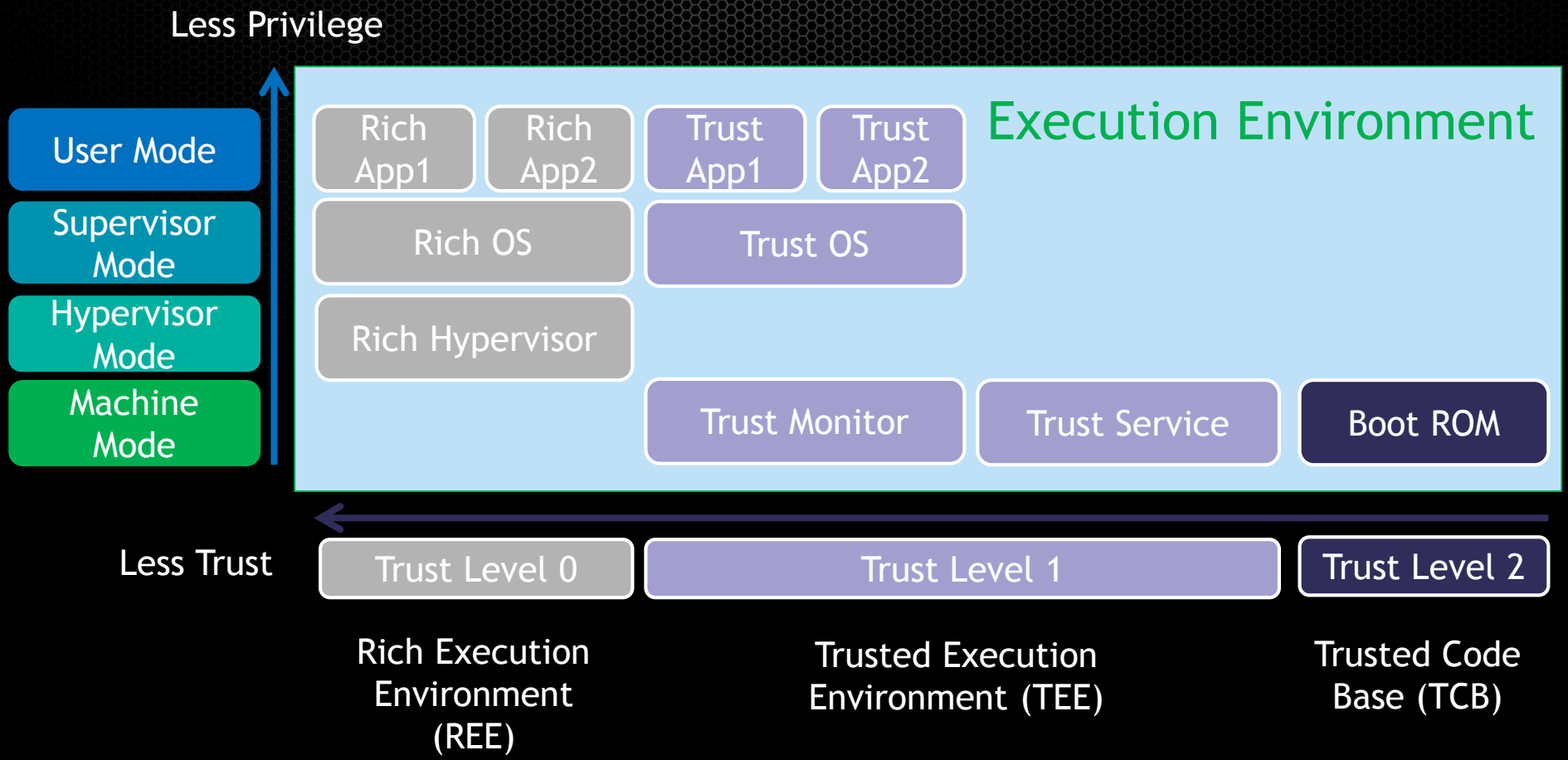
What is the problem?

- Attack surface is growing
- Systems become more open
- Consequences are more severe



http://www.autosec.org/pubs/cars-usenixsec2011.pdf

Add trust levels



Reduce attack surface of security sensitive software & provide isolation

Security Final Remarks

- Other proposals under discussion
  - Crypto ISA extensions to accelerate common algorithms (MicroSemi)
  - Security metatags
- Excellent overview of all proposals in

  Richard Newell, Sr Principal Product Architect, Microsemi Corp.

  Escrypt Security Class, Embedded World Conference, Germany

  March 15, 2017
- Hardware attacks (e.g. differential power attacks) also a concern; not preventable by architecture

Final thougths

- NVIDIA will use RISC-V processors in many of its products
- We are contributing because RISC-V and our interests align
- **Contribute to the areas that you feel passionate about!**