

#### About me

- + GPU Compute software architect at Arm
  - Working on Mali GPUs
- + Participating in Khronos working groups since 2016
  - Mostly OpenCL, SPIR, Vulkan



# A little quizz

- Who has heard of Arm?
- → Who has heard of Arm Mali GPUs?
- → Who has programmed an Arm Mali GPU?



#### Mali GPUs

Arm Mali is the most widely used GPU in the world

\*B\*illions of units!



### Agenda

- + A very brief history of Mali GPUs
- + Introduction to the architecture of Mali GPUs
- + OpenCL in mobile and embedded
- + Some challenges for OpenCL in mobile/embedded space



#### Mali GPUs – A brief history

- + 2005: Falanx announces Utgard architecture and Mali-200 GPU
- + 2006: Arm acquires Falanx
- → 2010-2014: Midgard architecture
  - Unified shaders, VLIW engine
  - https://www.anandtech.com/show/8234/arms-mali-midgard-architecture-explored
- + 2016-2018: Bifrost architecture
  - Warp engine, thread-level parallelism
  - https://www.anandtech.com/show/10375/arm-unveils-bifrost-and-mali-g71
- + 2019 onwards: Valhall architecture
  - See next slide
- + To learn more
  - https://en.wikipedia.org/wiki/Mali (processor)
  - https://www.youtube.com/watch?v=DO\_68Hjs2UI&list=PLKjl7IFAwc4QUTejaX2vpIwXstbgf8Ik7&index=5

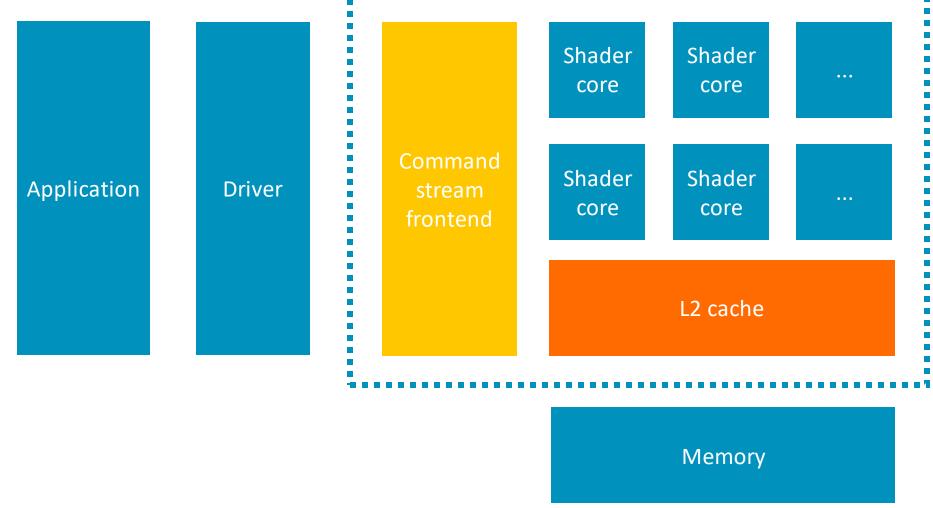


## Mali Valhall 5th gen architecture

- + Command-stream frontend
  - Flexible, good fit for modern APIs
- + 16-wide warp engine
  - Superscalar, matrix multiplication acceleration, etc
- Compression
  - o AFBC, AFRC, etc
- Ray tracing
- ... and a lot more

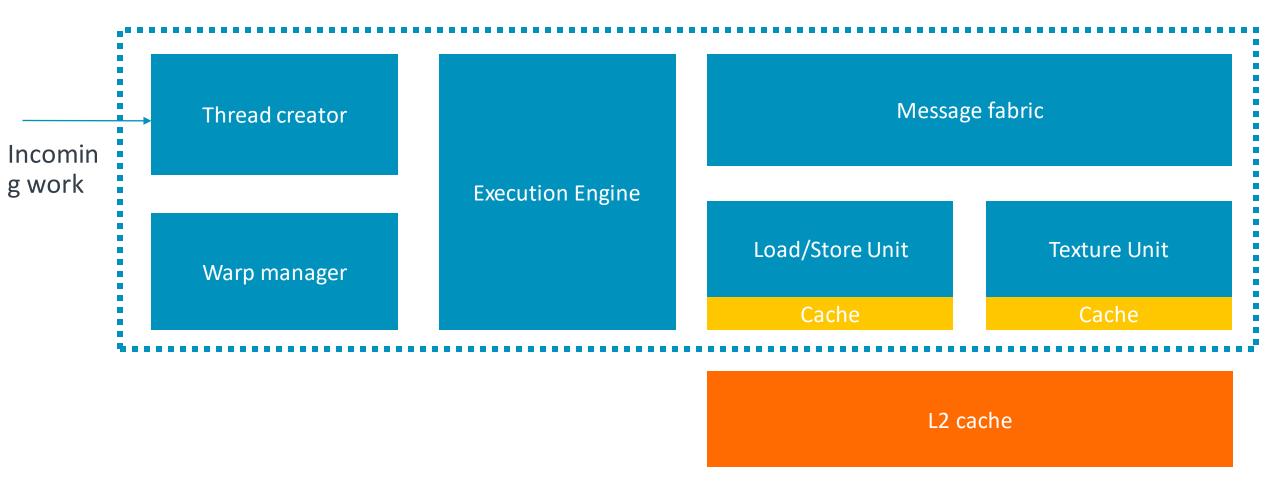


# Anatomy of a modern Valhall GPU





### Zooming into a shader core





## OpenCL in mobile and embedded

- + People often ask: where/how is OpenCL used in the mobile and embedded space?
- → The short answer: it's everywhere!
- From the phone you take your holiday pictures with
- ... to your electric car
- ... with a detour by your TV and/or set top box
- ... not forgetting your security cameras
- ... who said Machine Learning?
- ... and more!





#### Some challenges for OpenCL: compression

- Mobile platforms have stringent power constraints (typically 1-3 W)
- Memory bandwidth costs power, say 100mW / GB/s as a rule of thumb
- Compression is important to reducing power
- Cleanly integrating image compression technologies will require overhauling some aspects of image support
  - OpenCL lagging behind in terms of application control
  - See <u>image tiling control</u>, <u>DRM format modifier</u> extensions



## Some challenges for OpenCL: latency

- Many use cases are very sensitive to end-to-end latency
  - OpenCL used in a longer processing pipeline
- OpenCL has provided good improvements recently
  - Command buffers
  - External semaphores
- ... but there's more to do.





### Some challenges for OpenCL: Android

- OpenCL is ubiquitous on Android phones
  - Many key product value adds in shipping devices
- ... but not supported officially
- This status-quo is removing opportunities for GPU Compute on mobile devices
- Steer application developers towards Vulkan (officially supported)
  - ... but Vulkan is complex
  - o ... and migration is not free
- Layering on top of Vulkan could help
  - Reduce cost for applications and reuse investments in OpenCL



# Any questions?

+ Looking forward to interesting discussions.

- + Join us :)
  - o <a href="https://careers.arm.com/GPU-at-Arm">https://careers.arm.com/GPU-at-Arm</a>
  - https://careers.arm.com/search-jobs/GPU?orgIds=34601&kt=1





Thank You Danke Gracias Grazie 谢谢 ありがとう Asante Merci 감사합니다 धन्यवाद Kiitos شکرًا

ধন্যবাদ

תודה ధన్యవాదములు

