GEI Imagination TECHNOLOGIES

Scaling Graphics Performance with Multiprocessing

Kristof Beets Business Development Manager – POWERVR Graphics

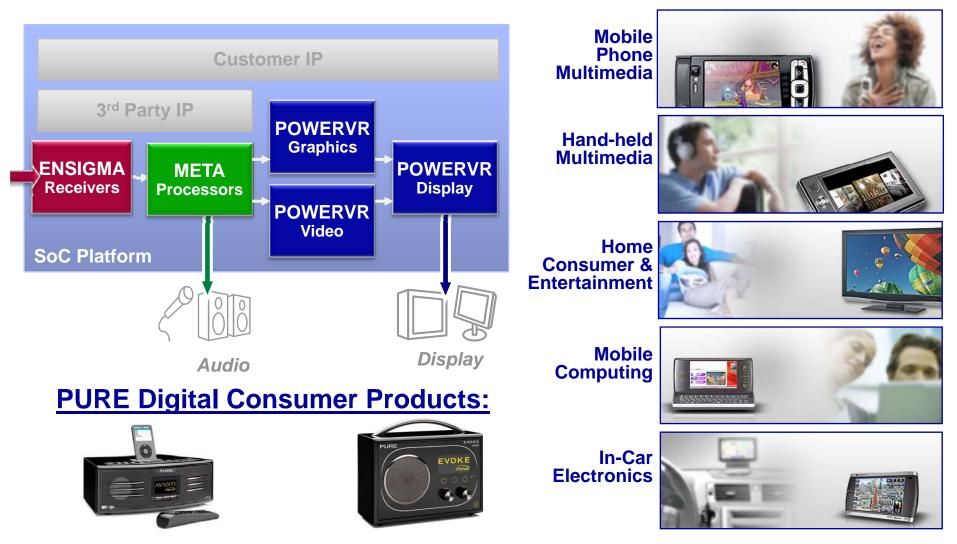
April 2nd 2009

Imagination – what we do



SoC Technologies & Solutions:

Markets:



© 2009 Imagination Technologies Ltd.

Our Licensees

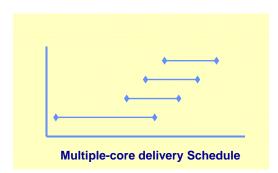




Graphics Acceleration The Scale of the Problem



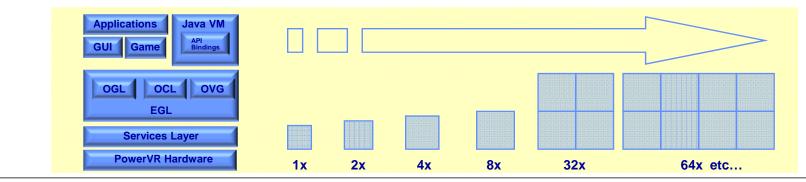
- The diversity of applications using embedded 3D requires an extremely broad range of performance, power and cost points.
 - Must track process nodes
 - Must be implementable using ASIC techniques



 It is essential to maintain a single, coherent software stack.



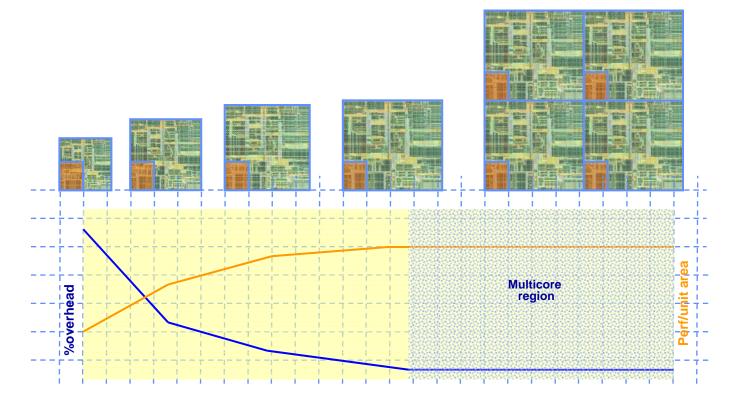
- Delivering those scalable solutions in a timely manner
 - Core variants must leverage base platform to enable rapid deliveries
 - Strike a balance between optimisation and time to market



© 2009 Imagination Technologies Ltd.

How to optimise scalability Multicore or optimised multi-pipelined cores?



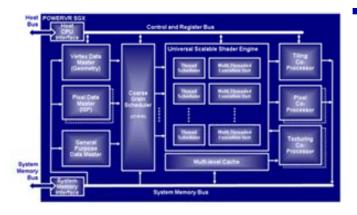


Infrastructure overhead as proportion of area varies with performance level

- Performance/unit area increases as absolute performance of multi-pipeline core increases
- Returns diminish as ratio of overhead to core area increases
- Balance point between efficiency and time to market determines the start of the Multicore Region.

SGX Overview & Roadmap



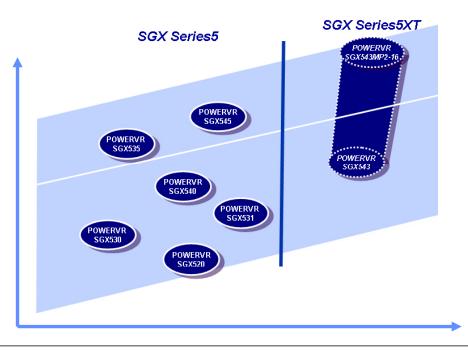


SGX architecture exploits the inherent fine-grained parallelism of graphics

 Task scheduling plus hardware thread management enables the problem to be distributed over multiple execution pipelines.

Screen Tiling

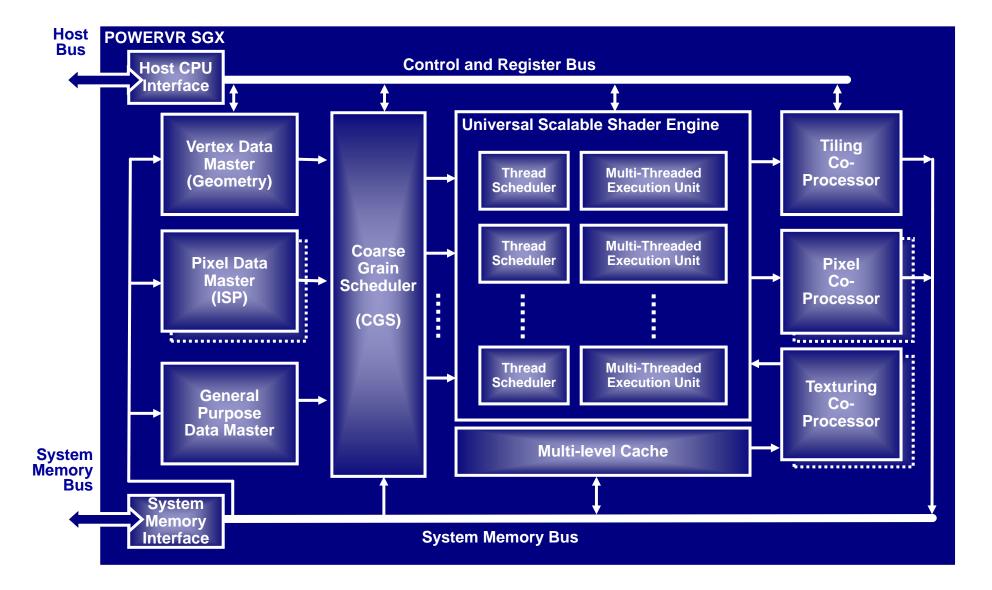




- As implementations move to multi-core, distributing the parallelism amongst cores becomes the key issue
 - TBDR offers a workable solution which can exploit parallelism without increasing latency.
 - Other solutions fail to distribute geometry workload or increase latency

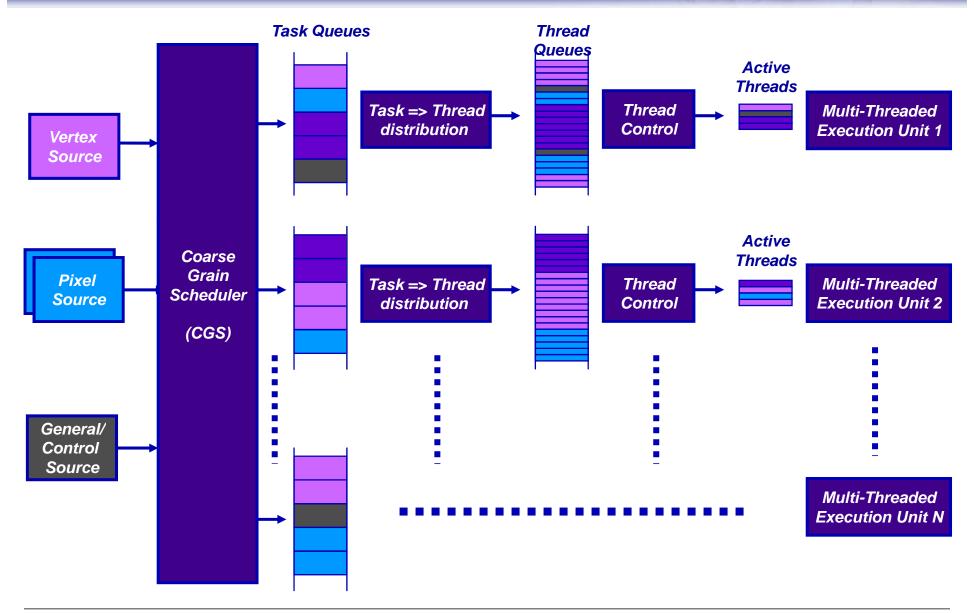
POWERVR SGX GP-GPU Architecture Data Flow Overview





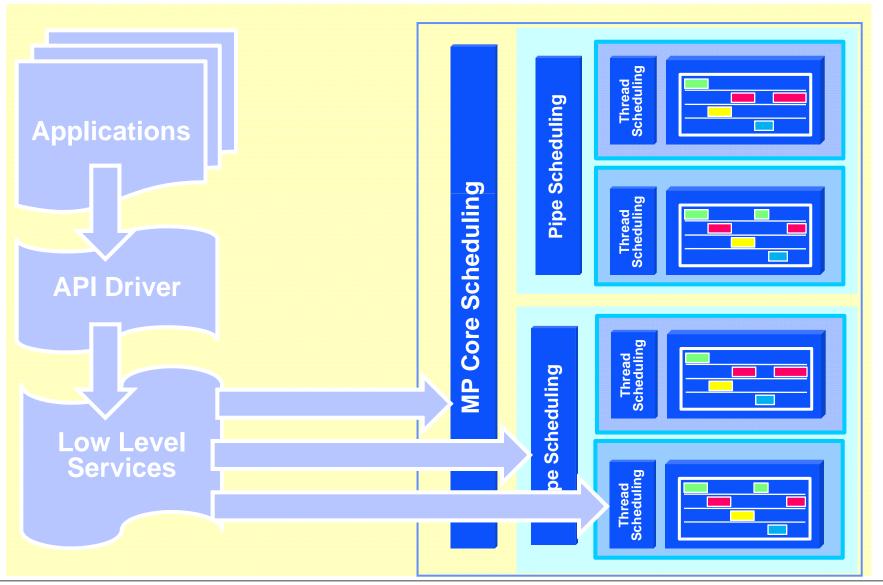
POWERVR USSE Thread Scheduling Overview





POWERVR SGX543 MP Architecture Hierarchical Scheduling of Tasks





© 2009 Imagination Technologies Ltd.



- An open standard for heterogeneous parallel programming
 - Enables applications to execute compute intensive tasks across one or more devices
 - Devices can include single/multi-core CPUs, GPUs or other general processing cores
 - Avoids the need to use custom or task-specific APIs to leverage a device's processing power - such as graphics APIs for computation using a GPU.

Components of OpenCL

- OpenCL consists of an API and C-like programming language
- User writes processing functions (kernels) in the OpenCL C language
 - A well defined subset of the C99 standard
 - Additional keywords and basic-types as appropriate for parallel & vector processing
 - Includes numerical accuracy requirements for mathematical operations
- The OpenCL API includes:
 - Device enumeration and querying
 - Kernel management and execution
 - Management of buffers and images for kernel input and output
 - Synchronisation and event handling
 - Interoperability with OpenGL and other graphics APIs
- For more details: <u>http://www.khronos.org/opencl/</u>

Conclusions



Tile based architectures are ideally placed for multicore scaling of performance

- But it has to be the right Tile Based Architecture: Scalable in all the correct dimensions.
 - Universal Load-balanced to handle Vertex/Pixel and General Instructions
 - The base unit has to be right inefficiencies are amplified by multicore
- Mobile Multicore solutions will be here within a year
 - Further erosion of the distinction between appliances and computers
 - Raw graphics power will empower UI and applications designers
 - GP-GPU already in R&D using SGX with all major OEMs e.g. Image Processing

The performance curve shows no signs of flattening out

In fact, we are just getting started...



GEI Imagination TECHNOLOGIES

Scaling Graphics Performance with Multiprocessing

Kristof Beets Business Development Manager – POWERVR Graphics

April 2nd 2009