# UK Systems Research Challenges Workshop:
## Fast, Unmodified, Full-system Mobile CPU/GPU Simulation

**Tom Spink**
tspink@inf.ed.ac.uk
Post-doctoral Researcher at the University of Edinburgh

Graphics Processing Units (GPUs) have seen a lot of attention over the past decade, and in particular have grown to support workloads that are not strictly graphical in nature. Their accelerated functions for computation, coupled with their high number of concurrent processing units enables fast computation for a wide range of computing domains.

GPUs are present in devices across the entire computing spectrum, from desktop and workstations, through to mobile phones and self-driving cars. There is a lot of development around GPU architectures, and this trend is growing rapidly.

But, a particular challenge is the actual simulation of these ubiquitous devices. The instruction set architecture is often proprietary, the device drivers are maintained by the companies themselves—in some cases shipped as binary blobs, and the inner workings of the GPU are opaque.

Although there is plenty of activity in the desktop GPU simulation space, there is a distinct lack of options for simulating the GPUs found on embedded platforms. These devices are highly proprietary, and so those simulators that do exist do not actually simulate the hardware platform itself, but only the high-level interface (e.g. OpenCL), missing out on crucial instrumentation opportunities at the architectural level.

Furthermore, these simulators either require modified user-space software stacks, or operate independently of an operating system, thus making true unmodified, like-for-like simulation impossible.

In this talk, I propose to discuss our recent work in the area of high-speed simulation for embedded GPUs, and how our full-system simulator for an Arm Mali G71 achieves it's high-speed. I will describe the techniques utilised in the simulator, how this compares to existing simulators, how we have validated our simulator against a reference platform, and what the future of our heterogeneous simulation techniques are in general.

**Notes** Our simulator is based on Captive [1], which was presented at the previous UKSRC workshop. We have also recently released our GPU simulator (`https://gensim.org/simulators/gpusim`), as part of our publication in ISPASS'19 [2].

If time permits, I can also host a live demonstration.

## References

[1] Tom Spink, Harry Wagstaff, and Björn Franke. Hardware-accelerated cross-architecture full-system virtualization. *ACM Trans. Archit. Code Optim.*, 13(4):36:1–36:25, October 2016.

[2] Kuba Kaszyk, Harry Wagstaff, Tom Spink, Bjoern Franke, Michael O'Boyle, Bruno Bodin, and Henrik Uhrenholt. Full-system simulation of mobile CPU/GPU platforms. In *IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2019, March 24-26*. IEEE, 2019.