The duration of this examination is 50 minutes

Please answer *all three* questions

This examination is partly based on the ARM Cortex A15 processor architecture, as described in the Cortex-A15 Technical Reference Manual (Revision r2p1), relevant parts of which are available to you in the examination. Where this material is incomplete, you are invited to speculate using your understanding of the underlying architectural principles.

1 The A15's write streaming no-allocate threshold is configurable (page 4-55). Its default value is 12; it can also be turned off.

a What is the purpose of this feature?

b What program behaviour would benefit from the default value?

c What program behaviour might benefit from turning it off?

The three parts carry, respectively, 35%, 30%, and 35% of the marks.

2 a Estimate the A15's branch misprediction penalty. Explain your reasoning.

b The A15 is designed for mobile, battery-powered applications where power efficiency is a key performance factor. What aspects of the out-of-order processor microarchitecture do you think reflect this concern?

The two parts carry equal marks.

3 a What do you think limits the number of A15 cores that can share memory in a cache-coherent way?

b Consider the following program fragment:

```
float A[N], B[N], C[N];
for (int i=0; i<N; ++i) {
  B[i] = max(A[i-1], A[i], A[i+1]);
}
for (int i=5; i<N-5; ++i) {
  C[i] = B[i-5]+B[i+5];
}
```

(i) Suppose `N` is large, so the arrays do not fit into the A15's L2 cache. What might be the performance advantage of *reversing* the order of execution of the second loop?

(ii) Explain how these two loops can be *fused* into a single loop.

(ii) Suppose the array `B` is not used in the remainder of the program. Is it necessary to store the array `B`?

The two parts carry equal marks.