

Lab 3 -Report

Brief description

We use the sniper simulator in combination with McPAT to analyse the energy consumption of our recent optimizations regarding cache and ILP. Therefore, we start with a base configuration close to LAB1's base config. While applying the optimizations of the last two weeks, we keep track of the peak and average power and calculate energy consumption. We will provide three different configurations for high performance, a maximal peak power of 11.5 watt and a third minimizing the energy delay product.

Results

Tuning for high performance

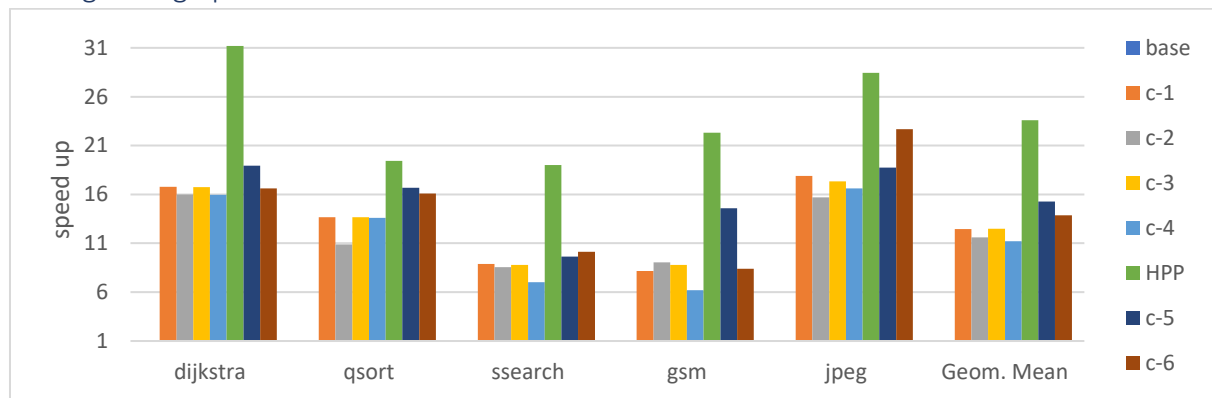


Figure 1: Different configurations to create a high-performance model

Like last lab we start setting all parameters to the maximum value (except for block size, c-1). This gives a mean speed up of 12.4. Reducing the number of RS entries to 16 (c-2) or setting L2-associativity to 2 (c-4) decreases performance while setting the associativity to four has almost no effect (c-3). The biggest impact has increasing the L2 block size to 64 (HPP). As for each cache the number of sets must be greater than block size, this requires increasing the L1 cache size. While we keep the cache size for L1 at that high level and only reduce the block size (c-5), we see that most of the speed up is indeed related to the block size. Reducing L2-cache size (c-6) to 16kB has a negative impact on performance.

Limit maximum power to 11.5 watt

Increasing the performance comes at a cost. Compared to the base configuration we reached a speed up of 31 while the used area increased by 307% and maximum power increased by a factor of 40 to 44 depending on the benchmark program. As too high peak power might cause thermal problems we limit the maximum power to 11.5 watt by reducing the available hardware.

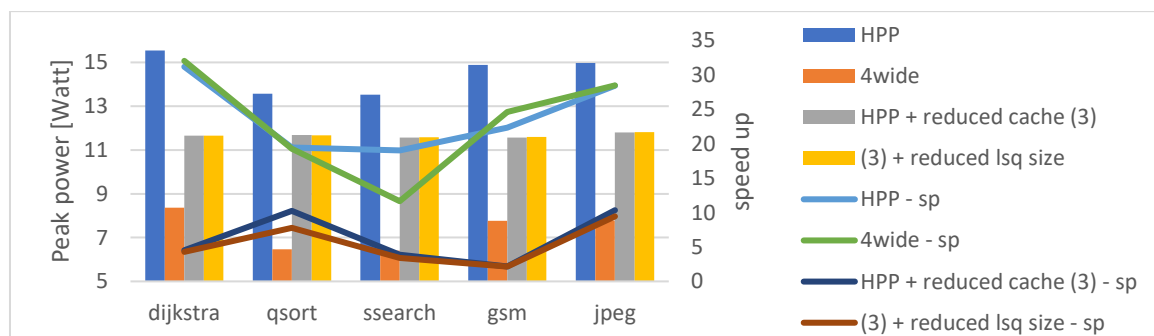


Figure 2: Exploiting different possibilities to cut down peak power usage

As we see reducing the decode and commit width to 4 offers less peak power while offering better performance than modifying cache and queue sizes.

Energy Efficiency

To improve the energy efficiency, we calculated the Energy-Delay-Product (EDP) for the base and HPP configurations. Comparing the EDP of this two configurations, it's possible to see that the EDP of base configuration is much higher than the EDP of HPP. The HPP is much more efficient (performance and energy improvement). Then, we try some new configurations (we change ROB to 64 and then we decrease width to 4). With both configurations, we had some improvements for some applications. But for the configuration with width equals to 4, we got an EDP lower for all apps, (comparing with HPP configuration). The E-eff design is closer to the base configuration (decreased ROB size to 64 and commit width to 4).

Design decisions

Firstly, we reached a configuration (HPP) with the higher performance. In this configuration we set all the parameters to the maximum value, except the associativity that we set it to 4. The biggest improvement was reached when, we change the L2 block size, because this also change the L1 cache.

After we studied the impact of the single parameters regarding energy efficiency, we try to improve it. For that, we study the trade-off between performance and energy consumption, using the EDP. We tried some different configurations and saw the impact of this changes in the EDP. In the end, we reach one configuration (E-eff), where the EDP was lower comparing with the HPP configuration. In E-eff, we set the ROB to 64 and commit width to 4.

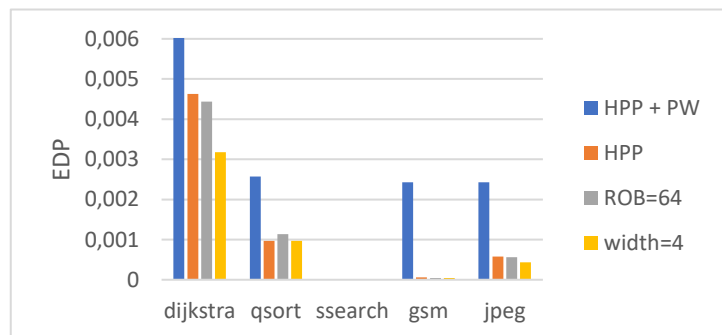


Figure 3: Improvements made to optimize EDP. Note that the value of dijkstra + powerwall is cut off (0.088). The last two entries are based on the HPP config.

Configuration	HPP	PW	E-EFF
In_order	false	false	false
LSQ size	16	16	8
RS entries	32	32	32
Timer window size	128	128	64
Width	8	4	4
L2 size	32	32	32
L2 associativity	4	4	4
Block size	64	64	64

Table 1: Final design after evaluation of simulation results

Conclusion

We have seen that increasing performance by adding hardware results in increased power usage. However, as the computation time is significantly reduced, higher peak power can still lead to less energy consumption – as long as we stay in some bounds while adding additional hardware components. There is a trade-off between tuning for high performance and energy efficiency e.g. if a processor is designed for mobile devices, but if one compares the HPP and E-EFF profile concerning performance, the last is much closer to HPP than to the power-wall or base profile.