

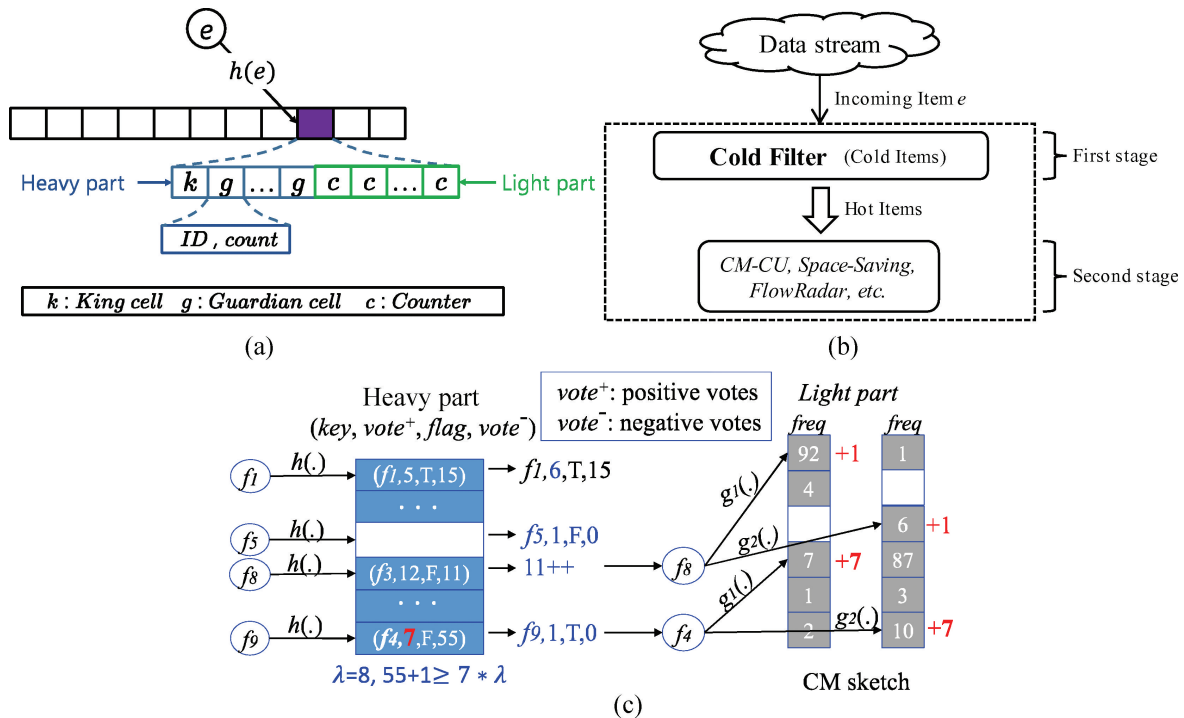
## Fast and accurate measurements in data streams

With the support by the National Natural Science Foundation of China, the research team directed by Dr. Yang Tong (杨全) at the School of Electronics Engineering and Computer Science, Peking University, recently proposed several novel sketching techniques for data streams, which were published on conferences of SIGCOMM'18, SIGMOD'18, and SIGKDD'18.

In data stream applications, accurate and fast measurement and mining are fundamental components, which provide various valuable information. Recently, sketch-based approaches have been considered as the best choice for these tasks. Previous studies mainly focus on a good trade-off among accuracy, speed and memory usage for one specific measurement or mining task. Yang's team focuses on multiple tasks while achieving high accuracy, high speed, and high memory efficiency.

All of their research is based on a key observation: the separation of hot items and cold items can improve time and space efficiency, as most of the real data streams are highly skewed. This means that most of the items are unpopular (called cold items), while a few items are very popular (called hot items). The separation has two advantages. On the one hand, real world applications have different requirements for hot items and cold items, so it is natural to separate them. On the other hand, hot items and cold items influence each other in traditional sketches, which leads to a poor accuracy for both of them. The key challenge is how to separate them accurately and efficiently. In ElasticSketch, they keep hot items and allow an item to evict a current hot item by voting. In ColdFilter, they use a one-direction filter to store cold items, and let only hot items enter the second stage. In HeavyGuardian, they keep and guard hot items by exponential decay.

Their work provides lightweight approaches for measurement and mining tasks in data streams, and has been implemented on advanced devices such as P4 switches and FPGA, which is significant for future larger and faster data stream applications.



**Figure** Basic structures of HeavyGuardian (a), ColdFilter (b), and ElasticSketch (c).