**INTRODUCTION**

External sorting is used primarily when data that needs to be sorted is unable to fit into RAM at one time, thus necessitating iterative access of blocks of this data to solve the problem. The algorithm consists of two main steps: first, the data is read into memory a block at a time, sorted, and written back out into temporary files and second, these temporary files are then merged two, three, four, ten, or etc. at a time, similar to merge sort.

**IMPLEMENTATION DETAILS AND ISSUES**

My implementation of external sort closely follows the specifications. Users can generate an input file of any size and name, choose page size (default 4K - 4096 bytes), choose a number of memory blocks B, choose input file name, and choose output file name.

1. During my implementation of PASS-0, I make the mistake of considering the size of a page equal to the size of a generated run equal to a page. For instance, if a page size is 4 KB and the user uses 20 blocks, than each temporary run generated will have a maximum capacity of 4 KB instead of 80 KB. For this reason, *more* runs are generated however, the space consumed and the number of reads and writes remains as it should because the amount of space and memory read/written is the same.
2. A second problem is that a small number of bytes are ignored during the merging procedure which means that step is imprecise, yet very close to being perfect.

**TIME PERFORMANCE (in milliseconds)**

The given table summarizes the observed time performance (in milliseconds) with varying file sizes, blocks, and page sizes. It illustrates, quite clearly, that external sorting cost is less when files are of lesser size, more ***B*** blocks are used, or the page size is larger. This can be explained by analyzing the equation representing the cost of external merge sort: 2N \*  logB-1 **where the term being multiplied on the right side represents the number of passes needed as a function of N and B, N being the number of pages and B being the number of blocks. An increase in B implies cost decreases and an increase in page size implies N decreases which again implies cost decreases.

