KVik metrics

$t := |\texttt{Hot Files}|/(|\texttt{Hot Files}| + |\texttt{Data File}|)$
S_k – size of key (in bytes)
S_v – size of value (in bytes)
N – number of records in `KVik`
M – number of keys `Log Hash Table` can hold (capacity)
f – size of offset (in bytes)
U_w – speed of write requests
E – number of write cycles (parameter of your SSD)
F – size of SSD
$w := S_k + 2 \cdot f$
c – `Log Hash Table` fill factor (by default, it is assumed to be 0.9)
s – `Index Hash Table` fill factor (by default, it is assumed to be 0.6)

Assumption: key and value can be placed on one memory page

`RAF` $= c \cdot M/N + (N - c \cdot M)/N \cdot (1 + 1) = 2 - c \cdot M/N$

Assuming, requests are distributed uniformly and will hit `Log Hash Table` with probability $c \cdot M/N$.

`WAF` $= 1 + 1/(c \cdot M) \cdot (2 \cdot c \cdot M) + 1/(N \cdot t) \cdot (2 \cdot N \cdot t) = 5$

First term is write operation to `Log File` that happens every time. Second term stands for situations when `Log Hash Table` is filled and needs to be merged into `Index Hash Table` (in fact it should be better because before performing modification of `Index Hash Table` we presort all keys by hash function used in there, so ones that hit same bucket will be written using only one write operation). Since it happens once in $c \cdot M$ operations we divide write operations it takes on it to take average. Each key will have two write operations (one for key written to `Index Hash Table` and one for value written to `Hot Files`). The third term is really similar to the second one but there we need one write operation to write value to `Data File` and one more write operation to update offset value in `Index Hash Table` (actually in case of update it can be done inplace so offset can remain the same but in general case it is so).

`SAF` $= (c \cdot M \cdot (S_k + S_v) + t \cdot N \cdot (S_v + S_k) + N \cdot S_v + 1.67 \cdot N \cdot w)/(N \cdot (S_v + S_k)) =$
$= c \cdot M/N + t + 1 - S_k/(S_v + S_k) + 1.67 \cdot (S_k + 2f)/(S_v + S_k) =$
$= t + 1 + c \cdot M/N + 0.67 \cdot S_k/(S_v + S_k) + 3.4 \cdot f/(S_v + S_k)$

Here everything is very simple: the first term in numerator is `Log File`, the second is `Hot Files`, the third is `Data File`, the fourth is `Index Hash Table` (fill factor was determined experimentally and $0.6^{-1} = 1.(6)$).

`MO` $= \frac{5 \cdot N/8 + M \cdot (S_k + f)}{N} = 5/8 + M/N \cdot (S_k + f)$

The first term in numerator is memory that we need to store `Filter`, the second term is `Log Hash Table`.