$\texttt{Index}(k_1/k_2/\ldots/k_P) = \sum_{m=1}^{P} \binom{k_m+m-1}{m}$

Examples:

- for P=2 and N=1, the ordering is 00,01,11
- for P=2 and N=2, the ordering is 00,01,11,02,12,22
- for P=3 and N=2, the ordering is 000, 001, 011, 111, 002, 012, 112, 022, 122, 222
- for P=1, the index of the genotype a is a
- for P=2, the index of the genotype "a/b", where $a \leq b$, is $b(b+1)/2+a$
- for P=2 and arbitrary N, the ordering can be easily derived from a triangular matrix

$b \setminus a$	0	1	2	3
0	0			
1	1	2		
2	3	4	5	
3	6	7	8	9

- HQ (Integer): Haplotype qualities, two comma separated phred qualities.
- LAA is a list of n distinct integers, giving the 1-based indices of the ALT alleles that are observed in the sample.

  In callsets with many samples, sites may grow to include numerous alternate alleles at the same POS. Usually, few of these alleles are actually observed in any one sample, but each genotype must supply fields like PL and AD for all of the alleles—a very inefficient representation as PL's size is quadratic in the allele count. Similarly, in rare sites, which can be the bulk of the sites, the vast majority of the samples are reference. To prevent this growth in VCF size, one can choose to specify the genotype, allele depth and the genotype likelihood against a subset of "Local Alleles". LAA is the 1-based index into ALT, defining the alleles that are actually in-play for that sample and the order in which they are interpreted. LAA is required when interpreting local-allele fields and must be present if any local-allele fields are neither omitted nor MISSING. Since BCF encodes zero length vectors as MISSING, a LAA containing the MISSING value should be treated as the empty vector (i.e. a REF-only site) if any local-allele fields are neither omitted nor MISSING. All specifications-defined A, R and G FORMAT fields have a local-allele equivalent that should be interpreted in the same manner as it's matching field except for the ALT alleles considered present and the order in which they are interpreted. For example, if REF is G, ALT is A,C,T,<*> and a genotype only has information about G, C, and <*>, one can have LAA=[2,4] and thus LPL will be interpreted as pertaining to the alleles [G, C, <*>] and not contain likelihood values for genotypes that involve A or T. GQ is still the genotype quality, even when the genotype is given against the local alleles. In the following example, the records with the same POS encode the same information (some columns removed for clarity):

POS	REF	ALT	FORMAT	sample
1	G	A,C,T,<*>	GT:LAA:LAD:LPL	2/4:2,4:20,30,10:90,80,0,100,110,120
1	G	A,C,T,<*>	GT:AD:PL	2/2:20,.,30,.,10:90,.,.,80,.,0,.,.,.,.,100,.,110,.,120
2	A	C,G,T,<*>	GT:LAA:LAD:LPL	0/3:3:15,25:40,0,80
2	A	C,G,T,<*>	GT:AD:PL	0/3:15,.,.,25,.:40,.,.,.,.,.,0,.,.,80,.,.,.,.,.
3	C	G,T,<*>	GT:LAA:LAD:LPL	0/0:3:30,1:0,30,80
3	C	G,T,<*>	GT:AD:PL	0/0:30,.,.,1:0,.,.,.,.,.,.,30,.,.,80
4	G	A,T,<*>	GT:LAA:LAD:LPL	0/0::30:0
4	G	A,T,<*>	GT:AD:PL	0/0:30,.,.,.:0,.,.,.,.,.,.,.,.,.

  Due to BCF encoding empty vectors as missing, implementation-defined Number=LA local-allele fields should not be used if distinguishing between zero-length data and missing data is required at REF-only sites.

  It is recommended that VCF libraries provide an API in which local allele encoding can be abstracted away from the API consumer and values accessed through their corresponding non-local key.

- LPL: is a list of $\binom{n}{\text{Ploidy}}$ integers giving phred-scaled genotype likelihoods (rounded to the closest integer; as per PL) for all possible genotypes given the set of alleles defined in the LAA local alleles. The precise ordering is defined in the GL paragraph.
- M[0-9]+[ACGTUN] (Float): Fraction of DNA or RNA bases modified with the given ChEBI ID.