There is an excellent description of loop and compiler optimizations available from the canonical text by Severance and Dowd. Unfortunately the text has not been updated for modern hardware and languages, but the principles are still sound. Chapter 4 sections 1 and 4 speak specifically to optimizing compilers and loop optimizations.

It is available as a community commons license.

https://cnx.org/contents/u4IVVH92@5.2:bEZZukPR@1/Introduction-to-the-Connexions-Edition

While memory hierarchy numbers are constantly changing, Wikipedia's entry on the memory hierarchy is surprisingly up to date.

https://en.wikipedia.org/wiki/Memory hierarchy

Other thoughts: https://wiki.gentoo.org/wiki/GCC_optimization