Refutation of three phase, all reduce algorithm across processing units for scalable deep learning

© Copyright 2018 by Colin James III All rights reserved.

Abstract: A three-phase algorithm to do an all-reduce across all GPUs is not tautologous and refuted.

We assume the method and apparatus of Meth8/VŁ4 with Tautology as the designated proof value, F as contradiction, N as truthity (non-contingency), and C as falsity (contingency). Results are a 16-valued truth table in row-major and horizontal, or repeating fragments of 128-tables for more variables. (See ersatz-systems.com.)

LET  p, q, r, s;  ~ Not;  &  And;  > Imply, greater than.


[G]roup k GPUs together, then use a three-phase algorithm to do the all-reduce across all GPUs ... Figure 5: ...

1. reduce within GPUs of the same group,
2. store the partial results to a master GPU in each group, then ...
3. launch Ring all-reduce across p/k groups: after each master GPU gets the final result, propagate the final result [back] to every GPU.

Fig 5: Three phase, all reduce algorithm for GPU aggregation.

We ignore the intra ring of phase one as trivial, and assign logic values to the four inter rings as row-major to map the data flow in both directions.

\((s>r)>(p>q))\&((s>q)>(p>r))\);
\(TFFF\ \ TFFT\ \ TTTF\ \ TFTT\) (2.2)

We map the discrete broadcast phase as:

\(((s>p)\&(s>q)\&(s>r)))\&(\(r>p)\&(r>q)\&(r>s)))\&\)
\(((q>r)\&(q>s)\&(q>p)))\&(\(p>q)\&(p>r)\&(p>s)))\);
\(TFFF\ \ FFFF\ \ FFFF\ \ FFFT\) (3.2)
Remark: Eq. 2.1 contains the "then" word as a connective meaning the implication operator applies to Eqs. 2.1 as implying 3.1. In other words, if Eq. 2.1, then Eq. 3.1.  

\[(p>q) > (s>r) \land (p>r)\]  
\[(((s>p) \land ((s>q) \land (s>r))) \land ((r>p) \land ((r>q) \land (r>s)))) \land (((q>r) \land ((q>s) \land (q>p))) \land ((p>q) \land ((p>r) \land (p>s))))\]  

\[
\begin{array}{cccc}
T & T & F & T \\
F & T & F & T \\
F & F & F & F \\
T & F & F & F \\
\end{array}
\]  

Eqs. 2.2, 3.2, and 4.2 as rendered are *not* tautologous. This means the three-phase algorithm to do the all-reduce across all GPUs is refuted.