



וואווי הציגה ב-Chips Hot מאיץ בינה מלאכותית שבכוונתה להרחיב מהפעלה של מכשירים לבישים לעבודות אימון ותפעול במרכזי נתונים. היא גם תיארה מערכות המבוססות עליהם, ממערכות על שבב לטלפונים חכמים, מכונות ותחנות בסיס סלולרית ועד לשרתים ואשכול של 512 פטה-פלופס. המצגת הראתה עבודה מאוד מתקדמת בשבבים, תוכנה ומערכות. בהיבטים רבים נראה שוואווי מקדימה מתחרים כמו אינטל או אפילו חברות הזנק נוטלות סיכונים כמו סרברס שמתמקדת באופן צר באימון במרכזי נתונים.

הדור הראשון של שבבים המבוססים על ליבות DaVinci תוכננו תוך 11 חודשים בלבד. החברה מצפה לפרוס 100 מיליון מכשירים שמשמשים בליבות השנה, אמר הנג ליאאו, מנהל מו"פ בוואווי שהציג את המצגת בווידאו לאור המתיחות הפוליטית בגלל מלחמת הסחר ארה"ב - סין, שחלק גדול ממנה מתמקד בחברה שלו.

"זה היה כנס בינ"ל במשך שנים, ואני שמח שאתה מציג את המצגת כאן", אמר מהנדס מאינטל בחלק של שאלות ותשובות, וקיבל מחיאות כפיים ספונטניות של תמיכה מהנוכחים. "יש לנו גרסאות לקוח [של DaVinci] לטלפונים סלולריים ומצלמות חכמות שמשמשות באותה ארכיטקטורה כמו מרכזי הנתונים. הבנו שיצירת ערימת תוכנה היא מאמץ עצום, ולכן רצינו שאותן תוכנה וארכיטקטורה ישמשו ממכשירים מאוד קטנים ועד למכשירים מאוד גדולים", אמר ליאאו. התוכנה כוללת את Mindspore, מסגרת הבינה המלאכותית של וואווי. החברה גם פיתחה ממנה שתי שכבות של תוכנה, Pytorch ו-TensorFlow, לתרגם ממנה לחומרה שלה. ברבעון הבא וואווי תפרסם בוחני ביצועים של שבבי DaVinci פרטניים ושל האשכול שלה עם 2,048 צמתים. התוצאות יכללו השתתפות בבוחני הביצועים MLPerf, אמר ליאאו. מוצר הדגל הוא Ascend +N7 910, פרוסה של 182mm² עם 32 ליבות DaVinci, שמספקת 256 טרה-פלופס בפעולות נקודה צפה של 16 סיביות ומפזרת 350W. וואווי תכננה צומת שרתים של 6,000W שכולל במארז שמונה מהשבבים ו-DRAM של 1.5 טרה-בייט. ליאאו תיאור את הדור הבא של 910 שמשמש בערימה תלת מימדית של זיכרון מטמון SRAM וגם בזיכרון ראשי HBM. גרסה לטלפונים חכמים תשתמש ב-DRAM תלת מימדי מותאם ספציפית עם ק"פ רחב שמחובר למעבד יישומים. "הניתוח שעשינו מראה שבתחום 300-400W אפשר לפתור את הבעיות, האתגר היותר גדול הוא לערום. ליאאו אמר, "יותר הדוק חום תקציב יש DRAM-ול, דק אותו לעשות הקושי בגלל, DRAM, אחד הנוכחים שאל את ליאאו על הוויכוח כיום לגבי האפשרות שערימות ה-HBM מהדור הבא יעברו אל מעל לשבב המאיץ כדי להתמודד עם טווח ההגעה הקטן. "נראה ש-HBM בצד היא עדיין הגישה הכי מעשית, אבל ננסה למקם גם בצד וגם מעל", אמר ליאאו.

"אנחנו מאמינים שצריך לשפר את טכניקת החיבור [הנוכחית של HBM] ואפשר להגדיל את צפיפות החיבור. זה יעזור להפחית משמעותית את ההספק", הוסיף.

{loadposition content-related}