Annexure A

Format of storing data in the system for PCR-Penta as opposed to storing in *interleaved* format for other algorithms

The difference between the two formats is better understood though an example. Let two 7×7 pentadiagonal systems be given by

$$\begin{pmatrix} c_1 & d_1 & e_1 & & & & \\ b_2 & c_2 & d_2 & e_2 & & & \\ a_3 & b_3 & c_3 & d_3 & e_3 & & \\ & a_4 & b_4 & c_4 & d_4 & e_4 & \\ & & a_5 & b_5 & c_5 & d_5 & e_5 \\ & & & a_6 & b_6 & c_6 & d_6 \\ & & & & a_7 & b_7 & c_7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix}$$

and

$$\begin{pmatrix} r_1 & s_1 & t_1 & & & & \\ q_2 & r_2 & s_2 & t_2 & & & \\ p_3 & q_3 & r_3 & s_3 & t_3 & & \\ & p_4 & q_4 & r_4 & s_4 & t_4 & \\ & & p_5 & q_5 & r_5 & s_5 & t_5 \\ & & & p_6 & q_6 & r_6 & s_6 \\ & & & & p_7 & q_7 & r_7 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \end{pmatrix}.$$

For solving the above two systems by PCR-Penta, system data should be stored in six arrays, each of length 14, in the device global memory. The first five arrays, call them d_a, d_b, d_c, d_d, and d_e respectively, contain the five diagonals and the last array, denoted by d_y, contains the right hand side vectors of both the systems. Then for successful execution of PCR-Penta the system data should be stored in the format given below.

$$\begin{aligned} d\_a &= (0, 0, a_3, a_4, a_5, a_6, a_7, 0, 0, p_3, p_4, p_5, p_6, p_7), \\ d\_b &= (0, b_2, b_3, b_4, b_5, b_6, b_7, 0, q_2, q_3, q_4, q_5, q_6, q_7), \\ d\_c &= (c_1, c_2, c_3, c_4, c_5, c_6, c_7, r_1, r_2, r_3, r_4, r_5, r_6, r_7), \\ d\_d &= (d_1, d_2, d_3, d_4, d_5, d_6, 0, s_1, s_2, s_3, s_4, s_5, s_6, 0), \\ d\_e &= (e_1, e_2, e_3, e_4, e_5, 0, 0, t_1, t_2, t_3, t_4, t_5, 0, 0), \\ d\_y &= (y_1, y_2, y_3, y_4, y_5, y_6, y_7, z_1, z_2, z_3, z_4, z_5, z_6, z_7). \end{aligned}$$

For other parallel pentadiagonal solvers such as cuPentaBatch, and *getrf* and *gpsv* functions corresponding to cuBLAS and cuSPARSE libraries, system data should be stored in a different format for **interleaved memory access** which is given by

$$
\begin{aligned}
d\_a &= (0, 0, 0, 0, a_3, p_3, a_4, p_4, a_5, p_5, a_6, p_6, a_7, p_7),\\
d\_b &= (0, 0, b_2, q_2, b_3, q_3, b_4, q_4, b_5, q_5, b_6, q_6, b_7, q_7),\\
d\_c &= (c_1, r_1, c_2, r_2, c_3, r_3, c_4, r_4, c_5, r_5, c_6, r_6, c_7, r_7),\\
d\_d &= (d_1, s_1, d_2, s_2, d_3, s_3, d_4, s_4, d_5, s_5, d_6, s_6, 0, 0),\\
d\_e &= (e_1, t_1, e_2, t_2, e_3, t_3, e_4, t_4, e_5, t_5, 0, 0, 0, 0),\\
d\_y &= (y_1, z_1, y_2, z_2, y_3, z_3, y_4, z_4, y_5, z_5, y_6, z_6, y_7, z_7).
\end{aligned}
$$