

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 June 2002 (06.06.2002)

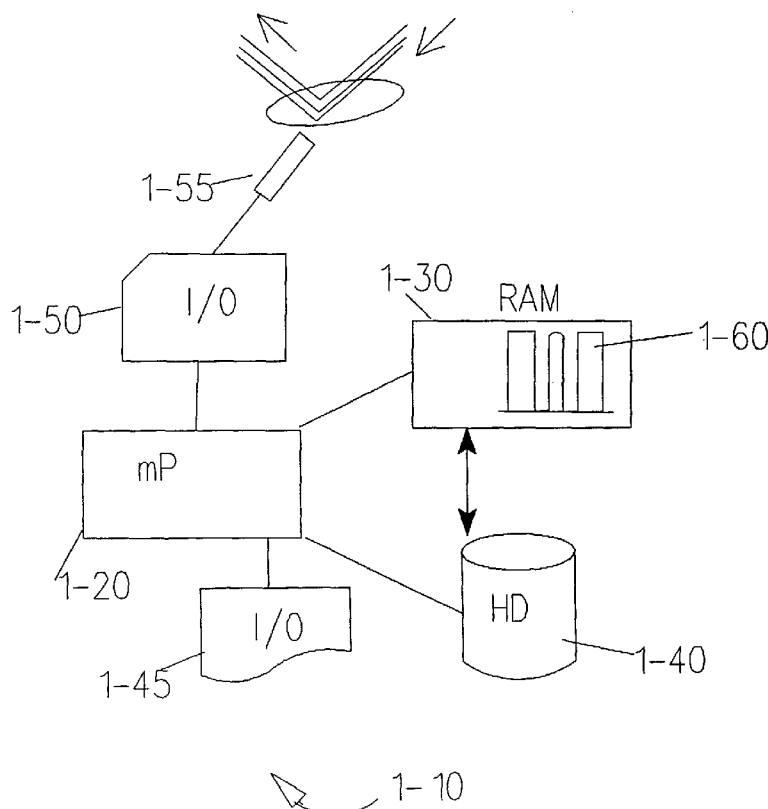
PCT

(10) International Publication Number
WO 02/44820 A2

- (51) International Patent Classification⁷: **G05B 19/00** (74) Agents: **YIM, Peter, J.** et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).
- (21) International Application Number: PCT/US01/44616
- (22) International Filing Date:
28 November 2001 (28.11.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/727,531 28 November 2000 (28.11.2000) US
- (71) Applicant (for all designated States except US): **TIMBRE TECHNOLOGIES, INC.** [US/US]; 5341 Randall Place, Fremont, CA 94538 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **DODDI, Srinivas** [IN/US]; 470 Boynton Avenue, #28, San Jose, CA 95117 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: CLUSTERING FOR DATA COMPRESSION



(57) Abstract: An improved profileometry data collection and analysis system employing software that performs clustering analysis on library data stored in memory that represent semiconductor chip wafer profiles, for use in matching real-time data signals from data collected by profileometry instruments. To better perform a match in real-time between the incoming real-time data signals and the profile library data into clusters, and to extract representative cluster data points of the clusters. The representative of the clusters are stored in primary memory (e.g. RAM), while the data forming the clusters are stored in secondary memory (e.g., a hard drive). A real-time data signal is then first compared to the representative cluster data points, and when a match is made with a particular representative cluster data point, the cluster associated with the representative cluster data point is loaded from secondary memory into primary memory. Next a further search is made with the incoming real-time data signal to find the closest match to the data in the cluster. In this way the entire library data does not have to be searched sequentially, and the entire library does not have to reside in primary memory in order to be quickly searched, which both conserves time and

primary memory. Techniques are disclosed to further refine special cases of points residing on the boundary of the cluster. In a preferred embodiment, the partitioning method for the cluster is based on the T. Gonzalez algorithm.



Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.