

A 75 nm 7 Gb/s/pin 1 Gb GDDR5 Graphics Memory Device With Bandwidth Improvement Techniques

Rex Kho, David Boursin, Martin Brox, Peter Gregorius, *Member, IEEE*, Heinz Hoenigschmid, *Member, IEEE*,
 Bianka Kho, Sabine Kieser, Daniel Kehrer, Maksim Kuzmenka, Udo Moeller, Pavel Veselinov Petkov,
 Manfred Plan, Michael Richter, Ian Russell, Kai Schiller, Ronny Schneider, Kartik Swaminathan,
 Bradley Weber, Julien Weber, Ingo Bormann, Fabien Funfrock, Mario Gjukic, Wolfgang Spirkl, Holger Steffens,
 Jörg Weller, *Member, IEEE*, and Thomas Hein

Abstract—Modern graphics subsystems (gaming PCs, mid-high end graphics cards, game consoles) have reached the 2.6–2.8 Gb/s/pin regime with GDDR3/GDDR4, and experimental work has shown per pin rates up to 6 Gb/s/pin on individual test setups. In order to satisfy the continuous demand for even higher data bandwidths and increased memory densities, more advanced design techniques are required. This paper describes a 7 Gb/s/pin 1 Gb GDDR5 DRAM and the circuit design and optimization features employed to achieve these speeds. These features include: an array architecture for fast column access, a command-FIFO designed to take advantage of special training/tracking requirements of the GDDR5 interface, a boosting transmitter to increase read eye height, sampling receivers with pre-amplification and offset control, multiple regulated internal voltage ($V_{INT} = 1.3$ V) domains to control on chip power noise, and a high-speed internal VINT power generator system. The memory device was fabricated in a conventional 75 nm DRAM process and characterized for a 7 Gb/s/pin data transfer rate at 1.5 V V_{ext} .

Index Terms—Graphics DRAMs, read and write training, low latency synchronization, boosted transmitter, sampling receivers with offset compensation, high speed core, high speed vint with multiple domains.

I. INTRODUCTION

MEMORY interfaces for graphics sub-systems have reached the 4 Gb/s/pin regime with the introduction of GDDR4 [1]–[3], but reliably achieve 2.6–2.8 Gb/s/pin in real world systems. The primary method used in order to achieve these speeds has been to simultaneously increase system and data clock frequencies. This puts pressure on designers and technology to increase core (array) frequencies well beyond

Manuscript received April 30, 2009; revised August 07, 2009. Current version published December 23, 2009. This paper was approved by Guest Editor Eugenio Cantatore.

R. Kho was with Qimonda, and is now with Bosch Sensortec, Germany (e-mail: climbob@lewermann.de).

D. Boursin, B. Kho, S. Kieser, I. Russell, K. Schiller, B. Weber, and M. Gjukic are with Qimonda AG, Munich, Germany.

M. Brox, H. Hoenigschmid, M. Kuzmenka, M. Plan, M. Richter, R. Schneider, K. Swaminathan, I. Bormann, F. Funfrock, W. Spirkl, J. Weller, and T. Hein are with Elpida Memory Europe GmbH, Munich, Germany.

P. Gregorius and D. Kehrer are with Infineon Technologies AG, Munich, Germany.

U. Moeller was with Siemens AG, Munich, Germany.

P. V. Petkov is with Qimonda AG, Neubiberg, Germany.

J. Weber is with EADS Astrium GmbH, Germany.

H. Steffens is with Ident Technology AG, Wessling, Germany.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSSC.2009.2034417

those of DDR2 or DDR3 memories, and create complex clocking systems for command and latency control [2], [3]. GDDR5 introduces a new training and tracking scheme, a novel clocking scheme, pre-fetch 8 and array bank grouping in order to break the speed barriers of GDDR3/4 [6]. In the GDDR5 clocking scheme, the command clock (CK) runs at half the frequency of the data clocks (WCK). Combined with a pre-fetch of 8 and bank grouping, this reduces array and command control speed requirements and allows design effort to be focused on the high speed (Data) portion of the GDDR5 system. Due to the introduction of training and tracking, the control of read/write latency can also be considerably simplified.

The new interface features of GDDR5 and higher data clock frequencies require special design techniques in order to realize the full potential of the GDDR5 specification. Various I/O techniques have been proposed and implemented in order to properly transmit and receive ever smaller data eyes. These include, but are not limited to DFE (decision feedback equalization) receivers, integrating receivers, multi-level signaling, and pre-emphasis [1], [4], [7], [8]. Previous work has also primarily concentrated on external noise sources such as Simultaneous Switching Noise (SSN), Inter Symbol Interference (ISI) and crosstalk, and shown various methods to control these [1]–[5], [7], [8]. In order to reliably achieve even higher speeds in real graphics systems, it is necessary to improve upon the I/O circuitry, as well as control internal noise sources. In this paper, we will address the following problems to achieving 7 Gb/s/pin in a GDDR5 system: Improvements to the array architecture in order to achieve sufficient data throughput, clocking and latency control issues, transmitter and receiver enhancements to improve data eyes, and internal power and noise control.

II. CHIP ARCHITECTURE

In order to achieve fast column-column access times, array improvements are necessary. In previous generations of graphics chips without bank grouping, the array was arranged such that the column select lines would extend from the Master CSL Decode areas (Fig. 1) over all banks. This forced the usage of large drivers/chip area in the Master CSL Decode section in order to achieve the necessary CSL-CSL activation speeds. In order to attain higher CSL-CSL Read and Write speeds, this implementation divides the column select architecture into Master and Local decode sections. An expanded view of one of the 256 Mb quadrants (QUAD) is shown in Fig. 1. Decode of

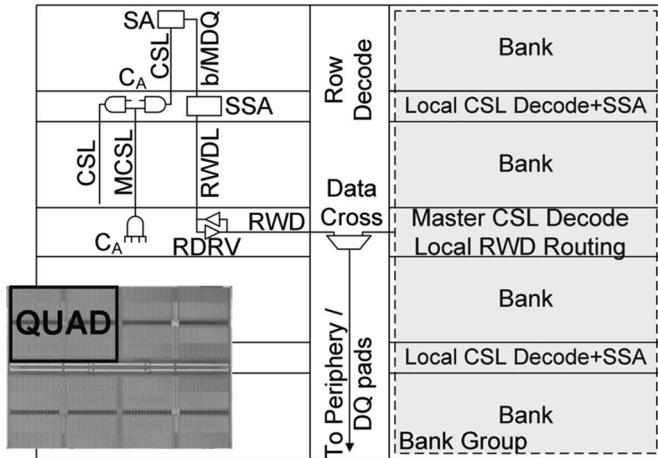


Fig. 1. Floorplan of expanded quad (photo + quad).

master column select lines (MCSL) is performed in the master CSL decode/local rwd (read write data) routing stripes. The MCSLs are routed to the Local CSL decode/secondary sense amp (SSA) stripes located between 2 banks. The MCSLs are further decoded to local CSLs which run over the banks. The local CSLs switch the local sense amp (SA) data onto short local data lines and then to the master data (MDQ) lines. The MDQ signals are re-sensed in the SSA stripes, and read-write data lines (RWDL) are transmitted to the MCSL decode/local rwd routing stripes and re-buffered. The re-buffered data lines are re-transmitted over the data crossing and then transferred to the periphery data circuitry and the data (DQ) pads. Fig. 1 shows the data transmission path for a read access. For a write access, the decode of the CSL path is unchanged, but the data path is simply reversed. All data lines are driven bidirectionally to minimize wiring area. The short CSL and MDQ lines enable fast random (2 clock cycle) column-column access times for read and write array accesses. In bank grouping mode [6], column-column access is only allowed by toggling between bank groups. In Fig. 1, a bank group is highlighted and represents a grouping of 4 independent banks. Each Quad has 2 bank groups. During back to back read-read or write-write accesses, toggling only between bank groups relaxes core speed requirements above the GDDR5 specified 5 Gb/s/pin (625 MHz core access speed) by allowing more time for the MCSL/CSL pulses. In this mode, the MCSL/CSL signal high pulselength is extended by an analog delay.

III. CLOCKING AND LATENCY CONTROL

The GDDR5 system introduces a new clocking scheme with two write clocks WCK01 and WCK23, and the conventional command clock CK [6]. One WCK clock (driven at double the frequency of CK) is provided per two bytes of data.

To minimize complexity of the GDDR5 memory, any active alignment of CK and WCK phases on the DRAM can be omitted, thus leaving open the possibility of PLL/DLL free operation. Alignment of CK and WCK phases is accomplished by system level initialization in GDDR5. To serve this purpose, an on-chip phase detector returns early/late phase information back to the controller (PD Fig. 2).

During system initialization, the controller must shift WCK relative to CK, until the phase feedback information from PD shows that the CK/WCK phases are aligned internally in the DRAM. Careful control of internal DRAM delays is needed in order to optimize timing for write data transport into the memory array, and ensure correct latency control for write and read. In our implementation, the delays Dwck, Dck, and Dcl have been chosen in order to fulfill the following requirements:

Referring to Fig. 2, the governing equations (PLL off) are

$$\tau_{WCKRcv} + \tau_{BypMux} + \tau_{Bx} = Na \quad (1)$$

$$\tau_0 + \tau_{CKRcv} + \tau_{Ax} = Nb \quad (2)$$

$$\tau_0 + \tau_{CKRcv} + \tau_{CMDFF} + \tau_{Ex} + \tau_{Dcl} = Nc \quad (3)$$

$$\tau_{WCKRcv} + \tau_{BypMux} + \tau_{S2P} + \tau_{Dx} = Nd \quad (4)$$

$$\tau_{WCKRcv} + \tau_{BypMux} + \tau_{Dwck} = Pwck \quad (5)$$

$$\tau_0 + \tau_{CKRcv} + \tau_{Ax} + \tau_{Dck} = Pck \quad (6)$$

where τ_0 = WCK-CK offset at DRAM pins.

After alignment of WCK/CK phases by the controller, $Pwck = Pck$, and it is also desired (in our implementation) that $Na = Nb$, and $Nc = Nd + t_{Setup}$ (Periphery FF). Combining (1)/(2) and solving for τ_0 , we arrive at (7), which is important for the circuitry which interfaces to a GDDR5 chip. It is important to control the terms in (7) in order to limit the skew seen at the WCK and CK pins of the DRAM. A large skew forces the user to have a deep FIFO on the CPU/GPU side. By setting $Pwck = Pck$ (5)/(6), substituting in τ_0 from (7), and solving for τ_{Dwck} , we arrive at (8). This governs the delay necessary at the phase detector on the DRAM side. If this equation is satisfied, then the phase detector will report 0 skew when the clocks to the CMD FIFO are aligned ($Na = Nb$). By substituting (3) and (4) into $Nc = Nd + t_{Setup}$, and solving for τ_{Dcl} , we arrive at (9). If the delay τ_{Dcl} satisfies this equation, we guarantee that the data at the periphery FF is aligned when the WCK and CK clocks are aligned. Equations (7), (8), (9) are summarized below:

$$\begin{aligned} \tau_0 &= \tau_{WCKRcv} - \tau_{CKRcv} + \tau_{Bx} \\ &\quad - \tau_{Ax} + \tau_{BypMux} \end{aligned} \quad (7)$$

$$\tau_{Dwck} = \tau_{Dck} + \tau_{Bx} \quad (8)$$

$$\begin{aligned} \tau_{Dcl} &= \tau_{Ax} - \tau_{Bx} + \tau_{S2P} - \tau_{CMDFF} \\ &\quad + \tau_{Dx} - \tau_{Ex} + t_{Setup}. \end{aligned} \quad (9)$$

In the case of PLL on, the term τ_{BypMux} will be replaced by the following:

τ_{BypMux} (“no” compensated path in PLL FB path);
 $\tau_{BypMux} - \tau_{Ipath}$ (copy of path delay to be compensated in PLL FB path).

Equations (1)–(9) must be re-calculated, and the delays Dwck, Dck, and Dcl adapted accordingly.

According to (7)–(9), the design requirements are clear. Equalization of DRAM internal WCK-CK paths are desired in order to limit the size of delays Dwck, Dck, and Dcl, as well as to limit the offset τ_0 as seen by the controller. A conflicting requirement is to maximize read and write data eye

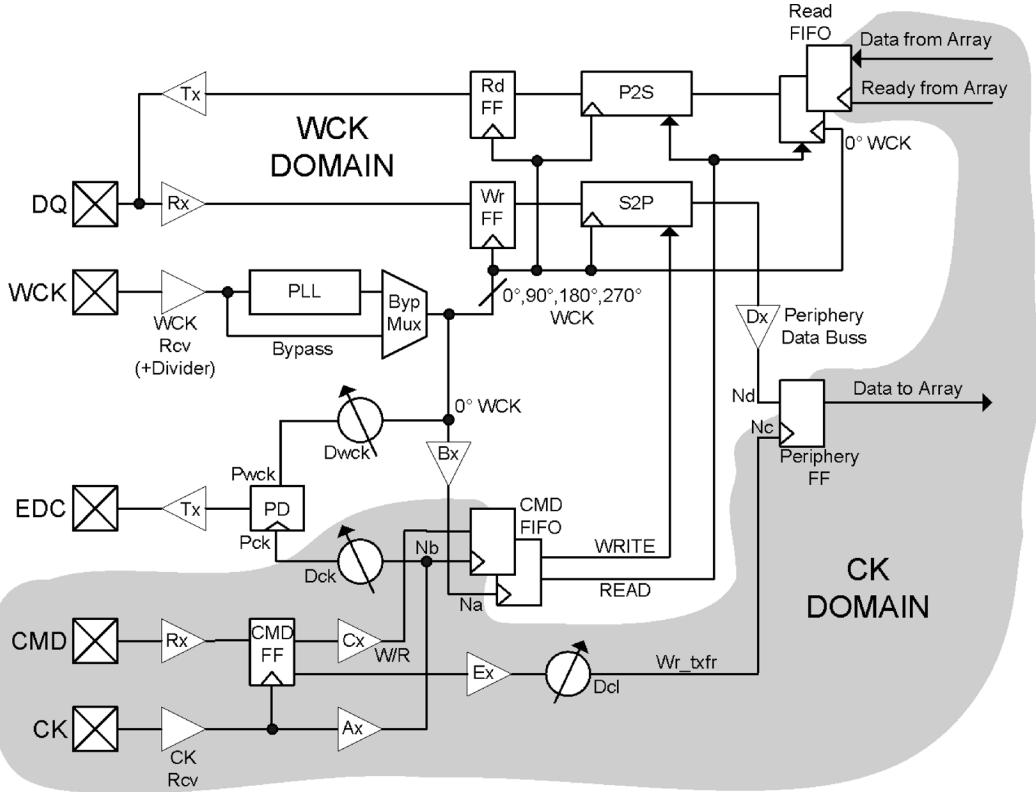


Fig. 2. Clocking and latency control.

by maintaining a short WCK path. Careful balance of these two requirements is necessary.

After the initial system alignment of WCK/CK to 0 phase offset as determined by PD, the clocks at the CMD FIFO are also aligned ($N_b = N_a$). The reset of the CMD FIFO is simplified considerably, avoiding any complicated FIFO reset/latency schemes which were previously necessary [3]. The 4-stage FIFO is reset, so that the output side (WCK domain) counter follows the input side counter (CK domain) by +1 tCK. This allows the FIFO to cover a $+2/-1$ tCK drift between WCK/CK after initial alignment. Write and read latency is calculated in the CK Domain with a simple counter (CMD FF), and the W/R signal is sent to the CMD FIFO and re-synchronized into the WCK domain. The WRITE/READ signals are then sent to the Serial-Parallel (S2P) and Parallel-Serial (P2S) circuitry to control write/read operations. This is necessary, since the GDDR5 WCK is a continuously running clock.

According to the equations above, after initial system alignment $N_c = N_d + t_{\text{Setup}}$ (Periphery FF). Due to this training requirement for GDDR5, and the proper control of delays in the WCK/CK control paths to satisfy the equation above, the transfer of input data from the WCK domain into the CK (DRAM Array) domain, and the subsequent latency control of the read/write command signals can be considerably simplified. The controller (CMD FF) must generate a signal $WR_{\text{txfr}} = \text{WRITE} + 1$ tCK. The data from the S2P will be present at the input of the periphery FF for 2tCK. If $WR_{\text{txfr}} = \text{WRITE} + 1$ tCK, then the WR_txfr will be aligned to the middle of the internal write data eye. This allows a drift

of external WCK/CK alignment by ± 1 tCK, enabling correct write operation without usage of a register based write-data FIFO. This saves both layout space and power (and reduces internal switching power noise).

IV. I/O CIRCUITRY

In order to fulfill the high speed GDDR5 interface requirements, it was necessary to improve on the GDDR3/4 style receiver and transmitter.

A. Transmitter (Tx)

The transmitter (Fig. 3) improves on the traditional time-domain architecture [5] in order to improve the target data-rate, and to enable operation at a down-converted internal Vint. The transmitter is driven with differential CMOS signals (INT_T, INT_C) to ensure skew-free voltage domain transitions. The input data voltage domain transition is performed in two stages. The domain transfer shifts the GND-referenced input into the VSSQ domain. The level shift then shifts the signals P_T/C, N_T/C from the VINT to VDDQ domain.

In order to preserve duty cycle at high frequency, the $V_{\text{INT}} - V_{\text{DDQ}}$ level shifter (Fig. 4) implements a CML transistor load (MP_1, MP_2), in parallel to the conventional cross coupled pair (MP_3, MP_4, MN_2, MN_3). To reduce current consumption at low operating frequencies, the additional CML stage can be disabled (VBP bias control).

A low-overhead programmable transmitter boosting technique was also added, to improve transmitter adaptability to alternate channel topologies. Boosting transistors MP_1, MN_1

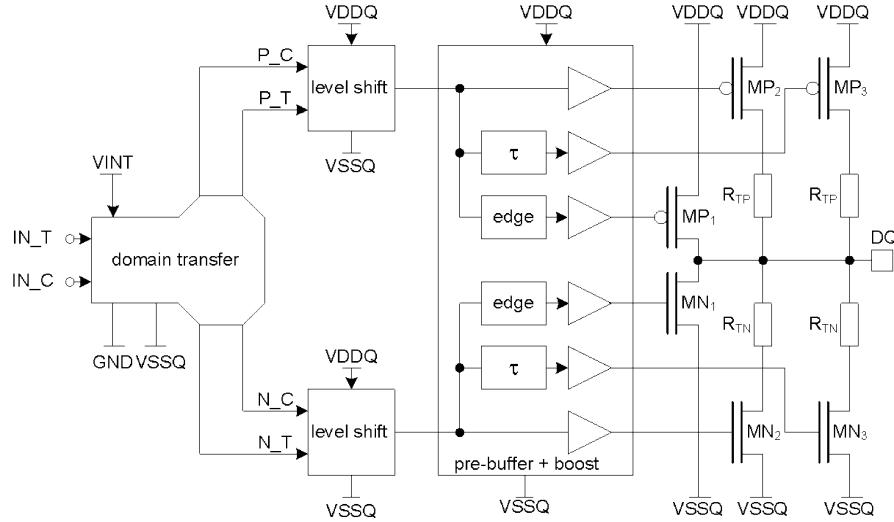


Fig. 3. Transmitter.

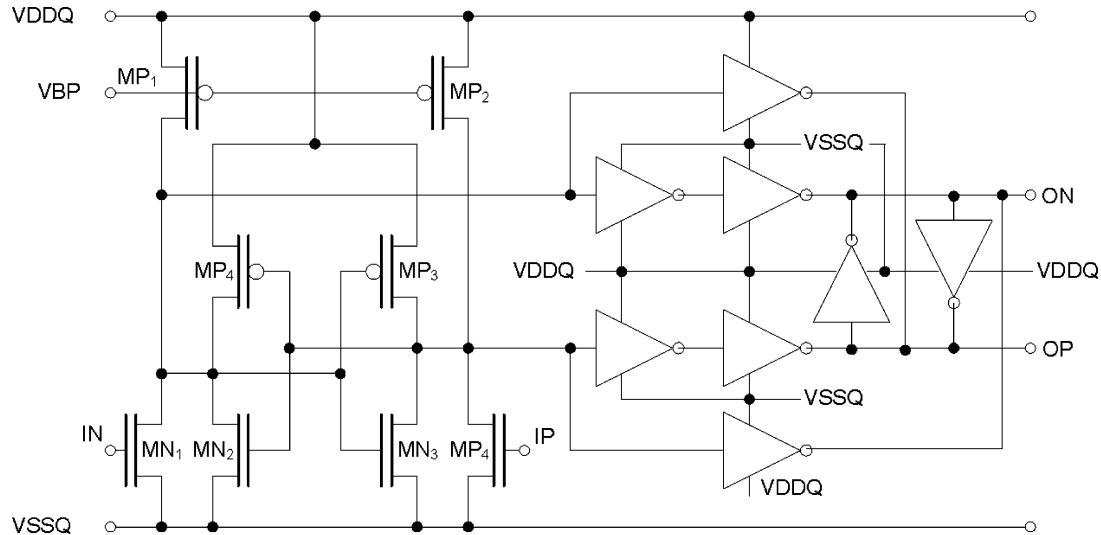


Fig. 4. level shifter.

(Fig. 3) can be activated for a short period following any signal transition. In order to keep the design of the transmitter sufficiently compact, a full implementation of synchronous pre-emphasis was avoided. The boosting transistors add a negligible 50 fF capacitance to the data pins. Simulation results (Fig. 5) show a vertical-data-eye improvement of 100 mV at a data rate of 7 Gb/s/pin for the transmitter eye at the GPU. A realistic system model including coupled PCB transmission lines was used.

B. Receiver (Rx)

GDDR5 specifies single-ended signaling with on-die termination (ODT), and an externally or internally generated receiver reference voltage (VREF). The ODT impedance and VREF level can be independently fine-adjusted through mode register commands for both input clock and command/address signals.

In order to improve noise rejection, VREF is locally decoupled to V_{DDQ} .

In our implementation, the classical DDR/GDDR style receiver with CMOS amplification before latching was abandoned in favor of a 2-stage latch style receiver. A basic block diagram is shown in Fig. 6. The first stage is a CML single-ended to differential converter with resistor loads, and decouples the input load from the following latching receivers. The second stage is comprised of four parallel latching receivers with offset-calibration [5]. The serial to parallel stage after the latching receivers places the data on the internal data bus in 2tCK intervals. The bias for the input differential pair (Fig. 7) is generated by an auxiliary replica circuit with output common mode equal to VREF. A standard differential input receiver suffers from non-zero common mode gain due to the finite impedance at node V_x . Capacitance at node V_x contributes to high frequency output distortion because of this output common mode shift. The

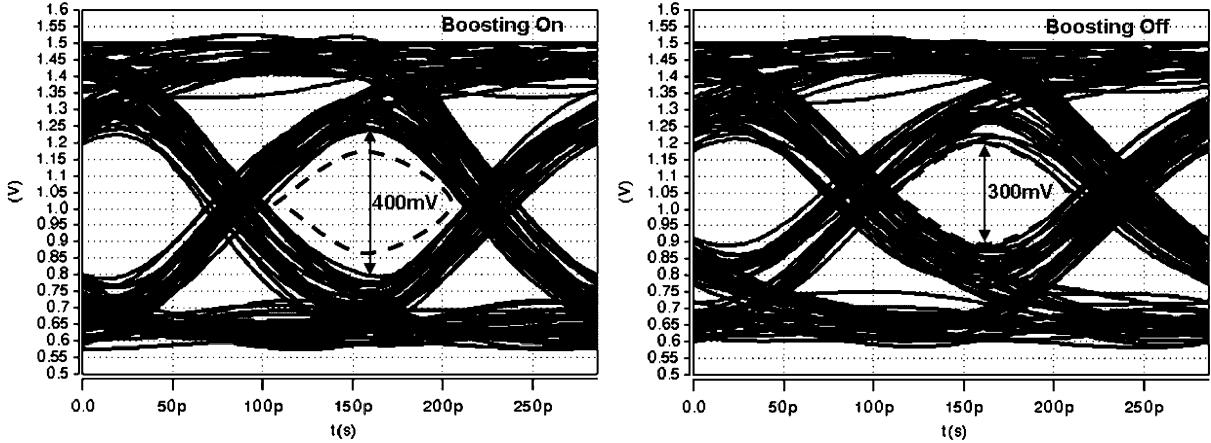


Fig. 5. Simulated transmitter boosting @ 7 Gb/s.

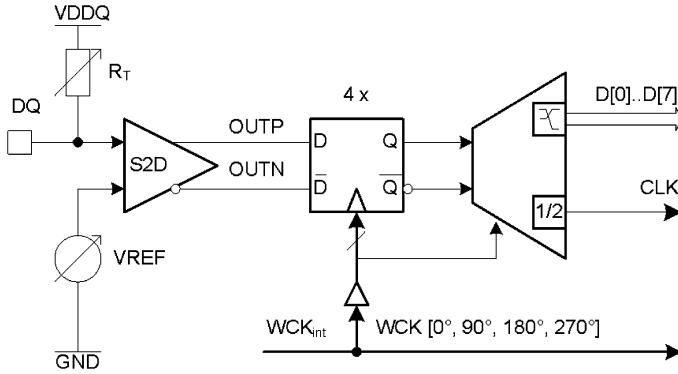


Fig. 6. Receiver block diagram.

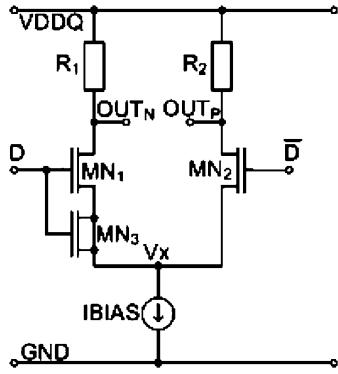


Fig. 7. Single-ended-to-differential converter.

additional capacitor MN3 (Fig. 7) compensates for the high frequency effects of the non-zero common mode gain [9].

V. POWER

A. Noise Control

Previous work has shown the importance of noise control on external I/O supply lines (V_{DDQ}/V_{SSQ}) in order to ensure robust data eyes. In this design, special attention was also paid to internal voltage generation and distribution to enable high frequency chip operation. Individual clocking/data path components were modeled and then numerically simulated (Fig. 8).

Noise of various frequencies and magnitudes was injected into the I/O circuitry (V_{DDQ}), the internal high speed data and clocking paths (VINT_SPEED), and the core and command circuitry (VINT_CORE—not shown). The simulation method is based on a phase vector in the time domain. Narrowband jitter phenomena can be modeled precisely in the time domain based on modulation schemes like frequency or phase modulation [10]. Building blocks within the clock and data path were described with time domain equations. Using this concept, it is also possible to verify system dynamics for nonlinear circuit and system behavior, and data-dependent jitter phenomena [11] can be predicted in more accurate manner. The system model was extended with physical models for building blocks such as the transmitter, receiver, buffers and phase-locked loops. The on-chip clock tree was modeled as an RLC-tree based on derivative equations, and both capacitive and inductive coupling as well as source and sink impedance were taken into account. The clock and data path buffers were modeled with a power supply modulated delay time (n_{sup} -factor).

The architecture for simulation of READ/WRITE noise paths (Fig. 9) shows the READ, the EDC (Error Detection Code), and the WRITE paths and the write clock (WCK) distribution. The pictured buffers and channel (PCB) as well as the phase-locked-loops (PLL) are supplemented with noise sources. Fig. 10 shows a basic buffer model including power supply noise dependency of the buffer delay. The buffer delay sensitivity Δt_{BUF} to supply noise is determined by the dimensionless n_{sup} -factor. For a more realistic model behavior the supply noise transfer function $H_s(\omega)$ can be included for band limiting effects.

This simulation concept [10], combining power and signal integrity analysis, enables the prediction of supply noise effects on overall jitter and timing budgets. The complete GDDR5 I/O section was verified for READ and WRITE operation with PLL on and off. In PLL-on mode no additional clock tree delay compensation was used, and data dependent jitter was taken into account for predefined READ and WRITE patterns. The simulation results (Figs. 11 and 12) show that a 2% VINT (26 mV) 400 MHz power-supply noise in PLL-off operation is critical enough to close the read data eye by 0.5UI. With PLL-on, we would expect

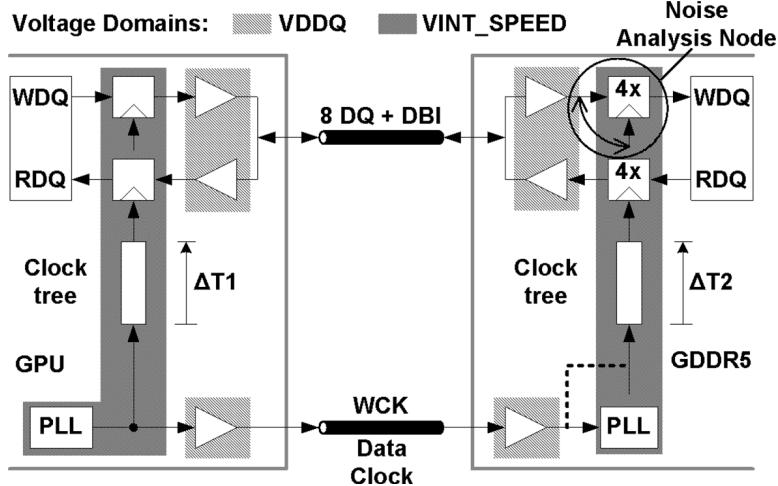


Fig. 8. Model for internal power noise sensitivity simulation.

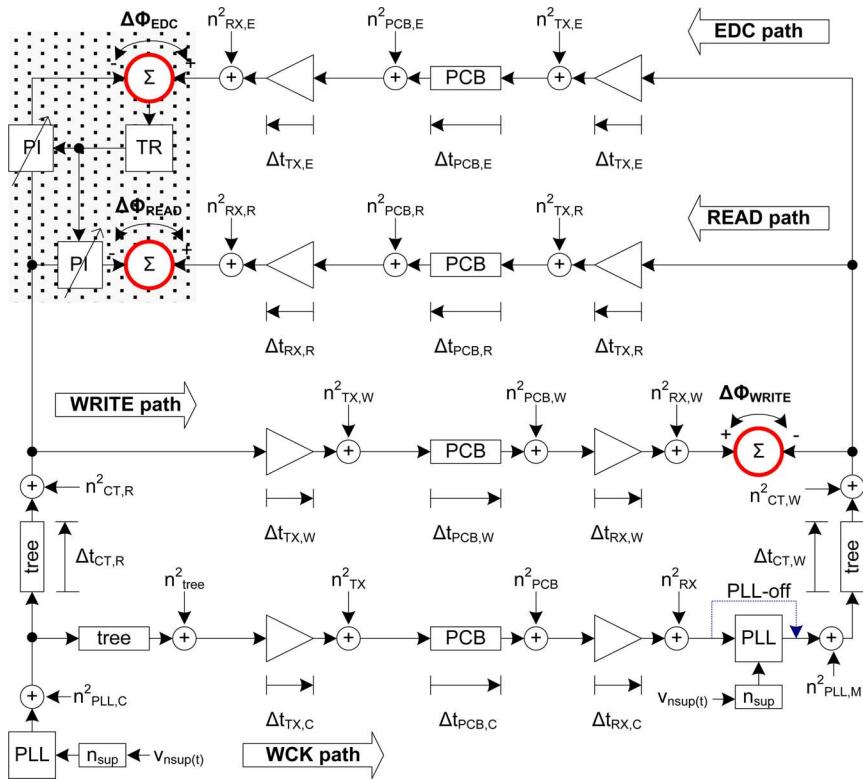


Fig. 9. Architecture for simulation of READ/WRITE noise paths.

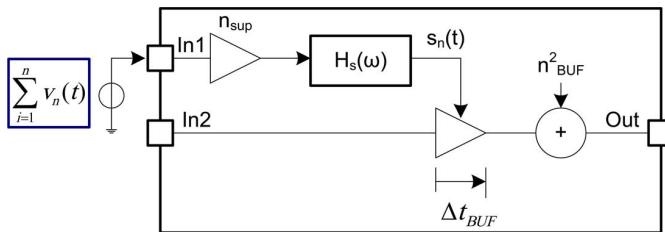


Fig. 10. Buffer delay model.

that uncorrelated high frequency noise components would be attenuated. Simulation results with PLL-on show reduced sensitivity to noise (0.5UI eye closure at 7% VINT, 400 MHz power

noise), as well as attenuation of higher frequency noise components (>1 GHz).

The sensitivity analysis shows the high speed data path and internal clocking system to be particularly sensitive to noise components created by command and core access. Therefore, separate (V_{INT} speed) voltage domains were created to isolate the write/read data circuitry and high-speed internal data clocking system from power noise generated during array access. In addition, a special push-pull V_{INT} generator was developed to improve the asymmetric overshoot/undershoot transients, reaction time and power supply noise rejection parameters of the push-only V_{INT} generator.

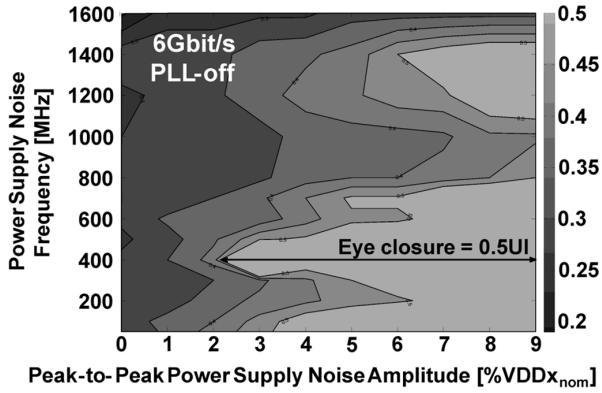


Fig. 11. READ supply noise sensitivity for PLL-off mode.

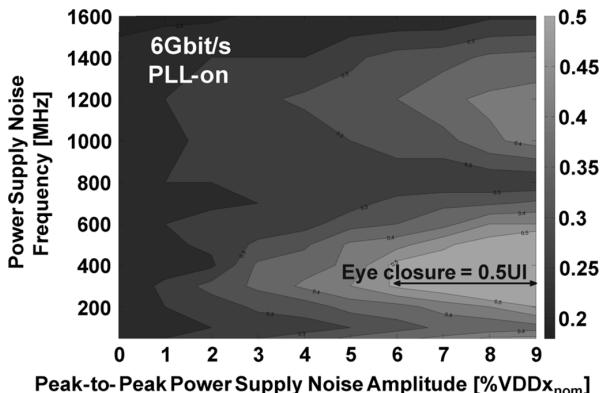
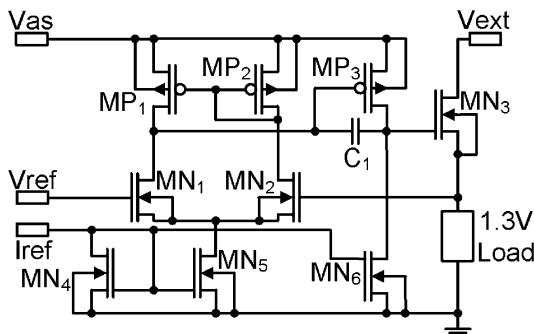


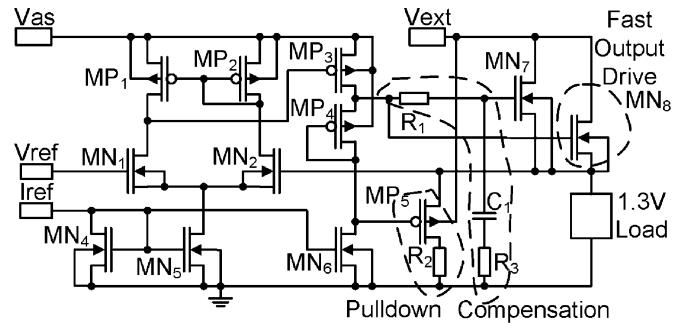
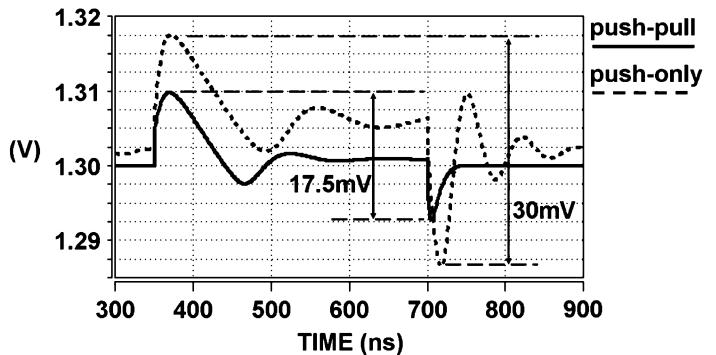
Fig. 12. READ supply noise sensitivity for PLL-on mode.

Fig. 13. Traditional V_{INT} generator.

B. V_{INT} Generation

In the traditional 2-stage op-amp V_{INT} generator (Fig. 13), the Miller compensation (MP₃, C₁) is Iref-bias dependent and offers a feed-forward (op-amp and boosted Vas supply) noise path, thus reducing the power supply rejection. Load response time is also relatively poor due to the simple compensation scheme. The output of the generator is push-only, leading to asymmetric over/under transients. A modified V_{INT} generator (Fig. 14) was used to overcome these problems.

The modified V_{INT} generator uses RC compensation (R₁, R₃, C₁) with no direct feed forward connection, thus reducing the Iref-bias dependency and simultaneously improving

Fig. 14. Modified V_{INT} generator.Fig. 15. Simulation result V_{INT} generator.

the power supply noise rejection (no Miller compensation). The output stage is divided into low frequency (R₁, R₃, C₁, MN₇) and high frequency (MN₈) paths, improving response time without overcompensation. Pulldown (MP₅, R₂) is added to compensate overshoot transients. Careful control of R₁-R₃ and C₁ is necessary to prevent excessive dynamic/static generator power consumption, and to maintain stability over all process corners. The generators were connected directly to the power pads to reduce on-chip parasitic wiring effects to Vext, and only 300 uA of Vas current is consumed per V_{INT} generator. If Vas pump efficiency is taken into account, this represents 800–900 uA of Vext current. This is an acceptable addition to Vext current draw in order to stabilize power delivery in the sensitive Vas, Vpp, and V_{INT} power domains. Simulation results comparing the push-only to push-pull V_{INT} generator is shown in Fig. 15. The simulation result shows over-under shoot of 30 mV is reduced to 17.5 mV, and response-time and DC-offset characteristics are also improved.

In order to further reduce noise generation on the V_{INT} power net, the boosted supply Vas(2.6 V) is delivered by an improved Vpp/Vas pump system. The basic topology of the two stage Vpp/Vas pump system is shown in Fig. 16. In the traditional “synchronously switching pump” system, the pump clocks are switched on or off dependent upon sensed internal power levels. The improved Vpp/Vas pump system uses a frequency controlled pump. The internal VCO runs constantly, with a frequency proportional to the error to Vref. The frequency of the VCO reacts within 20–50 ns to a step load change for Vas and Vpp. This allows a smooth control of Vpp/Vas, reducing output

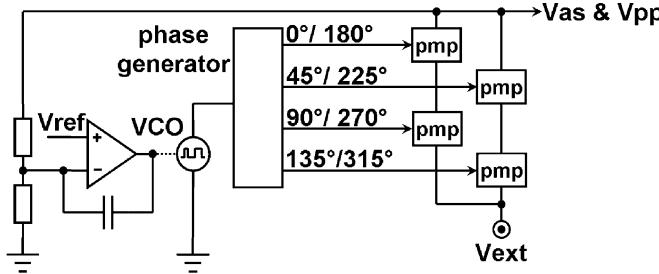


Fig. 16. Basic topology: two stage Vpp/Vas pumps.

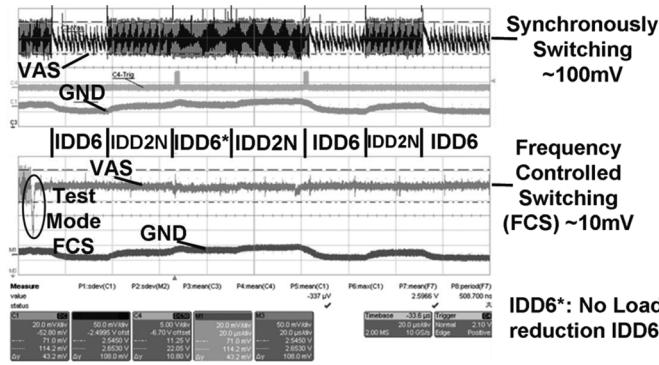


Fig. 17. Synchronously versus frequency controlled switching.

ripple by factor 10× compared to a synchronously switching pump methodology.

VI. MEASUREMENT RESULTS

Wafer measurements (Fig. 17) of the internal Vas pumped voltage were made to compare the Synchronously Switching Pump to the Frequency Controlled Switching (FCS) methodology. IDD6 (Self Refresh) and IDD2 (Active Power Down) states were chosen for the comparison, as opposed to IDD4/7 (active writes/reads) in order to prevent I/O effects from disturbing the results. IDD6* is a “No Load reduction” Self Refresh, where the generators are not switched off. The resultant Vas disturbance for IDD6* should match that of IDD2N, and this is verified in the measurements. In Fig. 17, it can be seen that Synchronously Switching of the pumps causes 100 mV disturbances on Vas, where Frequency controlled switching has only 10 mV Vas noise. Since this voltage is used on all first stages of the V_{INT} generators, any noise reduction here improves the V_{INT} generator performance.

Internal VINT speed–GND measurements were made in order to quantify the effects of switching noise on the internal VINT speed power nets. Local VINT speed was measured relative to local GND at the internal clock buffers. As shown in Fig. 18, the effect of array activates (act) or precharges (pre) on VINT speed are negligible, supporting the concept of separated internal VINT power domains. Continuous writes and reads have a minor 10 mV effect on the VINT speed power domains. The effect of this power disturbance on the internal data-path clock is ± 35 ps period track for writes and ± 30 ps for reads. This is the equivalent of ± 8.75 ps per data eye for writes and ± 7.5 ps per data eye for reads. The measurements were performed on wafer (active probes on internal test points) at a

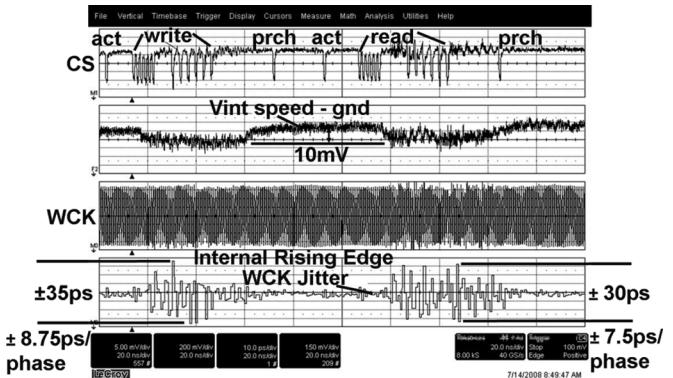
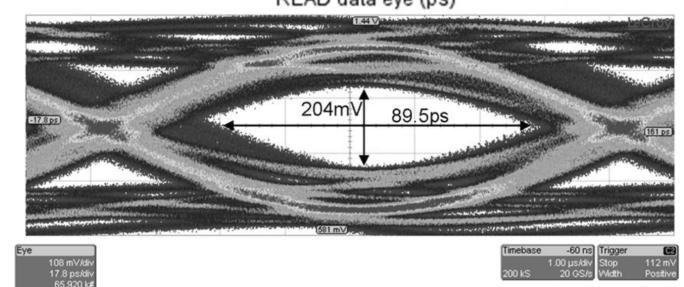
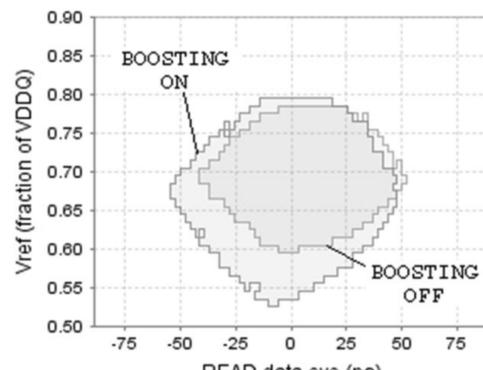
Fig. 18. Internal V_{INT} –GND chip measurement.

Fig. 19. Read data eye with/without boosting, and oscilloscope eye.

data rate of 5 Gb/s/pin (data rate limit due to cantilever probe card setup on wafer). The test pattern maximized toggling on both the DQ pads and internal data bus (DBI off, Address bus inversion off).

Fig. 19 shows measurement results of the read data eye with and without boosting, and an oscilloscope measurement of read data eye at a 7 Gb/s/pin data rate. The measurement was performed on a tester board with 40 Ohm line trace impedance, and 60 Ohm termination at the receiver side. The boosted eye is 80 mV larger than the unboosted eye, but offset (GDDR5 system is V_{DDQ} terminated). In order to achieve full benefit of the boosting, an adjustable V_{REF} system is built into the GDDR5 device. The oscilloscope read data eye is a composite eye over all bits in a burst.

Fig. 20 shows a composite (all DQs toggling) read data eye shmoos. The pattern used in the shmoos and oscilloscope measurement consists of a DC DBI compliant pattern with approximately 13×10^6 write and read memory array accesses. The pattern was constructed in order to maximize ISI (inter-symbol

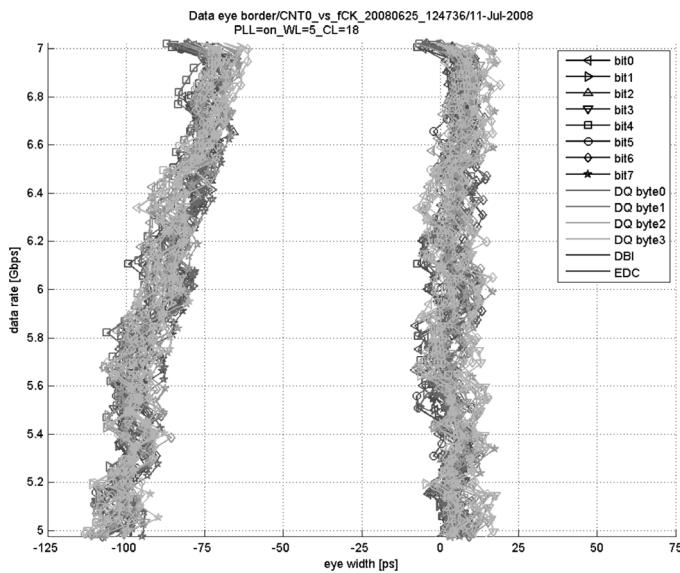


Fig. 20. Read data eye plot (all DQs toggling).

interference) and crosstalk stress on a per-pin basis. A data rate of 7 Gb/s/pin can be seen in the plots. Write data eye shmoos plots are comparable with an opening of 50–55 ps at 7 Gb/s/pin. The number of writes and reads used in generation of this plot corresponds to a BER (bit error rate) of approximately 10^{-8} .

Fig. 21 shows a composite (all DQs toggling) read BER plot (4.8–7 Gb/s/pin). The pattern used in the plot is a DC DBI compliant pattern and was constructed in order to maximize ISI (inter-symbol-interference) and crosstalk stress on a per-pin basis. A 60 ps read window at 10^{-8} BER can be seen in Fig. 21, and corresponds well to the value shown in Fig. 20. The Write BER plot (4.8–7 Gb/s/pin) is shown in Fig. 22 for comparison. A 44 ps read and 30 ps write eye is present at BER of 10^{-12} . All composite shmoos (Figs. 20–22) were made without the benefit of per pin training. It is expected that the composite eyes would enlarge by the amount represented by the static skew between the individual DQs.

Fig. 23 shows measured PLL attenuation of high frequency noise components in the V_{INT} speed domain. The PLL closed loop bandwidth was set to $f - 3 \text{ dB} = 20 \text{ MHz}$, and a 200 mV peak-to-peak sinusoidal supply disturbance was applied to the V_{INT} speed domain at the PLL. The ref sig is measured directly at the output of the PLL, and the mxd sig is measured after the first clock mixing stage following the PLL. It can be seen that high frequency noise components in the V_{INT} speed domain above the bandwidth of the PLL are attenuated.

The chip microphotograph can be seen in Fig. 24. The location of the MCSL (Master Column Select Line) decode and data redrivers (RDRV), as well as the LCSL (Local Column Select Line) decode and secondary sense amps (SSA) is shown. The location of the V_{INT} speed domains and PLLs is highlighted. All circuitry outside of the V_{INT} speed domains is supplied by the V_{INT} core domain. A summary of device features is given in Table I.

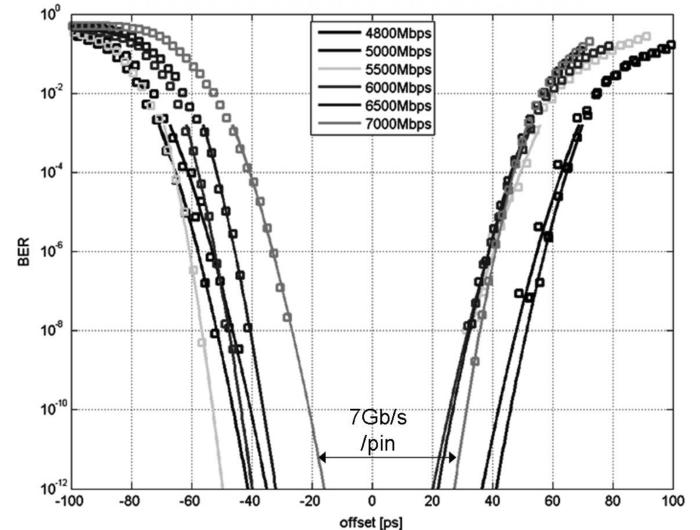


Fig. 21. Read BER plot (all DQs toggling).

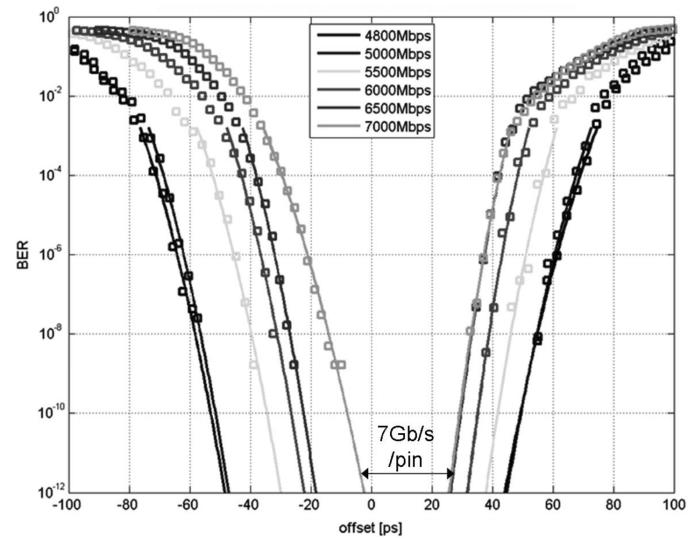


Fig. 22. Write BER plot (all DQs toggling).

TABLE I
DEVICE FEATURES

Process	75nm CMOS 3 Metal
Package	170 ball BOC
Chip size	130mm ²
Data rate	7Gb/s/pin (CL 18)
Bank	16(1G)
Organization	32M x 32
Burst Length	8
I/O	32DQ + 4EDC + 4DBI
Power Supply	1.5V +/- 3%

VII. SUMMARY AND CONCLUSION

A 1.5 V 1 Gb GDDR5 compliant memory has been fabricated in 75 nm 3 metal layer DRAM process. Special emphasis

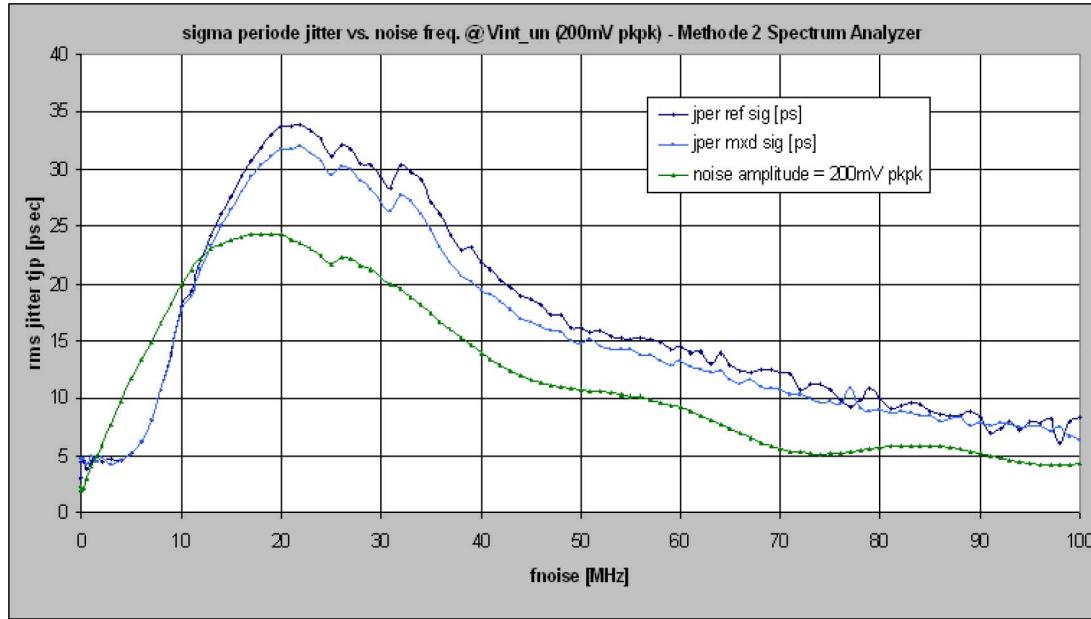


Fig. 23. PLL attenuation of high frequency power supply noise.

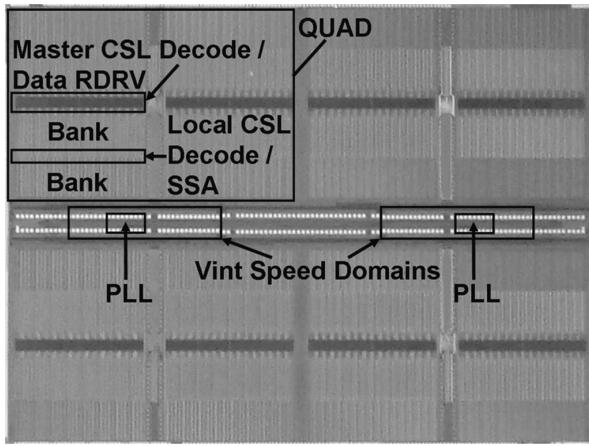


Fig. 24. Chip microphotograph.

was placed on transceiver and power system design and layout throughout the entire design phase. High effort was placed into chip level power noise control. During the system analysis phase, it was also decided to abandon the typical amplifier-CMOS delay-latch type DQ receiver in favor of a latching receiver with no wck path matching delay. WCK-DQ offset was increased in order to favor minimal noise sources in the DQ path. With these circuit design techniques, we demonstrate a 7 Gb/s/pin GDDR5 DRAM device.

REFERENCES

- [1] J.-D. Ihm *et al.*, "An 80 nm 4 Gb/s/pin 32 b 512 Mb GDDR4 graphics DRAM with low-power and low-noise data-bus inversion," in *IEEE Int. Solid-State Circuits Conf. (ISSCC 2008) Dig. Tech. Papers*.
- [2] K.-W. Lee *et al.*, "A 1.5-V 3.2 Gb/s/pin graphic DDR4 SDRAM with dual-clock system, four-phase input strobing, and low-jitter fully analog DLL," in *IEEE Int. Solid-State Circuits Conf. (ISSCC 2007) Dig. Tech. Papers*.
- [3] B. Johnson, B. Keeth, F. Lin, and H. Zheng, "Phase-tolerant latency control for a combination 512 Mb 2.0 Gb/s/pin GDDR3 and 2.5 Gb/s/pin GDDR4 SDRAM," in *IEEE ISSCC 2007*, Session 27.5.

- [4] M. Brox, H. Fibranz, M. Kuzmenka, F. Lu, S. Mann, M. Markert, U. Muller, M. Plan, K. Schiller, P. Schmolz, P. Schrogmeier, A. Tuber, B. Weber, P. Mayer, W. Spirk, H. Steffens, and J. Weller, "A 2 Gb/s/pin 512 Mb graphics DRAM with noise reduction techniques," in *IEEE Int. Solid-State Circuits Conf. (ISSCC 2006) Dig. Tech. Papers*, pp. 537–546.
- [5] Z. Gu, P. Gregorius, D. Kehrer, L. Neumann, T. Rickes, H. Ruckerbauer, R. Schledz, and M. Streibl, "Cascading techniques for a high-speed memory interface," in *IEEE Int. Solid-State Circuits Conf. (ISSCC 2007) Dig. Tech. Papers*, Session 12.7.
- [6] GDDR5. [Online]. Available: http://www.qimonda-news.com/download/Qimonda_GDDR5_whitepaper.pdf
- [7] J. Y. Sim *et al.*, "Multilevel differential encoding with precentering for high-speed parallel link transceiver," *IEEE J. Solid-State Circuits*, vol. 40, no. 8, pp. 1688–1694, Aug. 2005.
- [8] S. Sidiropoulos and M. Horowitz, "A 700-Mb/s/pin CMOS signaling interface using current integrating receivers," *IEEE J. Solid State Circuits*, vol. 32, no. 5, pp. 681–690, May 1997.
- [9] H. Partovi *et al.*, "Single-ended transceiver design techniques for 5.33 Gb/s graphics applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC 2009) Dig. Tech. Papers*.
- [10] M. Li, A. Martwick, G. Talbot, and J. Wilstrup, "Transfer functions for the reference clock jitter in a serial link: Theory and applications in PCI express," in *DesignCon 2005*.
- [11] B. Analui, J. F. Buckwalter, and A. Hajimiri, "Data-dependent jitter in serial communications," *IEEE Design and Test of Computers*, 2004.



Rex Kho received the B.S. degree in electrical engineering from Michigan State University in 1989, and completed an internship at IBM where he was responsible for circuit simulation and layout for a 16 Mb DRAM. In 1991, he finished one year of a Masters program and then joined EDS (Electronic Data Systems) where he developed database, object oriented, and microcontroller programs for the automotive industry in assembly, C, Gupta SQL and X-Windows. In 1994, he joined IBM and was involved in all aspects of Embedded DRAM development, from circuit design and layout, to production test and characterization. He completed his M.S.E.E. from National Technical University in 1997 during his employment at IBM.

In 2000, he joined Infineon/Qimonda, where he designed and developed RLDRAMI, RLDRAMII, DDR2, DDR3, and GDDR5 products. He was design team leader for the first GDDR5 product at Qimonda. In 2009, he joined Bosch Sensortec, where he currently develops sensor ASICs for consumer applications. He holds over 30 U.S. and foreign patents, and has published four papers and conference contributions.



David Bourdin was born in Dijon, France, in 1981. He received the M.S. degree in electrical engineering from CPE-Lyon University, Lyon, France, in 2005.

He joined the GDDR5 (graphics double data rate 5) design team at Infineon Technologies AG/Qimonda AG, Munich, in January 2006, where he was responsible for system and block level verilog modelling during the concept and design phase of the GDDR5 project. In April 2008, he took on a field application engineering assignment at Qimonda North America, and participated in the successful bring-up of the first commercial application using GDDR5 DRAM.



Martin Brox received the Dipl. degree in physics and the Dr. degree in applied physics from the University of Münster, Germany, in 1988 and 1992, respectively.

In 1992, he joined the IBM/Siemens/Toshiba development project in Essex Junction, VT. He initially designed mini-DRAMs used for line-monitoring and switched afterwards to a DRAM design department where he developed the memory core for a 64 M-0.35 μm DRAM. In 1997, he moved to Siemens, Germany, and worked as a design team leader for multiple high-speed or high-density DRAM devices: notably, the first Rambus-DRAMs and the 512 Mb-GDDR3 which was designed into all current generation game consoles (Xbox 360, PS3, Wii). Currently, he is technically responsible for the design development of graphics DRAMs at Elpida. His special expertise is in memory architectures and in the design of high-speed circuits (e.g., DLL, receiver/transmitter). He holds more than 30 U.S. patents and has published about 30 papers and conference contributions.

Dr. Brox has been a member of the technical program committees of ISSCC and ESSCIRC. He was awarded a scholarship by the Studienstiftung des Deutschen Volkes, and received the ITG Award of the VDE in 1992.



Peter Gregorius (A'99–M'07) was born in Kirn, Germany, in 1969. He received the M.S.E.E. degree in electrical engineering and the Ph.D. degree in electronics from the University of Bremen, Germany, in 2004 and 2007, respectively.

From 1990 to 1995, he was with IC-Haus, an ASIC company in Frankfurt, Germany. In 1995, he joined Micronas Semiconductors, Munich, Germany, where he focused on BiCMOS design. In 1996, he joined Melexis Micro-electronic Systems (formerly Thesys GmbH), Erfurt, Germany. As head of the BiCMOS

Design Department, he was in charge of concept and design of high-frequency products. In 2000, he joined Infineon Technologies AG, Munich. As Principle Engineer within the Memory Product Division, he is responsible for concept and design of high-speed serial links in the area of personal computer applications. He holds several patents in the area of high-speed communications.



Heinz Hoenigschmid (M'09) received the B.S. and M.S. degrees in electrical engineering from the Technical University Munich, Germany.

During 18 years in the semiconductor industry, he was engaged in circuit design and management for DRAMs and emerging memories. He is currently Executive Manager at Elpida Memory, responsible for the design of graphic DRAMs. He holds 72 U.S. patents, and has authored 15 publications.

Dr. Hoenigschmid is a member of the IEEE ISSCC technical program committee.



Bianka Kho received the diploma degree in electrical engineering from Kiel Polytechnic in 1995, with focus on high frequency telecommunication electronics.

She joined Siemens/Infineon Semiconductor in 1996, where she was responsible for development of VHDL behavioral models, full chip simulation and full custom circuit development for embedded DRAM macros. In 1999, she joined a cooperative embedded DRAM development team of Infineon/IBM in Burlington, Vermont, and was responsible for the design transfer and customer and test support. After a parental leave pause in 2000, she returned to Infineon/Qimonda in 2006, and assumed the position of modelling engineer. Her duties included the conception and implementation of parameterizable verilog simulation models for GDDR3 and GDDR5 (Graphics DDR5) memory devices, debug of simulation models, full chip simulation, circuit analysis, PERL scripting for automated verification routines, and creation of verilog replacement models for mixed level simulation.



Sabine Kieser was born in Munich, Germany, in 1971. She received the diploma degree in physics from the Technical University of Munich, Munich, Germany, in 1996.

She joined the Siemens Semiconductor Memory Department in 1996, and was engaged in various DRAM design activities, with special focus on technology related rowpath and sense amplifier design topics. As a member of the Graphics DRAM design group at Siemens, Infineon and Qimonda, she worked on simulation and characterization of new

sense amplifier circuit concepts, and row and array design and verification for specialized high speed Graphic DRAM requirements.



Daniel Kehrer was born in Austria in 1976. He received the diploma (M.S.) degree in communications engineering in 2001 and the Ph.D. degree in electrical engineering from Vienna University of Technology, Austria, in 2003.

He joined the high-frequency circuit corporate research division of Infineon Technologies AG, Munich, Germany, in 2003. From 2003 to 2005, he was engaged in the design of high-speed CMOS electronics for wireline communications, with focus on high speed trancieving at low power consumption.

In 2005 he joined the Memory Products division where he was involved in research activities for post DDR3 products. From 2006 to 2009, he worked with the graphics DRAM development team on GDDR5 memory devices, and managed graphics DRAM design projects. In 2009 he joined Infineon Technologies, where he currently works on RF-circuitry for GPS and UMTS applications.



Maksim Kuzmenka was born in Minsk, Belarus, in 1972. He received the M.S. degree in electrical engineering from the Belarusian State University of Informatics and Radio-electronics in 1993.

In 1993 he joined BELOMO (Belarusian Optical & Mechanical Association) where he performed switched mode power supply design. In 2001 he joined Infineon Technologies AG in Munich, where he was responsible for DDR2 and DDR3 memory system signal integrity simulations and system definition. In 2006 he joined Qimonda AG in Munich, where he has been involved in chip design, specializing in: design of I/O circuitry for bitrates up to 7 Gbps, ultra-short response LDOs, and clock distribution circuitry. In 2009, he joined Elpida Memory (Europe) in Munich. His current interests include analog and mixed signal chip design, high speed interfaces and power electronics.



Udo Moeller was born in Duisburg, Germany, in 1960. He received the Diplom-Ingenieur der Elektrotechnik (M.S.E.E.) from Duisburg University in 1985.

He joined Siemens AG in 1988, and developed microcontrollers for consumer entertainment electronics until 1995. In 1996 he joined the embedded DRAM department, and technically managed and directed several embedded DRAM projects. In 2001 he moved to the graphics DRAM design department, where he was involved in the development of GDDR3 and GDDR5 graphics DRAMs. He is currently working for a small telecommunications company.



Ian Russell received the B.Sc. degree in natural philosophy (physics) and the M.Sc. degree in instrument design from Aberdeen University, Scotland, U.K., in 1974 and 1976, respectively.

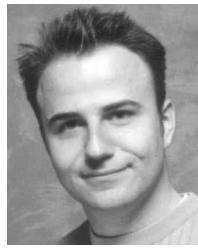
In 1975 he joined Sperry Gyroscope, Bracknell, U.K., where he was involved in developing various sensor systems. In 1980 he moved to National Semiconductor, Fürstenfeldbruck, Germany, to work in product application and development. From 1983 to 1987 he designed digital CMOS ASICs for telecommunication applications at AT&T Micro-

electronics, Munich, Germany. Between 1987 and 2005 he designed analog and mixed-signal system-on-chip ASICs and full-custom IP blocks, for various market applications at Eurosil Electronic, Motorola Semiconductor Products, and Philips Semiconductors in Munich. His special interests included ADCs and DACs, PLLs and DLLs, and low-power crystal oscillators. He joined Infineon Technologies Memory Products, later Qimonda, Munich, in 2005 and was responsible for the design of high-speed clocking systems for GDDR5 graphics DRAM.



Pavel Veselinov Petkov was born in V. Tarnovo, Bulgaria, in 1974. He received the B.S. and M.S. degrees in physics from Sofia University "St. St. Kliment Ochridski", Bulgaria, in 1997 and 1998, respectively.

From 1998 to 2002 he worked for Melexis, Bulgaria. From 2002 to 2006 he was employed by the high-speed video interface department of Philips, Starnberg, Germany. In 2006, he joined the graphics department of Qimonda, Neubiberg, Germany, and was involved in design of high speed graphics memory products (GDDR5) with special focus on offchip drivers and receivers.



Kai Schiller was born in Cottbus, Germany, in 1977. He received the Diplomingenieur der Elektrotechnik degree (M.S.E.E.) in microelectronic engineering from the University Duisburg-Essen, Germany, in 2003.

He joined Infineon Technologies/Qimonda AG in 2004, where he was involved with GDDR2, GDDR3, DDR3 and GDDR5 DRAM designs. His main areas of responsibility were analog circuit design for internal power, reference systems, and high speed I/O trancievers.



Manfred Plan was born in Augsburg, Germany, in 1960. He received the diploma in electrical engineering (M.S.E.E.) from the Technical University Munich, Germany, in 1987, with focus on semicustom electronics.

After receiving his diploma, he first performed integrated biosensor research, and then joined a two-year trainee program (SGP) at Siemens AG in 1988. He developed MOS circuitry in several different product development divisions, in close cooperation with design departments in Villach/Austria and the University of California, Berkeley. From 1990 until 1998, he was involved in chip and memory design for the Consumer (TV/Video/Image) Electronics Development Division of Siemens AG. In 1998, he joined the Chipcard Development Division at Siemens/Infineon AG and focused on microcontroller design. In 2000, he moved to the Memory Product Development Division of Infineon/Qimonda AG where he developed several dram designs for embedded, server and high speed graphic applications (RLDRAM, GDDR3, GDDR5). In 2009 he joined the Elpida Memory Design Center in Munich, where he is currently working on new GDDR5 memory designs.



Ronny Schneider was born in Soemmerda, Germany, in 1976. He received the M.S. degree in physics from Friedrich-Schiller-University, Jena, Germany, in 2001.

He joined Infineon Technologies Corp., Burlington, VT, in 2001 where he was involved in circuit design and failure analysis for DDR and DDR2 SDRAMs. From 2004 to 2006 he was the team leader for a DDR2 SDRAM design in Xi'an, China. Since 2006, he was responsible for the full-chip verification of the GDDR5 SDRAM products for Qimonda AG, Munich, Germany. In 2009 he joined Elpida Memory (Europe) GmbH, Munich, Germany.



Michael Richter was born in Frankfurt, Germany. He received the M.S. degree in electrical engineering from the Technical University Munich, Germany.

He joined Siemens Semiconductors in 1984 where his work included ASIC design and design support for ASIC customers. Later, he joined Infineon Technologies AG where he was responsible for the product definition of smart card ICs and served as a program manager in the development of a high-speed crypto IC. With Qimonda AG he was responsible for product definition and standardization of high-speed Graphics DRAMs, which he now continues with Elpida Memory in Munich, Germany.



Kartik Swaminathan was born in New Delhi, India, in 1979. He received the B.S. degree in electrical engineering from Arizona State University, Tempe, AZ, in 2001, and the M.S. degree in electrical engineering from the University of Mysore, India, in 2004.

He joined Infineon Technologies AG, Munich, Germany, in 2004. He was a member of the design for test (DfT) group from 2004–2005 and worked on memory array test features. After Qimonda AG was spun off from Infineon Technologies in 2006, he joined the graphics memory design group and worked on GDDR5 and GDDR3 datapaths. In 2009 he joined Elpida Memory (Europe) GmbH, Munich, Germany.



Bradley Weber was born in Kitchener, Canada, in 1976. He received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, Ontario, Canada, in 2001.

He joined Infineon Technologies AG in 2001 and later Qimonda AG, Munich, Germany, in 2006. Since then, he has been involved with Graphic DRAM design, focusing on high-speed data paths, clocking and generator design.



Mario Gjukic was born in Munich, Germany, in 1972. He received the diploma in physics and the Ph.D. degree from the Technische Universität München, Munich, Germany, in 2002 and 2007, respectively.

From 2002 to 2006 he worked at the Walter Schottky Institut, Munich, Germany, in the field of basic semiconductor research, with special focus on magnetic semiconductors and crystalline SiGe thin films. After joining Qimonda AG, Munich, Germany, in 2006, he has been involved in the high speed testing of DRAMs.



Julien Weber was born in Colmar, France, in 1980. He received the M.Sc. degree in electrical engineering from the Cardiff School of Engineering, Cardiff, U.K., and the Ingénieur des Grandes Ecoles degree in Génie Electrique from the Institut National des Sciences Appliquées de Strasbourg, Strasbourg, France.

He joined Infineon Technologies AG, Munich, Germany in 2004 as test engineer in the Memory Product Department. He was responsible for the worldwide support of fabrication sites and the test program development. After the spin off of the Memory Product Department into Qimonda AG in 2006, he joined the Graphics Design group as a circuit designer and Design for Testability Engineer. In this position, he was mainly responsible for the definition, implementation and documentation of the design for test features of GDDR5 products. In 2009, he joined EADS Astrium GmbH, where he is responsible for parts project engineering.



Wolfgang Spirkl received the M.Sc. and Ph.D. degrees in physics from the Ludwig-Maximilians-University, Munich, Germany, for research in solar energy physics.

In 1992 he spent a research year on X-ray topography of semiconductor dislocations. After 1993, he completed his habilitation (post lecture qualification) at LMU Munich. He worked on non-imaging optics for solar energy collection, and together with Prof. Feldmann's team, performed research on ultra short laser pulse experiments. In 1998 he joined Qimonda AG (formerly Siemens/Infineon AG) where he lead the electrical memory device verification and design analysis group, with focus on the development of new verification algorithms based on Monte Carlo methods as well as on high speed data transmission. He is currently with Elpida Memory (Europe) GmbH, and involved in the development of graphic DRAMs.



Ingo Bormann was born in Braunschweig, Germany, in 1970. He received the diploma in physics from the University of Göttingen, Germany, in 1997.

From 1998 to 1999 he worked as a consultant for Gauss Software AG. Starting in 1999, he worked as a researcher with focus on semiconductor optoelectronics at the Technical University of Munich and received the Ph.D. in physics. In 2004 he joined Infineon Technologies AG, and was responsible for DRAM core and interface component test topics.

From 2006–2009, he was involved with the development of GDDR5 interface test at Qimonda. In 2009, he switched to Elpida Memory (Europe) GmbH, and is currently involved in DRAM design analysis.



Holger Steffens was born in Nordrhein-Westfalen, Germany, in 1976. He received the electrical engineering degree from the University of Applied Sciences of Aachen, Germany, in 2003.

He then joined Infineon Technologies AG (currently Qimonda AG), in Munich, Germany, where he was involved in design analysis, and application engineering for GDDR2, GDDR3 and GDDR5 memories, as well as technical lead for system engineering knowledge development. His interests include verification methodology, and signal integrity testing for high-speed digital/analog interfaces. He is currently Project Manager for Ident Technology AG.



Fabien Funrock was born in Strasbourg, France, in 1974. He received the M.S. degree in electrical engineering from the Physics School of Grenoble, France (École Nationale Supérieure de Physique de Grenoble), and the M.S. degree in physics from the University of Karlsruhe, Germany, in 1998.

He joined the product engineering department of the memory division of Infineon Technologies (now Qimonda) in 2000. He is involved in test, analysis and optimization of Commodity and Graphic DRAMs (SDR, DDR, DDR2, GDDR3 and GDDR5).

In 2009 he joined Elpida Memory (Europe) GmbH, Munich, Germany.



Jörg Weller (M'07) was born in Munich, Germany, in 1968. He received the diploma in electrical engineering (M.S.E.E.) from the Technical University of Munich in 1994.

He joined Siemens AG in 1994, where he was responsible for design and layout of a 4 M-bit DRAM, continuing with production wafer testing for RDRAMs in 1998. In 2001, he started design analysis for graphics DRAMs with Infineon AG. Following the spin-off of Qimonda AG in 2006, he continued design analysis work on GDDR3 and GDDR5 interfaces. In 2009, he joined Elpida Memory Europe and is involved in graphic DRAM development.



Thomas Hein was born in Leisnig, Germany, in 1968. He received the Dipl.Ing. degree in information technology from the Technical University Dresden, Germany, in 1995.

In 1995, he joined the Semiconductor Division of Siemens AG in Munich, Germany, where he was involved in circuit design and product development of Multi-Bank MDRAM for graphics applications. Since then, he has led the design of several graphic memory projects such as SGRAM, GDDR1, GDDR3, GDDR4 and GDDR5. He is currently

member of the JEDEC GDDR5 Task Group defining GDDR5. His interests in DRAM design include new DRAM architectures, next-generation DRAM, chip packaging and high-speed interfaces. In 2009, he joined Elpida Memory (Europe) GmbH, Munich, Germany.