The Effects of Preheating in Millisecond Anneals on Dopant Activation in Silicon

Paul Timans
Mattson Thermal Products GmbH
Dornstadt, Germany
paul.timans@mattson.com

Christian Pfahler
Mattson Thermal Products GmbH
Dornstadt, Germany
christian.pfahler@mattson.com

Markus Hagedorn
Mattson Thermal Products GmbH
Dornstadt, Germany
markus.hagedorn@mattson.com

Alexandr Cosceev
Mattson Thermal Products GmbH
Dornstadt, Germany
alexandr.cosceev@mattson.com

Manuela Zwissler
Mattson Thermal Products GmbH
Dornstadt, Germany
manuela.zwissler@mattson.com

***Abstract*—MOS transistor scaling faces major challenges from the rapid rise in the parasitic resistance as the source/drain (s/d) contact area is reduced. Millisecond annealing (MSA) can provide efficient dopant activation at s/d regions and reduce contact resistivity, but process temperatures are increasingly limited by device integration requirements. Optimization of annealing conditions requires understanding of how different stages of the heating profile affect the process outcome. This paper uses the effective process time concept, where experimentally measured temperature-time data are weighted according to Arrhenius kinetics, to explore the effects of multi-stage MSA recipes on dopant activation. The approach was applied to self-amorphized Si wafers implanted with high doses of P or As at low energies. For heavy As doping, slow cooling cycles strongly affected the electrical activation, whereas for P the hottest stage of the anneal was most significant. The effect of the preheating stage was isolated by using heating cycles where the preheat conditions varied while the slow cooling stages were very similar. Significant differences in dopant activation were found for both MSA processes with different preheat stages and also for low temperature spike anneals with different ramp up characteristics. The results suggest that the final dopant activation depends on more than just peak temperatures and cooling cycles, and that optimization of preheating provides new opportunities for improved activation.**

Keywords— doping, annealing, millisecond annealing, spike anneal, activation, deactivation, solid phase epitaxy, phosphorus, arsenic, n-type doping, kinetics

I. INTRODUCTION

Scaling of MOS transistors is increasingly challenged by parasitic resistances in the s/d regions and contacts [1]. Increasing the electrically active dopant concentration reduces resistance, but this is limited by solid solubility, electrical deactivation, dopant loss and segregation at interfaces, combined with the need for strict control of dopant diffusion. Earlier studies identified how various MSA recipe conditions can affect electrical activation for heavily-doped n-type layers formed by ion implanting As or P into silicon [2,3]. High-dose P doping provided better activation than As doping, where activation degraded for higher doses. Higher peak temperatures helped activation, but temperatures are limited by integration of materials such as SiGe. Once the MSA pulse length exceeds 0.7ms, it has a weak influence on activation. Higher preheat temperatures degraded dopant activation for heavy As doping, probably because of increased dopant deactivation during the relatively slow cooling stage after the pulsed heating, although deactivation was only significant at temperatures >725ºC. For heavy P doping there was little evidence of deactivation effects, but there was a strong link with the high temperature pulsed heating stage. The concept of a dopant solid solubility limit defined by a process temperature was found to be of limited value; it is essentially an equilibrium concept that does not address kinetics, which are important for low thermal budget methods like MSA [3]. Furthermore, the electrical activation can be dominated by the solid-phase epitaxy (SPE) that regrows the initial amorphous layer and may finish during the preheat stage or the early part of the pulsed heating [2]. This paper extends the study of MSA process conditions, especially by exploring the role of the preheating stage.

II. INITIAL EXPERIMENTS

300mm p-type (100) Czochralski Si wafers with resistivity of 1-50Ωcm were preamorphized by Si^+ implantation to create an amorphous silicon (a-Si) layer ~50nm thick, which was then implanted with either As^+ at 2 keV or P^+ at 1 keV, with doses of either $10^{15}/cm^2$ (samples As1 and P1) or $3x10^{15}/cm^2$ (samples As3 and P3). The wafers were annealed in Mattson Technology's Millios® MSA tool, which provides real-time measurement and control of the temperatures of both the front and the back surfaces of the wafer. The details of implant and anneal conditions have been described previously [2].

Initially, four sets of eight wafers with each type of implant were processed with the MSA anneal conditions summarized in Table I. These experiments included variations in peak temperature (T_{peak}), intermediate temperature (T_i) and heating pulse lengths (t_P), defined by the time spent at temperatures > T_{peak}-50K. In some cases the cooling cycle was delayed by adding a 1s-duration "soak" step at 725ºC, in order to explore the role of the post-pulse cooling. Doping distributions were measured by secondary ion mass spectrometry (SIMS) and electrical properties were determined by sheet resistance (R_S) measurements. The peak values for carrier activation, N_A, were calculated by combining the R_S and SIMS profile data. This approach has previously been found to give N_A data broadly consistent with results from differential Hall effect measurements [2].

TABLE I. INITIAL MILLISECOND ANNEAL RECIPES

Anneal Type	Anneal Conditions			
	T_{peak} (ºC)	T_i (ºC)	t_P (ms)	Cooling
Normal Pulse (NPMSA)	1175	725	0.7	Normal
High Peak T	**1200**	725	0.7	Normal
High Intermediate T (HTIMSA)	1175	**775**	0.7	Normal
Short Pulse	1175	725	**0.2**	Normal
Long Pulse	1175	725	**1.0**	Normal
Short Pulse, Delay Cool	1175	725	**0.2**	**Delay**
Normal Pulse, Delay Cool	1175	725	0.7	**Delay**
Long Pulse, Delay Cool	1175	725	**1.0**	**Delay**

978-1-5386-6829-0/18/$31.00 ©2018 IEEE

III. Effective Time Analysis

One challenge comes from how to interpret the effect of a complex, multi-stage heating profile on the final electrical activation, as reflected by N_A, after annealing. N_A depends on many phenomena, including SPE, dopant diffusion, segregation effects and the atomic-scale interactions between atoms and point defects that are responsible for electrical activation and deactivation. For a physical effect with Arrhenius kinetics, an effective time, t_{eff}, can be used to summarize the impact of a temperature cycle as the time required at constant reference temperature to produce the same effect. t_{eff} is given by

$$t_{eff} = \int_{t1}^{t2} \exp\left\{-\frac{E_A}{k}\left(\frac{1}{T(t')}-\frac{1}{T_{ref}}\right)\right\}dt', \qquad (1)$$

where T(t) describes the absolute temperature as a function of time, T_{ref} is an arbitrary reference temperature, E_A is an activation energy, k is Boltzmann's constant and t1 and t2 are the times at the start and end of the period of interest, respectively [3].

Fig.1 shows the various stages considered in the analysis, including preheating (PH) to T_i, the fast pulsed heating (FH) to T_{peak}, and then a 30ms-long period of very fast cooling (FC), during which the wafer's surface rapidly loses heat by thermal conduction to the cooler substrate. At the end of FC the wafer has a uniform temperature, T_{post}, and further isothermal cooling (SC) happens by the much slower process of heat loss from the wafer's surfaces.

For the calculation of t_{eff}, E_A should characterize the kinetics of the main physical phenomenon affecting the process result. However, several phenomena could be involved and continuous changes in the microstructure during the anneal may make the best choice of E_A unclear. In this study, t_{eff} values were calculated for each of the 32 wafers processed, for each stage of the heating cycle, for a range of E_A values between 0.5 and 6eV, with T_{ref}=725°C. When N_A was plotted against t_{eff} for each combination of recipe stage and E_A, some datasets showed an interesting degree of correlation. For example, Fig.2 shows N_A for each type of implant plotted against t_{eff} for the SC stage, calculated with E_A=1.9eV. The figure includes the square of the correlation coefficient for linear fits, R^2, which is zero when there is no correlation and unity if there is perfect correlation. For the heavily-As-doped wafers, As3, there is a good correlation as N_A falls with increasing t_{eff} (R^2=0.83). The E_A of 1.9eV is consistent with existing theories about As deactivation kinetics [4]. One might expect that N_A for As3 would also correlate with the highest temperature in the SC stage, T_{post}, but that correlation was weaker (R^2=0.52), illustrating the benefit of looking at effective process times rather than specific temperatures in the heating cycle. For the other implants, there is no correlation. On the other hand, Fig.3, with N_A plotted against t_{eff} for the FH stage, calculated with E_A=6eV, only shows a strong correlation for P3 (R^2=0.91), suggesting that N_A for P3 is dominated by the hottest part of the heating cycle. Indeed, plotting N_A against T_{peak} also gave a reasonable correlation (R^2=0.73), although not quite as strong as that seen in Fig.3, underlining the value of including the dynamics of the temperature cycle, rather than relying on specific temperature values.

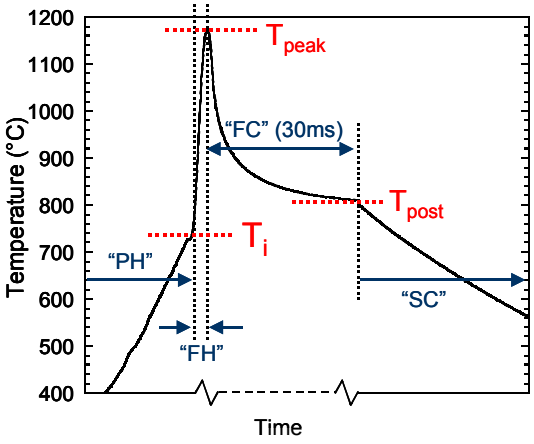

Fig. 1. Definitions of the stages of MSA recipes.

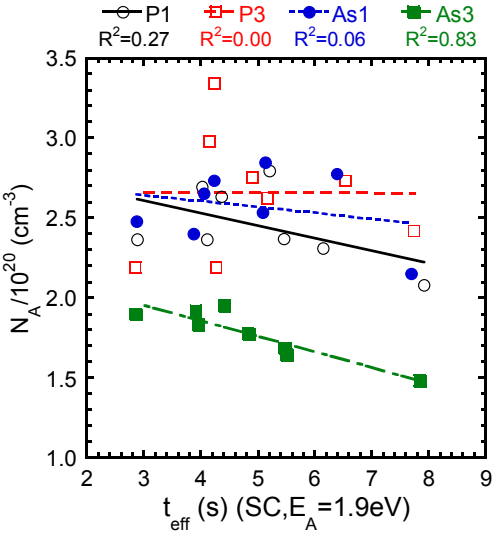

Fig. 2. Linear fits of peak carrier concentration, N_A, against the effective time, t_{eff}, during the slow cooling stage (SC), with E_A=1.9 eV. For the As3 implant, the relatively high correlation coefficient (R^2=0.83) suggests that the slow cooling after the pulse is responsible for electrical deactivation.

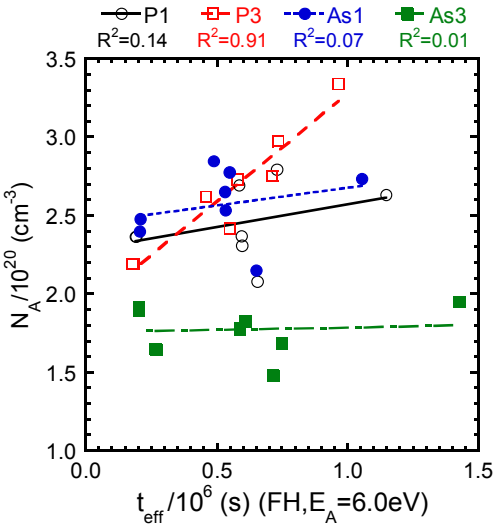

Fig. 3. Linear fits of peak carrier concentration, N_A, against the effective time, t_{eff}, during the fast heating stage (FH), with E_A=6.0 eV. For the P3 implant, the high correlation coefficient (R^2=0.91) suggests that the fast pulsed heating is the main factor affecting carrier activation.

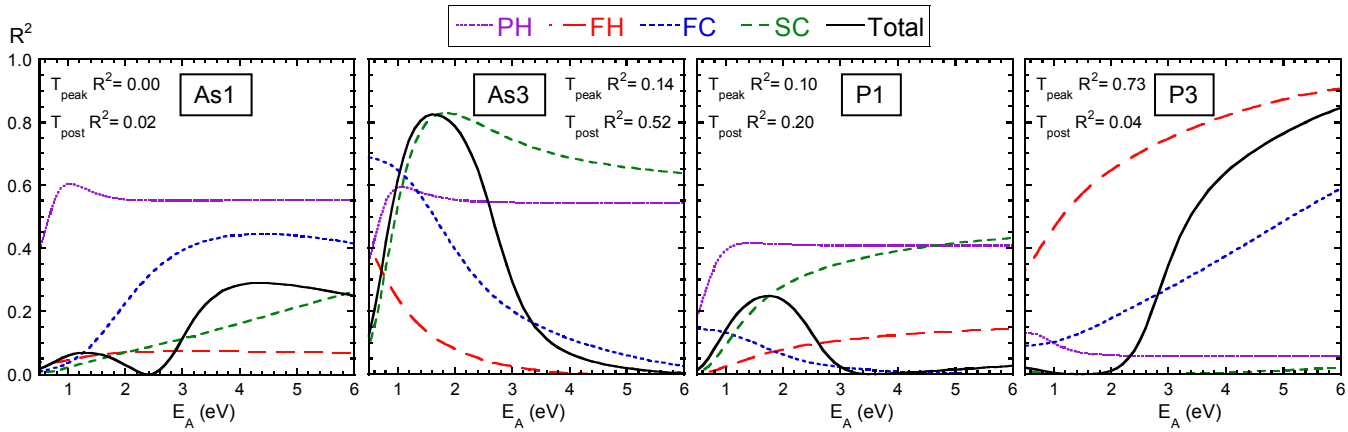

Fig. 4. The square of the correlation coefficients (R^2) for linear fits of the maximum carrier concentrations (N_A) to effective times, t_{eff}. R^2 is shown for the fits to t_{eff} calculated in each of the four heating stages shown in Fig.1 (PH, FH, FC and SC) and for the whole anneal, for E_A between 0.5 and 6eV. The analysis highlights differences in the factors that affect the carrier activation in the four implant types (As1, As3, P1 and P3). The figure also includes R^2 for fits to the peak temperature, T_{peak} and to the temperature immediately after the pulsed heating, T_{post}.

Fig.4 summarizes the complete t_{eff} analysis by showing R^2 as a function of E_A, for each implant type and recipe stage, and also for the whole heating cycle. Each figure includes the R^2 values for fits against T_{peak} and T_{post}, two key temperatures that characterize the heating cycle. Caution is necessary in assessing the significance of R^2; correlation is not causation. t_{eff} values for different recipe stages may not be independent of each other and the experimental conditions in Table I did not vary t_{eff} much for some recipe stages. Nevertheless, the analysis suggests factors to investigate in more detail. Fig.4 demonstrates the significance of the slow cooling for the As3 samples through the peak in R^2 at ~1.9eV, and that of the fast pulsed heating for P3, where R^2 rises with E_A. For the lower dose cases, P1 and As1, N_A is insensitive to variation in t_{eff} for the pulsed heating and cooling stages. Here, electrical activation could be dominated by the initial SPE, which is likely to finish during either the preheat stage or the early part of the pulsed heating [2].

IV. Isolating the Preheat Effects

Since the initial experiments did not vary the preheat t_{eff} significantly, a second set of experiments was devised to explore this aspect. An interesting question arises about the role of SPE in dopant activation, as opposed to the effect of the very high temperature stage of the MSA anneal on dopant solubility. Increasing the temperature of the SPE regrowth may improve dopant activation [5], but this is difficult to uncouple from the effects of dopant deactivation during cooling; the approach taken here was to employ four recipes with very different preheating/ramp-up stages but nearly the same slow cooling cycle. They include the "normal pulse" condition of the previous MSA cycles (NPMSA) and a recipe similar to NPMSA but with a 30s soak at 625°C before the 150K/s ramp to the T_i of 725°C (PSMSA). This 30s "pre-soak" is long enough for SPE to fully crystallize 50nm of undoped a-Si. Two other annealing methods were used; one was a backside pulsed heating anneal (BSH), where the recipe was identical to NPMSA but the wafer was loaded upside-down so that the pulsed heating was delivered to the back of the wafer, rather than the implanted side. The implanted region experiences the same "slow" heating cycle as in NPMSA but not the fast high-temperature heating pulse; the result is a spike anneal with a very fast ramp from T_i to the peak temperature (=T_{post}). The other anneal was a conventional spike anneal, ramping up at 150K/s to the T_{post} temperature for NPMSA. Fig.5 shows examples of the four types of anneal, focusing on the preheating/ramp up and the slow cooling stage.

These anneals have very similar slow cooling stages, confirmed by the closely matched t_{eff} values in Table II, calculated for E_A=1.9eV to emphasize that they should cause similar deactivation. The table also includes predictions of the extent of a-Si regrowth, d_{reg}, possible during the preheat stages of the MSA recipes or the ramp-up stages of the BSH and spike anneals. d_{reg} is calculated based on well-known SPE kinetics, and is directly proportional to t_{eff} for E_A= 2.68eV [2]. The large differences in the preheat/ramp-up stages lead to a large variation in d_{reg}. For NPMSA, the initial 50nm-thick a-Si layer will not fully crystallize during the preheat, whereas for PSMSA, d_{reg} is >100nm. For the spike anneal a 50nm a-Si layer should regrow before reaching T_{peak}, but for BSH, regrowth would only finish after the ramp up. One major caveat here comes from the fact that these d_{reg} calculations do not take account of the effect of the heavy doping on SPE, which could have a large effect on the last 10nm of regrowth [6].

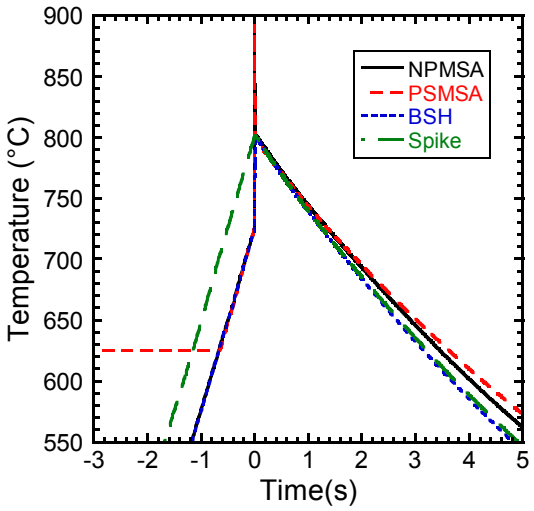

Fig. 5. Temperature-time profiles for the "slow parts" of the anneals that were used to explore the effects of different preheating/ramp-up schemes. The slow cooling starts from ~800°C for all four anneals, so that the effective time is closely matched in the cooling stage. The BSH cycle is like NPMSA, but without the pulsed surface heating to ~1175°C, which is only visible as a vertical line at 0s in this graph.

TABLE II. COMPARISON OF THE CHARACTERISTICS OF THE ANNEALS USED TO EXPLORE THE EFFECT OF PREHEATING. t_{eff} WAS CALCULATED FOR E_A=1.9eV IN THE SLOW COOLING STAGE; d_{reg} WAS CALCULATED FOR THE PREHEAT/RAMP-UP STAGES

	As1			As3			P1			P3		
Recipe	T_{peak} (°C)	t_{eff} (s)	d_{reg} (nm)	T_{peak} (°C)	t_{eff} (s)	d_{reg} (nm)	T_{peak} (°C)	t_{eff} (s)	d_{reg} (nm)	T_{peak} (°C)	t_{eff} (s)	d_{reg} (nm)
NPMSA	1175.0	4.1	19	1176.8	4.0	19	1176.3	4.0	19	1182.4	4.1	19
PSMSA	1173.3	4.0	105	1167.7	4.0	105	1173.9	4.0	105	1168.3	4.0	105
BSH	800.6	3.7	36	800.1	3.7	33	801.1	3.8	33	801.4	3.8	34
Spike	801.5	3.8	204	801.4	3.8	203	-	-	-	-	-	-
HTIMSA	1177.0	7.7	88	1177.4	7.8	90	1178.8	7.9	90	1169.1	7.7	90

Table II also includes data for the high-T_i MSA (HTIMSA). This case was included for comparison here, because of all the recipes in Table I, HTIMSA gave the lowest N_A. This might be expected from the greater deactivation expected from a higher T_{post}, but the preheating condition is also different, which complicates interpretation [2]. This can be seen from the numbers in Table II; the t_{eff} is ~7.8s for HTIMSA, but only ~4s for NPMSA, while d_{reg} is ~90nm and 19nm in the two cases, respectively.

Fig.6 compares R_S for the anneals in Table II. SIMS data were not available in all cases, so the R_S data are used as an approximate guide to the activation trends. For all the implants, PSMSA gave higher R_S than NPMSA. Although the peak temperatures in the PSMSA, shown in Table II, tended to be slightly lower than those for the NPMSA, such large rises in R_S suggest that increasing the t_{eff} in the preheat stage has a significant effect on dopant activation. For P1, As1 and P3, the R_S from HTIMSA was slightly lower than that from PSMSA, suggesting that the long pre-soak stage may be just as significant as the higher T_{post} in the HTIMSA case. For the As3 case, deactivation effects are expected to dominate, which probably explains why the HTIMSA gave a slightly higher R_S than the PSMSA.

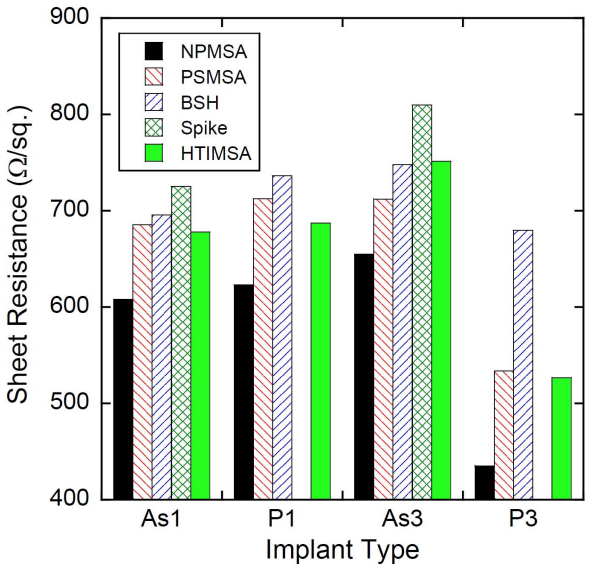

Fig. 6. Sheet resistances after the anneals in Table II (NPMSA, normal pulse MSA; PSMSA, pre-soak MSA; BSH, backside pulsed heating; Spike, spike anneal; HTIMSA, high-T_i MSA). The first four recipes had matched slow-cooling curves, but different preheating/ramp-up stages.

Comparing the NPMSA with the BSH anneal, we see that the high temperature MSA stage leads to much lower R_S. However, for P1, As1 and As3 the R_S from PSMSA is only slightly lower than for BSH, and it is not clear whether the R_S reduction in NPMSA comes from the difference in the SPE temperature, or from increased dopant solubility at the peak temperature of the MSA. A comparison of results from the low-temperature spike anneals is instructive; for the As implants studied here, the R_S values from BSH are lower than for the spike anneals. Table II shows that the lower thermal budget of the BSH anneals during the ramp-up to the peak temperature would cause less regrowth before the peak temperature is reached than in the spike anneals. For low-temperature spike anneals the final result also depends on more than just the peak temperature and the cooling curve; the nature of the ramp-up stage is also significant.

V. CONCLUSIONS

A series of experiments was devised to isolate the effects of dopant deactivation during cooling from effects introduced during preheating in MSA recipes. For both low temperature spike anneals and MSA, the ultimate activation is affected by more than just the peak temperature and the cooling behaviour, underlining the need for fuller understanding of kinetic effects, especially for the early stages of annealing. Variations in the preheating conditions produced significant changes in sheet resistance and it seems that dopant activation could be improved by reducing the thermal budget in the preheating. However, further work will be needed to establish the physical phenomena involved and identify an optimal preheating strategy.

REFERENCES

[1] Y. C. Yeo, 10th International Workshop on Junction Technology (IWJT), 2010, (Shanghai, 2010), p.55.

[2] P. J. Timans, M. Hagedorn, C. Pfahler, A. Cosceev, M. Zwissler, L. Rubin, and A. Joshi, 2016 21st International Conference on Ion Implantation Technology (IIT), (Tainan, 2016), p.272.

[3] P. J. Timans, 18th International Workshop on Junction Technology (IWJT), (IEEE, 2018) p. 28.

[4] D. Nobili, S. Solmi, and J. Shao, J. Appl. Phys., vol.90, (2001) p.101.

[5] S. H. Jain, P. B. Griffin, J. D. Plummer, S. McCoy, J. Gelpey, T. Selinger, and D. F. Downey, IEEE Trans. Electron. Dev., vol.52 (2005) 1610.

[6] S. Ruffell, I. V. Mitchell and P. J. Simpson, J. Appl. Phys., vol.98 (2005) 083522.