

Maciej Frankiewicz

Streszczenie pracy doktorskiej pt.:

**„Generator taktujący TCO sterowany temperaturą  
jako nowy element systemu zwiększania mocy obliczeniowej i redukcji energii zasilania”**

Technologia fabrykacji współczesnych układów scalonych podlega ciągłemu procesowi skalowania w celu sprostania rosnącym wymaganiom związanym ze zwiększaniem zapotrzebowaniem na moc obliczeniową. Oznacza to, że razem ze zmniejszaniem rozmiarów charakterystycznych elementów zawartych w układzie, znacznemu wzrostowi ulega gęstość mocy wydzielanej w obudowie tego układu. W rezultacie podwyższona zostaje temperatura pracy układu scalonego, co powoduje zaistnienie problemów termicznych, które muszą być brane pod uwagę przez inżynierów w procesie projektowania. Zjawiska termiczne występujące w układach scalonych powodują zmniejszenie niezawodności ich pracy, a tym samym prowadzą do błędów w pracy lub, w najgorszym przypadku, do uszkodzenia układu.

Głównym celem pracy jest przedstawienie metody dynamicznej kontroli mocy wydzielanej w układzie scalonym oraz implementacja tej metody w prototypowym układzie scalonym. Zaprezentowany układ stanowi generator sygnału zegarowego przestrajany temperaturą - TCO (ang. *Temperature-Controlled Oscillator*), którego zadaniem jest utrzymanie jak najwyższej wydajności pracy taktowanego układu, zwykle procesora, przy czym konieczne jest aby jego temperatura była niższa od dozwolonej maksymalnej wartości. Rozważana metoda została oparta na technice dynamicznego skalowania częstotliwości, a sygnał sterujący oscylatorem wytwarzany jest bezpośrednio na podstawie pomiaru wewnętrznej temperatury układu scalonego. W pracy tej została przedstawiona i udowodniona następująca teza:

***Możliwe jest zaprojektowanie scalonego generatora przestrajanego w sposób ciągły i natychmiastowy temperaturą układu scalonego, gwarantującego bezpieczeństwo termiczne pracy cyfrowego układu scalonego (np. procesora) przy jednoczesnym zapewnieniu możliwie wysokiej wydajności pracy tego układu.***

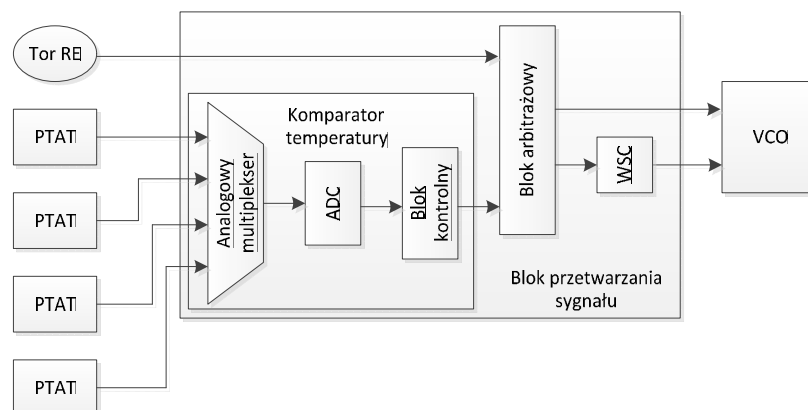
Struktura pracy składa się z kilku zasadniczych części ujętych w siedmiu rozdziałach. Pierwszy z nich stanowi wprowadzenie do problematyki pracy, definiuje jej tezę oraz streszcza zawartość dalszej części pracy. Treść drugiego rozdziału stanowi przegląd literatury dotyczącej istniejących metod kontroli wydajności pracy układu scalonego z uwzględnieniem jego temperatury (DPM – ang. *Dynamic Power Management*). Koncentruje się on na fizycznych realizacjach metod, pozwalających na ocenę ich przydatności. Metody te najogólniej można podzielić na realizowane w sposób programowy lub sprzętowy. Pośród rozwiązań sprzętowych można wyróżnić metody przestrzenne (ang. *spatial*), tj. wykorzystujące zasoby sprzętowe architektury procesorów wielordzeniowych (np. migrację wątków pomiędzy rdzeniami procesora) oraz czasowe (ang. *temporal*), tj. wpływające na wydajność pracy procesora w trakcie jego działania poprzez regulowanie parametrów fizycznych. W ostatniej grupie należy wymienić techniki dynamicznego bramkowania zegara (DCT – ang. *Dynamic Clock Throttling*), dynamiczne skalowania napięcia zasilania (DVS – ang. *Dynamic Voltage Scaling*) oraz dynamiczne skalowania częstotliwości taktowania zegara (DFS – ang. *Dynamic Frequency Scaling*). Szczególną uwagę w tym rozdziale poświęcono rozwiązaniom predykcyjnym, mającym potencjalnie zagwarantować większą efektywność pracy układu. Zauważono, że wszystkie rozwiązania usiłujące przewidzieć temperaturę układu scalonego są metodami z grupy metod programowych. Tymczasem układ będący przedmiotem pracy doktorskiej w tej klasyfikacji należy zaliczyć do grupy metod sprzętowych – jest on oparty na technice DFS z pewnymi elementami techniki DCT oraz predykcją temperatury metodą sprzętową.

Kolejne części pracy dotyczą badań przeprowadzonych bezpośrednio przez autora oraz opisują prototypowe wersje układu zaprojektowane i przetestowane przez autora. Rozdziały następujące po przeglądzie literatury oraz ich przedmiot stanowią oryginalny wkład własny autora pracy. Przed przystąpieniem do prac projektowych wykonano szereg działań mających na celu dokładne poznanie zjawisk termicznych mogących zachodzić w układach scalonych oraz mających wpływ na ich

działanie. Zagadnieniom tym poświęcono rozdział trzeci pracy. Jego pierwsza część dotyczy pomiarów parametrów termicznych układu scalonego pracującego w różnych konfiguracjach, włączając w to sposób chłodzenia układu scalonego, jego położenie oraz zastosowaną obudowę. Zaproponowana konfiguracja pracy układu ma istotny wpływ na temperaturę osiąganą przez układ scalony. Z tego powodu konieczna jest dokładna znajomość wartości tych parametrów. Wiedza ta pozwala również na precyzyjne konstruowanie modeli termicznych układów scalonych. Zebrane wyniki zostały wykorzystane w drugiej części rozdziału trzeciego. Opisuje on szczególny przypadek bezwładności termicznej układu scalonego, mogący mieć duże znaczenie przy projektowaniu układów dynamicznie zarządzających własną temperaturą. Badane zjawisko zostało najpierw wyjaśnione symulacyjnie na podstawie modelu termicznego układu scalonego, a następnie eksperymentalnie dowiedziono prawdziwości wyciągniętych wniosków oraz udowodniono możliwość wystąpienia tego zjawiska w rzeczywistym układzie scalonym. Ostatnia część rozdziału trzeciego rozważa wpływ temperatury na różne konstrukcje generatorów przestrajanych napięciem. Taki układ stanowi jeden z najważniejszych elementów projektowanego układu TCO. Kluczowy element przygotowania do prac projektowych związany był z wyborem właściwej struktury generatora oraz znajomością jego reakcji na zmianę temperatury pracy.

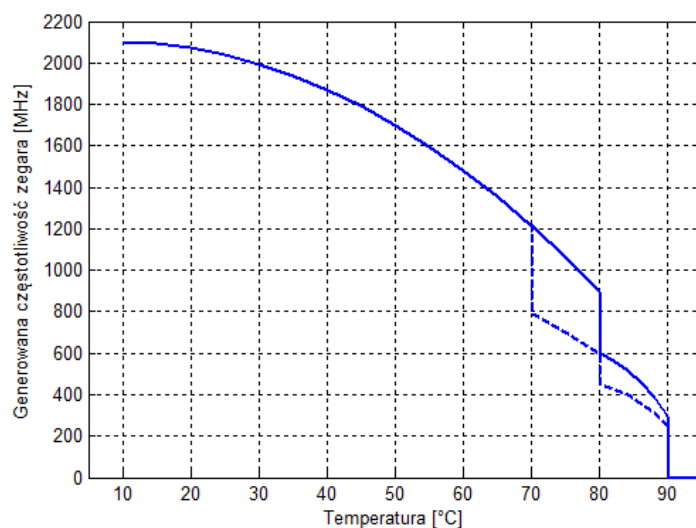
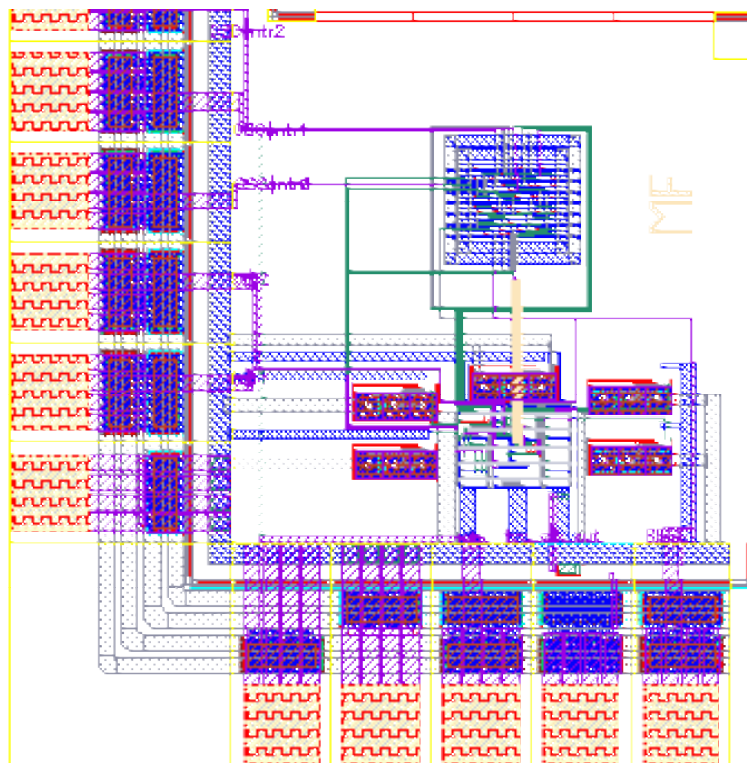
Po zapoznaniu się z możliwymi zjawiskami termicznymi występującymi w układach scalonych oraz przeanalizowaniu wpływu temperatury na układy scalone możliwe było rozpoczęcie prac stanowiących teoretyczne podwaliny pod projekt układu TCO. W rozdziale czwartym zawarto analityczne rozważania na temat działania projektowanego układu oraz wpływu temperatury na częstotliwość generowaną przez układ TCO. Dzięki usystematyzowanemu podejściu matematycznemu możliwe były również analizy parametryczne, pozwalające na przebadanie wpływu kształtu charakterystyk czujnika temperatury oraz przestrajanego generatora na punkt pracy układu TCO. Rozdział piąty przedstawia analizy numeryczne badające w sposób behawioralny możliwe do zaimplementowania struktury układu TCO. Przebadane modele pozwoliły na ocenę efektywności różnych metod ochrony przed przegrzaniem układu scalonego z zastosowaniem generatorów przestrajanych temperaturą. Rozdział ten pokazuje również przydatność metod predykcyjnych pozwalających na osiągnięcie lepszej wydajności pracy układu scalonego. Wyniki przedstawione w rozdziałach czwartym i piątym miały bezpośrednie przełożenie na wybór zaimplementowanej struktury układu scalonego.

Najistotniejszą część pracy stanowi rozdział szósty opisujący zaprojektowany prototypowy układ generatora TCO, będącego wynikiem prac zawartych we wszystkich poprzedzających rozdziałach. Przedstawione zostały dwie zaprojektowane, wyprodukowane i przetestowane wersje układu: uproszczone, realizująca podstawowe funkcjonalności układu TCO, oraz pełną, zawierającą całość przewidywanych do realizacji funkcji. Podstawową cechą wyróżniającą zaprojektowany układ spośród innych istniejących rozwiązań jest jego ciągła (tj. realizowana w sposób analogowy, nie dyskretny) oraz natychmiastowa (tj. bez zbędnego opóźnienia) reakcja na temperaturę układu scalonego. Zrealizowaną strukturę układu przedstawia poniższy rysunek.



Z powyższego rysunku wynika, że możliwe są dwa podstawowe tryby kontroli częstotliwości generowanej przez układ, a tym samym mocy wydzielanej przez pracujący procesor: programowy RE (ang. *Required Efficiency*), realizujący wydajność wymaganą przez blok nadrzędny (np. system

operacyjny) lub sprzętowy, na podstawie informacji z czujników temperatury PTAT. Czujniki te zostały w sposób równomierny rozmieszczone na powierzchni układu scalonego. Informacja pochodząca z czujników ulega przetworzeniu w celu odnalezienia maksymalnej wartości temperatury występującej chwilowo w układzie scalonym. Maksymalna odnaleziona wartość temperatury posłuży następnie, po przetworzeniu przez blok kształtowania charakterystyki (WSC, ang. *Wave Shaping Circuit*), do bezpośredniego przestrajania generatora sygnału taktującego procesor (VCO, ang. *Voltage-Controlled Oscillator*). Blok kontrolny zawiera również układ predykcyjny, wskazujący na kierunek zmian temperatury, a tym samym pozwalający uwzględnić informację o przyszłej wartości temperatury w procesie przestrajania generatora. Czas przetwarzania sygnału pochodzącego z czujników temperatury jest pomijalnie mały w porównaniu z termicznymi stałymi czasowymi układu scalonego, tym samym uzasadnione jest stwierdzenie natychmiastowej reakcji układu TCO. Poniższe rysunki przedstawiają finalną wersję topografii układu wyprodukowanego w technologii CMOS UMC 180 nm 1,8 V oraz zmierzoną zależność pomiędzy aktualną maksymalną temperaturą układu scalonego oraz częstotliwością sygnału zegarowego generowane przez układ TCO.



Zaprojektowany układ może współpracować w zasadzie z dowolnym procesorem zdolnym do pracy ze zmienną częstotliwością zegara, jednak w celach testowych zrealizowana została architektura procesora OctaLynx D zawierającego jednostkę odpowiedzialną za obsługę zdarzeń termicznych. Procesor ten oprócz m.in. dedykowanych rejestrów zawierających informację o aktualnie maksymalnej wartości temperatury dostarczonej przez układ TCO oraz nastawy dla generatora VCO (w przypadku sterowania programowego) zawiera również zestaw przerw termicznych pozwalających na programową kontrolę termiki procesora. Przerwania te zostały wykorzystane w porcie systemu operacyjnego FreeRTOS na opisywany procesor. Procesor razem z dedykowaną wersją systemu operacyjnego stanowią elementy toru RE..

Własnym, oryginalnym wkładem autora w stan wiedzy technicznej wniesionym przez powyższą rozprawę są:

- koncepcja układu TCO jako generatora przestrajanego temperaturą w sposób ciągły, przy jednoczesnym zastosowaniu sprzętowej metody predykcji temperatury,
- próba klasyfikacji istniejących metod zarządzania mocą wydzielaną w układzie scalonym z uwzględnieniem metod predykcyjnych oraz umiejscowienie projektowanego układu w zaprezentowanej klasyfikacji,
- analiza wybranych zjawisk termicznych występujących w układach scalonych, a zwłaszcza model i pomiar laboratoryjny szczególnego przypadku bezwładności termicznej układu scalonego polegającego na lokalnym wzroście temperatury powierzchni układu po ograniczeniu wydzielania w nim mocy,
- analiza działania, modelowanie, projekt oraz testy układu TCO wyprodukowanego w dwóch wersjach: o podstawowej i pełnej funkcjonalności.

Przeprowadzone prace, a zwłaszcza końcowy projekt i testy układu scalonego potwierdzają i udowadniają tezę zawartą na początku pracy. W trakcie realizacji pracy opublikowanych zostało ponad trzydzieści artykułów, które ukazały się w specjalistycznych czasopismach (w tym cztery artykuły w czasopismach posiadających współczynnik „Impact Factor”) oraz materiałach znaczących konferencji międzynarodowych i krajowych. Prezentacje konferencyjne dotyczące treści przedstawianych w pracy były nagradzane przez komitety naukowe konferencji: prezentacja zatytułowana „Thermal models for dynamic clock control” została nagrodzona „Outstanding Paper Award” podczas 19-tej międzynarodowej konferencji Mixed Design of Integrated Circuits and Systems mającej miejsce w Warszawie w 2012 roku, natomiast prezentacja „Overheat protection circuit for high frequency processors” została wyróżniona podczas X Krajowej Konferencji Elektroniki mającej miejsce w Darłównie w 2011 roku. Autor pracy był stypendystą w ramach projektu „Doctus - Małopolski fundusz stypendialny dla doktorantów” współfinansowanego ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego. Większa część prac badawczych oraz projektowych została sfinansowana z grantu Narodowego Centrum Nauki „Generator taktujący TCO sterowany temperaturą jako nowy element systemu zwiększania wydajności oraz redukcji energii zasilania procesorów” o numerze NCN N N515 500340. Część dotycząca ekstrakcji parametrów termicznych układu scalonego została sfinansowana z grantu dziekańskiego pod tytułem „Badanie własności termicznych układu scalonego wraz z modelami termicznymi procesora dla generatora taktującego sterowanego temperaturą” o numerze 15.11.230.091.